DATA-INDEPENDENT MASS SPECTROMETRY STRATEGIES FOR THE IDENTIFICATION OF ATRA-MEDIATED PROTEIN SIGNATURES OF DIFFERENTIAL CELLULAR RESPONSE

JOHN D CHAPMAN

A dissertation submitted in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy
University of Washington
2013

Reading Committee:
Dr. David R. Goodlett, Chair
Dr. Nina Isoherranen
Dr. Thomas Baillie

Program authorized to offer degree:
Department of Medicinal Chemistry

University of Washington

Abstract

Data-independent mass spectrometry strategies for the identification of atRA-mediated protein signatures of differential cellular response

John D Chapman

Chair of the Supervisory Committee:
Professor David R. Goodlett
Department of Medicinal Chemistry

Mass spectrometry is a powerful proteomics tool. Advancements in instrumentation and data acquisition techniques allow researchers to identify and quantify thousands of proteins from cellular samples in a high throughput fashion. Here, I examine the field of data-independent acquisition strategies. More specifically, I illustrate the development of a novel CSI PAcIFIC approach to reduce the total sample analysis time from 4.2 days to 12 hours without deleteriously affecting the quality and quantity of protein identifications. This CSI PAcIFIC method is then used to expand the understanding of the divergent cellular response of MCF-7 and HepG2 cells when treated with all-*trans* retinoic acid.

In this dissertation, I also propose a new technique for identifying poly-ADP-ribosylation via CID mass spectrometry. Enzymatic simplification of the heterogeneous post-translational modification proves valuable for the efficient identification of covalently modified amino acids.

Contents

Chapter 1

A REVIEW OF MASS SPECTROMETRY BASED PROTEOMICS TOOLS AVAILABLE TO STUDY RETINOID-MEDIATED ALTERATIONS IN CELLULAR FUNCTION WITH A FOCUS ON ONCOLOGICAL APPLICATION

1.1 Retinoids and their use in clinical applications

Retinoids are a family of structural analogs of vitamin A important in mediating transcriptionally activated cellular responses.[1] Here, I will introduce naturally present retinoids and routes of dietary consumption, absorption, and cellular trafficking before I further define retinoids, structural analogs and their role in oncology as antineoplastic agents. Vitamin A, also known as retinol, is consumed via numerous dietary sources;[2] and post-consumption, several metabolites are enzymatically synthesized that are necessary in a diverse array of biological processes including differentiation of epithelial tissues, embryonic development, growth, immune response, reproduction, and vision.[3, 4] Vitamin A is a naturally occurring compound that is sourced in the diet from two families of structurally related compounds: carotenes and retinoids.[2] Compounds within the family of carotenes are found in vegetables and fruits such as carrots, spinach, and cantaloupe melon as well as animal sources including eggs and milk. In the brush border of the small intestine, one pathway for retinoid absorption involves the carotene dioxygenase catalyzed break down of carotenes to retinal, the aldehyde analog of retinol. Retinal is further reduced to retinol by retinaldehyde reductase. Subsequently, retinol is esterified with palimitic acid to form retinyl palmitate, which is readily incorporated into chylomicrons that are transported through the blood to the liver.[5] Alternatively, vitamin A is ingested from animal sources in the form of retinol and retinyl esters, a common storage form of vitamin A in the liver of animals. For example, liver cod oil contains an extremely high dose of vitamin A in the form

of retinyl ester. Consumed retinyl esters are de-esterified in the small intestine and join the pathway of absorption into the blood stream by re-esterification to retinyl palmitate. Upon delivery by chylomicrons, the retinyl esters are stored in hepatic stellate cells. As needed, retinyl esters are de-esterified to retinol, which binds to apo-retinol binding protein (apo-RBP). The ligand bound retinol binding protein (halo-RBP) is the primary form of circulating vitamin A.[6, 7]

Cellular delivery of vitamin A occurs when halo-RBP interacts with the transmembrane protein, stimulated by retinoic acid gene 6 protein (STRA6), which facilitates transport of retinol across the cell membrane. On the cytoplasmic membrane interface, STRA6 releases retinol to cellular retinol binding protein (CRBP), which further facilitates the deliver to enzymes responsible for the conversion of retinol to retinoic acid.[8] Retinol is oxidized by alcohol dehydrogenases (ADH) and retinol dehydrogenases (RDH) to retinal. Further oxidation to the acid form of vitamin A, retinoic acid, occurs enzymatically via retinaldehyde dehydrogenase (RALDH).[9-11] Upon formation of retinoic acid, cellular retinoic acid binding protein 2 (CRABP-II)[12] or fatty acid binding protein 5 (FABP5)[13] will transport retinoic acid to the nucleus and deliver the compound to specific nuclear transcription factors (discussed in more detail in the following section). Alternatively, retinoic acid can be metabolized by a family of cytochrome P450 enzymes including CYP26A1, CYP26B1, and CYP26C1 to form polar metabolites including 4-hydroxy-retinoic acid and 4-oxo-retinoic acid.[14] These polar metabolites are then excreted from the cell via conjugation elimination.[15]

Vitamin A has long been known for its role in healing. For more than 3500 years it has been used as a factor that can cure deficiency diseases.[16] More recently, the ability of retinoids to influence the cell cycle, differentiation and apoptosis in a cell-type dependent manner has led to the prescribed use of retinoids as human chemotherapeutic agents.[17] Pharmaceutically, retinoid

compounds can be categorized into three generations of analogs (Figure 1).[18] First generation compounds constitute naturally occurring compositions comprising three currently used in medical applications: all-*trans* retinoic acid (tretinoin), 9-*cis* retinoic acid (alitretinoin) and 13-*cis* retinoic acid (isotretinoin). The high activity of retinoids in keratinocytes has led to their prescribed use in the treatment of skin health issues such as acne vulgaris, cystic acne, keratosis pilaris and psoriasis.[19] Additionally, all-*trans* retinoic acid and 13-*cis* retinoic acid are also used in chemotherapeutic regiments. More specifically, atRA has demonstrated efficacy in the therapy of acute promyelocytic leukemia,[20] and 13-*cis* retinoic acid is used in the treatment of neuroblastoma.[21] Second generation retinoids include two aromatic retinoids: etretinate, and the acid metabolite, acitretin. Both second generation compounds are not used for antineoplastic effects, but alternatively have proven effective in the treatment of severe psoriasis.[22] Third generation retinoid compounds are tazarotene, adapalene, and bexarotene. Of the third generation compounds, only bexarotene displays antineoplastic qualities. As illustrated in Figure 1, bexarotene shows high structural similarity to 9-*cis* retinoic acid, and its selectivity as an retinoic X receptor (RXR) agonist has proven beneficial in the treatment of cutaneous T cell lymphoma.[23] Additional off-label uses of bexarotene include lung and breast cancer. The following embodied work will focus on the use of all-*trans* retinoic acid, as it is the most potent agonist of the retinoic acid receptor (RAR) in the retinoid family. Documented oncology cases of the application of atRA include the treatment of acute promyelocytic leukemia,[20, 24] peripheral T-cell lymphoma,[25] neuroblastoma,[26] and hormone-refractory prostate cancer.[27] Despite wide spread use, the exact molecular mechanism, and more specifically, the global alterations in protein pathways and activity that lead to the phenotypic effects of atRA are poorly understood.

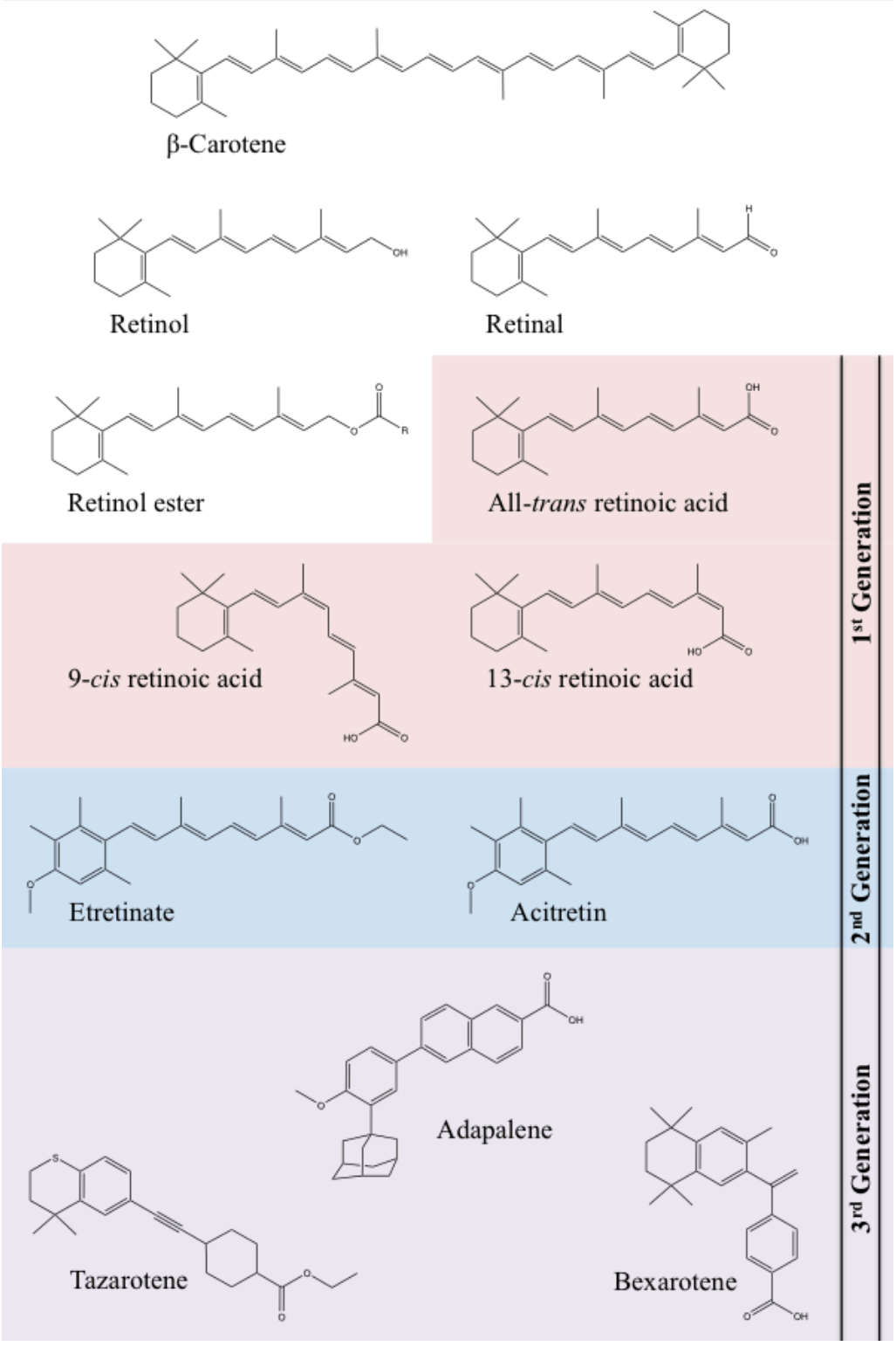


Figure 1. Family of retinoid compounds including a representative carotene, β-carotene, and the three pharmaceutically relevant generations of retinoid analogs.

1.2 Pleiotropic cellular response of all-*trans* retinoic acid

The challenges associated with understanding atRA-mediated changes in protein expression are a result of the pleiotropic effects of the compound. It is well established that atRA mediates gene expression and the resulting inhibition of cellular growth through interactions with ligand-activated transcription factors. The most prominently affected transcription factors are the family of retinoic acid receptors: -alpha,[28] –beta,[29] and -gamma[30] (RAR-α, RAR-β, RAR-γ). RAR forms a heterodimer with retinoid X receptor (RXR)[31] when the two transcription factors are bound to specific sequences of DNA defined as retinoic acid receptor elements (RARE).[32] RAREs are characterized by direct repeats (DR) of two hexanucleotides with a defined base pair gap between the protein binding of RAR and RXR. The RAR/RXR heterodimer will bind to sequences with 2 (DR2) and 5 (DR5) base pair gaps. When atRA binds to RAR, the protein undergoes conformational changes that weaken protein-protein interactions and allow for dissociation of transcription complex corepressors. Subsequently, coactivators can bind to the atRA activated RAR/RXR heterodimer and activate downstream gene expression. Transportation of atRA to the nucleus via binding proteins is vital in the process of atRA-mediated gene expression. For example, cellular retinoic acid-binding protein 2 (CRABP-II), plays a role in the process of atRA-mediated gene expression and a subsequent decrease in proliferation as CRABP-II delivers atRA directly to RAR.[12] At chemotherapeutically active doses, atRA will activate a series of protein pathways involved in apoptosis through de novo protein expression. For example, in the breast carcinoma cell line, MCF-7, atRA has been shown to induce expression of intrinsic apoptosis pathway factors such as initiator caspase-9 and executioner caspase-7. These proteins subsequently activate secondary protein pathways resulting in

impaired mitochondrial function and release of pro-apoptotic mitochondrial factors including cytochrome C.[33]

Paradoxically, it has been observed that atRA can also promote cell growth and proliferation in a tissue-specific manner, as documented in keratinocytes.[34] In recent years, the peroxisome proliferator-activated receptor beta/delta (PPAR-β/δ) has been shown to interact with and be activated by atRA; thereby providing an alternate pathway by which the pleotropic effects of atRA can be more fully defined. Much like the RAR/RXR heterodimer, PPAR-β/δ forms a heterodimer with RXR. Retinoic acid binds to PPAR-β/δ to promote the expression of a subset of genes that differs from RAR-mediated gene expression.[35] Whereas the RAR/RXR heterodimer binds to DR2 and DR5 RARE sequences, PPAR-β/δ promotes the expression of a unique proteome by preferential interaction with a pair of hexanucleotides separated by a 1 (DR1) base pair gap.[36] Additionally, delivery of atRA, and thus activation of the PPAR-β/δ/RXR heterodimer, is accomplished by a unique transport protein, fatty acid-binding protein 5 (FABP5). This alternative route of atRA-mediated gene expression activates a multitude of compounding anti-apoptotic and pro-survival signal pathways.[13] Categorically, PPAR-β/δ-mediated gene expression leads to changes in fatty acid transportation, fatty acid oxidation, adipocyte differentiation, adaptive thermogenesis, ubiquitination, and cell survival.[37-39] More specifically, atRA-activated PPAR-β/δ/RXR increases transcription of a family of phosphatidylinositol 3-kinases (PI3K), which subsequently activate the serine/threonine-specific protein kinase, Protein Kinase B (PKB or Akt). In turn, activated Akt: 1) impedes apoptotic pathways by inhibiting proteins such as Bcl-2-associated death promoter (BAD); 2) activates cell proliferation mechanisms including mammalian target of rapamycin (mTOR); and 3) induces the means for increased protein synthesis necessary for cell signaling and growth.[40, 41]

The diverse range of cellular response attributed to atRA has long been of concern to the research and clinical community. Increasing interest in using atRA as a chemotherapeutic for its potential to promote differentiation, cell cycle arrest and apoptosis requires that the scientific community also understands the diverging consequence of atRA as an activator of cell proliferation[42]. Recently, Schug et al. demonstrated that the balance of CRABP-II and FABP5 are vital to the fate of the cell as they control transport of atRA to their respective nuclear transcription factors. Basal levels of CRABP-II and FABP5 are critical in partitioning atRA between RAR and PPAR-β/δ and activating the respective pro- and anti-apoptotic cellular responses. CRABP-II and RAR interact with atRA with high affinity, in the range of 0.1-0.2 nM. Both FABP5 and PPAR-β/δ associate with atRA at an affinity of 10-50 nM. Reasonably, it has been noted that activation of PPAR-β/δ will only be displayed in cells that have a high basal FABP5/CRABP-II ratio due to the noted differences in atRA affinity. For example, in keratinocytes, levels of FABP5 are much greater than CRABP-II and therefore cell proliferation and growth are the dominantly activated cellular mechanisms.[35] To more fully understand the complex cell signaling cascades generated by RAR and PPAR-β/δ, there is need to further characterize the transcribed protein alterations that occur following atRA treatment. Discovering secondary protein activity that contributes to the delicate balance of pro- and anti-apoptotic activation will generate a more thorough comprehension of the tissue-specific responses to atRA and the utility of this compound in the therapy of cancer, skin care, and obesity.

1.3 Applied proteomics approaches

Mass spectrometry (MS) is an increasingly powerful tool in proteomics and is available to be an essential technique employed in understanding the global effects on the proteome mediated by atRA.[43] Unfortunately, the widespread adoption of new MS technologies and methodologies often lags far behind development. Much of the current literature elucidating atRA-induced differences in proteome profiles is founded on techniques that resolve proteins using two-dimensional gel electrophoresis (2-DE).[44, 45] 2-DE is capable of simultaneously resolving thousands of proteins.[46] But, the fundamentals of the technique impose several limitations that negatively impact the ability of researchers to observe the global alterations of the proteome necessary in understanding the complete molecular mechanism of atRA.[47] In addition to the caveats pertaining to gel spot excision, efficiency of protein digestion, peptide extraction and MS characterization, it has been well characterized that both highly acidic and basic proteins as well as large molecular weight proteins are vastly underrepresented in 2-DE.[48] While improvements such as two-dimensional difference gel electrophoresis (2D-DIGE) have improved the quantification, reproducibility, resolution and sensitivity of 2-DE,[49] the loss of previously mentioned protein groups still confines the scope and application for which this method is suitable. To expand the current understanding of atRA-mediated changes in protein expression and protein activity we must turn to mass spectrometry-based shotgun proteomics. As a highly refined technique, it is possible to confidently categorize a cellular proteome in a rapid, reproducible and quantitative manner.[50] In the following sections I will discuss advances in proteomics acquisition techniques and means of protein quantification as available tools to characterize atRA-mediated proteome alterations.

1.4 Mass spectrometry acquisition techniques

Following the genomic revolution, proteomics, the study of the protein complement of a biological system, has gained a stronghold in modern life sciences. This happened in spite of the intricate nature of proteomes, which makes their analysis far less efficient than genomics. The complexity of proteomic studies centers on the extremely dynamic nature of protein expression and degradation. Nevertheless, the knowledge of the proteins present in a given biological state is vital for our understanding of cellular functions at the molecular level. The means by which we generate detailed lists of proteins and their relative abundance levels with respect to environmental or patho-physiological conditions are therefore critical in solving challenging questions in biology and medicine.[51, 52] To derive such information, experimental approaches termed "shotgun proteomics" have been developed, in which liquid chromatography (LC) and mass spectrometry are used in combination to analyze highly complex peptides mixtures obtained by proteolytic digestion of protein extracts from biological or clinical samples.[43] An essential aspect of this analysis is the ability to individually sequence ideally all tryptic peptides present in such highly complex mixture, with the best analytical throughput possible.

To overcome this challenge, MS instrumentation has demonstrated remarkable improvements in data acquisition rates with a concurrent increase in sensitivity over the past decade.[53, 54] In parallel, advances in chromatography have yielded ever greater peak capacity, providing enhanced proteome coverage by further lowering detection limits.[55, 56] Yet, despite how far the field has evolved since the beginning of proteomics in the early 1990's,[57, 58] modern technologies still fall short of the goal of defining a whole proteome in a single analysis. Michalski et al. have thus recently demonstrated what most researchers in the area had observed or surmised, i.e. that the number of peptides sequenced during a liquid chromatography-tandem

mass spectrometric (LC-MS/MS) analysis is dwarfed by the total number of detected species in the associated precursor ion trace.[59] Hence, in spite of continuing advances, mass spectrometers still appear challenged to individually derive sequence information for each peptide observed in the MS trace during an LC separation, prompting the quest for ever more evolved technologies to do so. The major challenges of such a quest lie in the large dynamic range of protein expression, the complexity of the sample mixture in terms of numbers of proteins, their relative stoichiometry, and the lack of methods to amplify proteins in a manner analogous to the polymerase chain reaction for DNA. As a consequence, for a particular peptide to be sequenced, the corresponding protein must be present in sufficient quantity following multiple steps of sample processing to yield an exploitable signal.

It is generally accepted that the dynamic range problem may be solved by sample fractionation and/or enrichment. Therefore, most laboratories engage in various strategies to separate proteins and/or peptides prior to mass spectrometric analysis using traditional methods from protein chemistry, such as anion or cation exchange, one- or two-dimensional polyacrylamide gel electrophoresis, isoelectric focusing and affinity capture or depletion.[60] Ideally, when coupled with each other and the reversed phase LC-MS/MS analysis most commonly used in shotgun proteomics, these methods should provide orthogonal separation in order to achieve the best peak capacity and ensure that each fraction is as unique as possible. In order to make the most of fractionation, LC-MS/MS acquisition methods should provide efficient sampling to thoroughly characterize all the peptides in each fraction at every moment in chromatographic time. In practice, however, none of the separation techniques at hand provides samples of sufficient simplicity to allow the currently available mass spectrometers to achieve this goal of complete characterization. Additionally, each separation or enrichment step produces

multiple fractions each of which must be characterized, which reduces the analytical throughput, and also introduces experimental variability that renders quantitative proteomics more difficult and challenging.[61]

One of the most important advances that empowered shotgun proteomics was the advent of automated data acquisition routines. While manufacturers tend to have their own trade name for these computer methods that control the mass spectrometer, I will refer generically to these as data-dependent acquisition (DDA). With DDA methods, the most abundant ionized species from each precursor ion scan are selected for subsequent isolation, activation and tandem mass analysis.[62] In addition, in an effort to limit redundancy, a companion method referred to as "dynamic exclusion" (DE) attempts to prevent reselection of the same precursor ion over and over again in successive dissociation events. This allows, in theory, the maximum number of unique ions to be selected by the DDA methods.[63] These processes have greatly increased proteome coverage and extended the dynamic range of detection for shotgun proteomics by maximizing the number of selected precursor ions most likely to lead to peptide identification, which are generally thought to be those of highest relative abundance.[63] However, as alluded to earlier, this combination of DE and DDA methods has generally failed to provide full coverage of peptides in a complex mixture. The deficit of these otherwise remarkable methods has been described in numerous studies showing the non-reproducible nature of peptides detected in replicate analyses of the same sample.[64, 65]

DDA methods have two principal inherent limitations: (1) precursor ion selection is strategically designed to pick ions in order of decreasing relative abundance while requiring an abundance threshold minimum for ion selection, and therefore DDA methods only sample a restricted dynamic range of the proteome; and (2) the serial nature of DDA, and use of DE,

makes it improbable that a tandem mass spectrum for a precursor ion will be acquired at the apex of its chromatographic elution profile, which adversely affects the data quality for purposes of both identification and quantification.[66] For these reasons, several groups have investigated ways to circumvent the "one peak at a time" selection process of DDA, giving rise to the concepts of multiplexed[67] and data-independent acquisition (DIA).[68] These approaches are aimed at fully utilizing the capabilities of mass spectrometers to maximize MS acquisition time and to address the need to expand the detectable dynamic range, lower the limit of detection, and improve on the overall confidence of peptide identifications and relative protein quantification measurements. In this review, I will inventory advancements made in the field of multiplexed and DIA methods and characterize how specific strategies seek to overcome the current limitations of DDA and expand the overall coverage of the proteome.

Like multiplexing tandem MS approaches, data-independent acquisition (DIA) methods are being developed to expand the detectable dynamic range of the proteome and increase the accuracy of protein quantification measured via mass spectrometry. The defining characteristic of DIA methods is that they do not use real-time ion selection based on precursor ion scans to obtain searchable spectra for peptide identification. Therefore data acquisition is in fact independent from prior information collected for the sample at hand (i.e. an MS survey or precursor ion scan). To accomplish this, predetermined *m/z* ranges are interrogated either: (1) by fragmenting all ions entering the mass spectrometer at a single point in chromatographic time (broadband DIA); or (2) by dividing the full *m/z* range into fixed smaller *m/z* isolation windows that are each independently and consecutively analyzed. In the analysis of complex proteomics samples, the *m/z* range of various DIA methods affects the inherent presence of multiplexed spectra. Therefore, DIA methods often differ in the way they deal with data analysis. Here, I will

trace the lineage of DIA techniques from in-source fragmentation to in-cell fragmentation as well as numerous further manipulations of in-cell fragmentation using static, predetermined *m/z* ranges for batch analysis.

Broadband DIA using in-source fragmentation

Data-independent acquisition for complex proteomic samples truly began with the development of methods involving nozzle-skimmer dissociation by Purvine et al. (*vide supra*).[69] Shotgun CID circumvented the need for ion selection by increasing the nozzle-skimmer voltage to induce in-source peptide fragmentation of all ions as they entered the mass spectrometer. It was inherent that multiplexed spectra would be produced when peptides co-eluted from a chromatrographic separation. Purvine et al. demonstrated that two unique peptide identifications could be assigned to a single "pseudo-tandem mass spectrum" by manually generating two .dta files containing identical tandem MS information, but different assignments of the precursor ion *m/z* values in each file. This information could then be analyzed using conventional search algorithms such as SEQUEST.[70] As admitted by the authors, at the time of publication, a serious limitation of shotgun CID was the fact that automated software for data interpretation of data-independent acquired multiplexed spectra was not available, and presented a bottleneck in the application of this DIA workflow for the analysis of complex proteomic samples.

Purvine's et al. work built on previous reports of in-source fragmentation for proteomics applications.[71, 72] In particular, van Dongen et al. had previously demonstrated that peptide fragment ions generated in the source region of mass spectrometers closely resembled those formed by conventional CID in the collision cell.[73] While improvements to in-source fragmentation for complex proteomic applications have been explored,[74] it should be acknowledged that, compared to in-cell fragmentation techniques, spectra recorded using nozzle-

skimmer fragmentation contain increased chemical noise in the low *m/z* region, which can be attributed, in part, to over fragmentation of peptides.[75] Regardless, the work of Purvine et al. was critical in the development of DIA for proteomic applications, as it constituted a proof-of-concept on how multiplexed spectra recorded on moderate resolution and mass accuracy instruments and independent from precursor ion workflows could be used to assign unique sequences to co-fragmented peptides.

Broadband DIA using in-cell fragmentation

Soon after shotgun CID was described, Waters Corporation put forth a related method they referred to as MS^E, developed for use on their Q-TOF instruments. The MS^E method constituted an enhancement of the shotgun CID approach that instead implemented peptide ion fragmentation in the more controlled environment of a collision cell.[76] Silva et al. utilized an LC/MS-TOF and cycled the collision cell energy voltage between a low energy value (10 eV) to capture MS information, and a high energy value (28-35 eV) to fragment all incoming precursor ions and produce a high mass accuracy multiplexed tandem mass spectrum (called a MS^E spectrum). The ability to rapidly switch between low and high collision energy in the collision cell allowed for alternating scans to be collected in rapid succession during a single LC-MS analysis. As a result, all precursor and fragment ion information could be captured at multiple points during the elution of a chromatographic peak – a clear advantage compared to serial CID acquisition methods. The generation of multiplexed spectra was a major data analysis challenge for the MS^E technique and a dedicated search algorithm termed "Ion accounting" was developed specifically for this purpose (*vide infra*). The original 2005 publication by Silva et al. also demonstrated that the relative quantification values assigned in the above process showed strong linear correlation over a concentration gradient of 0.1 pmol to 5 pmol of exogenous protein

spiked into a human serum sample, illustrating for the first time the utility of broadband DIA for quantitative purposes.

Since its inception, MS^E has become a routine acquisition method on several Waters instruments and has been successfully utilized in studies including: the drug-induced changes in Mycobacteria Tuberculosis,[77] a quantitative analysis of the Escherichia coli proteome,[77] the proteomic analysis of the root-not nematode *Meloidogyne hapla*[78] as well as metabolomics applications.[79, 80] More recently, the technology was successful in characterizing the secretome of human embryo prior to implantation[81] revealing differences between positive and negative-implantation groups, and thus providing putative secreted biomarkers of competent embryos. Another interesting study using MS^E yielded a novel potential target to control cisplatin resistance in neuroblastomas.[82]

In several reports, DIA approaches have demonstrated significant increases in the number of identifications and much improved sequence coverage for the analyzed proteins.[83-86] For instance, Blackburn et al. reported dramatic improvement in individual protein sequence coverage and overall proteome coverage obtained by the DIA MS^E approach for three samples ranging in complexity from a simple mixture of four commercially available proteins to a highly complex in-solution digest of the protein content of a whole tomato leaf.[85] For the four-protein mixture, although all four proteins could be readily identified by both methods, the sequence coverage for individual components was increased by a factor of 2- to 32-fold using the DIA-derived data. For the more complex tomato leaf proteome sample, the average number of peptides per protein increased from 1.4 to 2.6 from the DDA to the DIA approach. In addition, the overall proteome coverage as determined by the number of unique identified proteins was substantially higher using the DIA dataset (576 proteins) than with the corresponding DDA

dataset (162 proteins). Moreover, the authors insisted that improvements were most pronounced for the lowest abundance sample components. In their analysis, they found many instances of precursor ions identified during DIA that were either sampled by DDA at unfavorable time during their chromatographic separation, thus resulting in poor quality MS/MS spectra, or not sampled at all due to dynamic exclusion of adjacent peaks.

One apparent exception to this trend of improved proteome coverage using DIA comes from an article by Bern et al. who stated that the DDA method yielded more identifications than the DIA.[66] In this case, however, it is important to note that the DIA method was performed on a fraction of the total *m/z* range used in the DDA experiment. When this element was factored in, the actual identification yield was higher by 12% for proteins and 20% for peptides in favor of DIA.

In analytical sciences, reproducibility is a synonym of reliability, and in this respect, DIA methods, and in particular MS^E and PAcIFIC (*vide infra*) have shown significantly better reproducibility than their DDA counterparts on the basis of identical chromatographic separations. For example, in the analysis of a simple protein mixture of only four components using DDA, Blackburn et al. reported that only 12.5% of the peptides were observed across all three replicate injections using, whereas nearly 70% of the peptides were observed in only one out of three replicate.[85] Conversely, in the MS^E analysis of the same sample, nearly 80% of the identified peptides were observed across all three replicate injections, with less than 7% of them occurring in only one of the three replicates. Similar results have also been reported by Geromanos et al.[84] and Bern et al.,[66] who announced 88% reproducibility of peptide identification in replicate DIA analyses of a rat blood plasma sample. In this latter article, they also pointed out increased reproducibility of spectral counts in DIA compared to DDA.

The implementation of in-cell fragmentation and the inherent benefits of data-independent acquisition have been demonstrated in numerous robust biological applications. Since conception of MS^E, this mode of data-independent acquisition has been adapted and implemented on other instrumental platforms for the analysis of complex samples, including those of interest in proteomics. In particular, it has been coupled with ion mobility in an effort to provide enhanced peak capacity. Building on an initial demonstration by Hoaglund-Hyzer et al.,[87] Baker et al.[88] thus demonstrated that the ramping of collisional energy with drift time yielded better quality tandem mass spectra across the whole ion mobility spectrum. They noted that this method had the potential to substantially reduce the under-sampling problem associated with DDA methods while preserving the specificity of peptide identifications, thus providing enhanced proteome coverage. It should be noted though that ramping collision energy over time in a collision cell was a common practice in use on triple quadruple mass spectrometers prior to the advent of ion trap mass spectrometers in proteomics.

A recent alteration of MS^E is the so-called "all-ion fragmentation" (AIF) which was developed specifically for the Thermo Exactive instrument, a benchtop Orbitrap equipped with a C-trap and an HCD collision cell. Unlike conventional Orbitrap hybrid instruments featuring linear ion traps,[89] the Exactive is not capable of precursor ion isolation as required for DDA. As a consequence, the only applicable MS/MS method was one that did not involve ion selection, hence the term AIF.[90] The authors demonstrated that confident qualitative peptide assignments and accurate quantitative measurements could be obtained with AIF using alternating scans between: (1) an accurate precursor ion mass measurements accomplished by accumulating ions in the C-trap prior to injection into the orbitrap; and (2) an all-ion fragmentation scan obtained by injecting accumulated ions from the C-trap into the HCD cell for fragmentation, and then

transferring fragment ions back to the C-trap before injection into the orbitrap for mass analysis. In addition, the authors implemented a three-step fragmentation technique that ramped the collision energy during ion accumulation to enhance peptide sequence coverage. The observation that peptides require varying degrees of collision energy to generate comprehensive peptide ion coverage for purposes of confident identification is a common issue with shotgun CID approaches. Ramos et al. also addressed this problem on a Q-TOF by performing parallel in-source fragmentation and in-cell fragmentation, termed p^2CID.[75] The resulting total ion fragmentation information was more comprehensive due to the presence of additional complementary fragment ions. The authors noted that this technique was useful for larger peptides for which additional fragment ion information can be necessary for confident identification. In AIF experiments, the speed of the Exactive allowed for ultra-high resolution scans (100,000 at 200 *m/z*) to be obtained every second, for an overall duty cycle of two seconds. Typically, precursor-fragment ion lineage is lost in data-independent acquisition due to a lack of ion selection. Like in the MS^E process, this information can be reconciled by considering the fact that a precursor ion and its respective fragment ions have the same elution profile since they both result from the same eluting peptide. In AIF, an elution profile is generated in which intensity values of precursor and fragment ions must precisely change in concert with one another. Based on a correlation coefficient of the determined elution profiles, AIF peaks are assigned to potential precursor ions and the generated tandem-mass spectra are searched just like conventional datasets.

DIA on selected m/z ranges

Up to this point I have discussed data-independent acquisition methods that revolve around fragmenting all ions entering a mass spectrometer at each single point in

chromatographic time. The other possibility is that all ions in a narrow, select *m/z* range are fragmented. Depending on the *m/z* window width and the scanning rate of the instrument, this may require multiple LC-MS/MS analyses to be conducted, each covering a select section of the entire *m/z* range of interest. In practice these contiguous *m/z* ranges of interest can either overlap or not as each successive *m/z* range is profiled.

In 2004, Venable et al. demonstrated this novel approach to data-independent acquisition by employing a relatively small isolation window (10 *m/z* wide).[68] The experimental sequence operated in a cycle that covered a desired total *m/z* region by sliding the initial isolation window. The use of a restricted window size had two distinct advantages: (1) a relative small window limits the *m/z* diversity of precursor ions that can contribute to a tandem mass spectrum, thereby simplifying downstream data analysis; and (2) this strategy can effectively increase the measured dynamic range since highly abundant species only affect a narrow portion of the *m/z* space. As this process occurred in a single LC analysis, it was critical to allow for repetitive sampling of each isolation width during a single chromatographic peak to ensure confident peptide identification and improved accuracy of quantification. In contrast to conventional DDA techniques that employ "dynamic exclusion" to minimize precursor ion oversampling in an effort to increase throughput and detectable dynamic range, Venable et al. sought to capitalize on the data-independent strategy to more thoroughly sample each chromatographic peak (i.e. sampling each peptide multiple times) effectively increasing identification confidence and, more importantly, the accuracy of quantification. This quantitative advantage was demonstrated with superior DIA measurements compared to a conventional DDA method when analyzing a 1:1 and 10:1 trypsin digested unlabelled and ^{15}N-labeled yeast-whole cell lysates. Additionally, the authors noted the advantages of performing quantification measurements from fragment ion

mass spectra as opposed to precursor ion scans: (1) increased signal to noise ratios (up to 350%), (2) increased sensitivity, and (3) increased specificity based on peptide fragmentation, similar to what is obtained using selected reaction monitoring experiments (SRM). This technique also has an efficiency advantage over conventional DDA methods in that no time is spent acquiring precursor ion information. By only acquiring tandem mass spectra, 20% more spectra were obtained in a single LC analysis over that of a comparable DDA method. Although this did not correspond with an increase in peptide identifications in the Venable study, it was noted that by utilizing a DIA strategy with an isolation width of 10 *m/z*, the acquired spectra were inherently multiplexed. The authors did not address the need for assigning identifications from multiple peptide fragments in a single spectrum and instead only assigned one peptide per spectrum, thereby negating some of the advantages of the proposed methodology. The latter point is likely due to the lack of dedicated software solutions to address multiplexed data at that time.

In developing this novel technique, Venable et al. investigated many parameters associated with selecting an optimal isolation width with respect to the resulting spectral noise and peptide fragmentation efficiency. The authors generalized that spectra generated with larger isolation width typically contained increased noise over those generated with smaller isolation widths. At the selected precursor ion isolation width of 10 *m/z*, the effect of this noise did not appear to impact peptide identifications. Due to the data-independent nature of this acquisition method, it was also not known at the time if the location of the precursor ion within the isolation window affected the degree of fragmentation. In other words, would a peptide ion with an *m/z* at the periphery of an isolation window fragment with the same efficiency as one with an *m/z* at the center of the isolation window? To answer this question, Venable et al. characterized the fragmentation efficiency of a precursor ion by shifting the center *m/z* of the isolation window

with respect to the actual precursor ion *m/z*. Again, no effect on fragmentation efficiency was noted for a 10 *m/z* isolation width.

Over the past few years, innovative methods have been proposed that extend the initial concept put forth by Venable in ways that may well transfigure the analysis of complex proteomic samples. These methods include: Precursor Acquisition Independent From Ion Count (PAcIFIC),[91] extended data-independent acquisition (XDIA),[92] SWATH-MS,[93] and Fourier transform-all reaction monitoring (FT-ARM).[94] All methods proceed in accordance with the following motif: (1) accumulate precursor ions from a selected *m/z* window; (2) collectively fragment the accumulated precursor ions; (3) record the resulting (multiplexed) tandem mass spectrum; (4) shift the *m/z* accumulation window and repeat beginning with step (1). These methods thus rely on selective precursor ion accumulation and subsequent fragmentation that is achievable only in instruments with trapping capabilities. Depending on the design of the acquisition method and the size of the sliding *m/z* isolation window, these data-independent methods may or not employ dedicated methods for deciphering inherently acquired multiplexed spectra.

PAcIFIC

Recently, Panchaud et al. proposed a novel data-independent acquisition technique referred to as Precursor Acquisition Independent From Ion Count (PAcIFIC)[91] that attempts to embody the best of DDA and DIA methods. Namely, data-independent tandem MS of peptides in narrow *m/z* ranges to maximize the detectable dynamic range. PAcIFIC was initially developed as the natural extension of the DDA method Gas-Phase Fractionation (GPF)[95] where the entire *m/z* range of a mass analyzer is divided up into 4-12 narrow regions from which DDA ion selection is done. Like the GPF experiment, PAcIFIC is iterative in the sense that a sample

must be analyzed *n* times for *n* designated *m/z* ranges. The GPF method has been shown to greatly increase the dynamic range by forcing the DDA process to examine ions of lower and lower relative abundance. In an attempt to maximize identifications, however, Scherl et al. observed that the efficiency of identifications reached a plateau beyond which further use of narrower *m/z* ranges (~ < 50 *m/z*) failed to provide additional identifications. This likely stemmed from the fact that, in DDA analysis, the total number of identification that could be made from a complex sample depends directly on the number of detected precursor ions in each survey scan. This observation led in turn to the development of PAcIFIC where peptides present in all *m/z* channels are systematically subjected to tandem MS regardless of the presence of a precursor ion, thus acquiring tandem mass spectra even when no precursor ion is present.

Specifically, each successive 2.5 *m/z* channel is subjected to MS/MS across an entire 1000 *m/z* range. As with GPF, a sample must be examined multiple times, with the number of injections required being directly proportional to the scan rate of the ion trap used and the chromatographic conditions. For example, the first analysis might cover the *m/z* range from 400-430 with multiple selection channels centered on contiguous 2.5 *m/z* segments, overlapping by 1.5 *m/z* in order to select isotopic distributions that would otherwise be missed (see Figure 2A). The use of an isolation width that is similar to that used by DDA methods means that the ion trap is filled to its maximum capacity with ion originating only from a narrow *m/z* range at a time. This is similar to selected reaction monitoring (SRM) and differs from the way most DDA methods work, where the ion trap is filled by ions from the whole *m/z* range of interest, and all ions are ejected, save for the ones of interest for tandem MS. Consequently, the PAcIFIC approach provides an effective gain in detectable dynamic range, while the analytical throughput is dictated by the scanning rate of the mass spectrometer employed. Initially, this constraint

limited PAcIFIC analysis to 10 adjacent *m/z* channels per LC-MS/MS acquisition, but subsequent developments of the method and faster scanning ion trap instrumentation have extended this to ~ 25 *m/z* channels per run, for a total analysis time of 2.5 days. As depicted in Figure 2B, to cover the entire *m/z* range of interest for proteomic analysis, e.g. 400-1400 *m/z*, the initial PAcIFIC method required 67 injections of the same sample, resulting in a total sample analysis time of 4.2 days. On the other hand, the design of the small 2.5 *m/z* isolation width allowed for several other distinct advantages over other DIA techniques that utilize relatively large *m/z* isolation widths. First, there is typically no need to reconstruct the lineage between precursor and fragment ions, a step that is often necessary with other DIA methods. By default, the fragment ions of each PAcIFIC generated tandem mass spectrum originate from a specific precursor ion within the narrow *m/z* range and, as depicted in Figure 2C, the assigned precursor ion *m/z* value for a scan event is represented by the center of the 2.5 *m/z* channel. Using standard database search algorithms, one can thus accommodate the presence of any possible precursor ion that may be in the trap after isolation by simply increasing the precursor ion mass tolerance. Second, as mentioned above in comparison to DDA methods, the use of the small 2.5 *m/z* isolation width greatly increases the measureable dynamic range of the overall PAcIFIC analysis by negating the effect of highly abundant species outside of the narrow *m/z* window being interrogated. The authors thus reported an exceptional detectable dynamic range of eight orders of magnitude at an FDR of 0.5%. Finally, a quantitative advantage is achieved with such DIA techniques because iterative methods such as PAcIFIC allow for each chromatographic peak to be interrogated multiple times for each unique *m/z* channel.

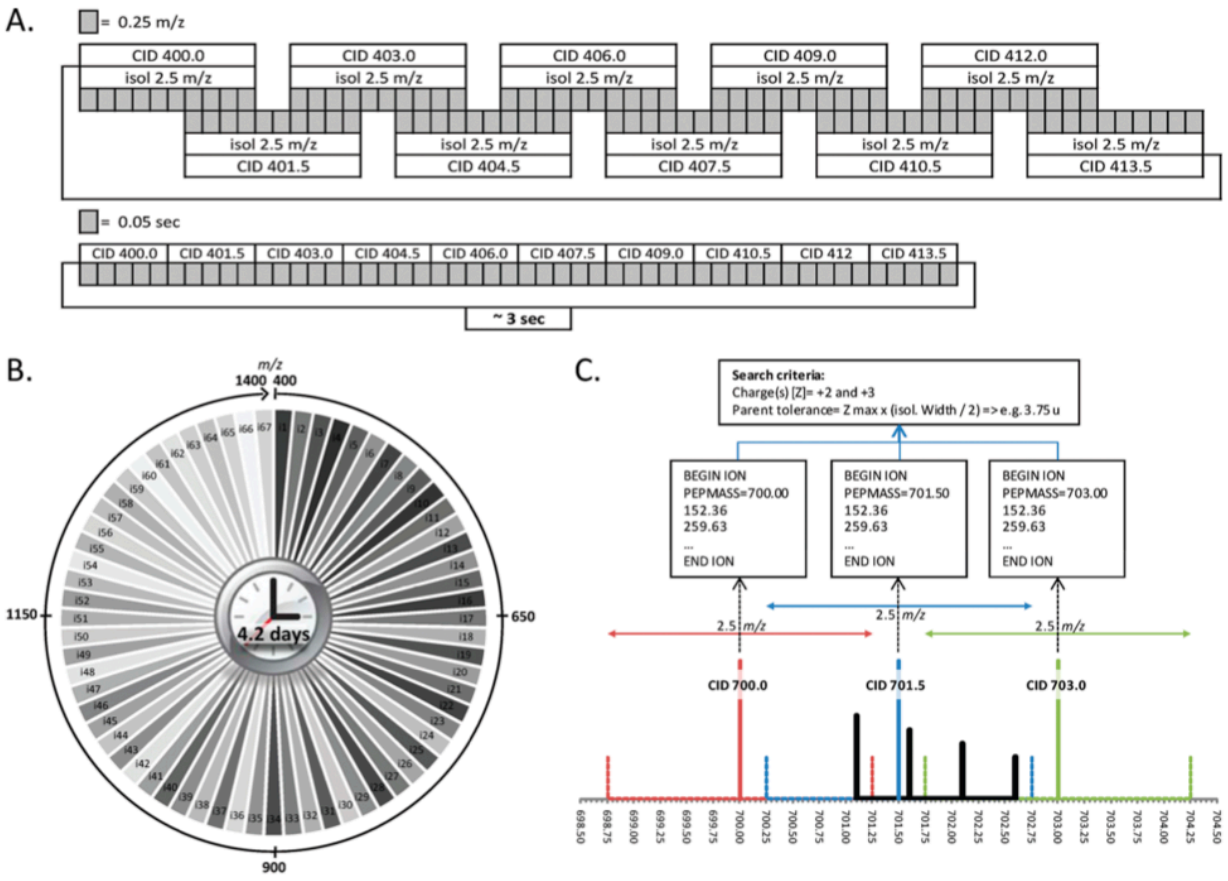


Figure 2. Reprinted from Panchaud et al. Anal. Chem. 2009, 81, 6481–6488: Principle of PAcIFIC. With this method, the sample is repeatedly analyzed by LC-MS/MS. (A) During each injection, a different precursor-ion m/z range is selected for tandem mass spectrometry independent of the precursor ion measurement. (B) In each narrow m/z range, all precursor ions are isolated in the ion trap and subjected to collision-induced dissociation until the desired m/z range is achieved. (C) Each spectrum is then searched against a database using the center of the window as the precursor mass and a tolerance corresponding to maximum charge state allowed times the isolation width divided by 2, e.g., 3(2.5/2) = 3.75 Da

One of the most popular schemes for quantitative proteomics using DDA-based mass spectrometry has been the convenience provided by tandem mass tags (TMT)[96] and the iTRAQ reagent kits.[97] Both approaches allow $n > 2$ samples to be combined after each is uniquely coded with a mass tag. The use of such mass tags has become very popular because it greatly increases throughput, provides quantitative analysis and can identify thousands of proteins from a complex sample in a single analysis. In 2011, Panchaud et al. published the results of coupling PAcIFIC with TMT and demonstrated increased throughput compared to DDA methods. In this report, the authors described three unique advancements to the PAcIFIC method:[98] (1) With the release of the LTQ Velos (Thermo-Fisher Scientific), a dual pressure ion trap capable of approximately twice the scan rate of the prior generation ion trap from the same manufacturer, PAcIFIC analysis of a single sample was decreased from the original 4.2 days to 2.5 days. Notably, this decrease in analysis, which provided a 68% increase in analytical throughput, did not affect the number of peptide identifications. Additionally, a reduction in the number of sample injections needed to cover a single PAcIFIC analysis decreased the total sample consumed during the experiment. (2) PAcIFIC was adapted to be compatible with the use of isobaric tags such as TMTs or iTRAQ. Release of quantitative information encoded in the tagged peptides when using an ion trap requires the use of pulsed Q-dissociation (PQD) fragmentation scan that forces the ion trap to scan out ions far below its preset cutoff. To accommodate these scans in the PAcIFIC method, a PQD scan was inserted after every CID scan so that each unique 2.5 *m/z* channel was consecutively analyzed with a CID scan to record qualitative peptide fragmentation and a PQD scan for quantitative information. Post-acquisition, this information was combined into a single searchable spectrum for every interrogated 2.5 *m/z* channel. This increased the quantitative capabilities of PAcIFIC, and also allowed for the simultaneous and comparative

analysis of several individual isobarically-labelled samples that were pooled prior to MS analysis. (3) Panchaud et al. devised an extension of the PAciFIC method that made it possible to expand the number of identifiable peptides by providing a means for deciphering multiplexed or chimeric tandem mass spectra. To accomplish this, precursor ions scans were recorded after every five DIA *m/z* channels. These survey scans were used to identify isotopic distributions during post-acquisition data processing. This information was processed much like many previously discuss DIA methods (e.g. shotgun CID, MS^E): software developed for PAcIFIC[98] creates multiple dta files for a multiplexed spectrum that replicates the MS/MS information of a given tandem mass spectrum differing only by assignment of the unique precursor monoisotopic *m/z* values. This allowed the authors to estimate the number of chimeric spectra in a PAcIFIC data set to be between ~ 8-16%, a number similar to that found by Scherl et al. for DDA-based shotgun proteomic data using similar HPLC gradient conditions.[99] Finally, the authors reported that PAcIFIC analysis of a yeast proteome, where protein copy number had been previously determined by non-MS methods, identified yeast proteins across the full proteome dynamic range. This is not surprising because the PAcIFIC method is conceptually similar to an SRM experiment: specifically, the ion trap is filled only with a precursor ion of interest (providing a gain in effective sensitivity) and only fragment ions from precursors residing in that narrow *m/z* window are recorded.

PAcIFIC analysis demonstrated great promise for biological applications due to the ability to identify a much larger dynamic range of proteins over DDA and other DIA methods.[91] While the cost in terms of instrument time and sample amount needed is increased, the promise of thoroughly cataloging a proteome and the ability to quantitatively monitor small dynamic changes in the vast complexity of cell culture and biological samples is of great interest for the

elucidation of potential biomarkers.[91, 95] PAcIFIC was thus used in the identification of plasma biomarkers for abdominal aortic aneurysms,[100, 101] urinary proteome analysis of irritable bowel syndrome patients,[102] and the identification of novel Tamoxifen agonists.[103] In the latter recent report by Hengel et al., comparative DDA analysis of identical samples did not uncover the same up-regulated proteins as PAcIFIC which provided higher individual protein sequence coverage than the DDA method and thus more accurate label-free quantitative data. Additionally, the added detectable dynamic range available from PAcIFIC has been used to characterize global changes in protein re-distribution, trafficking, expression and degradation in yeast peroxisomes grown on different carbon sources without the need to purify peroxisomes (Jung, S. et al, MCP, 2013 [Epub ahead of print])). In conjunction with this work, a more thorough investigation of orphan peptides (i.e. identified by tandem mass spectra from *m/z* channels where no precursor ion could be detected) originally reported by Panchaud et al. suggests that orphan peptides predominantly originate from low copy number yeast proteins (Jung, S. et al., in revision, JPR) again pointing a gain in detectable dynamic range.

XDIA

With the aforementioned successes of DIA techniques to provide more comprehensive proteome characterization, additional variations of the fundamental acquisition theme were developed. Carvalho et al. adapted DIA to middle-down proteomics;[92] a process of shotgun proteomics wherein the targeted proteolytic peptide size is much larger than that resulting from a trypsin digestion. As noted, the analysis of larger protein fragments leads to the possibility of increasing overall protein sequence coverage and therefore increasing confidence of correct peptide-to-protein assignment. The XDIA duty cycle is as follows: (1) a single survey scan is acquired; (2) followed by sequential 20 *m/z* DIA scans, spaced with 1 *m/z* overlap between

adjacent windows, covering the region of 400-1000 *m/z*. Each 20 *m/z* region of space is analyzed in two consecutive and distinct scan events; the first being an electron transfer dissociation (ETD) scan, and the second a scan that combined ETD and CID fragmentation for increased peptide ion fragmentation. The resulting data was process by deconvoluting the precursor ion scan and correlating, based on *m/z*, the monoisotopic mass of possible precursor peptide ions with the respective DIA ETD and ETD-CID fragment ion information. Carvalho et al. demonstrated for the XDIA analysis of a yeast cell lysate, that an increase in the number of peptide identifications as well as the number and accuracy of spectral counts per protein was achievable compared to standard DDA experiments. Additionally, the development of XDIA in compatibility with middle-down proteomics increases protein sequence coverage and therefore is an asset to monitoring post-translational modifications.

SWATH

Quite recently, AB Sciex and the Aebersold laboratory reported on the development of a new DIA method referred to as SWATH.[93] In SWATH, multiplexed tandem mass spectra (called MS^{ALL}) are collected over predefined precursor ion windows, and the resulting signals are interpreted using prior knowledge of the peptides present in the mixture and their experimental fragmentation spectra. Developed for use with a QqTof, SWATH performs data-independent fragmentation of all precursor ions entering the mass spectrometer by breaking down the *m/z* field into 25 *m/z* isolation windows, termed swaths. Analogous to the other methods discussed in this section, 32 consecutive 25 *m/z* swaths are used to cover the 400-1200 *m/z* region in the span of 3.2 seconds. This allows numerous repeat analyses of each window during the elution of a single chromatographic peak and results in a complete fragment ion map of the sample. The true novelty of this method lies in how the data is processed. The authors insisted that data analysis

capabilities for modern proteomics applications should cover two essential tasks: (1) identify as many proteins as possible in a given sample; (2) provide accurate quantitative measurements. To accomplish this, Gillet et al. used a peptide spectral library (PeptideAtlas, see www.peptideatlas.org) to parse the fragment ion spectra generated during DIA, and in accordance with retention time and relative intensities, to positively link ion fragmentation patterns to peptide identifications. Thus, the SWATH process is somewhat related to the accurate mass and time tag concept[104] that relies on a predefined library of identified peptides for making subsequent signal assignments. The data reduction technology of the SWATH method is also quite similar to that of an SRM experiment in that the specificity of peptide fragmentation is exploited to make positive peptide identification and provide accurate quantification. The major difference with SRM being that SWATH entirely circumvents the often complicated task of target peptide selection and transition optimization, since all precursors are fragmented during the experiment and all fragments detected. Also, all assays in a given swath are conducted in parallel instead of sequentially when using SRM. The use of SWATH for the analysis of *S. cerevisiae* demonstrated the capability to identify proteins over a detectable dynamic range of 4 orders of magnitude, which was still less than that achieved for an equivalent SRM experiment, but more than using a classical precursor extracted ion chromatogram method.

FT-ARM

Fourier transform-all reaction monitoring or FT-ARM is an independent DIA strategy that shares some common features with SWATH, although it was developed independently.[94] In particular, FT-ARM also relies of upon the selectivity of high mass accuracy data and the specificity of peptide fragmentation patterns to perform peptide identification and quantification. In the FT-ARM experimental sequence, all ions from a broad *m/z* range (typically 100 *m/z*) are

accumulated together and subject to simultaneous activation using various means (CID, IRMDP). Fragment ions are then analyzed on a high resolution/ ultra high mass accuracy instrument. It should be noted that the experiment does not incorporate precursor mass determination. The originality of the FT-ARM strategy lies in the matching of the multiplexed data with a database of theoretical or empirical peptide fragmentation patterns. These patterns are used to calculate a dot product score for each peptide against each DIA scan event across an LC separation, resulting in a score chromatogram for each peptide in the database. Thus, contrary to conventional database searches where the actual precursor masses are used as a filter for putative matches in the database, in FT-ARM, all tryptic peptides within the acquired *m/z* range are scored against all the recorded spectra. Interestingly, this method does not explicitly make use of precursor/fragments co-elution as an initial filter of putative lineage. As a matter of fact, the simultaneous presence of multiple fragments matching a particular peptide is a requisite for the detection of a peak in the score chromatogram for the corresponding sequence. Not unlike SWATH, FT-ARM relies on the specificity in peptide fragmentation and high mass measurement accuracy in a process that is essentially independent of precursor mass. In addition, it provides an effective means to wrest all the information from an acquired multiplexed dataset. Additionally, this search method is able to seamlessly identify numerous peptides from a single tandem mass spectrum that are inherently recorded when using such relatively large DIA isolation widths. Simulations indicate that sufficient mass accuracy and number of matching fragment ions permits low false discovery rates (FDR), even when considering large databases. Finally, the authors also demonstrated that the area under the peak of the score chromatogram, in addition to providing information on the quality of the match between the sequence and the observed spectra, scales linearly as a function of concentration of the corresponding peptide.

1.5 Quantitative proteomics techniques

Shotgun proteomics is not an inherently quantitative technique largely due to the diverse range of responses exhibited by peptides in the mass spectrometer. These responses are determined by the physicochemical properties of proteolytic peptides (amino acid composition, size, hydrophobicity, gas basicity, etc.); all of which, in combination, have an effect on the accuracy of quantification.[105] To moderate differences in peptide characteristics and to minimize quantitative deviation, the proteomics community has developed various techniques for absolute, semi-, and relative protein quantification. These techniques avoid a quantitative comparison across all peptide species contained in a sample, and instead allow for a direct comparison of one species across multiple samples or introduced standards. Each mass spectrometry-based protein quantification approach is defined by a unique sample preparation method and can be broken down into four distinct categories: metabolic incorporation, enzymatic and chemical alteration, introduced standards, and label-free.[50]

Metabolic incorporation of stable isotopes is a multiplexed quantitative technique that uses the mass difference between 'light' and 'heavy' stable isotopes to measure and differentiate the relative intensities of each uniquely labeled sample contributing to the total peptide abundance. The metabolic incorporation of such stable isotopes is accomplished by providing a growing organism with ^{12}C, ^{13}C, ^{14}N, or ^{15}N labeled amino acids so that the labeled amino acids are incorporated into protein composition.[106] A very common approach to metabolic incorporation of stable isotopes is termed stable isotope labeling by amino acids in cell culture (SILAC). Here, a 'heavy' labeled arginine and/or lysine is provided in the culture media of one sample and a 'light' arginine and/or lysine is provided in the second. Following treatment or exposure to the stimulus, protein samples are collected and combined for proteomics workup.

This one pot workup is an advantage of SILAC that eliminates the variability observed when samples are processed separately. Trypsin is commonly used to proteolytically digest protein samples. Therefore, SILAC methods typically employ the use of labeled arginine and lysine as it will result in all tryptic peptides containing at least one labeled amino acid, excluding the C-terminal peptide of each protein. Identically sequenced peptides, containing either a 'light' or 'heavy' arginine and/or lysine, will elute from the chromatography system at the same time due to the fact that the different isotopes of carbon and nitrogen will not change amino acid affinity with the chromatography material. The m/z shift due to incorporation of 'heavy' labeled amino acids is known and therefore the precursor ion information of identified peptides can be used to calculate the relative difference in abundance of uniquely labeled peptides. The information for each peptide is then computed to a difference in relative abundance of the entire protein to gain an understanding of protein expression and degradation that occurs due to the stimulus.[107]

Alternatively, the incorporation of traceable stable isotopes can occur following proteolytic digestion. These approaches involve chemical alteration of proteolytic peptides with isobaric tags designed to fragment into unique m/z reporter ions based on the distribution of stable isotopes within the tag. Common examples of this approach are: isotope coded affinity tags (ICAT);[108] isobaric tags for relative and absolute quantitation (iTRAQ);[97] and tandem mass tags (TMT).[96] In each of these techniques, multiple proteolytically digested samples are independently labeled with a unique isobaric tag and then pooled into one sample prior to MS analysis. A precursor ion represents a specific peptide sequence, composed of species from various sample origins labeled with unique isobaric tags. When fragmented, a distinct reporter ion specific to each isobaric tag will be released. The intensity of the reporter ions is then used to calculate the relative level of the specific peptide present in each sample. The advantage of this

technique over metabolic incorporation methods is the ability to multiplex up to eight samples, thereby drastically improving analysis throughput.

It is also possible to quantify peptides via mass spectrometry without biosynthetic or chemical manipulation using method such as absolution quantification (AQUA)[109] and multiple reaction monitoring (MRM).[110] Both of these methods represent a targeted quantification approach wherein user defined proteins are tracked, via selected proteolytic peptides, utilizing the introduction of 'heavy' labeled peptides of identical sequence. As 'heavy' labeled peptides will elute at the same time as their 'light' counterparts, the chromatographic peak of 'heavy' and 'light' labeled peptides can be compared to calculate an absolute value of quantification (i.e., mg).[111] While effective in targeted proteomics, this method is laborious to apply to global proteomics studies as each additional peptide or protein of interest requires a new standard and the necessary assay development.

Label-free quantitative proteomics methods are more suitable to global proteomics studies. These methods do not require the addition of biosynthetic or chemical manipulation of samples and rely on two demonstrated theories for measuring relative protein quantities. The first, termed spectral counting, is simply the summation of spectra contributing to the identification of a protein. Liu et al. demonstrated that the spectral counting method was valid was for a standard set of proteins and provided a linear dynamic range over 2 orders of magnitude.[112] When compared across samples acquired in a parallel fashion, spectral counting can define the magnitude of change in protein abundance. The second method for label-free quantification associates peptide identification information with the area under the curve of the precursor ion over time. Much like spectral counting, the area value can be used to compare relative abundances of peptides and proteins across sample analyses.[113]

Worked embodied in this dissertation will utilize the data-independent PAcIFIC method as the main technique of data-acquisition. As noted by Gillet et al. in their publication of the SWATH method,[93] the PAcIFIC method is capable of identifying proteins over the largest dynamic range (eight orders of magnitude) of other currently available DDA and DIA schemes.[91] Sample analysis completed in this dissertation is done with the focus of identifying proteins over the widest range of abundance as possible. With this in mind, I will utilize a quantitative approach that is fully compatible with the PAcIFIC method. The quantitative PAcIFIC approach is spectral counting. Spectral counting only counts the tandem mass spectra contributing to a protein identification and therefore no manipulation to the original PAcIFIC scheme is necessary.

1.6 Conclusions

Systems biology is a quickly growing field of science. The narrow focus of targeted studies, while informative, is giving way to the desire of researchers to be able to more quickly gain an understanding of environmental responses at the system level. The power of cataloging a cellular state (RNA, DNA or protein) holds the promise of defining the holistic role of cellular drug response. For proteomics, this is a daunting task. The human genome is composed of roughly 25,000 genes. By including RNA splicing and proteolysis events, these genes will translate to roughly 50,000 to 500,000 uniquely expressed proteins. Factor in post-translational modifications, and defining the complete protein complement of a human cell will require the analysis of millions of proteins.

Advances in faster, high resolution/high mass accuracy instrumentation and more thorough acquisition techniques are extending our capabilities to explore the proteome. While complete proteome coverage may never be feasible, we must place an emphasis on ourselves and our colleagues to use the technological advances have developed to gain a scientific edge in our research endeavors. Quicker adoption of modern instrumentation and data acquisition schemes is paramount to maximizing the potential data we can attain from every proteomics research study.

1.7 Notes to Chapter 1

Acknowledgements

Special thanks to Christophe Masselon at the CEA (Commissariat à l'énergie atomique et aux énergies alternatives) for his review of the discussion of data-independent mass spectrometry featured in this chapter. To be published as: Chapman et al. *Mass Spectrom. Rev.* **2013**.

Chapter 2

COUPLING CAPTIVE SPRAY IONIZATION WITH THE PACIFIC METHOD FOR IMPROVED RATES OF DATA-INDEPENDENT MASS SPECTROMETRY ACQUISITION

2.1 Introduction

The race to sequence the human genome captivated the scientific and medical community during the late 1990's.[114] The promise of a better understanding of human health and a figurative roadmap to improve medical therapies appeared to hinge on the knowledge locked inside the genes that encode our very being. In 2000, the first-draft of the nearly complete genome was published by the publically funded, Human Genome Project, and industry rival, Celara Genomics.[115-117] It was not until 2006 that the final pieces to the puzzle were completed when the annotated sequence of chromosome 1 was published; the largest chromosome of the human genome containing over 3,000 genes.[118] Over a decade after the first-draft, it is apparent that information beyond the genome is needed to truly understand dynamic cellular responses and causative factors of human illness. Naturally, this scientific pursuit extended to the products of gene expression – proteins.[119] Distinctive from the relatively static genome, proteomes – the entire protein complement of the cell – are living, breathing and dying networks that control the outcomes of cellular response.[120] The inherent intricacies of proteomes make their analysis far less efficient than genomics. Nevertheless, the knowledge of the proteins present in a given biological state is vital for our understanding of cellular functions at the molecular level. The means by which we generate detailed lists of proteins and their relative abundance levels with respect to environmental or patho-physiological conditions are therefore critical in solving challenging questions in biology and medicine.[51, 52]

Advancements in mass spectrometry-based proteomics and acquisition techniques are the critical driving forces that will lead to the ability to more completely characterize a proteome. 'Shotgun proteomics', the analysis of highly complex peptide mixtures obtained by proteolytic digestion of protein extracts from biological or clinical samples, is an efficient means by which we can categorize whole proteomes.[43] Techniques of 'shotgun proteomics' can be delineated by two unique approaches: data-dependent acquisition (DDA) and data-independent acquisition (DIA). DDA relies on precursor ion information from a survey scan of ions entering the mass spectrometer. Precursor ions are subsequently selected, typically based on descending abundance, for the fragmentation and tandem-mass spectrum acquisition necessary for peptide sequencing.[121] While DDA has been a cornerstone of proteomics, recent advancements in DIA show great promise for extending the dynamic range of peptide identifications that can be obtained during a single analysis.(Chapman et al. *Mass Spectrom Rev*, in press) DIA techniques differ from DDA in that ion selection is not based on information captured in concurrent survey ion scans and thus is truly independent from measured sample signal. Conceptual approaches of DIA have become quite diverse with techniques such as XDIA,[92] SWATH,[93] FT-ARM,[94] MS^E,[76] and PAcIFIC.[91] Here, I will focus on the approach developed by Panchaud et al., Precursor Acquisition Independent From Ion Count (PAcIFIC).[91, 98] As cited by Gillet et al.,[93] PAcIFIC is capable of recording the largest dynamic range of all the currently available DIA techniques.

PAcIFIC is a data-independent acquisition method that operates by successively interrogating predefined 2.5 m/z isolation widths, or channels, by tandem-MS across a user-defined m/z range (typically 1000 m/z). Batches of n continuous 2.5 m/z channels, overlapping by 1 m/z to allow for the selection of isotopic distributions, are analyzed per sample injection and chromatographic gradient. The n number of channels comprises the PAcIFIC duty cycle and

based on the scan rate of the instrument, n is adjusted to allow for repetitive sampling of each chromatographic peak. Each successive sample injection and chromatographic gradient will analyze the next n channels; and the PAcIFIC analysis continues until a total desired m/z range is complete.[91] Current optimized methods have been established for the LTQ Orbitrap and LTQ Velos (Thermo Scientific, San Jose, CA, USA): 27 sample injections each covering 25 2.5 m/z channels (2.5 days), and 44 sample injections each covering 15 2.5 m/z channels (4.2 days), respectively. As opposed to a precursor ion scan used in data-dependent methods, the use of the small 2.5 m/z isolation width greatly increases the measureable dynamic range of the total PAcIFIC analysis by negating the effect of highly abundant precursor ions outside of the narrow m/z window being interrogated. As a result, Panchaud et al. reported a detectable dynamic range of eight orders of magnitude at an FDR of 0.5%.[91] The data-independent acquisition method, PAcIFIC, has significantly increased the detectable dynamic range of complex proteomics samples over other data-dependent and data-independent acquisition schemes.[93] The PAcIFIC method has proven an invaluable tool for researchers looking to increase proteome coverage of complex biological matrices such as serum, tissue culture and urine samples.[100-103, 122] Additionally, it has been cited that the PAcIFIC method provides a quantitative advantage over other data-dependent acquisition techniques due to the repetitive sampling of each m/z channel during a single chromatographic peak.[98, 101] The current drawbacks of the PAcIFIC method are the time intensive nature of and the amount of sample consumed for each sample analysis.

Here, I aim to reduce the overall analysis time of the PAcIFIC method to 24 hours by coupling this approach with the captive spray ionization (CSI) source (Bruker, Billerica, MA, USA). The CSI source is a high-voltage non-tapered capillary emitter situated precisely in front of the mass spectrometer inlet to allow for the vacuum of the mass spectrometer to pull ambient

air into the source, directing a high percentage of sample ions into the mass spectrometer. The CSI source provides (1) axial gas flow that keeps the capillary emitter clean for prolonged spray stability; (2) gas flow directed at the spray tip to aid in desolvation and to nebulize the solvent into a Taylor cone; and (3) gas flow to create a vortex around the Taylor cone to funnel all of the sample into the mass spectrometer inlet.[123] In turn, it is possible to achieve a higher solvent flow rate with comparable sensitivity to nano-electrospray ionization.[124] The higher flow rate will allow for the efficient use of larger bore nano-columns, increased gradient control, as well as faster trapping, peptide elution, and column equilibration times. Initial work to optimize the CSI PAcIFIC method will be completed using Saccharomyces cerevisiae harvested during log-phase growth, a condition where comprehensive protein quantification using non-mass spectrometry techniques has been performed and can be applied to the mass spectrometry-derived protein identifications.[125] By working to decrease the overall sample analysis time, I hope to make the PAcIFIC acquisition method a more reasonable time investment for thoroughly cataloging proteomics samples and therefore expand its use in the scientific community.

2.2 Experimental procedures

Unless otherwise noted, all chemicals and solvents were purchased from Sigma Aldrich (St. Louis, MO, USA).

Saccharomyces cerevisiae growth

A single colony of strain BY4742 was grown in yeast extract peptone dextrose (YEPD) media overnight. The following day, the media was diluted and the yeast were grown to mid-log phase (OD600 = 0.56) at conditions matching those of Ghaemmaghami et al.[125] Saccharomyces cerevisiae were harvested by centrifugation at 5,000 x g for 10 minutes at 4 °C. The supernatant was removed and the cells were washed twice with ice cold PBS, each time quickly repelleting the cells at 5,000 x g for 2 mins. After the final wash, the PBS was removed and the resulting Saccharomyces cerevisiae cell pellets were stored at -80 °C until protein extraction and mass spectrometry sample preparation.

Protein extraction and quantification

Saccharomyces cerevisiae pellets were suspended in 1 mL of 100 mM ammonium bicarbonate by gently, but repetitively, vorexting and then lysed using probe sonicator. In short, suspended Saccharomyces cerevisiae were sonicated for roughly 5 seconds and then placed on ice to cool the suspension. This process was repeated twice more for each aliquot. Subsequently, lysed Saccharomyces cerevisiae suspensions were centrifuged at 5,000 rpm at 4 °C for 10 minutes to pellet cellular debris. The supernatant, containing cellular proteins, was removed and placed in a clean eppendorf on ice.

Each aliquot of Saccharomyces cerevisiae protein suspension was quantified using a BCA protein assay according to the manufacturer's instructions (Thermo Scientific/Pierce, Rockford, Il, USA). Sample protein concentrations were normalized by making aliquots of 200 μg/100 μL Saccharomyces cerevisiae protein with the dilution of a necessary amount of 100 mM ammonium bicarbonate.

Protein digestion and peptide preparation

Each 200 μg/100 μL Saccharomyces cerevisiae protein sample was denatured with the addition of urea to 6M. Subsequently, samples were buffered with the addition of 7 μL 1.5 M Tris pH 8.8, reduced with 2.5 μL of 200 mM tris(2-carboxyethyl)phosphine (TCEP) for 1 hour at 37 °C, alkylated with 20 μL of 200 mM iodoacetamide for 1 hour at room temperature in the dark, and then quenched with 20 uL of 200 mM dithiothreitol at room temperature. Prior to addition of sequencing-grade porcine trypsin (Promega, Madison, WI, USA) at a protein to enzyme ratio of 50:1, samples were diluted with 900 μL of 50 mM ammonium bicarbonate and 200 uL of MeOH. After an overnight incubation, peptides were desalted on a Vydac C18 macrospin column (The Nest Group, Southborough, MA, USA) according to the manufacturer's protocol. Resulting eluent was concentrated to near dryness on a SPD 111V SpeedVac (Thermo Savant, San Jose, CA, USA) and either stored at -80 °C for future use or reconstituted with a solution of 95% water/5% acetonitrile/0.1% formic acid peptide for immediate mass spectrometry analysis.

Nano-electrospray ionization mass spectrometry

Nano-HPLC was performed using a Waters NanoAquity (Milford, MA, USA). A homemade trapping column was made from 100 μm inner diameter (ID) capillary (Polymicro Technologies, Phoenix, AZ, USA) packed with 2 cm of 200 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). Successive analytical separation was performed on a homemade laser-pulled 75 μm ID (Polymicro Technologies, Phoenix, AZ, USA) column packed with 15 cm of 100 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). For each sample injection, approximately 1 μg, unless otherwise noted, of the peptide sample was loaded on the trapping column at 4 μL/min with 95% water/5% acetonitrile/0.1% formic acid for 5 minutes. Trapped peptides were then eluted from the trapping column onto the analytical column using a variable gradient with a flow rate of 0.25 μL/min. The gradient utilized two mobile phase solutions: A, water/0.1% formic acid; and B, acetonitrile/0.1% formic acid. The variable gradient used is as follows: 0 minutes, A (95%), B (5%); 55 minutes, A (65%), B (35%); 65 minutes, A (15%), B (85%); 75 minutes, A (15%), B (85%); 80 minutes, A (95%), B (5%); 80-100 minutes, A (95%), B (5%). Peptide digests were analyzed on both a LTQ Oribtrap and a LTQ Velos (Thermo Fisher, San Jose, CA, USA) by nano-electrospray ionization in positive ion mode. Ion source conditions were optimized using the tuning and calibration solution suggested by the instrument provider. For full range data-dependent acquisition, precursor ion scans were performed from 400-2000 m/z and data-dependent ion selection for MS/MS analysis operated within that m/z range. For gas-phase fractionation (GPF), precursor ion scans and data-dependent ion selection for MS/MS analysis were performed in the following four restricted m/z ranges: 400-520 m/z; 515-690 m/z; 685-970 m/z; or 965-2000 m/z. For tandem MS analysis, the isolation window was set to 2 Thompsons and a normalized collision

energy of 35% was used for all fragmentation events. Singly charged ions were excluded from analysis and dynamic exclusion was enabled with a repeat count of 1, a repeat duration of 30 seconds, an exclusion duration of 45 seconds and an exclusion list of 250. For PAcIFIC acquisition, the mass spectrometer parameters were set as defined by Panchaud et al.[91, 98]

Captive spray ionization mass spectrometry

Captive spray ionization was performed using a Waters NanoAquity (Milford, MA, USA). A manufactured fused silica 200 μm ID trapping column purchased from New Objective (Woburn, MA, USA) and packed with 2 cm of 200 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). Successive analytical separation was performed on a manufactured fused silica 200 μm ID trapping column purchased from New Objective (Woburn, MA, USA) and packed with 15 cm of 100 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). For each sample injection, approximately 1 μg, unless otherwise noted, of the peptide sample was loaded on the trapping column at 6 μL/min with 95% water/5% acetonitrile/0.1% formic acid for 2.5 minutes. Trapped peptides were then eluted from the trapping column onto the analytical column using a variable gradient with a flow rate of 2.5 μL/min, unless otherwise noted. The gradient utilized two mobile phase solutions: A, water/0.1% formic acid; and B, acetonitrile/0.1% formic acid. The variable gradient used for data-dependent acquisition methods is as follows: 0 minutes, A (95%), B (5%); 62 minutes, A (58%), B (42%); 65 minutes, A (15%), B (85%); 65.7 minutes, A (15%), B (85%); 65.75 minutes, A (95%), B (5%); 65.75-68 minutes, A (95%), B (5%). Peptide digests were analyzed on both a LTQ Oribtrap and a LTQ Velos (Thermo Fisher, San Jose, CA, USA) by captive spray ionization in positive ion mode. Ion source conditions were optimized using the tuning and calibration solution suggested by the

instrument provider and under conditions recommended by the CSI manufacturer (Bruker, Billerica, MA, USA). For full range data-dependent acquisition, precursor ion scans were performed from 400-2000 m/z and data-dependent ion selection for MS/MS analysis operated within that m/z range. For gas-phase fractionation (GPF), precursor ion scans and data-dependent ion selection for MS/MS analysis were performed in the following four restricted m/z ranges: 400-520 m/z; 515-690 m/z; 685-970 m/z; or 965-2000 m/z. For tandem MS analysis, the isolation window was set to 2 Thompsons and a normalized collision energy of 35% was used for all fragmentation events. Singly charged ions were excluded from analysis and dynamic exclusion was enabled with a repeat count of 1, a repeat duration of 30 seconds, an exclusion duration of 45 seconds and an exclusion list of 250.

Captive spray ionization PAcIFIC mass spectrometry

For CSI PAcIFIC acquisition, the mass spectrometry parameters employed are as defined by Panchaud et al.[91, 98] To optimize the number of injections per total analysis, the following conditions were tested on the LTQ Orbitrap. 44 sample injections using the variable gradient as follows: 0 minutes, A (95%), B (5%); 26 minutes, A (58%), B (42%); 29 minutes, A (15%), B (85%); 29.7 minutes, A (15%), B (85%); 29.75 minutes, A (95%), B (5%); 29.75-32 minutes, A (95%), B (5%). 33 sample injections using the variable gradient as follows: 0 minutes, A (95%), B (5%); 34 minutes, A (58%), B (42%); 37 minutes, A (15%), B (85%); 37.7 minutes, A (15%), B (85%); 37.75 minutes, A (95%), B (5%); 37.75-40 minutes, A (95%), B (5%). 22 sample injections using the variable gradient as follows: 0 minutes, A (95%), B (5%); 56 minutes, A (58%), B (42%); 59 minutes, A (15%), B (85%); 59.7 minutes, A (15%), B (85%); 59.75 minutes, A (95%), B (5%); 59.75-62 minutes, A (95%), B (5%). To optimize the number of

injections per total analysis, the following conditions were tested on the LTQ Velos. 44 sample injections using the variable gradient as follows: 0 minutes, A (95%), B (5%); 26 minutes, A (58%), B (42%); 29 minutes, A (15%), B (85%); 29.7 minutes, A (15%), B (85%); 29.75 minutes, A (95%), B (5%); 29.75-32 minutes, A (95%), B (5%). 27 sample injections using the variable gradient as follows: 0 minutes, A (95%), B (5%); 42 minutes, A (58%), B (42%); 46 minutes, A (15%), B (85%); 47.7 minutes, A (15%), B (85%); 47.75 minutes, A (95%), B (5%); 47.75-50 minutes, A (95%), B (5%). 19 sample injections using the variable gradient as follows: 0 minutes, A (95%), B (5%); 64 minutes, A (58%), B (42%); 68.5 minutes, A (15%), B (85%); 70.2 minutes, A (15%), B (85%); 70.25 minutes, A (95%), B (5%); 70.25-72.5 minutes, A (95%), B (5%). 15 sample injections using the variable gradient as follows: 0 minutes, A (95%), B (5%); 85 minutes, A (58%), B (42%); 89 minutes, A (15%), B (85%); 90.7 minutes, A (15%), B (85%); 90.75 minutes, A (95%), B (5%); 90.75-93 minutes, A (95%), B (5%).

Database search and protein quantification

Data acquired on the LTQ Orbitrap and LTQ Velos was converted from Thermo's RAW format to the universal mzXML format and searched against a database containing all known Saccharomyces cerevisiae proteins using SEQUEST.[70] For the data-dependent acquisition techniques (400-2000 m/z and 4-window GPF), search parameters for SEQUEST included a precursor ion tolerance of 2.1 Da, trypsin enzyme specificity, cysteines modified with iodoacetamide and the variable option for methionines in reduced or oxidized form. PAcIFIC search parameters for SEQUEST were modified to increase the precursor ion tolerance to 3.75 Da. Results were analyzed with Peptide Prophet[126, 127] ensuring that peptide hits with a probability of >0.99 were accepted and linked to protein entries. Protein identifications were

linked to the protein molecule per cell value as quantified by Ghaemmaghami et al.[125] in order to assess a quantification value for each identified protein.

Hardklör analysis

Isotopic distributions were identified using software developed in the MacCoss laboratory (University of Washington, Seattle, WA, USA) by Hoopmann et al.[128] All settings were used in default mode. Subsequent to Hardklör analysis, summation of possible precursor ions was completed on the assumption that potentially activated peptides (i.e. fragmentation and tandem-MS scan) were represented by isotopic distributions where $z \geq 2+$.

2.3 Results

Captive spray ionization gradient optimization

The directed ambient gas flow established by the CSI source allows for sensitivity comparable to nESI at increased solvent flow rates[129]. To capitalize on this feature of the CSI source, I in turn aim to use the higher solvent flow rate to increase sample throughput. Initially, I intend to establish a more rapid chromatographic gradient for use with the CSI source that will produce a similar quality of peptide and protein identifications in comparison to a data-dependent nESI method used in our laboratory.

A generic nESI chromatography column system used in our laboratory is a 100 μm inner diameter (ID) trapping column packed with 2 cm of Magic C18 packing material, linked through a four-way fitting to a 15 cm 75 μm ID analytical column situated in proximity to the mass spectrometer inlet. The trapping flow rate is 4 μL/minute and maintains a pressure of 1800 psi, and the analytical flow rate is 0.25 μL/minute and reaches a maximum pressure of approximately 900 psi. Following a trapping period of 5-minutes, peptides are eluted from the analytical column and enter the mass spectrometer using a 100-minute sample analysis that contains a 65-minute gradient optimized peptide elution (Figure 1-A). As indicated by the red triangles in Figure 1-B, peptides from 1 μg of a trypsin-digested yeast cell lysate elute from the column after 10-minutes and continue to elute till approximately 67-minutes into the total chromatographic gradient. Due to the low flow rate of nESI, it takes approximately 17-minutes for a complete change in solvent to fully equilibrate the HPLC and chromatography column system, from HPLC pump to analytical column tip. Therefore, following the 85% ACN wash step, an equilibration time of 20-minutes is allowed to ensure that before the start of the next sample injection the solvent in the

HPLC and chromatography system has returned to trapping conditions of 5% ACN/0.1% FA (Figure 1-A).

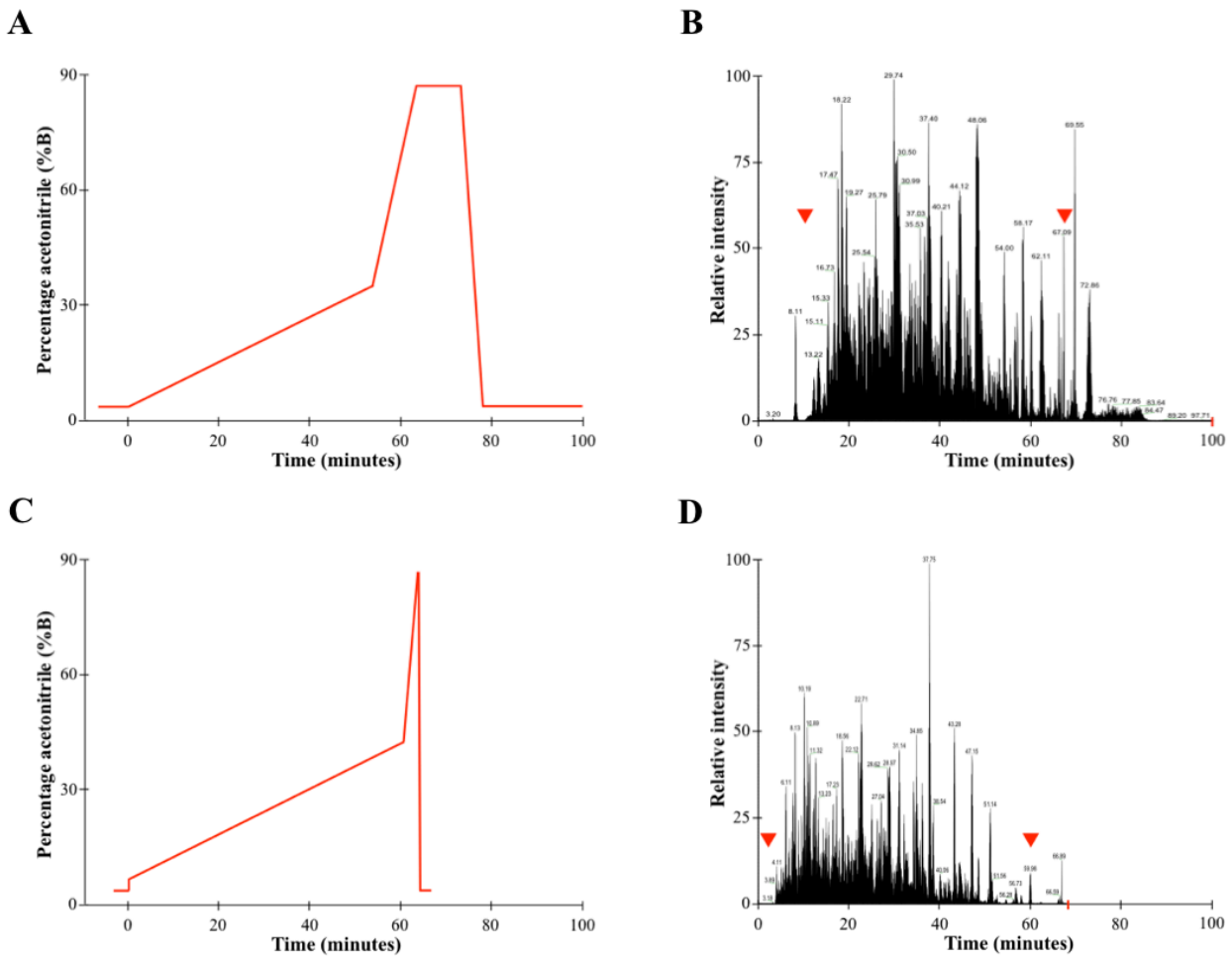


Figure 1. Chromatographic gradients used for **(A)** nESI and **(B)** CSI methods, with the time indicated before 0-minutes as the period allowed for solvent flow through the trapping column only. Total ion chromatograms from the analysis of 1 µg of trypsin-digested yeast cell lysate using the above chromatographic gradients for both **(C)** nESI and **(D)** CSI analysis. The end time of each analysis is denoted by a red x-axis tick mark. Additionally, the beginning and end time of peptide elution is denoted by red triangles.

Working from the above nESI gradient conditions, I created a CSI-compatible chromatographic gradient to increase the throughput of sample analysis while maintaining a similar signal sensitivity and peptide identification profile to our nESI results. To modulate the increase of pressure associated with an increase in solvent flow rate, I elected to use 200 μm ID trapping and analytical columns, both packed with the same length of Magic C18 packing material described in our nESI system. I tested a spectrum of solvent flow rates and observed that at 2.5 μL/minute I reached a maximum level of sensitivity based on recorded intensities of a standard. In addition, the flow rate of 2.5 μL/minute greatly increases the possibility for improved sample throughput. Using the same gradient slope as indicated in the nESI anslysis, it is observed that with a 1 pmol sample injection of angiotensin II the CSI signal intensity is 0.1 to 0.5 orders of magnitude lower than recorded with nESI (data not shown). The increase in capillary column ID for the CSI chromatography system increases the total volume within the trapping and analytical columns 5.42 times that of the described nESI system. But, the ten-fold increase in solvent flow rate used with the CSI source allows for an equilibration of the entire HPLC and chromatography column system, from pump to analytical column tip, in roughly 2.5-minutes. Using the higher solvent flow rate of 2.5 μL/minute, it is possible to drastically shorten both the lapse time for peptides to elute from the analytical column at the beginning of the chromatographic gradient and the equilibration period at the end of each sample analysis. The result is a 68-minute chromatographic gradient with a similar gradient slope and time to the nESI sample analysis - a reduction of 32-minutes (Figure 1-C). As shown in Figure 1-D, using the CSI chromatographic gradient, peptides start to elute at approximately 3-minutes and continue to elute till 60-minutes into the sample analysis. Additionally, the larger ID trapping column allows for an increase in the solvent flow rate during the trapping period to 6 μL/minute; effectively

reducing the needed trapping period to 2.5-minutes while maintaining a similar back pressure to that of the nESI system.

A primary goal of optimizing the chromatographic gradient for use with the CSI source is to maximize throughput by improving the percentage of time the mass spectrometer spends identifying peptides during a sample analysis. As illustrated in Figure 1, the total ion current (TIC) is depicted for a sample injection of 1 μg trypsin-digested yeast cell lysate for both nESI (Figure 1-B) and CSI (Figure 1-D). While all steps of the gradient are essential, for this comparison I equate the amount of time that peptides elute from the analytical column as the MS time utilization for the sample analysis. MS time utilization is the period of time during the sample analysis where the mass spectrometer can be productive in attaining spectra for positive peptide identification. As a baseline, the 100-minute gradient of the nESI analysis has a percent time utilization of 57.1%. Using the established CSI conditions, I was able to reduce the time it takes for peptides to elute from the analytical column and minimize the equilibration period at the end of the sample run, shortening the total gradient to 68-minutes. For the standard data-dependent acquisition method, 2702 peptides were identified using the nESI configuration and 2620 peptides were identified using the CSI setup. This calculates to an improved percent time utilization of 82.4%. This in turn can be calculated to an increase in the rate of peptide identifications; 38.5 peptides per minute using the CSI method established here, compared to 27.0 peptides per minute with the nESI method. Additionally, the same analysis can be done on the data collected from the gas-phase fractionation (GPF) analyses. The total gradient time for the nESI GPF method is 400-minutes and results in 4459 peptide identifications. The total gradient time for the CSI GPF method is 272-minutes and results in 4408 peptides. In recognition of multi-injection sample analyses (i.e. GPF and PAcIFIC) it is worthy to note that

the time saved per sample injection and chromatography gradient is additive and the CSI GPF is more than a 25% reduction in analysis time.

Furthermore, I used the assigned copy number information available for yeast proteins[125] to compare the protein molecules per cell value of identified proteins from each ionization source (Figure 2). This was done for both a standard data-dependent (400-2000 m/z) and a four-window (GPF) acquisition technique.[95] As mentioned, peptide and protein quantities are very close in quantity between the nESI and CSI methods. Additionally, as illustrated in Figure 2, the frequency distribution of peptide identifications for nESI and CSI follow a similar pattern. Although, it should be noted that the median value of the assigned protein molecule per cell, denoted by the black triangle (▼), is lower in both of the CSI analyzes.

Coupling CSI to the PAcIFIC acquisition technique

By virtue of altering the chromatographic gradient, it is necessary to review and optimize the PAcIFIC acquisition method for use with CSI. As designed by Panchaud et al., to maximize peptide identifications and benefit from improved quantitative measurements, the PAcIFIC duty cycle (i.e., interrogating each designated PAcIFIC m/z channel for a sample injection) must occur multiple times during a single chromatographic peak. Changing the gradient slope and solvent flow rate will affect the rate at which a chromatographic peak will elute from the analytical column into the mass spectrometer.[130] Therefore, I must optimize the acquisition technique to mirror the increase or decrease of peptide elution time based on changes in the CSI chromatographic gradient. To analyze the m/z region of 400-1400 m/z, there are a total of 667, 2.5 m/z PAcIFIC channels to interrogate. Based on this fixed value of m/z channels and the fact that I want to complete the total sample analysis in 24-hours, a relationship develops between the

number of m/z channels that are interrogated per sample injection and the allowable chromatographic gradient time. For example, the more m/z channels interrogated in a single PAcIFIC duty cycle reduces the total number of sample injections necessary to analyze all 667 PAcIFIC channels. In turn, based on a fixed total analysis time of 24-hours, a reduction in number of sample injections needed permits each individual chromatographic gradient to be longer.

Here, using this relationship I intended to optimize a CSI PAcIFIC method based on the total number of injections needed per PAcIFIC duty cycle. Empirically, I estimated the time of the PAcIFIC duty cycle on the LTQ Orbitrap when interrogating 15, 20 and 30 channels to be 3.9-5-seconds, 5.3-6.7-seconds, and 7.9-10.0-seconds, respectively (data not shown). Additionally, I selected these conditions to span a spectrum of PAcIFIC duty cycle lengths to determine the maximum number of peptide and protein identifications that can be made while ensuring that each chromatographic peak is interrogated repetitively for qualitative and quantitative purposes. As a baseline reference, I analyzed a trypsin-digest yeast cell lysate sample with the previously optimized nESI LTQ Orbitrap PAcIFIC method.[91] As seen in Figure 2, the 4.2-day nESI PAcIFIC method identified 13,468 peptides contributing to 2,120 proteins. As depicted by the frequency distribution of assigned protein molecules per cell, the median copy number value of the identified proteins is 5,040. The CSI PAcIFIC analysis of the same yeast cell lysate sample was performed with the following three methods: 1) 44 sample injections each interrogating 15 2.5 m/z channels with a single sample analysis time of 32.7-minutes; 2) 33 sample injections each interrogating 20 2.5 m/z channels with a single sample analysis time of 43.6 minutes; and 3) 22 sample injections each interrogating 30 2.5 m/z channels with a single sample analysis time of 65.5-minutes. Based on results illustrated in Figure 2, in comparison

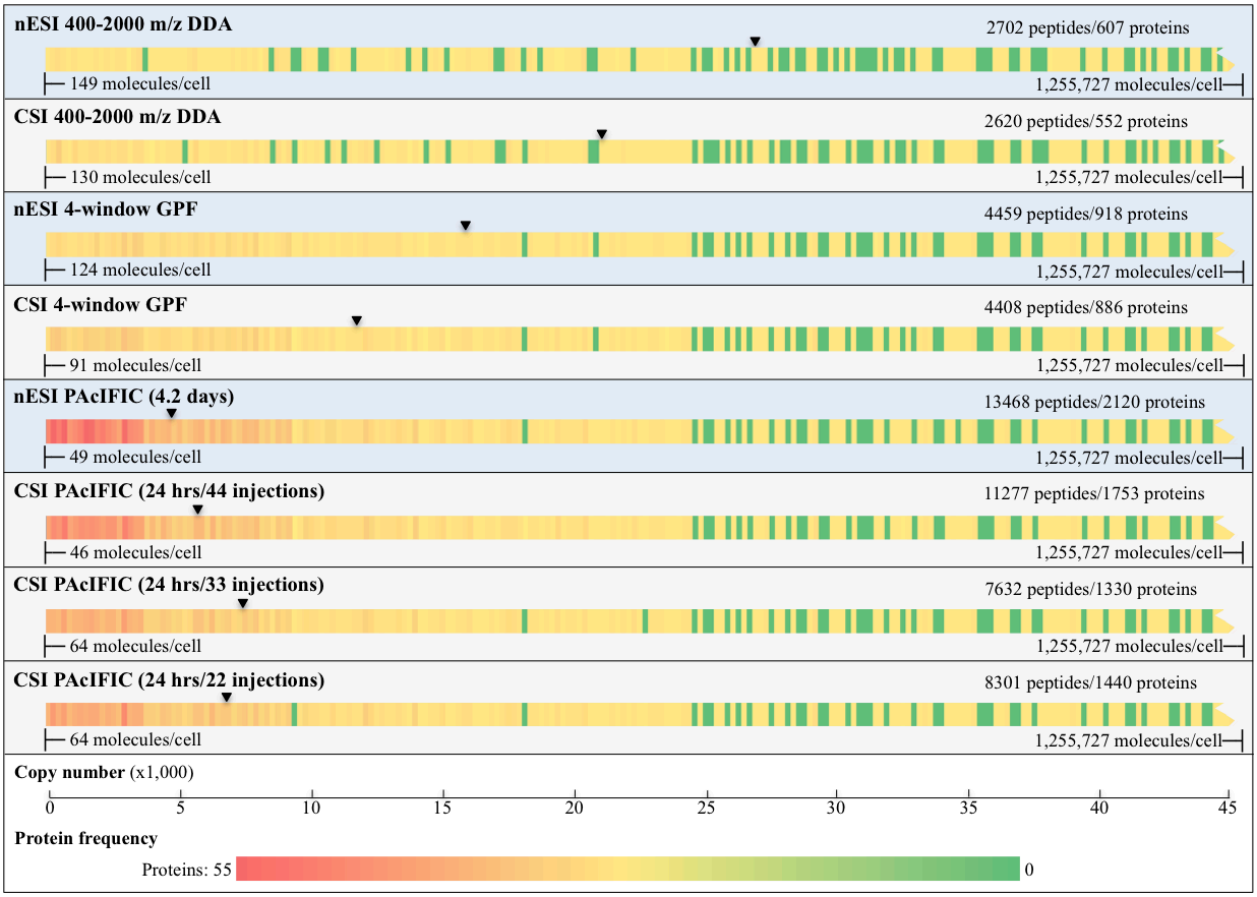


Figure 2. Frequency distribution based on protein molecules per cell, as assigned from values published in Ghaemmaghami et al.,[125] of trypsin-digested yeast cell lysate mass spectrometry derived protein identifications from a variety of ionization source and mass spectrometer acquisition methods. Each pixel represents a spread of 200 protein molecules per cell, starting from zero. Listed for each technique is the: total number of peptide and protein identifications, a minimum and maximum value assigned from the protein molecules per cell, and the median value for protein molecules per cell assignment, as denoted by the black triangle (▼).

between the CSI PAcIFIC methods analyzed, it can be observed that the conditions associated with 44 sample injections allows for a maximum number of peptide and protein identifications (11277 peptides, 1753 proteins), lowest assigned copy number median value (5941 protein molecules per cell), and the shortest PAcIFIC duty cycle that will ensure repetitive chromatographic peak sampling.

Increased sample loading and application to CSI PAcIFIC

To modulate the increase in pressure associated with higher solvent flow rates in our CSI system, the bore size of the trapping column was increased from 100 μm to 200 μm. Doubling the ID of the capillary quadruples the volume of Magic C18 material that can be packed in a fixed 2 cm trapping column length. This, in turn, increases the binding capacity of the trapping column; and I consequently recognized the potential of loading more sample per injection to effectively increase the amount of each peptide species available for analysis. First, I compared the peptide and protein identifications from a series of analyses completed with the 44 injection CSI PAcIFIC method to assess the effect of increasing the per sample injection amount from 1 μg to 2 μg. With no effect on the overall sample analysis time, this technique resulted in 7.2% more peptide identifications, and a 7.6% boost in protein identifications. Notably, the mean copy number value for the 2 μg sample load analysis decreased by 408.9 protein molecules per cell. By identifying proteins unique to the 2 μg analysis, it was found that these identifications originate from lower copy number per cell proteins with a mean and median value of 4062.6 and 2537.2 protein molecules per cell, respectively.

Next, I investigated the effect of further increasing the sample amount from 2 μg to 3 μg per sample injection. Here, it was observed that the number of peptide and protein identifications

decreased by roughly 8%. I attribute this effect to the possibility that the trapping column is no longer binding the peptide sample discretely and the possibility that I am producing an increased number of chimeric spectra[131] that are not properly handled with our current proteomics analysis pipeline.

Comparatively, I tested the effect of increased sample loading on nESI data-dependent acquisition (400-2000 m/z) using the 200 μm trapping column. Using Hardklor, available precursor ions were counted based on identified $z \geq 2+$ isotopic distributions. While a marked increase in precursor ions occurs with greater sample loading this did not result in a greater number of peptide or protein identifications (data not shown). Presumably, with our relatively short DDA nESI method, I do not spend enough time selecting available precursor ions for tandem-MS to benefit from the increase in the number of precursor ion availability. To do so, I would need to increase the chromatographic gradient time and/or use a faster scanning instrument. This highlights a characteristic difference in the design of data-dependent acquisition versus that of the data-independent technique, PAcIFIC. With PAcIFIC, a fixed amount of time is allotted to interrogate every m/z channel over a designated m/z range. Therefore, as seen from the data, it is possible to see great benefit by increasing the presence of peptide species by increased sample loading.

12-hour CSI PAcIFIC analysis

To further decrease the time per total sample analysis, I decided to reduce the overall m/z range analyzed. Data collected using CSI PAcIFIC with 44 sample injections was reviewed to select the m/z range with the highest rate of peptide identification. I discovered that a bulk of peptide identifications originate from a limited m/z range and elected to focus our efforts on

analyzing the 285 2.5 m/z channels located between 670 and 1096 m/z. With the trapping period and pauses in sample analysis (i.e., waiting for contact closure signals from the liquid chromatographer) this shortened analysis can be done in less than twelve hours. To validate this approach the original data set was reanalyzed to only include the above mentioned 19 sample injections. With a reduction in more than 50% of the time required to complete the PAcIFIC analysis, I still identify 65.3% of the peptide identifications and a staggering 90.5% of the number of protein identifications. Additionally, I observed no effect on the dynamic range and only a modest increase in the median value of protein molecules per cell (a rise from 5491.2 to 6210 protein molecules per cell).

CSI PAcIFIC optimization on the LTQ Velos

The PAcIFIC method has also been optimized for use with the faster scanning LTQ Velos (Thermo Scientific) with a nESI source using 27 sample injections each covering 25 2.5 m/z channels.[98] The advancements of the LTQ Velos include a front end S-Lens that helps to funnel a greater number of ions that enter the mass spectrometer into a dual-pressure ion trap that aids in both sensitivity and scan rate speed.[53] I have previously acknowledge the dependency of PAcIFIC method parameters on the scan rate of the instrument.[98] To couple with CSI with the PAcIFIC method on the LTQ Velos, it was necessary to evaluate and optimize the acquisition and chromatography parameters. Preliminary work coupling CSI to the LTQ Velos was done using trypsinized cell lysates from HepG2 cells. Using the chromatography gradients from the LTQ Orbitrap development, I first compared nESI and CSI protein identifications using a standard data-dependent acquisition technique (400-2000 m/z). As seen in Table 1, protein and peptide identification numbers are lower in the CSI analysis (4647 peptides identified with nESI

versus 3969 peptides identified with CSI). This is in agreement with observations from the work done on the LTQ Orbitrap, although counter to those experiments the average and median copy number of proteins identified increase on the LTQ Velos with the CSI. To establish an optimized PAcIFIC method with the CSI source on the LTQ Velos, I tested 15, 19, 27, and 44 injection PAcIFIC analyses to span a range of duty cycle time lengths. Previous work by Panchaud et al. established an optimized setting for nESI PAcIFIC on the LTQ Velos of 27 sample injections covering 2.5 days of total analysis time. Here, reviewing the data presented in Table 1, I observed a similarity of protein and peptide identifications qualities between 15, 19 and 27 sample injection PAcIFIC analyses. The most proteins were identified at 15 sample injections (1529 proteins). The largest number of peptides was identified at 19 sample injections (10054 peptides). In addition, the 19 sample injection identified the protein with the lowest assigned copy number as defined by Ghaemmaghami et al.[125] and shared the lowest median copy number value with the 15 sample injection PAcIFIC analysis (6301.7 protein molecules per cell).

Table 1.

	# Protein IDs	# Peptide IDs	Max Copy Number	Min Copy Number	Average Copy Number	Median Copy Number
nESI DDA(400-2000 m/z)	911	4647	1255722.3	105.5	40717.2	11662.8
CSI DDA (400-2000 m/z)	774	3969	1255722.3	105.5	45298.2	12937.9
nESI PAcIFIC (2.5 days)	1844	11284	1589836.9	64.3	23727.0	4948.2
CSI PAcIFIC (24 hrs/15 injections)	1529	9888	1255722.3	64.3	27695.4	6301.7
CSI PAcIFIC (24 hrs/19 injections)	1525	10054	1255722.3	49.2	27814.4	6301.7
CSI PAcIFIC (24 hrs/27 injections)	1523	9794	1255722.3	64.3	27458.3	6465.4
CSI PAcIFIC (24 hrs/44 injections)	1304	8678	1255722.3	98.9	31239.6	6278.4

2.4 Discussion

Recent advances in data-independent acquisition methods, especially PAcIFIC, have demonstrated encouraging advancements in the ability to expand the detectable dynamic range of proteomics samples.[93] Hengel et al. illustrated the effective application of the PAcIFIC method in comparison to a data-dependent 4-window GPF mass spectrometry technique. The study determined that the pro-metastatic protein anterior gradient 2 (AGR2) was significantly up regulated following tamoxifen treatment and increased expression of AGR2 was implicated in tamoxifen drug resistance. This protein was not identified in the data-dependent GPF technique or a transcriptomic screen, highlighting the necessity of being able to more thoroughly catalog biological and clinical samples.[103] Panchaud et al. calculated that the PAcIFIC method could identify peptides over a precursor ion dynamic range of eight orders of magnitude.[91] But, this comes at the cost of intensive time investment as a single PAcIFIC analysis takes 4.2 days on the LTQ Orbitrap and 2.5 days on the LTQ Velos. Here, I demonstrated that by coupling the PAcIFIC method with the CSI source, I was able to reduce the total analysis time to 24 hours without deleteriously affecting the quality of peptide and protein identifications. This is a reduction of 76.8 and 36 hours for the LTQ Orbitrap and LTQ Velos, respectively. In turn, I hope this method optimization makes PAcIFIC a more reasonable time investment for proteomics analysis of biological and clinical samples.

The fundamental acquisition features of the PAcIFIC technique rely on instrument scanning speed. As designed by Panchaud et al., the PAcIFIC duty cycle (i.e. all 2.5 m/z channels predetermined for a sample injection and chromatographic gradient) must repetitively interrogate each chromatographic peak for qualitative and quantitative purposes.[91] Repetitive precursor ion interrogation is a key advantage of the PAcIFIC method when compared to

standard DDA techniques. To increase peptide identifications in DDA, dynamic exclusion is employed to eliminate oversampling of the most abundant precursor ions. But, this technique also inhibits the probability of sampling a specific precursor ion when it reaches the apex of its respective chromatographic peak thereby sacrificing the qualitative and quantitative properties of a dataset. Qualitatively, improved signal to noise ratios (S/N) of fragment ions are observed when precursor ions are interrogated at the apex of the chromatographic peak leading to increased confidence of peptide identifications. Quantitatively, the repetitive sampling of each precursor ion correlates to a more accurate assessment of peptide abundance when using label-free quantitative methods such as spectral counting. Designing the PAcIFIC duty cycle to be short enough to capture multiple tandem-MS spectra across each chromatographic peak will ensure that some of those spectra are captured closer to the chromatographic peak apex than is typically achieved with DDA techniques.

As demonstrated here, by fixing the overall sample analysis time (i.e. 24 hours) it is necessary to optimize the number of m/z channels that can be analyzed per PAcIFIC duty cycle in relation to the corresponding rate at which a peptide will elute from the chromatographic system into the mass spectrometer. As illustrated in the comparison of CSI PAcIFIC analyses (Figure 1), the results indicate that the method with 44 sample injections allows for the most peptide and protein identifications as well as the lowest median protein molecule per cell value. As programed, the 44 injection CSI PAcIFIC method interrogates 15 2.5 m/z channels and has the shortest duty cycle (approximately 4 seconds) of the all approaches studied. Empirically, I estimated that many peptides elute from the chromatographic gradient created for the 44 injection CSI PAcIFIC method for 9-13 seconds. Therefore, this approach allows for the necessary repetitive interrogation of each m/z channel during the elution of a single peptide

species to generate quality data. It is our recommendation that use of the CSI PAcIFIC method with a LTQ Orbitrap follow the established protocol for 44 sample injections.

In this study, I also investigated the effects of increasing the sample load per injection on the number of identified peptides and proteins. The capability of the CSI source to generate and direct a larger percentage of ions into the mass spectrometer from a high flow rate nano-chromatography system allows the use of larger inner diameter capillaries for the trapping and analytical columns. This in turn increased the volume of C18 packing material within a fixed length trapping and analytical column and results in the ability to bind a greater quantity of peptide sample. I observed that increasing the peptide sample per injection from 1 μg to 2 μg resulted in 7.2% more peptide identifications, and a 7.6% boost in protein identifications. It was noted that the newly identified proteins originate from lower copy number per cell proteins with a mean and median value of 4062.6 and 2537.2 protein molecules per cell, respectively. Therefore, I postulate that by increasing the sample amount per injection I can effectively increase the presence of each peptide species allowing for low abundance peptides to overcome a threshold of precursor ion abundance necessary for identification. This highlights an additional fundamental advantage of the PAcIFIC method over other DDA techniques. Since a fixed number of m/z channels are interrogated evenly over the entire m/z region in each PAcIFIC analysis, it is possible to see great benefit in increasing the presence of peptide species by increased sample loading. But, this is a correlation that consists of a spectrum of both beneficial and adverse effects. On the other hand, by increasing the number of precursor ions of adequate ion abundance for identification I increase the probability that chimeric spectra, a single spectrum containing the fragment ions of more than one precursor ion species, will be generated. This is an especially important consideration for the short chromatographic gradient used in the

44 injection CSI PAcIFIC method developed here. The proteomics data analysis pipeline I used does not allow for deconvolution of chimeric spectra and therefore chimeric spectra can have a deleterious affect on SEQUEST scoring and the increased identification of false positives. This effect was noted when the sample amount per injection was increased from 2 μg to 3 μg and the corresponding number of peptide and protein identifications was reduced by 8%.

I also investigated further reducing the analysis time of the PAcIFIC method by reducing the total number of m/z channels to be interrogated. A PAcIFIC analysis of 400 – 400 m/z has 667 2.5 m/z channels. To reduce the overall analysis time to less than 12 hours, I designed the method to interrogate 285 2.5 m/z channels over the span of 670 – 1096 m/z. This m/z region was selected for the high density of peptide identifications. By reducing the overall m/z region interrogated, I identified roughly 65% of the peptide identifications. I anticipated a substantial decrease in number of peptide identification by the mere justification that I am producing fewer tandem-MS spectra for SEQUEST interrogation. But, unexpectedly, I maintained 90.5% of the number of protein identifications. The quality of remaining peptide identifications did not differ from those of the 24-hour PAcIFIC method, as the interrogation process for the peptides identified remains unchanged. Therefore, I find this approach to be a reasonable means for further reducing the PAcIFIC analysis time if peptide sequence coverage is not critical component of the analysis (i.e. searching for covalent peptide modifications).

A significant portion of the CSI analysis was to determine the practical application of the CSI source for the PAcIFIC method on both the LTQ Orbitrap and the LTQ Velos. Nano-spray ESI is a highly used technique with proven spray stability for long-term analysis. In this regard, I found that the CSI source operated exceedingly well with the LTQ Orbitrap. Peptide standards were used to evaluate the condition of the chromatography system and CSI spray stability during

the course of multiple PAcIFIC analyses. As done with our nESI experiments, the chromatography system and CSI source performed well and could be used for several PAcIFIC analyses before requiring replacement of the trapping and analytical column. CSI use with the LTQ Velos was more problematic. I found that the high flow rate and increased sample loading necessitated frequent cleaning of the S-lens. A 'dirty' S-lens leads to decreased ion transmission to the ion trap, effectively reducing the observed signal of our standard peptide. With this experimental information, I decided not to pursue further optimization of the CSI PAcIFIC method on the LTQ Velos. Additionally, I favor use of the CSI on the LTQ Orbitrap due to the ability to capture high mass accuracy and resolution precursor ion scans. This will allow for future coupling of AUC quantitative approaches and will also allow users to track the presence of precursor ions.

2.5 Conclusions and future directions

As proteomics mass spectrometrists, we are often asked how to *best* analyze samples for global proteome characterization. This question can start to be answered by understanding the 'economies of shotgun proteomics' or the cost advantages of increased sample analysis time. In mass spectrometry, time is money. Therefore, understanding a researcher's budget constraints is critical in presenting options for how to *best* analyze a sample. Presented here is an advancement of the PAcIFIC method that drastically reduces the overall analysis time without deleteriously reducing the number of peptide and protein identifications. Originally presented by Panchaud et al. as a 100.8 hour technique (4.2 days),[91] I have significantly reduced the PAcIFIC analysis time to 24 hours without significantly reducing the number of peptide and protein identifications. Additionally, I demonstrate further routes to reduce the analysis time to less than 12 hours by constraining the total m/z region to be analyzed. These approaches reduce the overall time and cost investment required, while achieving a similar proteome analysis. As we continue to advance the field of mass spectrometry, it is important that we keep the concept of 'economies of shotgun proteomics' in mind as a guideline for creating practical analysis techniques that can be quickly adopted by the scientific community. Additionally, as instruments capable of increased scanning speeds continue to become available, it will be necessary to continue to reevaluate DIA techniques that have already been developed.

2.6 Notes to Chapter 2

Acknowledgements:

Special thanks to the Aitchinson's laboratory at the Institute for Systems Biology (ISB, Seattle, WA), especially Jennifer Smith and Song Li, for preparation and donation of yeast cell pellets.

Special thanks to Robert Moulder at the University of Turku/Åbo Akademi Centre for Biotechnology for invaluable advice on selecting PAcIFIC channels most likely to result in peptide identifications.

Special thanks to the Isoherranen laboratory (Department of Pharmaceutics, University of Washington, Seattle, WA, USA) for preparation and donations of HepG2 cell pellets.

Chapter 3

DATA-INDEPENDENT MASS SPECTROMETRY TECHNIQUES FOR THE IDENTIFICATION OF PUTATIVE ALL-TRANS RETINOIC ACID-MEDIATED SECONDARY PATHWAYS AND ALTERATIONS IN PROTEIN EXPRESSION THAT DICTATE THE FATE OF A CELL

3.1 Introduction

All-*trans* retinoic acid is a highly active metabolite of the vitamin A family required by numerous cell types for a diverse array of biological processes including differentiation of epithelial tissues, embryonic development, growth, immune response, reproduction, fatty acid metabolism and vision.[3] Sought for the potential to promote differentiation, cell cycle arrest, apoptosis, and metabolism; atRA is used at pharmaceutically active doses in the therapy of cancers,[3, 26] treatment of skin issues,[22] and is being investigated as an anti-obesity medication.[132] It is well established that atRA mediates gene expression and the resulting inhibition of cellular growth through interactions with ligand-activated transcription factors. The most prominently affected transcription factors are the family of retinoic acid receptors: -alpha,[28] –beta,[29] and -gamma[30] (RAR-α, RAR-β, RAR-γ). RAR forms a heterodimer with retinoid X receptor (RXR)[31] when the two transcription factors are bound to specific sequences of DNA defined as retinoic acid receptor elements (RARE).[32] RAREs are characterized by direct repeats (DR) of two hexanucleotides with a defined base pair gap between the protein binding of RAR and RXR. The RAR/RXR heterodimer will bind to sequences with 2 (DR2) and 5 (DR5) base pair gaps. When atRA binds to RAR, the protein undergoes conformational changes that weaken protein-protein interactions and allow for dissociation of transcription complex corepressors. Subsequently, coactivators can bind to the atRA activated RAR/RXR heterodimer and activate downstream gene expression. Transportation of atRA to the nucleus via binding proteins is vital in the process

of atRA-mediated gene expression. For example, cellular retinoic acid-binding protein 2 (CRABP-II), plays a role in the process of atRA-mediated gene expression and a subsequent decrease in proliferation as CRABP-II delivers atRA directly to RAR.[12] At chemotherapeutically active doses, atRA will activate a series of protein pathways involved in apoptosis through de novo protein expression. For example, in the breast carcinoma cell line, MCF-7, atRA has been shown to induce expression of intrinsic apoptosis pathway factors such as initiator caspase-9 and executioner caspase-7. These proteins subsequently activate secondary protein pathways resulting in impaired mitochondrial function and release of pro-apoptotic mitochondrial factors including cytochrome C.[33]

Paradoxically, it has been observed that atRA can also promote cell growth and proliferation in a tissue-specific manner, as documented in keratinocytes.[34] In recent years, the peroxisome proliferator-activated receptor beta/delta (PPAR-β/δ) has been shown to interact with and be activated by atRA; thereby providing an alternate pathway by which the pleotropic effects of atRA can be more fully defined. Much like the RAR/RXR heterodimer, PPAR-β/δ forms a heterodimer with RXR. Retinoic acid binds to PPAR-β/δ to promote the expression of a subset of genes that differs from RAR-mediated gene expression.[35] Whereas the RAR/RXR heterodimer binds to DR2 and DR5 RARE sequences, PPAR-β/δ promotes the expression of a unique proteome by preferential interaction with a pair of hexanucleotides separated by a 1 (DR1) base pair gap.[36] Additionally, delivery of atRA, and thus activation of the PPAR-β/δ/RXR heterodimer, is accomplished by a unique transport protein, fatty acid-binding protein 5 (FABP5). This alternative route of atRA-mediated gene expression activates a multitude of compounding anti-apoptotic and pro-survival signal pathways.[13] Categorically, PPAR-β/δ-mediated gene expression leads to changes in fatty acid transportation, fatty acid oxidation, adipocyte

differentiation, adaptive thermogenesis, ubiquitination, and cell survival.[37-39] More specifically, atRA-activated PPAR-β/δ/RXR increases transcription of a family of phosphatidylinositol 3-kinases (PI3K), which subsequently activate the serine/threonine-specific protein kinase, Protein Kinase B (PKB or Akt). In turn, activated Akt: 1) impedes apoptotic pathways by inhibiting proteins such as Bcl-2-associated death promoter (BAD); 2) activates cell proliferation mechanisms including mammalian target of rapamycin (mTOR); and 3) induces the means for increased protein synthesis necessary for cell signaling and growth.[40, 41]

The diverse range of cellular response attributed to atRA has long been of concern to the research and clinical community. Increasing interest in using atRA as a chemotherapeutic for its potential to promote differentiation, cell cycle arrest and apoptosis requires that we also understand the diverging consequence of atRA as an activator of cell proliferation[42]. Recently, Schug et al. demonstrated that the balance of CRABP-II and FABP5 are vital to the fate of the cell as they control transport of atRA to their respective nuclear transcription factors. Basal levels of CRABP-II and FABP5 are critical in partitioning atRA between RAR and PPAR-β/δ and activating the respective pro- and anti-apoptotic cellular responses. CRABP-II and RAR interact with atRA with high affinity, in the range of 0.1-0.2 nM. Both FABP5 and PPAR-β/δ associate with atRA at an affinity of 10-50 nM. Reasonably, it has been noted that activation of PPAR-β/δ will only be displayed in cells that have a high basal FABP5/CRABP-II ratio due to the noted differences in atRA affinity. For example, in keratinocytes, levels of FABP5 are much greater than CRABP-II and therefore cell proliferation and growth are the dominantly activated cellular mechanisms.[35] To more fully understand the complex cell signaling cascades generated by RAR and PPAR-β/δ, there is need to further characterize the transcribed protein alterations that occur following atRA treatment. Discovering secondary protein activity that contributes to

the delicate balance of pro- and anti-apoptotic activation will generate a more thorough comprehension of the tissue-specific responses to atRA and the utility of this compound in the therapy of cancer, skin care, and obesity.

Much of the current literature elucidating atRA-induced differences in proteome profiles is founded on techniques that resolve proteins using two-dimensional gel electrophoresis (2-DE).[44, 45] 2-DE is capable of simultaneously resolving thousands of proteins.[46] But, the fundamentals of the technique impose several limitations that negatively impact the ability of researchers to observe the global alterations of the proteome necessary in understanding the complete molecular mechanism of atRA.[47] In addition to the caveats pertaining to gel spot excision, efficiency of protein digestion, peptide extraction and MS characterization, it has been well characterized that both highly acidic and basic proteins as well as large molecular weight proteins are vastly underrepresented in 2-DE.[48] While improvements such as two-dimensional difference gel electrophoresis (2D-DIGE) have improved the quantification, reproducibility, resolution and sensitivity of 2-DE,[49] the loss of previously mentioned protein groups still confines the scope and application for which this method is suitable.

Mass spectrometry based proteomics has been vital to expand the understanding of global protein alterations in cellular systems.[43] Data-independent acquisition techniques have extended the detectable dynamic range capabilities of shotgun proteomics[76, 91, 93, 94] and have been fruitful in the application to clinical research.[102, 103, 122] More specifically, the data-independent acquisition method termed Precursor Acquisition Independent From Ion Count (PAcIFIC) has significantly increased the detectable dynamic range of complex proteomic samples over other data-dependent and data-independent acquisition schemes.[91] The PAcIFIC method operates by successively interrogating predefined 2.5 m/z isolation widths, or channels, by tandem-MS

across a 1000 m/z range. Batches of *n* continuous 2.5 m/z channels, overlapping by 1 m/z, are analyzed per sample injection and chromatographic gradient. The *n* number of channels comprises the PAcIFIC duty cycle and each successive sample injection will analyze the next *n* channels until a total desired m/z range is complete.[91] As opposed to a precursor ion scan used in data-dependent methods, the use of the small 2.5 *m/z* isolation width greatly increases the measureable dynamic range of the total PAcIFIC analysis by negating the effect of highly abundant precursor ions outside of the narrow *m/z* window being interrogated. As a result, Panchaud et al. report a detectable dynamic range of eight orders of magnitude at an FDR of 0.5%.[91] The PAcIFIC method has proven an invaluable tool for researchers looking to increase proteome coverage of complex biological matrices such as serum, tissue culture and urine samples.[100-103, 122] Additionally, a quantitative advantage is achieved with such DIA techniques due to the repetitive sampling of each m/z channel during the elution of a single chromatographic peak.[101]

Here, I intend to use data-independent acquisition (DIA) mass spectrometry methods to profile atRA-mediated protein signatures of two different global cellular responses. I will use two model cell lines: 1) the breast carcinoma cell line, MCF-7, which displays classical RAR-mediated cell cycle arrest followed by apoptosis;[133] and 2) the hepatocellular carcinoma cell line, HepG2, which is less characterized but has shown indications of alternative pathway activation. Employing the technique, Precursor Acquisition Independent From Ion Count (PAcIFIC), with means of semi- and relative-quantification, I will compare the proteomes of MCF-7 and HepG2 cell lines to search for signatures of atRA-induced protein alterations. In addition, I will discuss advances to the PAcIFIC method that drastically increase sample throughput.

3.2 Experimental procedures

Cell viability

Cell growth was measured by adding WST-1 reagent (Roche, Indianapolis, IN, USA). Briefly, cells were grown in a 96 well plate cultured in a 5% carbon dioxide environment in a humidified incubator at 37 °C. Using 6 replicates per treatment and a DMSO control, the effect of atRA treament (0.001 – 50 μM) on cell proliferation was measured after 72 hours of treatment. Every 24 hours, the media was removed from each treatment group and cells were washed with PBS, and media containing atRA was replaced. At 72 hours, 10 μL of WST-1 reagent was added to each well and incubated at 37 °C for 1 hour to measure cell viability and growth according to manufactures instructions. The absorbance was measured at 450 nm with background subtraction at 680 nm. Each treatment was normalized to the control cells and expressed as a percentage of the control.

Cell culture

Human breast carcinoma (MCF-7) and hepatocellular carcinoma (HepG2) cell lines were cultured in a 5% carbon dioxide environment in a humidified incubator at 37 °C. The growth medium used was Dulbecco's Modified Eagle Medium (DMEM) with pyruvate and low glucose. The medium was supplemented with 10% FBS and 1% penicillin/streptomycin. 10 cm^2 dishes were grown to 70% confluence prior to starting atRA and DMSO vehicle treatments. The cells were treated with either at-RA (10 μM) or DMSO (control) for 4, 24, and 72 hours. The media on the cells grown longer than 24 hours was replaced every 24 hours with fresh treatment media till harvest. Cells were harvested by washing twice with ice cold PBS and scraping the cells into 0.5 mL of 100 mM ammonium bicarbonate. The resulting cell suspension was sonicated to lyse

the cells and subsequently centrifuged at 5,000 g force to pellet cellular debris. Supernatant from each sample was extract and placed in a clean eppendorf on ice. Protein quantification was performed with a BCA protein quantification kit following the manufacturer's protocol (Thermo Scientific, Pierce). Protein quantity was normalized for each sample prior to further proteomics sample preparation.

Protein digestion and peptide preparation

200 μg/100 μL protein sample were denatured with the addition of urea to 6M. Subsequently, samples were buffered with the addition of 7 μL 1.5 M Tris pH 8.8, reduced with 2.5 μL of 200 mM tris(2-carboxyethyl)phosphine (TCEP) for 1 hour at 37 °C, alkylated with 20 μL of 200 mM iodoacetamide for 1 hour at room temperature in the dark, and then quenched with 20 uL of 200 mM dithiothreitol. Prior to addition of sequencing-grade porcine trypsin (Promega, Madison, WI, USA) at a protein to enzyme ratio of 50:1, samples were diluted with 900 μL of 50 mM ammonium bicarbonate and 200 μL of MeOH. After an overnight incubation, peptides were desalted on a Vydac C18 macrospin column (The Nest Group, Southborough, MA, USA) according to the manufacturer's protocol. Resulting eluent was concentrated on a SPD 111V SpeedVac (Thermo Savant, San Jose, CA, USA) and stored until further use.

Nano-electrospray ionization mass spectrometry

Nano-HPLC was performed using a Waters NanoAquity (Milford, MA, USA). A homemade trapping column was made from 100 μm inner diameter (ID) capillary (Polymicro Technologies, Phoenix, AZ, USA) packed with 2 cm of 200 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). Successive analytical separation was performed on a homemade laser-pulled 75 μm ID (Polymicro Technologies, Phoenix, AZ, USA) column packed

with 20 cm of 100 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). For each sample injection, approximately 1 μg, unless otherwise noted, of the peptide sample was loaded on the trapping column at 4 μL/min with 95% water/5% acetonitrile/0.1% formic acid for 5 minutes. Trapped peptides were then eluted from the trapping column onto the analytical column using a variable gradient with a flow rate of 0.25 μL/min. The gradient utilized two mobile phase solutions: A, water/0.1% formic acid; and B, acetonitrile/0.1% formic acid. The variable gradient used is as follows: 0 minutes, A (95%), B (5%); 55 minutes, A (65%), B (35%); 65 minutes, A (15%), B (85%); 75 minutes, A (15%), B (85%); 80 minutes, A (95%), B (5%); 80-100 minutes, A (95%), B (5%). Peptide digests were analyzed on both a LTQ Oribtrap and a LTQ Velos (Thermo Fisher, San Jose, CA, USA) by nano-electrospray ionization in positive ion mode. Ion source conditions were optimized using the tuning and calibration solution suggested by the instrument provider. For PAcIFIC acquisition, the mass spectrometer parameters were set as defined by Panchaud et al.[91, 98]

Captive spray ionization mass spectrometry

Captive spray ionization was performed using a Waters NanoAquity (Milford, MA, USA). A manufactured fused silica 200 μm ID trapping column purchased from New Objective (Woburn, MA, USA) and packed with 2 cm of 200 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). Successive analytical separation was performed on a manufactured fused silica 200 μm ID trapping column purchased from New Objective (Woburn, MA, USA) and packed with 15 cm of 100 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA, USA). For each sample injection, approximately 1 μg, unless otherwise noted, of the peptide sample was loaded on the trapping column at 6 μL/min with 95% water/5% acetonitrile/0.1% formic acid for

2.5 minutes. Trapped peptides were then eluted from the trapping column onto the analytical column using a variable gradient with a flow rate of 2.5 μL/min, unless otherwise noted. The gradient utilized two mobile phase solutions: A, water/0.1% formic acid; and B, acetonitrile/0.1% formic acid. The variable gradient used for the optimized PAcIFIC acquisition method is listed in Chapter 3: Section 1.2. Peptide digests were analyzed on a LTQ Oribtrap (Thermo Fisher, San Jose, CA, USA) by captive spray ionization in positive ion mode. Ion source conditions were optimized using the tuning and calibration solution suggested by the instrument provider and under conditions recommended by the CSI manufacturer (Bruker/Michrom).

Database search and protein quantification

Data acquired on the LTQ Orbitrap and LTQ Velos was converted from Thermo's RAW format to the universal mzXML format and searched against the IPI human database v3.49 (http://www.ebi.ac.uk/IPI/IPIhuman.html) using SEQUEST.[70] the precursor ion tolerance was set to 3.75 and additional search parameters for SEQUEST included trypsin enzyme specificity, cysteines modified with iodoacetamide and the variable option for methionines in reduced or oxidized form. Results were analyzed with Peptide Prophet[126, 127] ensuring that peptide hits with a probability of >0.99 were accepted and linked to protein entries.

Protein quantification and statistical filtering

For the CSI PAcIFIC dataset, protein identifications were assigned relative quantification values using in-house spectral counting software.[134] The relative quantification value for each protein is determined by summing all peptide tandem MS spectra correlated to a respective

protein.[112] Following spectral counting, the data was filtered using a Student's t-test and a p-value cutoff of 0.05 or less to ensure that protein changes were consistent throughout the biological triplicate and technical duplicate analyses.

Putative identification of RARE sequences within promoter regions

The DNA sequences of the identified proteins were scanned as described by Laperriere et al. to determine if retinoic acid response elements (RAREs) existed within the respective promoter regions.[135] In-house Perl scripts were written to collect and parse stretches of genomic information for potential RARE promoter sequences in a two-step process.[136] The first step uses the BioMart API (http://www.biomart.org) to access Ensembl (http://www.ensemble.org) and record the user-defined gene information to a text file in FASTA format. Genes of interest are listed in FASTA format and are accessed by the Perl script. For each gene in the input file, 10,000 DNA base pairs are captured directly upstream of the gene and amended to an output file. The second step uses an independent Perl script to parse the aforementioned collection of gene information for RARE sequences. Perl string matching is employed to match sequences with the following requirement: (A|G)G(G|T)T(C|G)A(A|G|T|C)n(A|G)G(G|T)T(C|G)A, where n is one, two or five bases as defined by DR1, DR2 and DR5 RARE sequences, respectively. As defined by Perl, bases listed within parentheses and separated by the character "|" allow the string matching algorithm to match only one of the listed bases. If a series of base pairs in a gene match the string representing a potential RARE sequence based on the above rule, the gene name and the determined RARE sequence found are listed in a FASTA format output file.

3.3 Results

Differential effect of atRA on cell proliferation

The variant cellular responses to atRA are well documented in the literature, although the cellular mechanisms of this divergence are not completely understood.[35] To further establish cellular outcomes mediated by atRA, cell viability assays were performed on two cell lines of interest to document possible changes in cell viability and ultimately the effect of atRA on cell proliferation.[137] The first cell line is from a human breast carcinoma, identified as MCF-7, and has been shown to have a pro-apoptotic response following atRA treatment.[33] The comparative cell line, human hepatocellular carcinoma (HepG2), was selected based on preliminary data indicating a potential atRA-mediated anti-apoptotic response. Cell viability assays were performed at a range of atRA concentrations (0.001, 0.01, 0.05, 0.1, 0.5, 1, 10, 30, 50 μM) due to the well-established atRA dose-dependent cellular response.[138, 139] The dose-dependent effects of atRA on MCF-7 and HepG2 cells at 72 hours are illustrated in Figure 1. It should be noted that at atRA concentrations between 0.001 and 1 μM, minimal effects are observed on cell proliferation at 72 hours. At atRA concentrations of 10, 30, and 50 μM, a continual decrease in MCF-7 proliferation with increasing atRA concentration is observed. Alternatively, in the HepG2 cell line, there is no observed decrease in proliferation over the tested atRA concentration span (0.001 – 50 μM). As shown in Figure 1, the lowest concentration where a significant difference between the proliferation of MCF-7 and HepG2 cells occurs is at 10 μM atRA. In order to quantitatively capture alterations in the proteome signatures responsible for the observed divergent cellular fates of MCF-7 and HepG2 cell lines at a pharmacologically relevant concentration, samples for proteomics analysis were collected after treatment with 10 μM atRA.

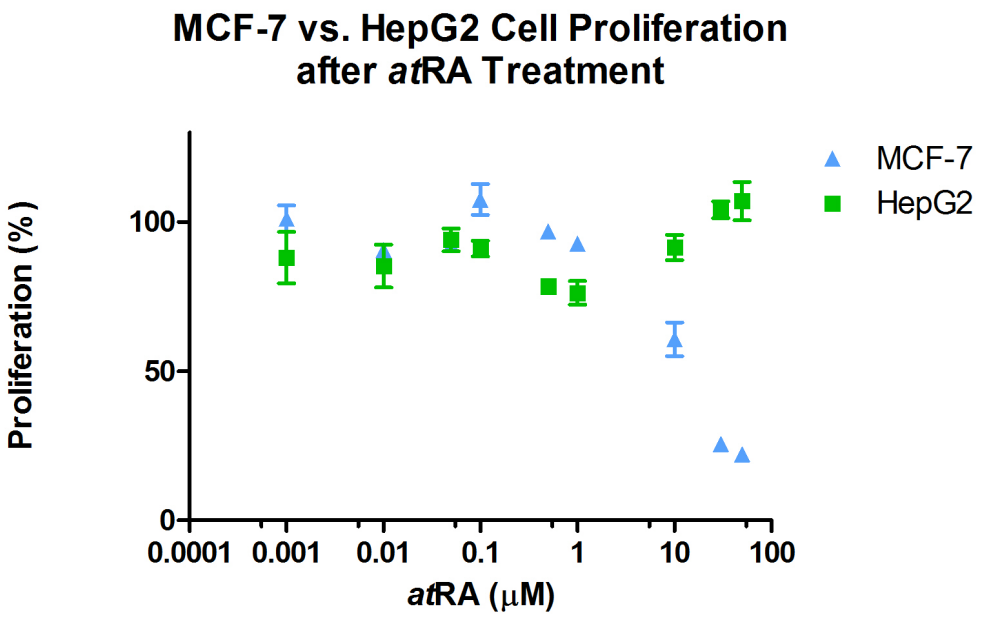


Figure 1. Effect of increasing concentrations of atRA on cell proliferation assay of the breast carcinoma cell line, MCF-7, and the hepatocellular carcinoma cell line, HepG2, treated with atRA concentrations ranging from 0.001 - 50 μM measured after 72 hours of treatment. *Figure courtesy of Sam Arnold.

Relative quantitative proteomics analysis of MCF-7 and HepG2 cells

I applied a recently developed 12-hour CSI PAcIFIC method (see Chapter 2) to study the proteome alterations of MCF-7 and HepG2 cells treated with 10 μM atRA. MCF-7 samples were analyzed over a treatment time course of 4, 24, and 72 hours and HepG2 samples were analyzed at 72 hours. Spectral counting was used to quantify relative changes in protein abundance between atRA treated samples and DMSO controls. Relative quantification values are reported in Table 1 as a ratio of the spectral counts recorded in the atRA treated samples over the spectral counts of the respective DMSO vehicle control. The biological triplicates were analyzed in technical duplicate and a Student's t-test was used to statistically assess significant differences

between the DMSO control and atRA treated cell samples. The p-value from the Student's t-test is reported in Table 1 and a threshold of ≤ 0.05 was used to select proteins of interest (underlined). The number of proteins found to be significantly changed in expression over the MCF-7 time course is as follows: 4 hours, no proteins; 24 hours, 12 proteins; and 72 hours, 44 proteins. While I did not expect to observe many changes at 4 hours, this time point was analyzed to ascertain if there is evidence of any proteasome-directed degradation following atRA treatment. To aid in the comprehension of the extensive list detailing significant changes in protein abundance mediated by atRA, proteins were placed in functional categories and discussed in the following sections.

Protein expression and integrity

Protein abundance and functionality is a tightly regulated process that is vital to cellular health. The life of a protein begins with transcription, mRNA processing, and translation; continues in light of environmental challenges with the help of protective agents and chaperone activity; and, ultimately ends with protease-catalyzed degradation.[140-142] Here, I will examine each stage of the protein life cycle in search of possible atRA-mediated proteome signatures that differentiate MCF-7 and HepG2 response pathways.

A clear deviation was observed between MCF-7 and HepG2 cell lines in identified proteins with transcriptionally related function that are measured with significant abundance changes. MCF-7 protein presence at 72 hours depicts a clear down regulation of proteins vital to transcriptional activation. For example, both chromatin target of PRMT1 protein (CHTOP; p-value = 0.038) and TATA-binding protein-associated factor 2N (TAF15; p-value = 0.038) are down regulated. These proteins are responsible for translational initiation at distinct promoters

and activation of target genes.[143, 144] Additionally, the proliferating cell nuclear antigen (PCNA; p-value = 0.040), protein machinery that aids in polymerase activity during elongation is down regulated.[145] In a dichotomy of cellular response, the HepG2 atRA-mediated proteome signature reveals a general up regulation of transcriptional activators in concert with a down regulation of transcriptional silencing machinery. Of note is the increase in presence of replication factor C subunit 2 (RFC2; p-value = 0.027) responsible for promotion of DNA polymerase elongation activity.[146] And, a reduction in the protein abundance of transcriptional silencing factor, chromobox protein homolog 3 (CBX3; p-value = 0.024) is responsible for repression of E2F- and MYC-responsive genes.[147]

A distinct signature of atRA-mediated proteome changes in the HepG2 cell line is the number of significantly regulated proteins involved in RNA processing. In HepG2 cells I observed an extension of the increased activity in the protein expression process with an up regulation of machinery involved in RNA processing, maturation, and splicing. There is one exception in the identified proteins with an observed decrease in abundance of heterogeneous nuclear ribonucleoprotein H2 (HNRNPH2; p-value = 0.013). HNRNPH2 aids in pre-mRNA conversion to mature translatable mRNA. The data provides evidence that additional HNRN protein family members were simultaneously up regulated in HepG2 cells at 72 hours (HNRNPH3; p-value = 0.008 and HNRNPUL2; p-value = 0.017).[148]

Translation is a critical step in protein expression that is primarily driven by ribosomal activity. The eukaryotic ribosome is an intricate 80S protein complex that can be divided into two subunits: 60S and 40S. Each subunit is comprised of both RNA components and a multitude of proteins.[149] As depicted in Table 1, quantified protein identifications reveal unique signatures for atRA-mediate changes in protein abundance for MCF-7 and HepG2 cells. In MCF-7 cells, it

is observed that there is a down regulation of a 60S protein component (RPL28; p-value = 0.013) with an increase in a 40S protein component (RPS20; p-value = 0.001). Alternatively, in HepG2 cells, there is a decrease in abundance of a 40S protein component (RPL13; p-value = 0.037). By reviewing protein identifications from both MCF-7 and HepG2 cells, it is noted that both cell lines elicit an atRA-mediated increase in abundance of proteins involved with initiating and promoting protein translation for MCF-7 (EIF6; p-value = 0.024) and HepG2 cells (EIF3A; p-value = 0.007, EEFIB2; p-value = 0.02, and EIF5B; p-value = 0.037).

Protein integrity and conformational maintenance is important to the life cycle of a protein. Changes in cellular environment and oxidative damage often require the need for chaperones and stress-induced protective measures.[150] In both the MCF-7 and HepG2 cells it is observed that there is a general increase in proteins capable of such function at 72 hours. There are several heat shock proteins up regulated in MCF-7 (HSPE1; p-value = 0.006) and HepG2 cells (HSPA1A; p-value = 0.031, and HSP90AB1; p-value 0.041). Additionally, several chaperones are observed in increased abundance (CALR; p-value = 0.030, and HYOU1; p-value = 0.038).

The final stage for proteins in their life cycle is often degradation by directed protease activity. As observed in Table 1, MCF-7 cells display an increased abundance of a proteasome component at 24 hours (PSMB7; p-value = 0.029) followed by a marked decrease in a complimentary proteasome component at 72 hours (PSMB3; p-value = 0.008). In HepG2 protein identifications I see a unique atRA-mediated protein signature of the induction of 26S proteosome components (PSMD3; p-value = 0.007, PSMC2; p-value = 0.025, and PSMD4; p-value = 0.047). The 26S proteasome is an ATP-dependent enzyme responsible for the

degradation of ubiquitinated proteins and increased presence of several components of this protein complex indicate targeted proteolysis of a specific subset of proteins.[151]

Cellular Trafficking

Numerous proteins identified with significantly relevant changes in abundance due to atRA treatment function in a capacity of trafficking cellular components. In the MCF-7 cells, atRA-mediated changes in protein abundance increase the presence of a nuclear import protein, importin-4 (IPO4; p-value = 0.047), at 24 hours. IPO4 is a receptor that recognized nuclear localization signals (NLS) and promotes transport across the nuclear membrane.[152] At 72 hours, there is no differential presence of IPO4 and additional identified proteins capable of cellular trafficking are down regulated. In HepG2 cells, I observed a unique atRA-mediated proteome signature as most identified proteins involved with cellular trafficking of proteins, amino acids and ions such as Ca^{2+} and Fe^{3+} are increased in abundance at 72 hours. Of particular importance for future discussion of cell cycle regulation is the protein, E3 SUMO-protein ligase RanBP2 (RANBP2; p-value = 0.013). RANBP2 is critical to the Ran-GTP-mediated import of proteins into the nucleus for control of the cell cycle.[153, 154]

Metabolism

From the proteins categorized in Table 1, it can be seen that metabolic pathways account for a large number of the quantified proteins (34 of 143) determined to be significantly altered in abundance following atRA treatment. In recently published studies, unique cellular responses to atRA appear to be determined by ratio of retinoic acid transport proteins, CRABPII and FABP5, and the respective activation of either RAR or PPAR-β/δ.[35] Of importance to the control of

metabolism, additional pathways activated by PPAR-β/δ involve fatty acid transport, fatty acid oxidation, and adipocyte differentiation.[155] This is in correlation with the proteins involved in fatty acid metabolism and quantified as up regulated in HepG2 cells at 72 hours (PLIN2; p-value = 0.001, FASN; p-value = 0.002, ECH1; p-value = 0.006, ACSL3; p-value = 0.012, GOT1; p-value = 0.014, Acyl-coenzyme A dehydrogenase; p-value = 0.019, ACSL4; p-value = 0.025, ALDH3A2; p-value = 0.030, CYB5R3, p-value = 0.036, and PPT1; p-value = 0.049).[156]

Cell cycle regulation

Cell growth and division is an intricate process in which the cell will pass through several defined stages and checkpoints to ensure the healthy production of a daughter cell. The cyclical process starts at the G_1 phase, wherein the newly divided cells will continue to grow, replicate organelles, and prepare for DNA synthesis. Following the G_1 phase, the cell will enter the S phase. Here, DNA is duplicated so that a new copy of the genome can be passed onto the daughter cell. Subsequently, in the G_2 phase the cell will continue to grow for a period of time and prepare for mitosis (M phase) wherein the cell will divide. Alternatively, the cell can enter or exit the G_0 phase, a resting or quiescent state, from the G_1 phase.[157] Tracking the protein factors that stimulate or repress the cell cycle is critical to understanding the unique response a cell line may have to atRA. Here, I will review the protein identifications linked to cell cycle regulation.

It has been demonstrated that MCF-7 cells exposed to atRA treatment undergo G_1/G_0 cell cycle arrest prior to apoptosis.[133] Furthermore, Zhu et al. demonstrated that atRA-mediated cell cycle control in MCF-7 cells leads to a decrease in the expression and activity of the transcription factor E2F1 (E2F1) and the retinoblastoma-associated protein (RB1). RB1 phosphorylation (pRB) is necessary for the cellular transition from G_0 to G_1 phase and pRB1 interrupts RB1 binding inhibition of E2F1. All-trans retinoic acid reduces phosphorylation levels

of RB1 thereby promoting G_0 phase and repressing RB1/E2F1 transcription activity resulting in cell cycle arrest.[158]

Here, the proteomics data presented in Table 1 implicates additional proteins and pathways that contribute to G_0/G_1 cell cycle arrest of MCF-7 cells. An increase in protein expression of the lamina-associated polypeptide 2, isoform alpha (LAP2; p-value = 0.013) was observed at 72 hours. This contributes to cellular quiescence since LAP2 will bind to RB1, inhibiting the necessary RB1 protein interactions with the E2F family of transcription factors to promote target gene expression critical for the transition to the S phase. Additionally, increased expression of LAP2 led to hyporphosphorylation of RB1 further inhibiting RB1/E2F1 interaction.[159-161] Tumor protein D53 (TPD52L1; p-value = 0.024), decreased in MCF-7 cells at 72 hours has been shown to play a critical role in cell cycle progression and decreased abundance is linked to cell cycle arrest.[162] Furthermore, it was observed that several members of the minichromosome maintenance (MCM) complex were down regulated in MCF-7 cells at 72 hours: MCM4 (p-value = 0.028) and MCM7 (p-value = 0.045). The MCM complex is a strictly regulated replicative helicase that plays a pivotal role in formation of DNA forks and elongation that occurs 'once per cell cycle' during the S phase.[163] The G_0/G_1 arrest of the cell cycle induced by atRA in MCF-7 negates the necessity of these proteins and therefore the observed decrease in expression of MCM complex proteins is inline with published studies.

In correlation with the above pathways, the proteomics data suggests involvement of additional cell cycle regulating pathways that halt progression past the G_1 phase in MCF-7 cells. All-trans retinoic acid decreases the expression of hematological and neurological expressed 1 (HN1; p-value = 0.032) by nearly 50% at 24 hours. Varisli et al. and Laughlin et al.

demonstrated that knockdown of HN1 prolongs and arrest cell cycle progression in the G_1 phase.[164, 165]

Comparatively, results from the cell viability assay indicate that HepG2 cells experience proliferative effects following atRA treatment. Accordingly, significantly regulated proteins identified in Table 1 indicate that HepG2 cells progress beyond the G_0/G_1 phase. Increased expression of MCM2 (p-value = 0.005), a member of the minichromosome maintenance complex, indicates that a population of HepG2 cells are actively preparing to duplicate DNA during the S phase.[166] Additionally, the SUMO-activating enzyme subunit 1 (SAE1; p-value = 0.029) shows increased expression at 72 hours. SAE1 mediates ATP-dependent activation of SUMO proteins, which are critical for DNA replication that occurs during the S phase of the cell cycle.[167]

Furthermore, proteomics data from HepG2 cells at 72 hours measures that the GTP-binding nuclear protein Ran (RAN; p-value = 0.042) is up regulated; a protein that is necessary for coordinating the onset of mitosis and the completion of the S phase.[168] RAN interacts with the protein, regulator of chromosome condensation (RCC1; p-value = 0.011), which is up regulated in HepG2 cells at 72 hours. The RCC1-RAN complex regulates chromosome condensation at the tail end of the S phase and simultaneously detects unreplicated DNA. RCC1 specifically recruits RAN to nucleosomes and by generating a RAN concentration gradient between the nucleus and cytoplasm activates RAN-mediated nucleotide exchange activity.[157]

Cell survival

Cell cycle, cell death, and cell survival pathways are highly interconnected. Measuring changes in expression of proteins involved in the stated pathways is critical to understanding the

activation and/or deactivation of these complex signaling cascades.[169] The MTT assay, as well as published literature, conclusively demonstrates the apoptotic effect of atRA in MCF-7 cells. Here, I will discuss the role of several identified proteins in the positive regulation of cell survival.

While the final cellular fate of MCF-7 cell following atRA treatment is apoptosis, I identified proteins that indicate a certain level of cellular protection is still occurring. The Ras GTPase-activating protein-binding protein 1 (G3BP1; p-value = 0.010) is increased in expression in MCF-7 cells at 72 hours. While the cellular localization of G3BP1 is critical to activity, changes in expression indicate that this downstream effector protein of Ras signaling leads to cell cycle progression.[170] Additionally, RuvB-like 1 protein (RUVBL1; p-value = 0.024) is a protein that is essential for cellular growth, chromatin remodeling and DNA replication.[171] The protein, isocitrate dehydrogenase (IDH1; p-value = 0.011) is increasingly up regulated from 24 to 72 hours. Responsible for conversion of $NADP^+$ to NADPH, IDH1 controls the redox balance of the cell and protects against oxidative damage.[172] Additional cytoprotective proteins were also significantly regulated. Stress-70 protein (HSPA9; p-value = 0.045) was measured as up regulated in both MCF-7 and HepG2 cells at 72 hours. HSPA9 is important in cell cycle progression and may act as protective protein chaperone.[173] Conversely, the down regulation of proteins that typically promote apoptosis is a sign of directed pathway control to restrict specific routes of apoptosis. BH3-interacting domain death agonist (BID; p-value = 0.013) is down regulated in MCF-7 cells at 72 hours. BID induces changes in Bax to promote the release of cytochrome C release during apoptosis.[174] Additionally, galectin-1 (LGALS1; p-value = 0.022), a protein that induces E2F1-mediated apoptosis in cooperation with p53 is down regulated.[175, 176] HepG2 cells continue to proliferate in the presence of atRA. Here, I review identified proteins that implicate possible atRA-mediated pathways involved in HepG2 cell survival. 3-ketoacyl-

CoA thiolase (ACAA2; p-value = 0.020) is up regulated in HepG2 cells at 72 hours. Cao et al. demonstrated that ACAA2 was an inhibitory binding partner of BNIP3, a protein that belongs to the BH3-only subset of the Bcl-2 family, and abolished BNIP3 mediated apoptosis and mitochondrial damage.[177] Supporting cell proliferation in HepG2 cells, the proteomics data quantified up regulation of the protein, 28 kDa heat- and acid-stable phosphoprotein (PDAP1; p-value = 0.017). PDAP1 enhances the PDGFA-stimulated growth observed in a variety of cell lines.[178] Similar in action to HSPA9, HepG2 cells express an additional protein to mediate oxidative stress caused on by atRA treatment. Glutaredoxin-3 (GLRX3; p-value = 0.020) is a protein that is up regulated in HepG2 cells at 72 hours. It is a component of cellular redox maintenance and uses cellular glutathione to promote cell survival during oxidative stress.[179] Thioredoxin domain-containing protein 12 (TXNDC12; p-value = 0.029), upregulated in HepG2 cells, has a functional similarity to TXN and enhances cell viability during oxidative stress.[180]

Apoptosis

Cell survival and apoptosis is a fine balance of protein signaling cascades. In contrast to the above cell survival mechanisms, here will I discuss both a decrease in cell proliferative proteins and an increase in apoptosis promoting networks. GDNF family receptor alpha-1 (GFRA1; p-value = 0.003) is down regulated in MCF-7 cells at 72 hours. GFRA1 is a receptor for GDNF which mediates GDNF-induced activation of RET. In turn, RET stimulates activation of MAPK/ERK and PI3K/Akt pathways for cell survival.[181] Additionally, cell proliferating pathways experience a down regulation of protein components such as thioredoxin (TXN; p-value = 0.013). TXN is a inhibitory binding partner of ASK1, a MAPKKK that is required for TNF-alpha induced apoptosis. Normally, TXN has anti-apoptotic effects as it inhibits activation

of this pathway by inhibition of ASK1. Down regulation reduces ASK1 inhibition and therefore can promote pro-apoptotic signaling.[182] Anterior gradient protein 2 homolog (AGR2; p-value = 0.048) promotes tumor growth and proliferation.[183] Down regulation, as shown by the proteomics data at 72 hours, was shown to promote cellular senescence and eventually apoptosis.[184] Voltage-dependent anion-selective channel protein 1 (VDAC1; p-value = 0.015) is up regulated in both MCF-7 and HepG2 cells at 72 hours. VDAC1 forms a channel through the mitochondrial membrane to allow the release of apoptotic molecules such as cytochrome C, Smac/Diablo and apoptosis inducing factors (AIF) up stream activators of caspases.

Structural Integrity

Numerous proteins were quantified with significant differences in abundance at 72 hours in MCF-7 and HepG2 cells. There is a mixture of up and down regulated structural proteins for both cell lines. During cellular treatments, it was observed that MCF-7 cell undergo morphological changes at 72 hours becoming separated, shrunken and elongated in appearance. Accordingly, structural proteins are decreased in abundance that control intercellular junctions (DSP; p-value = 0.001) and cell structure (AHNAK; p-value = 0.008). HepG2 cells continue to proliferate without noticeable changes in morphological appearance.

Table 1. (Pt. 1 of 3)

		MCF-7						HepG2				
		4 hrs		*24 hrs*		*72 hrs*		*72 hrs*				
	UniProt	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**Protein Name**	**Gene Name**	**RARE**
Transcription	Q9Y3Y2	-	-	0.667	0.038	-	-	-	-	Chromatin target of PRMT1 protein	CHTOP	
	Q92804	-	-	-	-	0.667	0.038	1.000	1.000	TATA-binding protein-associated factor 2N	TAF15	DR2
	P12004	0.698	0.288	1.092	0.606	0.828	0.040	0.933	0.233	Proliferating cell nuclear antigen	PCNA	
	O14562	-	-	-	-	-	-	1.310	0.003	Ubiquitin domain-containing protein UBFD1	UBFD1	
	Q9BQG0	-	-	-	-	-	-	1.486	0.010	Myb-binding protein 1A	MYBBP1A	
	P20290	-	-	1.063	0.758	1.000	1.000	1.217	0.024	Transcription factor BTF3	BTF3	DR2
	Q13185	-	-	0.909	0.435	1.097	0.139	0.884	0.024	Chromobox protein homolog 3	CBX3	DR2
	P35250	-	-	-	-	-	-	2.300	0.027	Replication factor C subunit 2	RFC2	DR2
	Q9Y224	-	-	-	-	-	-	1.409	0.029	PF0568 protein C14orf166	C14ORF166	DR2
	Q08211	-	-	1.258	0.427	1.022	0.918	1.510	0.040	ATP-dependent RNA helicase A	DHX9	
RNA Processing	O43776	-	-	1.438	0.135	1.520	0.001	1.156	0.331	Asparagine--tRNA ligase, cytoplasmic	NARS	DR2
	Q99714	-	-	1.000	1.000	0.369	0.016	0.639	0.152	3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	
	O75643	-	-	1.087	0.729	1.154	0.538	1.700	0.004	U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	DR2
	P51114	-	-	-	-	-	-	2.083	0.006	Fragile X mental retardation syndrome-related protein 1	FXR1	
	P31942	1.000	1.000	1.063	0.779	0.882	0.387	1.917	0.008	Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	DR2
	Q1KMD3	-	-	1.196	0.091	1.108	0.283	1.364	0.017	Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	DR2
	P55795	-	-	1.036	0.811	0.974	0.878	0.700	0.013	Heterogeneous nuclear ribonucleoprotein H2	HNRNPH2	
	P62316	-	-	-	-	-	-	1.357	0.024	Small nuclear ribonucleoprotein Sm D2	SNRPD2	DR2
	P23246	0.773	0.655	0.949	0.672	1.082	0.547	1.200	0.034	Splicing factor, proline- and glutamine-rich	SFPQ	
	Q16630	-	-	-	-	0.833	0.184	1.304	0.036	Cleavage and polyadenylation specificity factor subunit 6	CPSF6	
	P49588	-	-	1.354	0.266	1.145	0.245	1.320	0.040	Alanine--tRNA ligase, cytoplasmic	AARS	DR2
	Q15637	-	-	1.154	0.493	0.909	0.519	1.400	0.047	Splicing factor 1	SF1	DR5
Translation	P60866	-	-	-	-	2.091	0.001	-	-	40S ribosomal protein S20	RPS20	
	P46779	-	-	0.806	0.013	-	-	0.522	0.010	60S ribosomal protein L28	RPL28	
	P56537	-	-	-	-	1.455	0.024	0.967	0.770	Eukaryotic translation initiation factor 6	EIF6	
	Q14152	-	-	-	-	-	-	1.781	0.007	Eukaryotic translation initiation factor 3 subunit A	EIF3A	
	P62241	0.470	0.157	1.073	0.544	1.134	0.146	0.804	0.025	40S ribosomal protein S8	RPS8	DR2
	Q9NR30	-	-	1.357	0.221	1.050	0.581	1.329	0.013	Nucleolar RNA helicase 2	DDX21	
	P24534	1.111	0.686	1.158	0.150	1.067	0.184	1.286	0.020	Elongation factor 1-beta	EEF1B2	
	Q08J23	-	-	-	-	-	-	1.818	0.029	RNA (cytosine(34)-C(5))-methyltransferase	NSUN2	
	O60841	-	-	-	-	-	-	1.512	0.035	Eukaryotic translation initiation factor 5B	EIF5B	DR2
	P26373	-	-	1.038	0.672	1.049	0.671	0.733	0.037	60S ribosomal protein L13	RPL13	DR2, DR5
	P50914	-	-	-	-	-	-	1.394	0.044	60S ribosomal protein L14	RPL14	DR2
Protien integrity	P61604	-	-	1.017	0.878	1.559	0.006	0.780	0.276	10 kDa heat shock protein, mitochondrial	HSPE1	
	Q13765	-	-	-	-	0.688	0.024	-	-	Nascent polypeptide-associated complex subunit alpha	NACA	
	Q9Y3C6	0.778	0.375	1.129	0.047	1.263	0.283	0.955	0.423	Peptidyl-prolyl cis-trans isomerase-like 1	PPIL1	
	P13667	-	-	-	-	1.231	0.379	1.181	0.028	Protein disulfide-isomerase A4	PDIA4	DR2
	Q08752	-	-	-	-	-	-	0.789	0.025	Peptidyl-prolyl cis-trans isomerase D	PPID	
	P27797	-	-	0.895	0.730	0.625	0.179	1.246	0.030	Calreticulin	CALR	DR2
	P08107	-	-	1.286	0.225	0.913	0.423	2.700	0.031	Heat shock 70 kDa protein 1A/1B	HSPA1A	DR2
	Q9NYU2	-	-	-	-	-	-	1.545	0.038	UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	
	Q9Y4L1	-	-	-	-	1.273	0.139	1.505	0.038	Hypoxia up-regulated protein 1	HYOU1	DR2
	P11021	0.595	0.252	1.074	0.529	0.991	0.230	0.938	0.040	78 kDa glucose-regulated protein	HSPA5	DR2
	P08238	0.733	0.387	1.060	0.573	1.066	0.502	1.132	0.041	Heat shock protein HSP 90-beta	HSP90AB1	
Protease activity	P49720	-	-	-	-	0.696	0.008	0.900	0.692	Proteasome subunit beta type-3	PSMB3	DR1, DR2
	Q99436	-	-	1.692	0.029	1.000	1.000	0.750	0.295	Proteasome subunit beta type-7	PSMB7	
	O43242	-	-	-	-	-	-	1.600	0.007	26S proteasome non-ATPase regulatory subunit 3	PSMD3	DR2
	Q06323	-	-	-	-	-	-	2.000	0.020	Proteasome activator complex subunit 1	PSME1	DR2
	P35998	-	-	0.970	0.873	1.143	0.298	1.200	0.025	26S protease regulatory subunit 7	PSMC2	
	P55036	-	-	0.850	0.420	0.720	0.296	1.211	0.047	26S proteasome non-ATPase regulatory subunit 4	PSMD4	
Cellular trafficking	Q9Y678	-	-	1.320	0.101	0.524	0.014	0.857	0.117	Coatomer subunit gamma-1	COPG1	DR2
	Q93084	-	-	1.200	0.469	0.533	0.042	-	-	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	ATP2A3	DR2
	Q8TEX9	-	-	1.211	0.047	1.000	1.000	1.286	0.039	Importin-4	IPO4	
	P05362	-	-	-	-	-	-	1.351	0.001	Intercellular adhesion molecule 1	ICAM1	
	Q9UNL2	-	-	-	-	-	-	0.619	0.040	Translocon-associated protein subunit gamma	SSR3	
	Q14974	-	-	1.313	0.259	1.028	0.839	1.154	0.004	Importin subunit beta-1	KPNB1	
	P02787	-	-	-	-	-	-	1.720	0.004	Serotransferrin	TF	
	O00410	-	-	-	-	-	-	1.768	0.007	Importin-5	IPO5	DR1, DR2
	Q8IV97	-	-	-	-	-	-	1.938	0.013	Solute carrier family 7	SLC7A5	
	P49792	-	-	-	-	-	-	1.316	0.013	E3 SUMO-protein ligase RanBP2	RANBP2	
	P28288	-	-	-	-	-	-	1.643	0.020	ATP-binding cassette sub-family D member 3	ABCD3	
	P51149	-	-	1.095	0.628	0.958	0.797	1.375	0.035	Ras-related protein Rab-7a	RAB7A	
	Q99829	-	-	-	-	-	-	1.313	0.038	Copine-1	CPNE1	
	O00151	-	-	1.364	0.387	1.500	0.187	1.308	0.040	PDZ and LIM domain protein 1	PDLIM1	

Table 1. (Pt. 2 of 3)

		MCF-7						HepG2				
		4 hrs		*24 hrs*		*72 hrs*		*72 hrs*				
	UniProt	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**Protein Name**	**Gene Name**	**RARE**
Metabolism	P23526	0.738	0.499	0.971	0.780	1.213	0.001	1.053	0.594	Adenosylhomocysteinase	AHCY	DR2
	O60701	0.605	0.332	1.018	0.745	0.751	0.012	0.934	0.163	UDP-glucose 6-dehydrogenase	UGDH	
	P51659	-	-	1.071	0.842	1.462	0.013	1.310	0.495	Peroxisomal multifunctional enzyme type 2	HSD17B4	
	P40925	0.674	0.329	1.375	0.019	1.063	0.447	1.015	0.725	Malate dehydrogenase, cytoplasmic	MDH1	DR2
	Q15181	-	-	0.938	0.875	1.933	0.026	1.229	0.264	Inorganic pyrophosphatase	PPA1	DR2
	Q9NR45	-	-	1.403	0.014	1.416	0.041	0.982	0.892	Sialic acid synthase	NANS	
	O43175	0.514	0.204	1.200	0.035	1.354	0.149	1.037	0.537	D-3-phosphoglycerate dehydrogenase	PHGDH	
	Q8NF37	-	-	-	-	0.682	0.036	1.000	1.000	Lysophosphatidylcholine acyltransferase 1	LPCAT1	
	P54819	-	-	1.435	0.042	1.529	0.058	1.031	0.649	Adenylate kinase 2, mitochondrial	AK2	
	Q06830	0.842	0.624	1.149	0.043	0.888	0.206	1.025	0.685	Peroxiredoxin-1	PRDX1	DR2
	P14324	-	-	-	-	0.476	0.052	1.524	0.010	Farnesyl pyrophosphate synthase	FDPS	
	Q99541	-	-	-	-	-	-	2.571	0.001	Perilipin-2	PLIN2	DR2
	P49327	0.542	0.287	1.080	0.482	1.071	0.459	1.271	0.002	Fatty acid synthase	FASN	
	P06737	-	-	-	-	-	-	2.462	0.006	Glycogen phosphorylase, liver form	PYGL	
	Q13011	-	-	0.979	0.891	1.061	0.535	1.275	0.006	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase	ECH1	DR5
	O95573	-	-	-	-	-	-	1.563	0.012	Long-chain-fatty-acid--CoA ligase 3	ACSL3	DR2
	P17174	-	-	-	-	-	-	2.118	0.014	Aspartate aminotransferase, cytoplasmic	GOT1	
	P49419	-	-	-	-	-	-	1.744	0.016	Alpha-aminoadipic semialdehyde dehydrogenase	ALDH7A1	DR2
	Q86SX6	-	-	-	-	-	-	2.267	0.016	Glutaredoxin-related protein 5, mitochondrial	GLRX5	
	P12268	-	-	1.097	0.535	0.690	0.152	1.500	0.019	Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	
	Q53HR2	-	-	0.722	0.283	1.500	0.365	1.682	0.019	Acyl-Coenzyme A dehydrogenase	-	
	P25705	-	-	-	-	1.188	0.639	1.560	0.022	ATP synthase subunit alpha, mitochondrial	ATP5A1	DR2
	Q8NBX0	-	-	-	-	-	-	1.357	0.024	Saccharopine dehydrogenase-like oxidoreductase	SCCPDH	
	O60488	-	-	-	-	-	-	1.389	0.025	Long-chain-fatty-acid--CoA ligase 4	ACSL4	
	P05091	-	-	-	-	-	-	1.247	0.033	Aldehyde dehydrogenase, mitochondrial	ALDH2	
	P48735	-	-	0.904	0.674	1.016	0.667	1.676	0.033	Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	
	P51648	-	-	-	-	-	-	1.889	0.030	Fatty aldehyde dehydrogenase	ALDH3A2	DR2
	P52209	-	-	1.214	0.326	1.253	0.141	1.449	0.035	6-phosphogluconate dehydrogenase, decarboxylating	PGD	DR2
	P00387	-	-	-	-	-	-	1.292	0.036	NADH-cytochrome b5 reductase 3	CYB5R3	
	P12277	-	-	-	-	-	-	1.380	0.040	Creatine kinase B-type	CKB	
	P80404	-	-	-	-	-	-	2.000	0.041	4-aminobutyrate aminotransferase, mitochondrial	ABAT	DR5
	P78417	-	-	-	-	-	-	1.348	0.047	Glutathione S-transferase omega-1	GSTO1	
	P50897	-	-	-	-	1.417	0.232	1.625	0.049	Palmitoyl-protein thioesterase 1	PPT1	
	P11586	0.403	0.208	1.289	0.363	0.889	0.558	1.282	0.050	C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	DR2
Cell cycle regulation	P42166	-	-	-	-	1.375	0.013	0.852	0.680	Lamina-associated polypeptide 2, isoform alpha	LAP2	DR2
	Q16890	-	-	1.100	0.538	0.688	0.024	-	-	Tumor protein D53	TPD52L1	
	P33991	-	-	1.192	0.298	0.690	0.028	1.457	0.135	DNA replication licensing factor MCM4	MCM4	
	Q9UK76	0.892	0.547	0.593	0.032	1.313	0.494	-	-	Hematological and neurological expressed 1 protein	HN1	
	P33993	-	-	0.973	0.770	0.681	0.045	0.941	0.680	DNA replication licensing factor MCM7	MCM7	DR2
	P49736	-	-	0.722	0.082	0.741	0.310	1.471	0.005	DNA replication licensing factor MCM2	MCM2	
	P18754	-	-	1.286	0.469	1.303	0.109	1.356	0.011	Regulator of chromosome condensation	RCC1	DR2
	Q9UBE0	-	-	-	-	-	-	1.818	0.029	SUMO-activating enzyme subunit 1	SAE1	DR2
	P62826	-	-	1.194	0.309	0.949	0.711	1.143	0.042	GTP-binding nuclear protein Ran	RAN	DR2
Cell Survival	Q13283	-	-	-	-	1.833	0.010	1.194	0.418	Ras GTPase-activating protein-binding protein 1	G3BP1	DR2
	O75874	-	-	1.390	0.040	2.304	0.011	1.210	0.087	Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	
	P55957	-	-	-	-	0.647	0.013	-	-	BH3-interacting domain death agonist	BID	
	P09382	-	-	0.773	0.430	0.615	0.022	0.952	0.649	Galectin-1	LGALS1	
	Q9Y265	-	-	-	-	1.357	0.024	1.000	1.000	RuvB-like 1	RUVBL1	
	P38646	0.741	0.504	1.054	0.576	1.149	0.045	1.212	0.049	Stress-70 protein, mitochondrial	HSPA9	DR2
	O00264	-	-	-	-	-	-	1.733	0.008	Membrane-associated progesterone receptor component 1	PGRMC1	DR2
	Q13442	-	-	0.913	0.727	1.030	0.815	1.786	0.017	28 kDa heat- and acid-stable phosphoprotein	PDAP1	DR2
	O76003	-	-	-	-	-	-	1.389	0.020	Glutaredoxin-3	GLRX3	
	P42765	-	-	-	-	-	-	1.727	0.020	3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	
	Q92597	-	-	-	-	-	-	1.415	0.024	Protein NDRG1	NDRG1	
	O95881	-	-	-	-	-	-	1.818	0.029	Thioredoxin domain-containing protein 12	TXNDC12	DR2
	P21741	-	-	-	-	-	-	1.308	0.047	Midkine	MDK	
Apoptosis	P56159	0.778	0.616	1.000	1.000	0.469	0.003	-	-	GDNF family receptor alpha-1	GFRA1	
	P10599	-	-	1.095	0.729	0.571	0.013	1.000	1.000	Thioredoxin	TXN	
	P21796	-	-	0.962	0.891	1.410	0.015	1.286	0.017	Voltage-dependent anion-selective channel protein 1	VDAC1	
	Q4JM47	-	-	0.955	0.553	0.661	0.048	-	-	Anterior gradient protein 2 homolog	AGR2	

Table 1. (Pt. 3 of 3)

		MCF-7						HepG2				
		4 hrs		*24 hrs*		*72 hrs*		*72 hrs*				
	UniProt	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**RA/Con**	**p-value**	**Protein Name**	**Gene Name**	**RARE**
Structural integrity	P15924	0.607	0.366	0.857	0.321	0.490	0.001	0.936	0.289	Desmoplakin	DSP	
	Q09666	0.732	0.653	0.753	0.192	0.116	0.008	1.649	0.191	Neuroblast differentiation-associated protein AHNAK	AHNAK	
	P37802	0.542	0.334	0.922	0.650	0.649	0.011	-	-	Transgelin-2	TAGLN2	DR2
	Q15149	-	-	0.907	0.533	0.580	0.014	0.857	0.328	Plectin	PLEC	DR1
	Q01082	-	-	1.081	0.708	1.712	0.016	1.028	0.895	Spectrin beta chain	SPTBN1	DR2
	P63313	-	-	0.765	0.345	0.679	0.033	0.727	0.101	Thymosin beta-10	TMSB10	
	Q9Y2B0	-	-	1.240	0.289	1.680	0.036	1.098	0.394	Protein canopy homolog 2	CNPY2	DR2
	Q8N392	-	-	-	-	-	-	0.519	0.006	Rho GTPase-activating protein 18	ARHGAP18	
	Q16658	-	-	-	-	-	-	2.643	0.011	Fascin	FSCN1	DR2
	P35579	-	-	1.125	0.414	1.964	0.058	2.619	0.028	Myosin-9	MYH9	DR2
	P67936	-	-	-	-	2.900	0.085	2.583	0.019	Tropomyosin alpha-4 chain	TPM4	DR2
	P46939	-	-	-	-	-	-	0.417	0.033	Utrophin	UTRN	
	Q9UDY2	-	-	-	-	-	-	1.375	0.033	Tight junction protein ZO-2	TJP2	DR2
	Q9HC35	-	-	-	-	-	-	0.698	0.039	Echinoderm microtubule-associated protein-like 4	EML4	
Misc.	P08758	-	-	1.222	0.070	0.806	0.013	1.122	0.311	Annexin A5	ANXA5	
	Q96CN7	-	-	1.098	0.742	0.489	0.024	0.722	0.232	Isochorismatase domain-containing protein 1	ISOC1	DR5
	Q9H3K6	-	-	1.042	0.819	1.526	0.043	1.385	0.303	BolA-like protein 2	BOLA2	
	P01024	-	-	-	-	-	-	1.400	0.005	Complement C3	C3	DR2, DR5
	Q99536	-	-	-	-	-	-	1.345	0.022	Synaptic vesicle membrane protein VAT-1 homolog	VAT1	

Identification of RARE sequences in gene promoters

To supplement the proteomics data, a Perl script was written to identify putative RARE sequences for each identified protein contained within their respective gene promoter region (defined as 10,000 base pairs upstream of the gene). The genomic data for each gene was captured and subsequently searched for direct repeats (DR) with a base pair spacing of 1 (DR1), 2 (DR2), and 5 (DR5) as defined by Laperriere et al.[135] The results, indicating putative RARE sequences, are listed alongside quantitative proteomics data in Table 1, Table 3 and Table 4.

3.4 Discussion

The underlying cellular mechanisms involved with atRA-mediated cellular response are of great interest to research and clinical communities. Prescribing retinoids in a broader application of chemotherapeutic regimens hinges on a more thorough comprehension of the divergent cellular response observed in specific tissue types. Additionally, understanding atRA-activated secondary protein pathways may lead to new drug targets. As established in the literature, it is apparent that the basal levels of CRABP-II and FABP5 determine the extent of PPAR-β/δ/RXR and RAR/RXR activated gene expression.[35] Detailing the resulting delicate balance of pro- and anti-apoptotic protein expression and activation will generate a more thorough comprehension of the tissue-specific responses to atRA. Here, I have demonstrated the utility of the data-independent acquisition method, PAcIFIC, and its ability to quantitatively categorize differences in protein abundance following atRA treatment. The CSI PAcIFIC method is well designed for high throughput proteomics analysis as they significantly reduce the overall analysis time of the PAcIFIC method.[91] Additionally, I find that this mass spectrometry-based approach is a much more thorough and sensitive technique for confidently and quantitatively cataloging changes in global protein abundance than the 2D-gel approaches that previous have been used to studied atRA-mediated changes in protein expression.[48, 49]

The CSI PAcIFIC dataset shown in Table 1 is the combination of results from three biological replicates, each run in technical duplicate. To qualify as a protein listed in Table 1, the identification must be found in all six proteomics analyses as well as have a Student's t-test p-value ≤ 0.05. The strict qualification and statistical filtering applied to this dataset ensure that only the most probable identifications are considered in pathway analysis. In addition, I developed a series of Perl scripts to identify putative RARE sequences. RAR/RXR heterodimers

have known affinity for DR2 and DR5 RARE sequences.[185] And, PPAR-β/δ/RXR heterodimers have known affinity for DR1 RARE sequences.[186] Although it should be noted that RXR/RAR heterodimers can bind to DR1 RARE sequences with reverse polarity to typical RAR/RXR binding, thereby inhibiting targeted gene expression.[187] The confounding possible effects of RAR/RXR and PPAR-β/δ/RXR binding require that the information presented in Table 1, Table 3, and Table 4 be evaluated with extreme care. While it is known that MCF-7 cells contain a greater amount of CRABP-II than FABP5 and therefore it can be assumed that PPAR-β/δ/RXR activation is minimal, the ratio of CRABP-II and FABP5 in HepG2 cells is unknown.[35]

In a life or death situation, the fine balance of pro- and anti-apoptotic signaling events mediated by atRA create a cellular environment where one route of this divergent response pathway must prevail. From the cell viability assay results illustrated in Figure 1, I observe that at 10 μM atRA MCF-7 cells exhibit a dramatic decrease in proliferation while HepG2 growth is not negatively impacted by the presence of a high concentration of atRA. The atRA-mediated effects in MCF-7 cells are well characterized in the literature.[133, 158] Here, our results confirm and expand the understanding of proteins linked to control of atRA-mediated MCF-7 cell cycle regulation, survival, and apoptosis. Following a treatment of atRA, MCF-7 cells undergo G_0/G_1 cell cycle arrest. Proteins involved in this process inhibit signaling mediated by the transcription factor E2F1 (E2F1), a critical protein pathway involved progression of the cell cycle. E2F1 and retinoblastoma-associated protein (Rb) form an activated complex that regulates initiation of DNA replication by controlling the expression of necessary protein machinery. Rb can be phosphorylated (pRb) by various kinases in the cell and pRb no longer can bind and activate E2F1 cell cycle progression. Our proteomics data suggests that LAP2, increased in abundance at 72 hours in MCF-7 cells, plays a key regulatory role in inhibition of the E2F1/Rb pathway.

LAP2 has a bi-functional inhibitory effect on Rb by directly binding to and inhibiting Rb by increased Rb phosphorylation. Both routes of Rb inhibition block E2F1 signaling and result in cell cycle arrest. The LAP2 gene promoter region also contains a DR2 RARE sequence and therefore increased expression at 72 hours may be a direct result to atRA-mediated gene expression. Additionally, the proteomics data presented here links two additional proteins of interest to atRA-mediated cell cycle control in MCF-7 cells. Both TPD53 and HN1 were observed in decreased abundance following treatment. HN1 was one of the few proteins to be found significantly down regulated at 24 hours and may suggest an early pathway involved in G_1 cell cycle arrest.[165]

Proteomic characterization of MCF-7 cells indicates a balance of pro- and anti-apoptotic factors involved in what appears to be a futile attempt at cell survival. Ras GTPase-activating protein (G3BP1) is up regulated in MCF-7 cells at 72 hours and contains a putative DR2 RARE sequence. Additionally, stress response proteins such as IDH1 are up regulated at both 24 and 72 hours and do not contain an RARE sequence indicating secondary pathway activation to the cellular environment triggered by atRA treatment. I also observed a decrease in the abundance of proteins involved in two important apoptosis pathways. First, BID, which activates BAX and induces ICE-like proteases that lead to the release of cytochrome C from the mitochondria is found in decreased abundance following atRA. Second, galectin is found in decreased abundance and is involved in promoting the E2F1/Rb pathway, a mechanisms that promotes advancement of the cell cycle.

These anti-apoptotic protein response pathways are counteracted by several quantified proteins that allow for pro-apoptotic signal cascades to result in the observed decrease in MCF-7 proliferation. Of note, I observe a decrease in abundance of three proteins normally important to

cell proliferation. For example, GRFA1, is protein that activates RET, which in turn activates pro-survival networks of MAPK/ERK and PI3K/Akt. Additionally, the proteomics data portrays a decrease in abundance of: TXN, which plays a role in TNF-α directed apoptosis; and AGR2, a protein that is found to correlate with cell growth and proliferation. The reduction in pro-survival proteins pathways is a factor that tips the balance of life and death in the direction of atRA-mediated apoptosis for MCF-7 cells.

In contrast, I observe a very different proteomics signature of quantified proteins found significantly changed in abundance due to atRA in HepG2 cells. While the levels of CRABP-II and FABP5 are unknown for this cell line, from the MTT assay and proteomics data I can assume that PPAR-β/δ/RXR gene expression is a prevalent means of cellular response. At the highest dose of 50 μM atRA, HepG2 cells show no change in cell proliferation (Figure 1). Additionally, PPAR-β/δ/RXR gene activation is known to induce expression of proteins involved with fatty acid metabolism. In the CSI PAcIFIC proteomics data, I observe increased abundance of 10 proteins involved in fatty acid metabolism and beta-oxidation.

I identified several proteins critical to understanding molecular mechanisms that enable HepG2 cells to continue to proliferate following atRA treatment. The proteomics data shows an increase in expression of 3-ketoacyl-CoA thiolase (ACAA2). ACAA2 inhibits pro-survival pathways by inhibiting BID and BAX. Additionally, TXNDC12 is increased in expression. Similar in functionality to TXN (identified in MCF-7 cells), increased expression of TXNDC12 allows for pro-survival signaling. It is apparent that atRA causes environmental stress to HepG2 cells. In conjunction with an up regulation of several stress response proteins and chaperones in Table 1, the TMT quantification data shown in Table 4 indicates that there are higher basal levels of several stress coping mechanisms and may indicate reasons why HepG2 cells are more apt at

avoiding atRA induced apoptosis. Additionally, qPAcIFIC analysis indicating higher basal levels of nucleophosmin (NPM1) and protein DJ-1 (PARK7) may provide reason for the maintained proliferation of HepG2 cell in high concentrations of atRA. Specifically, NPM1 protects cells from stress-induced apoptosis via negative regulation of ARF-p53 pathways. And, PARK7 promotes Akt-mediated cell survival signaling.

In summary, I demonstrate the application of the CSI PAcIFIC and qPAcIFIC method in the quantitative analysis of atRA-mediated protein alterations of MCF-7 and HepG2 cells. Time course analysis of MCF-7 provides information on early disparities in protein abundances between control and treated samples. And, protein quantification results obtained at 72 hours is a rich source of protein regulation data paramount to the cellular fate of MCF-7 and HepG2 cells following atRA treatment.

3.5 Conclusion and future directions

Mass spectrometry analysis is a key tool in discovery experiments for the global proteomics analysis of drug-induced cellular response. Unfortunately, application of cutting-edge proteomics acquisition techniques and instrumentation lags far behind development. Here, I have shown how data-independent mass spectrometry techniques can expand the understanding of molecular response mechanisms. This dataset provides great insight into protein components that are vital to atRA-mediated cellular response. In addition to bolstering published results, this data links numerous novel proteins with both RAR/RXR and PPAR-β/δ/RXR retinoid response pathways.

The advantage of mass spectrometry datasets over the dated electrophoresis techniques that founded our understanding of atRA-mediated changes in protein abundance is that they are a living database. As information about the functionality of proteins is discovered, a more detailed understanding of the identified pathways can be drawn. Additionally, it is possible to re-search the mass spectrometry data for additional protein identifications and post-translational modifications. The full meaning of signaling cascades is often reliant on the state of the protein (i.e. phosphorylated). Biochemical approaches should also be implemented to complement the proteomics data in understanding the activity of many proteins pathways discussed.

3.6 Notes to Chapter 3

Acknowledgements

Specials thanks to the Nina Isoherranen laboratory for their assistance with producing atRA treated cell samples, especially Ariel Topletz and Sam Arnold.

Chapter 4

MAPPING PARP-1 AUTO-ADP-RIBOSYLATION SITES BY LIQUID CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY

A portion of this body of work was done in collaboration with Jean-Philippe Gagné and Guy G. Poirier at the CHUQ Research Center (Laval University, Québec, QC, Canada) and is accepted for publication. (Chapman et al. *J Proteome Res* **2013**. Link: http://pubs.acs.org/doi/abs/10.1021/pr301219h)

4.1 Introduction

Poly(ADP-ribose) polymerase-1 (PARP-1) is an abundant chromatin-bound enzyme that catalyzes the assembly of polymeric ADP-ribose units linked by glycosidic ribose 1'–2' bonds.[188] Upon sensing DNA damage, activated PARP-1 cleaves nicotinamide adenine dinucleotide (NAD+) into nicotinamide and ADP-ribose, which in turn initiates a series of transfer reactions onto proteins involved in a variety of DNA transactions.[189] The initial ADP-ribosylation step is followed by the subsequent addition of ADP-ribose moieties to form the elongated and branched poly(ADP-ribose) (pADPr) chains. PARP-1 is a major target of the poly(ADP-ribosyl)ation reaction as it automodifies itself by intra molecular (*cis*) and inter molecular (*trans*) ADP-ribose transfer reactions.[190-192] According to the current model, the accumulation of pADPr at DNA strand breaks causes local remodeling of chromatin structure and promotes the recruitment of DNA repair and signaling factors.[193-196] In this regard, pADPr is a key regulator of the DNA damage response and cell signaling pathways.

For more than 40 years, the establishment of an ester linkage between the carboxyl group of either a glutamate or an aspartate residue has been the leading model for covalent attachment of pADPr to target proteins.[197] The underlying model for understanding covalent poly(ADP-ribosyl)ation by PARP-1 was based on the instability of the ester linkage to neutral hydroxylamine treatment and its alkali lability. However, the exact modification sites have never been rigorously demonstrated.[197] Through a series of experiments, it was demonstrated that

hydroxylamine-labile ADP-ribose bonds of PARP-1 were of a similar type of carboxyl ester as in mono-ADP-ribosylated histones.[198-201] According to a detailed study of the automodification biochemistry, PARP-1 was estimated to bear as many as 15 attachment sites for pADPr polymers.[202] Recently, this carboxyl linkage model has been called into question with studies reporting the identification of lysine residues as predominant acceptors of pADPr on PARP-1.[191] The identification of lysine as an acceptor site on PARP-2 and histone tails also challenges the concept of poly(ADP-ribosyl)ation via an ester linkage.[203, 204] In view of these unexpected results, we need to reconsider the assumption that the pADPr acceptor is solely the carboxyl side chain of a glutamate/aspartate residue. However, the other possibility for covalent linkage, a lysine-directed poly(ADP-ribosyl)ation mechanism, is inconsistent with several studies published by leading research groups over the last decades. Still, heterogeneous hydroxylamine sensitivity was reported for histones, i.e., hydroxylamine-resistant bonds were detected, suggesting that poly(ADP-ribosyl)ation may occur on amino acid acceptors other than glutamate and aspartate. In addition, one cannot exclude the possibility that the histones used at that time, which were extracted from animal tissues, could have been lysine-acetylated, a modification that would have impaired the subsequent poly(ADP-ribosyl)ation on those sites, thus creating a bias towards glutamate and aspartate residues.[199] To add to this confusion over which amino acid may accept pADPr modification, Adamietz & Hilz reported that hydroxylamine treatment liberates only part of the long and the short pADPr residues from protein conjugates, indicating two types of bonds – both of which are alkali-labile, but only one of which is susceptible to cleavage via neutral hydroxylamine.[205] Altogether, these studies have led to confusion in the literature as to which amino acid side-chains are actually targeted by poly(ADP-ribosyl)ation reactions. The identification of post-translationally modified residues is not only important to clarify the nature

of poly(ADP-ribosyl)ation on nuclear substrates and PARP-1 itself, but is also crucial to understanding the crosstalk between poly(ADP-ribosyl)ation and other post-translational modifications.[206, 207]

To begin to answer the question of which amino acid residue or residues PARP-1 targets for post-translational modification, it is clear that definitive evidence demonstrating the amino acid site of pADPr protein modifications is needed. Seeking clarification, many mass spectrometry methods have aimed to solve the problem of localizing pADPr modified amino acids, but given the size and branched nature of pADPr this has proven to be an extremely challenging task for multiple reasons. For example, several groups reported there is a loss of peptide sequencing information when collision induced dissociation (CID) is used as the peptide activation method for fragmentation.[208-212] The resulting tandem mass spectrum, which still contains evidence of the presence of pADPr, is virtually devoid of peptide fragment ions that are required for sequence assignment and determination of the modified residue. Additionally, the heterogeneity of the structure of pADPr produces a vast number of different pADPr species, complicating the task of monitoring this protein modification based on mass assignments. However, there are few successful reports of mass spectrometry methods capable of localizing pADPr protein modifications. These reports minimize the complexity of fragmentation in the tandem mass spectrum using the PARP-1 mutant E988K, which is incapable of catalyzing elongation or branching polymerization reactions found in pADPr.[213, 214] The result is a single ADP-ribose unit covalently linked to an amino acid residue on PARP-1. Both CID and electron transfer dissociation (ETD) activation have shown this to be a successful strategy, but these methods have little relevance to native PARP-1 which is fully capable of polymerizing extended and heterogeneous pADPr.[211, 215] On a similar theme, Messner *et al*. recently documented an

enzymatic approach to simplify pADPr prior to mass spectrometry analysis utilizing ADP-ribosylhydrolase 3 (ARH3).[204] This enzyme was used in combination with ETD activation to successfully identify sites of histone modifications, but as will be discussed herein, this approach leaves room for further optimization for the robust high-throughput identification of pADPr sites.

Here a novel approach is described to sequence peptides modified with pADPr and locate the site of modification. Using tandem mass spectrometry and additional biochemical techniques, it is shown that it is possible to determine the nature of the pADPr linkage found in PARP-1 and to identify key residues participating in this linkage. Specifically, samples were enriched for the ribose-5'-phosphate-containing peptides, which are remnants of phosphodiesterase-treated automodified PARP-1, to provide direct evidence that lysine, glutamate and aspartate residues are actual targets of the ADP-ribose linkage. The simplified pADPr remnant allows for identification using CID activation and widely available search algorithms, meaning that this method is readily amenable to almost all mass spectrometry laboratories with an established proteomics pipeline.[216] A significant challenge that lies ahead will be the global identification of sites of pADPr from cell culture samples or tissue homogenates. The proof-of-concept demonstrated here is immediately applicable to such complex biological samples in a high-throughput manner.

4.2 Experimental procedures

Unless otherwise noted, all chemicals and solvents were purchased from Sigma Aldrich (St. Louis, MO, USA).

Peptide polymer-blot analysis

The 1014 amino acid sequence of human PARP-1 (Uniprot accession code P09874) was used to generate an overlapping peptide library. The library is composed of 20-mers peptides with 5 amino acid off-set (68 peptides). Peptides were synthesized on four identical cellulose membranes by SPOT technology using a MultiPep RS peptide arrayer (Intavis, Chicago, IL). The membranes were first rinsed with a small volume of methanol for 5 min to avoid precipitation of hydrophobic peptides during the subsequent procedures. The membranes were pre-conditioned for 1h in a solution composed of 10 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.1% Tween-20. The first membrane was incubated for 1h in a reaction buffer [100 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10 mM DTT, 100 μM NAD and 10 μg/mL activated DNA (calf thymus DNA nicked with DNase, Sigma)] containing 50 μCi of [^{32}P]-NAD (800Ci/mmol, Perkin Elmer). The second membrane was incubated in the same reaction buffer containing 50 μCi of [^{32}P]-NAD and 10 μg of purified human recombinant PARP-1 E988K mutant (Enzo Life Sciences) while the third was incubated with 10 ug of human recombinant wild-type PARP-1 (Enzo Life Sciences). The last membrane was incubated in the reaction buffer containing 250 nM of Dihydroxyboronyl Bio-Rex (DHBB)-purified [^{32}P]-pADPr synthesized as described in [217]. The membranes were rinsed several times with the pre-conditioning solution containing 1M NaCl until no radioactivity could be detected in the washes. Subsequently, the membranes were air-dried and subjected to autoradiography.

Mono-ADP-ribosylation of a prototypic peptide standard

A peptide standard (AVKKLTVNPG) was mono-ADP-ribosylated using the PARP-1 mutant E988K with the following buffer: 100 mM Tris-HCl (pH 7.4 or 8.2), 10 mM $MgCl_2$, 10 mM DTT, 1 mM NAD, 10 % ethanol and 25 µg/mL activated DNA. The peptide was purchased from AnaSpec (Fremont, CA) with amide C-terminal protection and acetyl N-terminal protection at an HPLC verified purity of at least 90%. For each peptide reaction, 30 µg of the peptide standard was reacted in reaction buffer for 30 min at 30 °C. For direct analysis of the mono-ADP-ribosylated peptide, 10% formic acid (FA) was used to acidify the reaction solution to stop PARP-1 activity. To prepare the PDE-treated peptide standard following mono-ADP-ribosylation, 5 mU of purified PDE from *Crotalus adamanteus* venom (Sigma) was added to the reaction and incubated for 3h at 37 ^{o}C.

Phosphodiesterase (PDE) purification

PDE was purified according to the method of Oka *et al*. [218] Briefly, 0.5 U of PDE (based on enzymatic bis-(p-nitrophenyl) phosphate hydrolysis) from *Crotalus adamanteus* (Sigma) was dissolved in 5 mM KPO_4 buffer (pH 7.5) containing 50 mM NaCl and applied to a Blue Sepharose (Cibacron Blue F3GA) (GE Healthcare Life Sciences, Piscataway, NJ) column (2 mL resin). The column was equilibrated with the PDE resuspension buffer and the PDE was allowed to enter the column by gravity. The column was washed twice with 10 mL of 10 mM KPO_4 buffer containing 50 mM NaCl. PDE was eluted by elevating the KPO_4 concentration to 30 mM.

PARP-1 automodification and PDE digestion

Ten micrograms of highly active human PARP-1 (Tulip Biolabs, West Point, PA) were automodified in 100 µL of a reaction buffer containing 100 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10 mM DTT, 25 µM NAD and 25 µg/mL activated DNA for 30 min at 30 °C. After the automodification reaction, 50 µL of purified PDE was added to digest PARP-1-bound pADPr (2h at 37 ^{o}C). Digests were resolved by SDS-PAGE using 4-12% Criterion™ XT Bis-Tris gradient gels (Bio-Rad) and stained with SYPRO Ruby protein gel stain according to the manufacturer's instructions (Bio-Rad). Images were acquired using the Geliance CCD-based bioimaging system (PerkinElmer). For Western blots, extracts were transferred onto nitrocellulose membrane and probed with anti-pADPr antibody, clone 96-10 (1:10 000).

Thin layer chromatography (TLC)

TLC assays were performed essentially as described in [219]. Briefly, 1 nmole of [^{32}P]-labeled pADPr was incubated with increasing amounts of PDE for 1h at 37^{o}C. PDE activity was measured by analyzing the production of PRAMP. PEI-F (polyethyleneimine F) cellulose (Macherey-Nagel) TLCs were developed in 0.3 M LiCl and 0.9 M acetic acid according to Ménard and Poirier to separate pADPr from PRAMP. [219] TLCs were air-dried and autoradiographied.

pADPr immunoprecipitation

The anti-pADPr antibody, clone 10H, coupled to Protein-G coated Dynabeads™ (Life Technologies) was used to isolate pADPr-containing protein complexes following PARP activation as described in [220]. After the final washes, the pADPr-containing beads were digested

with 5 mU of PDE in the presence of Complete™ protease inhibitors (Roche) for 2h at 37°C. Digests were resolved by SDS-PAGE using 4-12% Criterion™ XT Bis-Tris gradient gels (Bio-Rad), transferred onto nitrocellulose membrane and probed with anti-pADPr antibody, clone 96-10 (1:10 000).

In-gel tryptic digestion

Excised gel bands were washed in 1 mL of 150 mM ammonium bicarbonate followed by 1 mL acetonitrile (ACN). These washing steps were repeated two additional times. Following the final wash, samples were evaporated in a SpeedVac (Thermo Savant, San Jose, CA) for 45 minutes. Twenty micrograms aliquots of sequencing grade porcine trypsin (Promega, Madison, WI) were reconstituted in 1 mL of 150 mM ammonium bicarbonate and placed on ice. Each dried gel band was submerged in the trypsin solution and placed on ice for 45 minutes. After swelling, it was ensured that each gel band was fully submerged in the trypsin buffer, and then gently mixed at room temperature overnight. Peptides were first extracted in 5% ACN/0.1% formic acid (FA) followed by 80% ACN/0.1% FA. Extracts were pooled and dried to near completion.

Titanium dioxide enrichment

Samples were reconstituted in 5 µL of 0.1% trifluoroacetic acid (TFA) and 25 µL of 6%TFA/80%ACN. Twenty microliters of a 10 mg/mL water solution of TiO_2 (ZirChrom Separations, Inc.) beads was added to a homemade gel-loading tip that contained a glass fiber frit. Gentle air pressure supplied from a syringe was applied to speed the flow of solvent and beads were always left wet. Ten microliters of 6%TFA/80%ACN was used to wash the beads. Sample

was passed through the column three times and the final elute was collected. Two wash steps were each performed twice: first with 10 μL of 6%TFA/80%ACN followed by 10 μL of 0.1% TFA. Each wash elute was collected. Ten microliters of 5% ammonium hydroxide was used to elute the capture peptides. Sixty microliters of 2% FA was used to immediately adjust the final elution pH. All samples were dried to near completion and reconstituted in 5% ACN/0.1% FA for mass spectrometry analysis.

Iron metal-affinity capture enrichment

Samples were reconstituted in 450 μL of 50% ACN/water/0.1% TFA. IMAC (Phos-Select, Sigma) resin was prepared by first resuspending in 50% ACN/water/0.1% TFA to make a 10% v/v slurry; spinning at 2000 rpm for 30 seconds and finally adding 50%/ACN/water/0.1% TFA to the pellet to make a 10% v/v slurry. Fifty microliters of resin was added to the sample and gently rocked for 30 minutes. Slurry was then transferred to a constricted gel-loading tip to capture the resin. Resin was washed with 50 μL of 50% ACN/water/0.1% TFA. Twenty microliters of 0.4 M NH_4OH was used to elute the peptides from the resin. All samples were then prepared for mass spectrometry analysis and reconstituted in 5% ACN/0.1% FA.

Mass spectrometry analysis

Nano-HPLC was performed using a Waters NanoAquity (Milford, MA). A homemade trapping column was made from 100 μm i.d. capillary (Polymicro Technologies, Phoenix, AZ) packed with 2 cm of 200 Å, 5 μm Magic C18AQ particles (Michrom, Auburn, CA). Successive analytical separation was performed on a homemade gravity-pulled 75 μm i.d. (Polymicro Technologies, Phoenix, AZ) column packed with 15 cm of 100 Å, 5 μm Magic C18AQ particles

(Michrom, Auburn, CA). Peptide samples were loaded on the precolumn at 4 μL/min with 95% water/5% ACN/0.1% FA. Trapped peptides were then eluted from the trapping column onto the analytical column using a variable gradient with a flow rate of 0.25 μL/min. The gradient utilized two mobile phase solutions: A, water/0.1% FA; and B, ACN/0.1% FA. The variable gradient used is as follows: 0 minutes, A (95%), B (5%); 55 minutes, A (65%), B (35%); 65 minutes, A (10%), B (90%); 75 minutes, A (10%), B (90%); 80-100 minutes, A (95%), B (5%). Peptide digests were analyzed on a LTQ Orbitrap (Thermo Fisher, San Jose, CA) by nano-electrospray ionization in positive ion mode. Ion source conditions were optimized using the tuning and calibration solution suggested by the instrument provider. Precursor ion scans were performed from 400-2000 *m/z* in the Orbitrap analyzer at a resolution of 30,000. Tandem MS analysis was performed in the linear ion trap using an isolation window of 2 Thompsons and a normalized collision energy of 35% for all fragmentation events. Seven tandem MS scan events were processed per precursor ion scan. Singly charged ions were excluded from analysis and dynamic exclusion was enabled with a repeat count of 1, an exclusion duration of 45 seconds and an exclusion list of 250. Each sample was analyzed in triplicate.

Search results and validation

Data acquired on the LTQ Orbitrap was converted from Thermo's RAW format to the universal mzXML format and searched against the IPI human database v3.49 (http://www.ebi.ac.uk/IPI/IPIhuman.html) using SEQUEST.[216] Search parameters for SEQUEST included a precursor ion tolerance of 2.1 Da, trypsin enzyme specificity, the variable option for methionines in reduced or oxidized form and the variable PDE cleavage product modification on glutamate, aspartate and lysine residues (212-Dalton ribose-5'-phosphate tag). Results were

analyzed with Peptide Prophet ensuring that peptide hits with a probability of >0.99 were accepted and linked to protein entries.[126, 221] All possible poly(ADP-ribosyl)ation sites were verified manually.

4.3 Results

Trans ADP-ribosylation assays on PARP-1 peptide arrays

The current prevailing model of human PARP-1 automodification is based on combined *cis* and *trans* activation at sites of DNA damage. As an initial screening strategy for the identification of PARP-1 *trans*-automodification sites, we used an autoradiography PARP-1 peptide array. Our intent was twofold: (1) to catalog potential *trans*-poly(ADP-ribosyl)ated peptides; and (2) to identify a peptide with a high stoichiometric ratio of ADP-ribosylation that can be used to generate a high yield, PARP-1 catalyzed mono-ADP-ribosylation standard for further mass spectrometry studies described below. To complete the peptide array, peptides were synthesized, twenty amino acids in length, along the full length PARP-1 sequence and designed to include a five amino acid overlap of each serial peptide (Figure-1A). Membranes were incubated with [^{32}P]-NAD and PARP-1 to reflect *trans*-poly(ADP-ribosyl)ation modifications. Although synthetic peptides are taken out of the context of a protein, peptide arrays have been shown to be valuable tools to localize a broad range of post-translational modification onto target proteins.[222] Notably, it has been shown that *trans*-poly(ADP-ribosyl)ation of different fragments of PARP-1 by wild-type PARP-1 occurs *in vitro*.[191] Thus, it was hypothesized that some peptides could mimic structural features of the PARP-1 protein and be recognized as *bona fide trans*-ADP-ribosylation sites.

Four identical PARP-1 peptide arrays were used to perform the assay. One membrane was incubated with the reaction buffer lacking PARP-1 to account for non-specific binding to [^{32}P]-NAD. As can be seen in Figure-1B it was confirmed that there was minimal non-specific binding to NAD throughout the membrane. Only peptide #67 strongly binds [^{32}P]-NAD, an observation that is consistent with the fact that this peptide corresponds to the NAD binding

pocket of PARP-1 in its active site (see Supplementary information for peptide listing). This result also suggests that the peptides on the membrane retain certain primary and secondary structure characteristics needed for biological activity. Next, the *trans*-ADP-ribosylation patterns on peptide arrays obtained with wild-type PARP-1 (Figure-1C) were compared with those of the E988K mutant (Figure-1D) that has enzymatic activity limited to producing mono-ADP-ribosylation. As expected, the modification patterns produced by wild-type PARP-1 and the E988K mutant PARP-1 were very similar. Specifically, we found a high level of qualitative correlation for the mono-ADP-ribosylated peptides for almost all positions that corresponded to poly(ADP-ribosyl)ation. Notably for several positions, the signal was stronger when using wild-type PARP-1, which can be explained by the incorporation of additional [^{32}P]-labeled ADP-ribose moieties in contrast to the E988K mutant that lacks elongation activity (e.g. peptides #16, #28 and #42). Thus, on a global basis, this *in vitro* assay demonstrates that the E988K mutant is capable of recapitulating poly(ADP-ribosyl)ation by wild-type PARP-1. However, in these assays, PARP-1 also undergoes automodification through mono and poly(ADP-ribosyl)ation. Thus, in order to rule out the possibility that radioactive labeling of specific peptides was a consequence of non-covalent interactions with PARP-1-bound pADPr, an additional membrane was probed with dihydroxyboryl Bio-Rex (DHBB)-purified (protein-free) [^{32}P]-pADPr (Figure-1E). Even though membranes were extensively washed under stringent conditions to reduce non-specific ionic and electrostatic interactions, several peptides strongly bound [^{32}P]-pADPr, generating some confusion over whether a peptide is covalently *trans*-poly(ADP-ribosyl)ated or interacting in a non-covalent fashion with pADPr. Only peptide #44 was strongly labeled in wild-type PARP-1 and E988K ADP-ribosylation assays while also showing very weak affinity for the pADPr itself in the non-covalent binding assay. Peptide #44 corresponded to the C-

terminus of the WGR domain of PARP-1. Although not yet clearly elucidated, this result shows that PARP-1 *trans*-ADP-ribosylation activity can minimally target a residue located within peptide #44. It was concluded, therefore, that the sequence contained in peptide #44 will serve well as a high yield peptide target for the synthesis of a PARP-1 catalyzed ADP-ribosylation standard. Importantly, these peptide array-based results underscored the need for more conclusive methods to localize PARP-1 regions targeted by its poly(ADP-ribosyl)ation activity based on the difficulty in interpreting the difference between covalent and non-covalent pADPr interactions.

CID fragmentation of a mono-ADP-ribosylated peptide - exploiting phosphodiesterase activity

Current literature demonstrates that there are a number of mass spectrometry fragmentation techniques that are not well suited for the analysis of poly(ADP-ribosyl)ated peptides.[208-212] Resulting tandem mass spectra of poly(ADP-ribosyl)ated peptides are often devoid of sufficient peptide fragment ion data to allow sequence assignment and determination of which residue(s) is modified.[209, 223] To explore the CID fragmentation efficiency of a PARP-1 catalyzed pADPr peptide standard, a prototypic peptide was synthesized that defines a PARP-1 region that is likely poly(ADP-ribosyl)ated based on the above results from the autoradiography peptide assay. We selected a peptide sequence that corresponds to the central region of peptide #44 on the polymer-blot (Figure-1C, D). Since peptide #44 is located between two overlapping peptides (#43 and #45) that are not substrates of PARP-1, the targeted region of peptide #44 can be restricted to its non-overlapping central region defined by the 10 amino acid peptide sequence AVKKLTVNPG (Supplementary information).

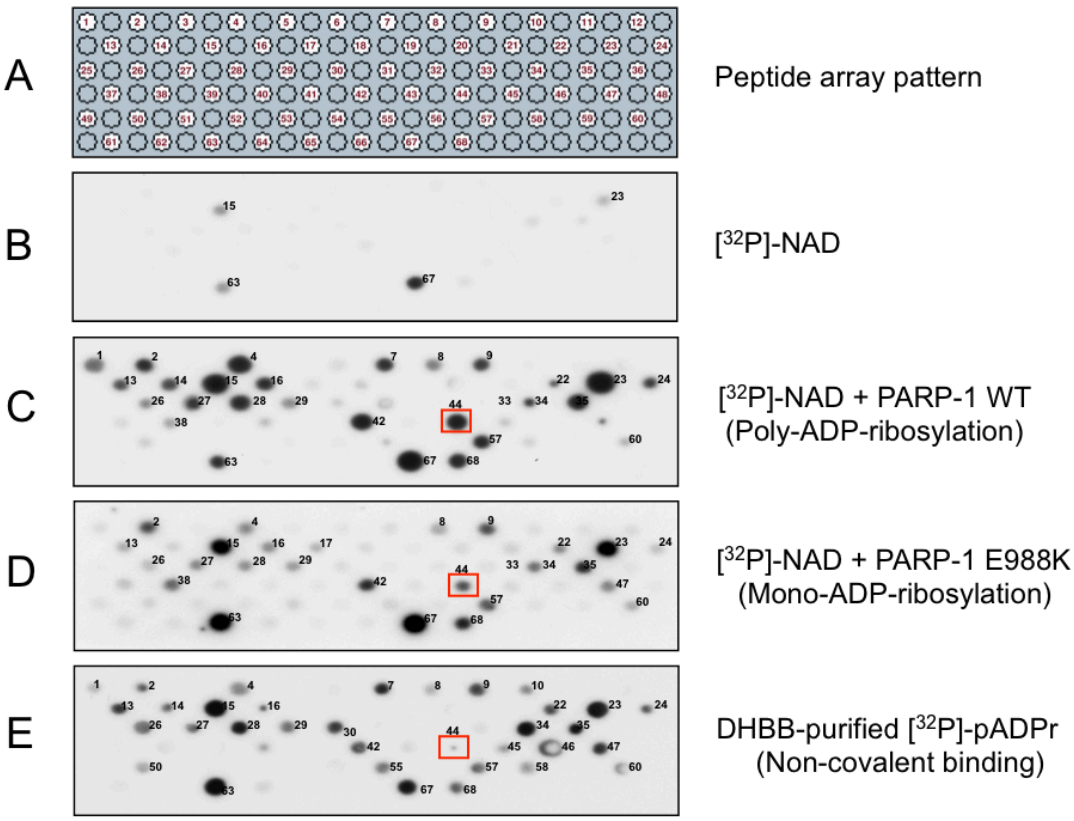


Figure-1. Peptide array-based identification of PARP-1 *trans*-automodification sites. (**A**) Overlapping PARP-1 peptides (20-mers peptides with five amino acid off-set) covering the full length human PARP-1 sequence were arrayed on a cellulose membrane using SPOT synthesis. An identical set of membranes were prepared according to the peptide array pattern shown in the top panel. (**B**) A membrane incubated with [^{32}P]-NAD to account for non-specific binding. (**C**) Poly(ADP-ribosyl)ated regions identified by incubating a membrane with [^{32}P]-NAD and the fully active wild-type PARP-1. (**D**) A screen for mono-ADP-ribosylated regions of PARP-1 was performed using [^{32}P]-NAD and the PARP-1 mutant E988K that lacks ADP-ribose elongation activity. (**E**) Finally, non-covalent pADPr-binding was assayed by incubating a peptide array with DHBB-purified [^{32}P]-labeled pADPr. For each experiment, membranes were extensively washed under stringent ionic conditions, air-dried and autoradiographed. The prototypic peptide #44 is boxed in red (see text).

Many research groups employ PARP-1 E988K to simplify biochemical and mass spectrometry analyses to mono-ADP-ribosylation modifications.[211, 215] Here, PARP-1 E988K was employed in the production of the mono-ADP-ribosylated version of peptide AVKKLTVNPG. Mass spectrometric analysis revealed that the prototypic peptide AVKKLTVNPG, which does not contain a glutamate or aspartate residue and is protected on the N- and C-terminus, is indeed covalently ADP-ribosylated by the PARP-1 E988K mutant, most likely on a lysine residue. First, as can be seen in Figure-2A, CID fragmentation of the unmodified peptide AVKKLTVNPG provides ample fragment ions for proper identification. However, after being mono-ADP-ribosylated by PARP-1 E988K, the CID tandem mass spectrum of the resulting peptide is virtually devoid of any peptide fragment ions and the two predominant fragment ions occur within the modification itself. These two fragment ions are represented by the notation p5 and m6 (Figure-2B), according to the nomenclature of Hengel et al.[208] In parallel, while many phosphopeptides exhibit facile loss of phosphate, they also often produce tandem mass spectra that allow for sequence assignment. Thus, it was hypothesized that phosphodiesterase (PDE) treated pADPr-modified peptides should behave similarly. Treatment of a pADPr-modified peptide with PDE should produce a mono-ADP-ribose modified peptide with a terminal ribose-5'-phosphate moiety thereby eliminating the most CID-labile bonds (Figure-2D and 3) from the pADPr structure. To test this hypothesis, the prototypic peptide was first modified by PARP-1 E988K and then treated with PDE to truncate the mono-ADP-ribose structure to ribose-5'-phosphate (Figure-2D). As demonstrated in Figure-2C, following CID fragmentation of the ribose-5'-phosphate modified peptide, the subsequent tandem mass spectrum contains the peptide fragment ions necessary for proper peptide identification.

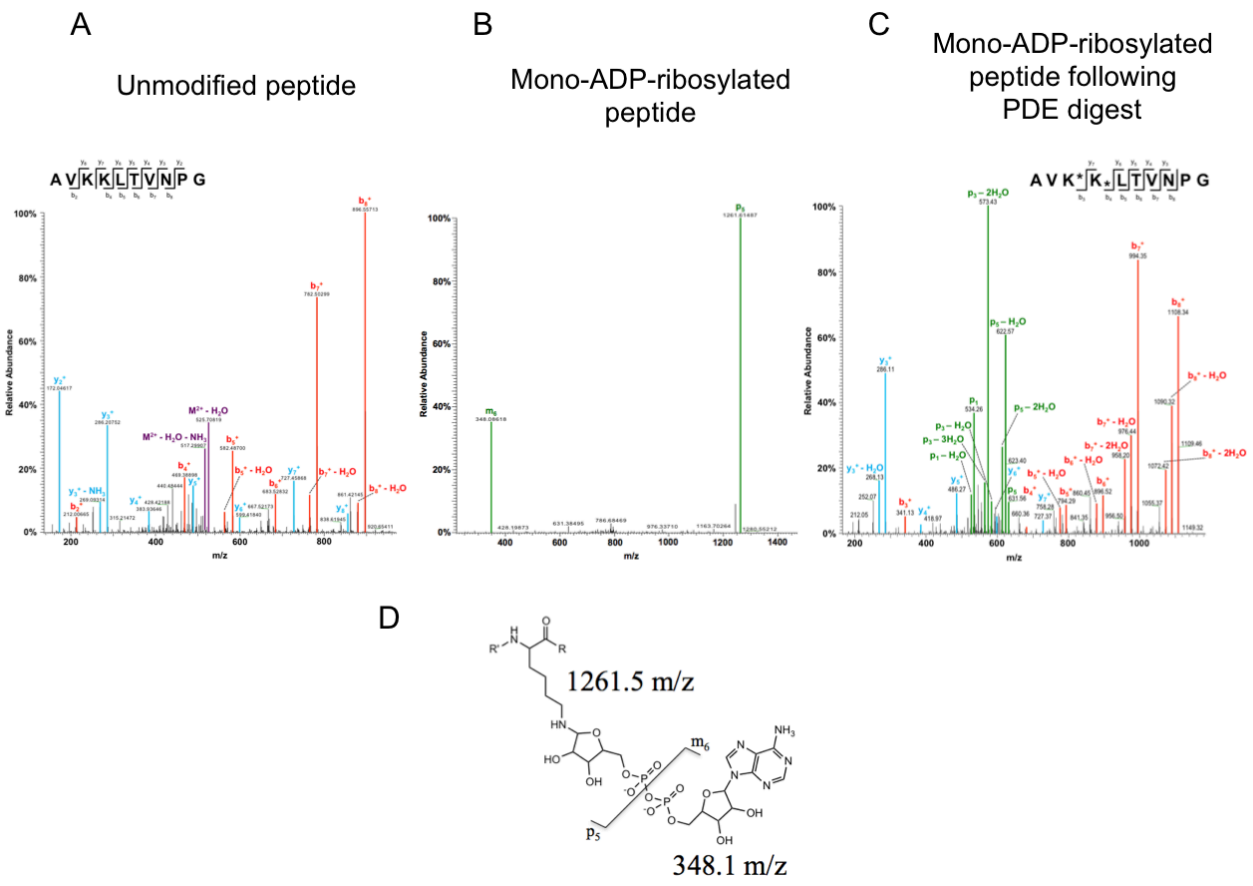


Figure-2. **Tandem mass spectral behavior of ribose-5'-phosphate peptides.** Tandem mass spectra of the prototypic peptide AVKKLTVNPG were acquired by collision induced dissociation (CID) as follows: (**A**) unmodified peptide (534.34 *m/z*, $z = 2^+$), (**B**) mono-ADP-ribosylated by PARP-1 E988K (804.87 *m/z*, $z = 2^+$) and (**C**) after PDE digestion of the mono-ADP-ribosylated peptide (640.34 *m/z*, $z = 2^+$). Matching *b*-ions are shown in red, *y*-ions are shown in blue, ion species representing fragmentation within the modification are shown in green,[21] and doubly charged ions of the peptide are shown in purple. The asterisk denotes the potential ribose-5'-phosphate-modified residues (see text). A structural schematic of the sources of fragmentation ions from the ADP-ribosylated peptide is shown in (**D**) with their respective *m/z* values. The line denotes the site of CID fragmentation and the structure of the product ions p_5 and m_6.

The fragmentation of the ribose-5'-phosphate-modified peptide AVKKLTVNPG yielded a tandem mass spectrum with a *b*-ion series pointing to a modification at one lysine, and a *y*-ion series pointing to a modification at the second lysine residue. For this peptide, non-ambiguous identification of the exact poly(ADP-ribosyl)ation site cannot be determined from the tandem mass spectrum to favor one site over the other because the fragment ions observed in the tandem mass spectrum are likely the total sum of fragment ions from both species. It should be noted that the tandem mass spectrum contains a population of product ions representing fragmentation of the ribose-phosphate modification. The series of ribose-phosphate fragment ions are flanked by ions representing fragmentation of a phosphate-oxygen bond (p5) and fragmentation of the entire ribose-5'-phosphate moiety resulting in an ion of the non-modified peptide (p1) (Figure-2C). Nevertheless, we confirm the basic hypothesis that converting a pADPr-modified peptide to a simpler ribose-5'-phosphate-peptide by PDE treatment allows for sequence assignment of the peptide. Here, we can conclude that PDE treatment of the ADP-ribose structure reduces the number of facile CID-labile bonds in the pADPr structure thereby allowing for adequate peptide fragmentation by CID to assign peptide sequence and localize the site of modification.

PARP-1 automodification and phosphodiesterase treatment

While the study of mono-ADP-ribosylated peptides by tandem mass spectrometry is complicated as described above, the study of poly(ADP-ribosy)lated peptides is nearly impossible. This is primarily due to the heterogeneity and size of the pADPr polymers that confounds the use of mass spectrometry. For example, the heterogeneous nature of pADPr and resulting complex CID fragmentation patterns hamper the traditional workflows of high-throughput proteomic analyses.[208] However, mass spectrometry-based proteomic approaches

have been successfully developed to identify structurally similar post-translational modifications such as glucuronidation and phosphorylation.[224, 225] Additionally, the low stoichiometric ratio of modified to unmodified peptides in a sample necessitates enrichment , as is the case for phosphorylation and acetylation.[226, 227]

The pADPr can be broken down into more simple constituents by specific glycohydrolases such as PARG [Poly(ADP-ribose) glycohydrolase] and ARH3. Both enzymes possess pADPr-degrading activity that yields monomers of ADP-ribose as final reaction products. While these enzymes can degrade complex polymers with high specificity, both also remove the last ADP-ribose moiety from proteins, erasing any evidence of pADPr. In contrast, when the pADPr is digested by a snake venom PDE, which targets the pyrophosphate linkage, one molecule of ribose-5'-phosphate is still attached to the acceptor protein (Figure-3).[228] Since we demonstrated that the ribose-5'-phosphate is compatible with CID by generating an interpretable mass spectrum (Figure-2C), this remnant was targeted as a signature to identify sites of poly(ADP-ribosyl)ation that occurs whether by *cis*- or *trans*-ADP-ribosylation.

In order to assess the efficiency of PDE to degrade complex pADPr, an experiment was conducted to monitor the release of phosphoribosyl-AMP (PRAMP) (Figure-3), the main reaction product generated following the digestion of DHBB-purified [^{32}P]-pADPr by PDE (Figure-3). Using a common TLC assay used to develop PARG reaction mixtures, one can observe a complete conversion of pADPr to PRAMP (Figure-4A)[219] Because the goal was to localize the poly(ADP-ribosyl)ation sites onto PARP-1, the differences in SDS-PAGE mobility of automodified human recombinant PARP-1 also were monitord before and after PDE treatment (Figure-4B, top panels). Automodification of PARP-1 causes severe protein smearing as the protein is covalently modified with polymers of ADP-ribose ranging from small oligomers to

long and branched molecules. The automodification was shown to be nearly complete because only traces of unmodified PARP-1 were observed in the presence of an excess of NAD. After incubation with PDE, most of these polymers are efficiently hydrolyzed as PARP-1 reappeared at the original position relative to the control unmodified PARP-1. To further examine the structural state of the polymer, the digestion products following PDE digests were analyzed by Western blot using the anti-pADPr polyclonal antibody 96-10. As expected, most of the PARP-1-bound pADPr is hydrolyzed (Figure-4B, lower right panel). Finally, the ability of the PDE to digest a broad range of pADPr polymers synthesized in cells after PARP activation (Figure-4C) was tested. HEK 293 cells were exposed to 100 μM of the DNA alkylating agent N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) for 5 minutes and pADPr-associated complexes were immunoprecipitated from whole cells extracts. As shown in Figure-4C, the immunoprecipitate is highly enriched in pADPr. The pADPr formed in a physiological context is also easily hydrolyzed by PDE. It is concluded that a wide range of pADPr molecules, either free or protein-bound, are readily hydrolyzed by the action of PDE. These Western blots also show that control PARP-1 is not poly(ADP-ribosy)lated as it will serve as a negative control in the mass spectrometry analysis. In addition, resolving the reaction products on SDS-PAGE serves as a filtering step to eliminate 5'-adenosine monophosphate (5'-AMP) and PRAMP from the modified PARP-1, containing the ribose-5'-phosphate remnant.

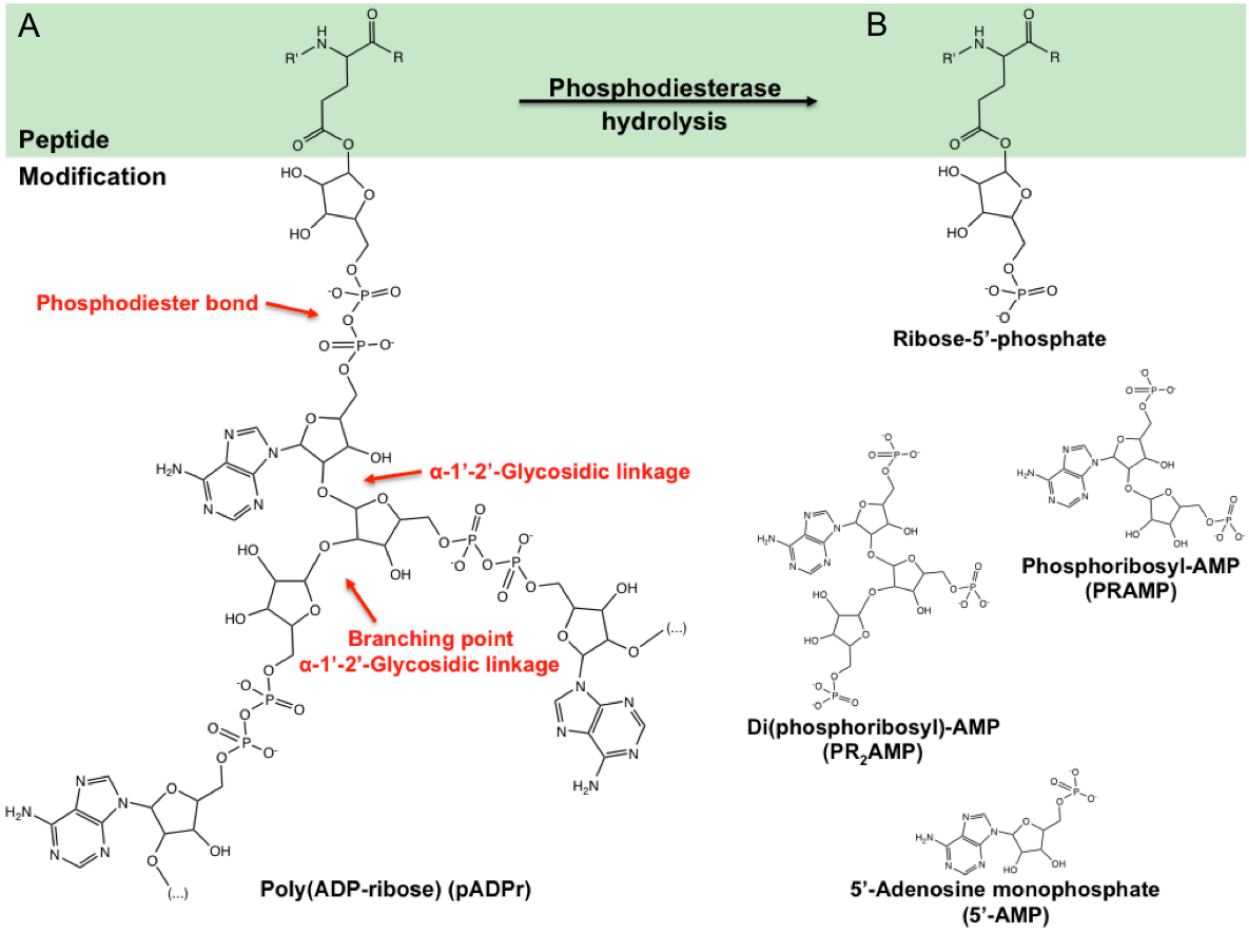


Figure-3. Schematic of a poly(ADP-ribosyl)ated peptide structure. The peptide backbone is shown over a green background. (**A**) A carboxyl ester ADP-ribose linkage is shown on the amino acid side chain (here, a glutamate (E) residue). The structure of the pADPr, including points of elongation, branching and the phosphodiester bond are shown in red. (**B**) The remnant ribose-5'-phosphate left adducted to the peptide after treatment with phosphodiesterase. Additional reaction products are shown (PRAMP, PR_2AMP and 5'-AMP).

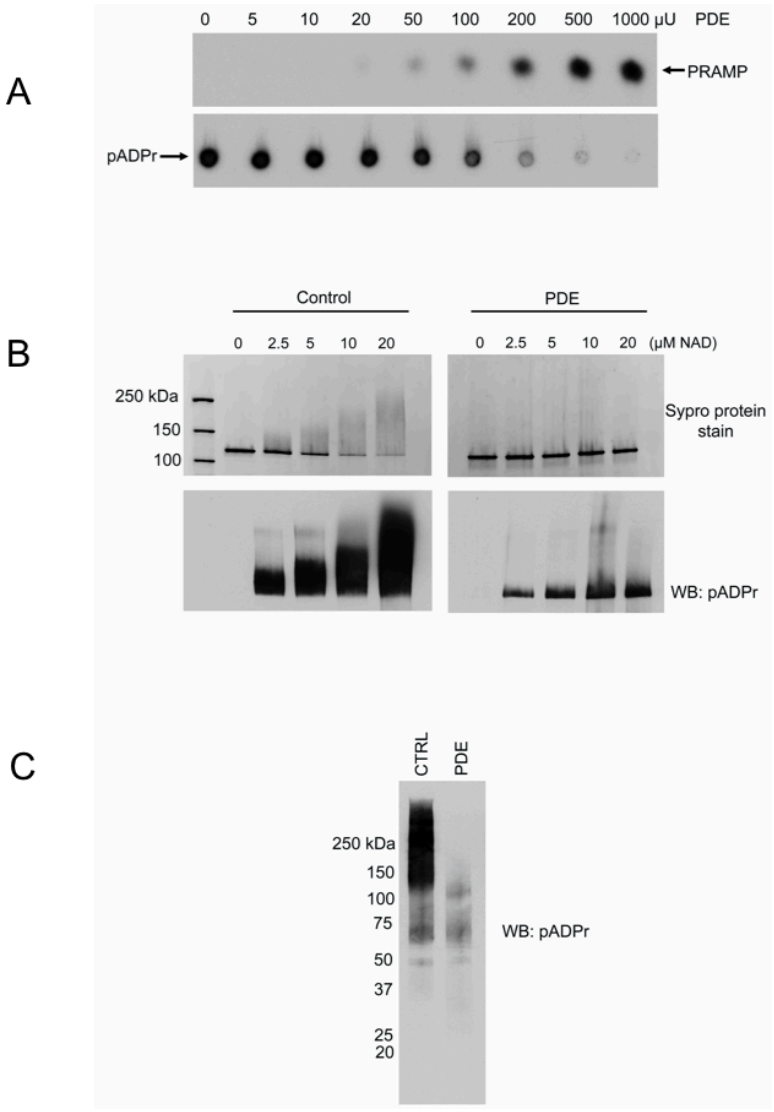


Figure-4. **Phosphodiesterase (PDE) digestion of *in vitro* and *in vivo* synthesized pADPr.** (**A**) [^{32}P]-labeled DHBB-purified pADPr synthesized *in vitro* from human recombinant PARP-1 was incubated with increasing amounts of PDE. Reaction products were hand-spotted, air-dried and separated by TLC. Plates were autoradiographied. (**B**) Upper panel: SYPRO Ruby stained SDS-PAGE showing the automodification of PARP-1. Lower panel: Western blot against pADPr (clone 96-10) showing the degradation of PARP-1-bound pADPr by PDE. (**C**) pADPr-containing protein complexes were affinity-purified by immunoprecipitation using anti-pADPr antibody (clone 10H) coupled to Protein G coated magnetic beads.

Localization of pADPr modifications by tandem mass spectrometry

Unfortunately, it was not possible to identify any pADPr modified peptides with tandem mass spectrometry analysis of the trypsinized gel band corresponding to PARP-1 containing ribose-5'-phosphate remnants. With an understanding of the low stoichiometric ratios of pADPr peptides, it seems likely that that once PARP-1 is proteolytically digested, non-modified peptides far outnumbered modified peptides. As shown in Figure-3, following treatment with PDE, the resulting ribose-5'-phosphate structure contains an exposed phosphate group. It was hypothesized that standard phosphopeptide enrichment strategies could be utilized to enrich for peptides containing ribose-5'-phosphate groups. Here, in parallel experiments, both TiO_2 and immobilized metal ion affinity chromatography (IMAC) were tested as enrichment strategies.

Following enrichment, peptides were analyzed using a standard data-dependent acquisition method. With a standard IMAC purification strategy, it was not possible to identify any pADPr sites on PARP-1. However, using the PDE and a TiO_2 enrichment strategy, nine automodification sites on PARP-1 were identified from triplicate biological preparations, each analyzed in duplicate by tandem mass spectrometry. Table-1 lists all of the identified ribose-5'-phosphate adducted peptides including their respective localization in PARP-1 functional domains. Two modifications were previously reported by Tao *et al*. and occur on the glutamate residues (underlined) in the peptide 487-A<u>E</u>PV<u>E</u>VVAPR-496.[215] These two modifications are located within a flexible loop that connects the BRCT and the WGR domains of PARP-1; a region particularly prone to intramolecular poly(ADP-ribosyl)ation upon PARP-1 activation.[192] One additional peptide (453-VVS<u>E</u>DFLQ<u>D</u>VSASTK-467) found with two modifications on acidic residues is located at the C-terminus of the BRCT domain, a protein module that is also predicted to flip over the catalytic domain of PARP-1 following its activation.[192] The additional

six sites of pADPr modification are located outside of the so-called automodification domain. One of these is E190, an N-terminal residue localized in the second zinc finger of PARP-1. An aspartic acid residue (D578) and a lysine (K579) are also targeted in the WGR domain of PARP-1, a domain that makes a contribution to the DNA damage binding interface.[192] The last three automodification sites were unexpectly found in the C-terminal ADP-ribosyl transferase (ART) domain of PARP-1: D807, E809 and E883. A representative peptide tandem mass spectrum is provided to illustrate fragment ion coverage and site-specific determination of poly(ADP-ribosyl)ated residue (Figure-5). Ambiguous assignment of the modification site occurred for some ribose-5'-phosphate-modified peptides. These peptides are also listed in Table-1 and MS spectra are provided as Supplementary information. A poly(ADP-ribosyl)ation site is located at the N-terminal boundary of the Zn2 domain (E116 or K119) and within the Zn2 domain (E147 or K148). Although interesting in the context of PARP-1 automodification studies, ambiguous modifications cannot be assigned to a single residue and it is not possible to discriminate between lysine and glutamate modifications. A three-dimensional structural model of full length human PARP-1 was reconstituted to display the automodification sites identified by mass spectrometry (Figure-6). Interestingly, all automodification sites were found to be solvent exposed and thus likely available for modification.

Table-1

Peptide sequences	Precursor *m/z*	Charge state (*z*)	PARP-1 domains	Previously Identified
Site-specific localizations*				
183-GFSLLAT**E**DK-192	646.28	2	Zn2	No
453-VVS**E**DFLQDVSASTK-467	918.40	2	BRCT	No
453-VVSEDFLQ**D**VSASTK-467	918.40	2	BRCT	No
487-A**E**PVEVVAPR-496	639.29	2	Interdomain BRCT - WGR	Yes [215]
487-AEPV**E**VVAPR-496	639.29	2	Interdomain BRCT - WGR	Yes [215]
572-LQLLED**D**KENR-582	792.35	2	WGR	No
572-LQLLEDD**K**ENR-582	792.35	2	WGR	No
803-VVDR**D**SEEAEIIR-815	871.39	2	ART	No
803-VVDRDS**E**EAEIIR-815	871.39	2	ART	No
879-IAPP**E**APVTGYMFGK-893	602.60	3	ART	No
Modified peptides with localization ambiguities*				
109-TLGDFAA**E**YA**K**-119	698.79	2	Zn2	No
143-MVDP**EK**PQLGMIDR-156	624.60	2	Zn2	No

* Peptides identified were classified as either *site-specific*, when CID mass spectral determinants were satisfactory to assign the post-translational modification to a single residue, or *ambiguous*, when the fragmentation pattern was insufficient to pinpoint the site of modification. Ribose-5'-phosphate modified amino acids are shown in bold and underlined. Refer to Supplementary information for complete MS data.

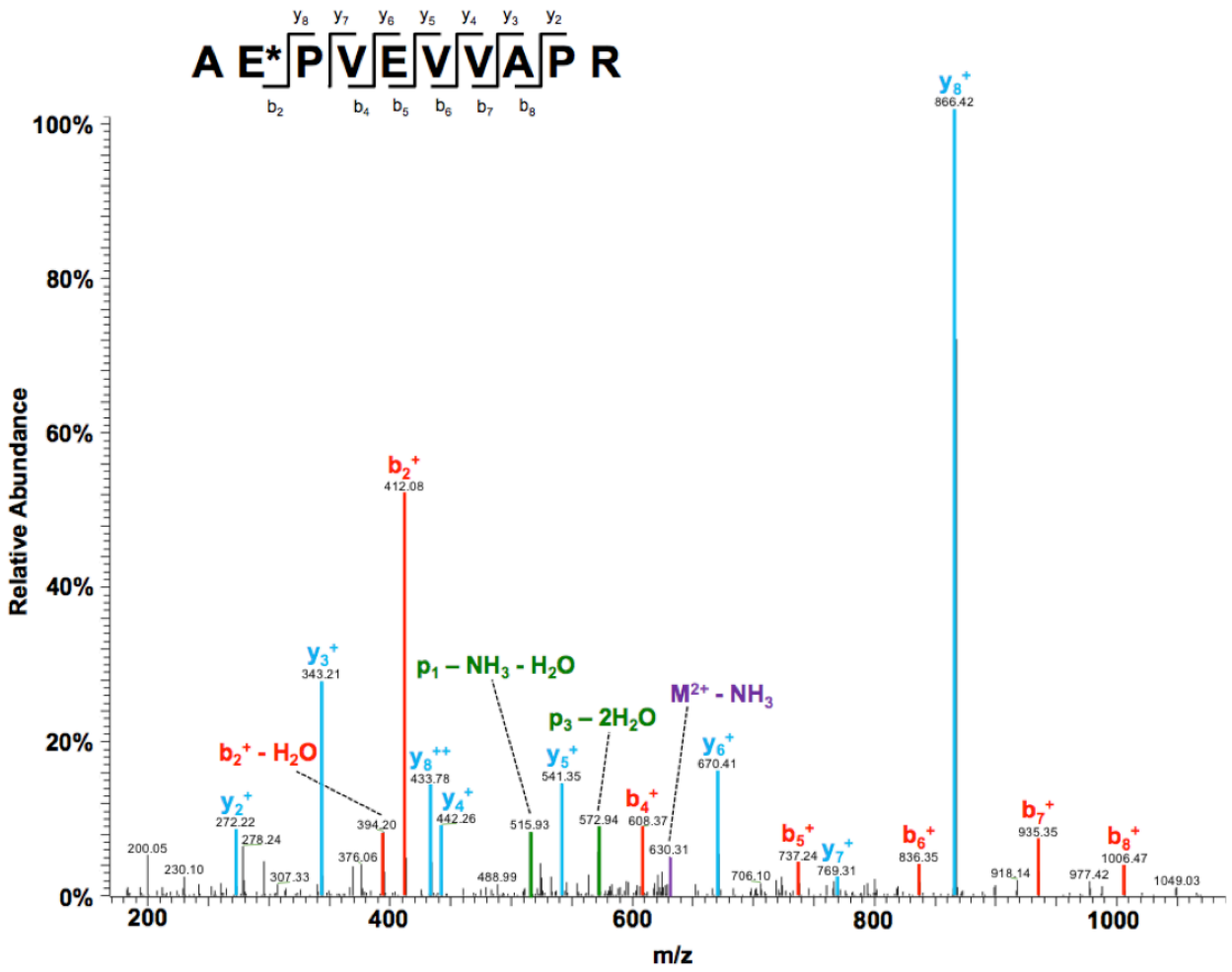


Figure-5. **Representative tandem mass spectum of a ribose-5'-phosphate-modified peptide from automodified human PARP-1.** The peptide AEPVEVVAPR is annotated with its matching *b*- (red), *y*-ion series (blue), fragmentation within the modification (green), and doubly charged ion of the peptide (purple). The asterisk denotes the actual site of the ribose-5'-phosphate signature, a remnant of the original poly(ADP-ribosyl)ation modification. Refer to Supplementary information for other MS spectra of ribose-5'-phosphate-modified peptides identified in this study.

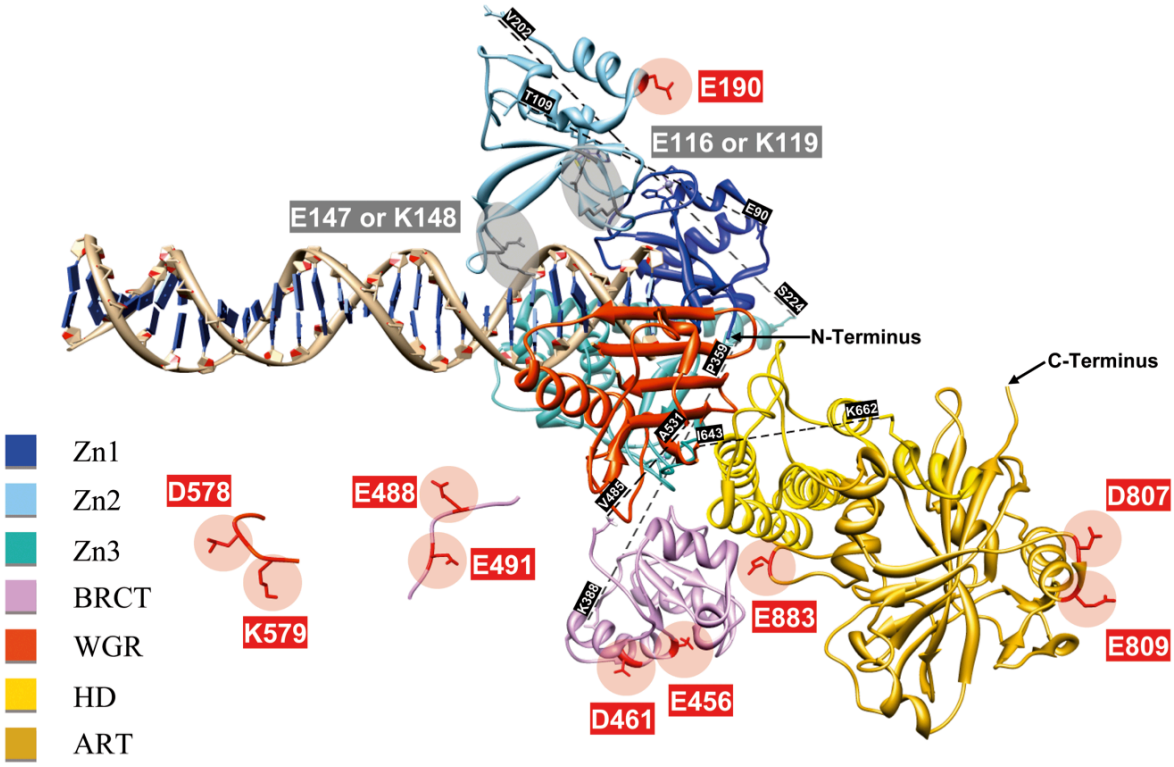


Figure-6. Three-dimensional model of human PARP-1 bound to a DNA break. The model was built upon the addition of the Zn2 (PDB code 4AV1) and BRCT (2COK) domains of PARP-1 to the PARP-1 structure published by Langelier and collaborators (i.e. Zn1, Zn3, WGR, HD and ART).[192] Domains can be distinguished by a color-coded legend. Inter-domain linkers, for which no structures are currently available, are depicted by a dashed black line with corresponding amino acid positions. Poly(ADP-ribosyl)ated residues are shown with their corresponding side chains over shaded circles. Tags indicate the residues targeted by PARP-1 automodification. Red color was used to depict site-specific automodification sites and grey for ambiguous identifications (see text). Residues E488, E491, D578 and K579 are shown unconnected to the PARP-1 as no corresponding PARP-1 structure is available for these regions. Zn: zinc finger; BRCT: BRCA1 C-Terminus domain; WGR: domain named after the conservation of Tryptophan (W) - Glycine (G) - Arginine (R); HD: helical subdomain; ART: ADP-ribosyl transferase (catalytic) domain.

As expected, a parallel analysis consisting of biological duplicates each analyzed in triplicate by tandem mass spectrometry of non-automodified PARP-1 did not allow detection of any modified peptides (data not shown). This finding is in agreement with the Western blot showing that non-activated PARP-1 does not have any residual modifications that may occur in either protein expression or purification (Figure-4B, lower panel). Additionally, the capture of ribose-5'-phosphate-modified peptides appeared to have been quantitative because modified peptides were detected in only the enriched fraction but neither the two wash fractions nor the flow through fraction. While the other three fractions did not contain modified peptides, they are valuable to the overall analysis of PARP-1 as they provided information about the whole protein and serve as a comparison of digestion efficiency between the control and automodified samples (data not shown).

4.4 Discussion

It has always been difficult to discriminate between the direct poly(ADP-ribosyl)ation of a target protein by PARP-1, usually referred to as the substrate, and the non-covalent interaction with pADPr. The fact that some proteins show strong affinity to free pADPr may lead to misinterpretation of experimental results. Further confusion can arise when considering a protein such as a PARP-1-interacting protein because PARP-1 itself is a major acceptor of pADPr with an ever-growing list of specific interactions.[229, 230] Here, a peptide array screening strategy was used to systematically locate covalent poly(ADP-ribosyl)ation sites in PARP-1. Although this approach is commonly used to identify other post-translational modifications, interactions with pADPr were shown to withstand stringent conditions, indicating that pADPr remains bound to peptides, once again bringing us back to the original problem. Still, by comparing the labeling patterns obtained as a result of PARP-1 activation and non-covalent polymer-binding assays, it was possible to demonstrate that a synthetic PARP-1 peptide can be specifically (ADP-ribosyl)ated in *trans*. However, few PARP-1 peptides could be specifically assigned as covalent *trans*-automodification sites. This rare finding may be linked to the strong preference of PARP-1 to perform autocatalytic poly(ADP-ribosyl)ation over heteromodification of target proteins (including itself), an enzymatic behavior also consistent with the DNA damage-induced compact domain rearrangement of PARP-1.[192, 231]

These results highlight the need for a reliable analytical method to unequivocally identify any particular poly(ADP-ribosyl)ation sites.To that end, a novel technique was developed for identifying site-specific locations of pADPr modification on PARP-1 via mass spectrometry. Utilizing the enzymatic cleavage of PDE to cleave phosphodiester bonds, the heterogeneous pADPr modification is reduced to a simple 212-Dalton ribose-5'-phosphate tag. This protein

modification signature is much more suitable to CID mass spectrometry analysis than pADPr or mono-ADPr because the bulk of fragment ions constitute *b*- and *y*-ions of the peptide. However, in order to provide sufficient signal for CID, a titanium dioxide enrichment proved to be essential for enrichment of the delta-212.094-Dalton ribose-5'-phosphate modified peptides. As with use of purification for phosphopeptides, strongly acidic peptides were also found in the TiO_2 enrichment fraction, but did not appear to obscure the signal of the ribose-5'-phosphate modified peptides.

In an effort to simplify the characterization of pADPr polymers previous researchers used various biological conditions to limit PARP-1 enzyme activity; e.g. by (i) drastically shortening the reaction time or limiting the NAD substrate concentration to minimize ADP-ribose incorporation into pADPr; (ii) using PARP-1 (E998K) so that only mono-automodification can occur; (iii) only modifying small domains of PARP-1 to reduce the overall peptide signal that results from digesting the entire protein. Here, PARP-1 was kept in its native form and used standard conditions so that the sites of pADPr formation could be linked with higher confidence to those that can occur *in vivo*. In this study, nine site-specific automodification sites on PARP-1 were localized in addition to three ambiguous assignments. Furthermore, the peptide 109-TLGDFAAE*YAK*-119 contains a possible modification on a terminal lysine. Trypsin is unlikely to possess the ability to recognize and cleave a ribose-5'-phosphate modified lysine, and therefore it is probable that this modification lies on E116. With that said, precise localization will require more thorough *b*- and *y*-ion coverage not obtained in this series of experiments. Considering that PARP-1 sequence coverage is typically limited to ~70% when using trypsin as the proteolytic enzyme, the identification of 12 automodification sites (Table-1) probably covers most of PARP-1 tryptic peptides accessible by common proteomics methods. The occurrence of

poly(ADP-ribosyl)ated residues in automodified PARP-1 found by mass spectrometry is also consistent with the biochemical ADP-ribose incorporation studies that indicated that each PARP-1 molecule had about 15 acceptor sites.[202]

In order to obtain usable mass spectrometry data for the localization of pADPr, two distinct requirements must be accomplished: (i) adequate fragmentation of the peptide ion is necessary for proper peptide identification; and (ii) *b*- and *y*-ion ion coverage must surround the modified residue for confident localization of the modification. As a first mass spectrometry-based approach, it was shown that a synthetic prototypic peptide can be mono-ADP-ribosylated by PARP-1 E988K *in vitro* and that the conversion of the modification to ribose-5'-phosphate is required for efficient peptide fragmentation. Although unambiguous site-specific localization of the ribose-5'-phosphate signature was impossible in this case, clues in the tandem mass spectrum point to modification of lysine residues. However, the molecular mechanisms responsible for formation of an ADP-ribose linkage to a lysine side chain remain unclear. The accumulation of chemically reactive ADP-ribose species, as a consequence of the NADase activity of PARP-1, has been suggested to explain a non-enzymatic glycation mechanism underlying the establishment of a lysine-ADP-ribose ketamine modification.[232, 233] It remains to be demonstrated if this happens under normal physiological conditions and if this phenomenon contributes to the automodification process.

There has been a long-standing debate as to which amino acid residues are modified by PARP-1. Initially, it was thought that only aspartate and glutamate residues were poly(ADP-ribosy)lated. Recently, conflicting evidence suggested that lysines were major targets of pADPr formation.[191] With this study of intact PARP-1, it is now possible to reliably assign the actual modified residues that are targets of PARP-1 automodification in its native form. Direct mass

spectrometry-based evidence is presented showing that aspartate, glutamate and lysine residues are acceptors of ADP-ribose on PARP-1. While the majority of automodification sites were mapped onto acidic glutamate and aspartate residues, which is consistent with the previous work of Tao *et al*, a lysine modification also was identified with high confidence.[215] Given the low-abundance of ribose-5'-phosphate-modified peptides, it is proposed that the consistent identification of glutamate and aspartate residues as PARP-1 automodification sites suggests that carboxyl ester linkages are the predominant acceptors. The lysine modification found in the context of intact PARP-1 automodification could be useful information to pursue further studies of lysine glycation by free ADP-ribose generated at high rates by PARP-1 following its activation.

An interesting feature of poly(ADP-ribosyl)ated residues is their common localization in regions of intrinsic sequence disorder. For example, the prototypic ADP-ribosylated peptide (K653 and/or K654) exactly matches a PARP-1 region that was not modeled due to apparent disorder (most likely a very flexible loop between residues 644-661).[192] This is also the case for residues identified in full length PARP-1. A series of modification sites were previously localized in a conformationally flexible surface exposed between the BRCT and WGR domains: K498, K521, K524[191]; E488 and E491 that was found in this work and by Tao and collaborators.[215] All three automodification sites found in the catalytic domain of PARP-1 (D807, E809 and E883) are also located on the tip of loop-shaped protein segments.

Historically, the automodification domain of PARP-1 has been reported to reside in a ~150 amino acid region that encompases the BRCT domain and the interdomain region that connects the BRCT to the WGR domain (region 373-524). Consistent with this description, two automodification sites in the BRCT domain of PARP-1 (E456 and D461) were identified in

addition to the E488 and E491 sites located in the BRCT-WGR interdomain region. However, poly(ADP-ribosyl)ation sites were found outside of the automodification domain. Notably, a glutamate residue in the second zinc finger of PARP-1 (E190) was identified. Two additional residues located within the Zn2 domain are also targeted (ambigously assigned to E116/K119 and E147/K148) suggesting that the second zinc finger of PARP-1 could be structurally and functionnaly altered by automodification. Finally, two modifications were identified in the WGR domain of PARP-1: D578 and K579. These modifications could also be functionally important given that the WGR domain provides a bridge between the DNA damage interface and the catalytic domain of PARP-1.[192] Collectively, 8 out of 12 poly(ADP-ribosyl)ation sites were mapped beyond the so-called automodification domain. According to the study of Tao *et al*., a complete deletion of the automodification domain of PARP-1 results in a modest reduction of automodification (~30%), suggesting that many other poly(ADP-ribosyl)ation sites exist outside the previously assigned automodification domain of PARP-1.[215]

PARP-1 automodification may arise from two scenarios.[234] In the model proposed by Langelier and collaborators, PARP-1 domains collapses on the DNA strand break in a compact monomeric conformation that can explain the strong preference for intramolecular automodification (*cis*-automodification model).[228] Given the compact arrangement of PARP-1 domains upon activation by DNA double-strand breaks, it is logical to think that only loops and disordered regions will have enough flexibility to contact the catalytic site and be poly(ADP-ribosyl)ated.[192] In the second scenario, two PARP-1 molecules dimerize at the DNA damage site and poly(ADP-ribosyl)ate each other in a *trans*-automodification process.[190] Under the conditions of the present study, both phenomena contribute to PARP-1 automodification and

might explain the localization of distant automodified residues from the active site in the ART domain.

4.5 Conclusions and future directions

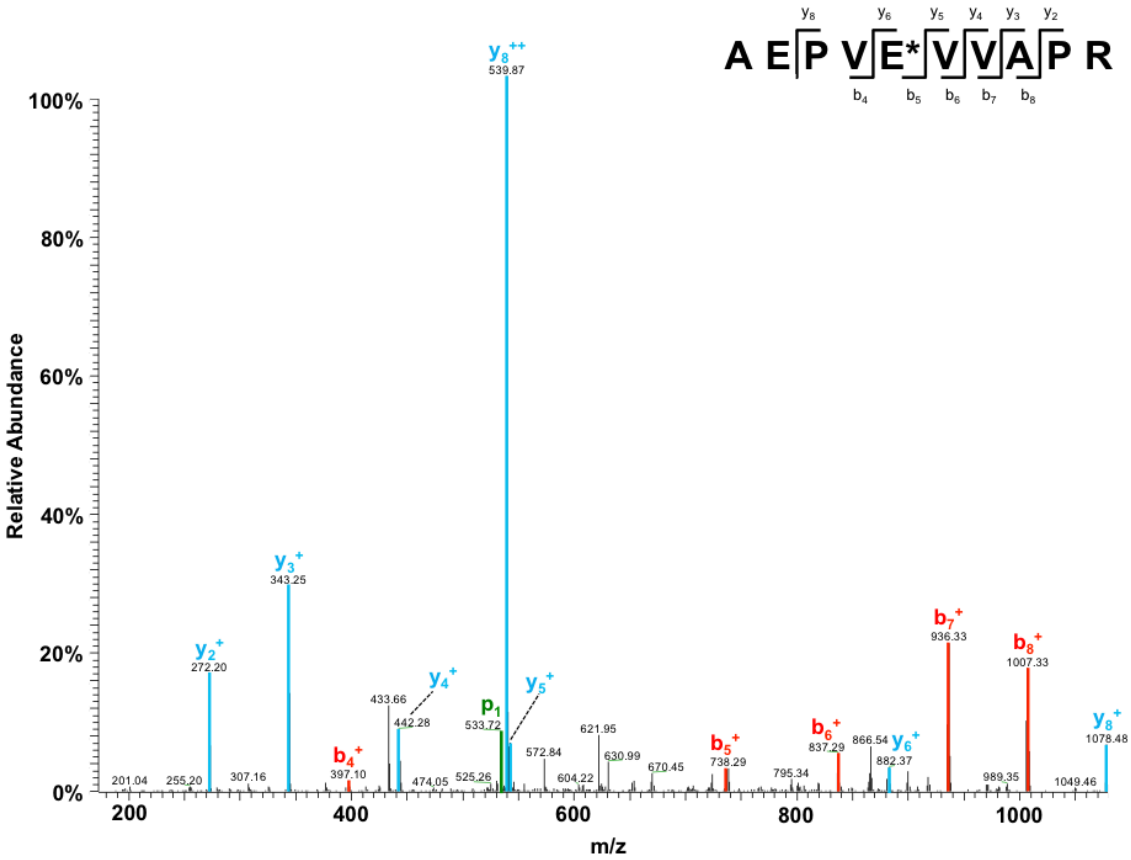
In conclusion, the method described here provides a valuable technique to better define the biological activity of PARP-1 and the consequences of poly(ADP-ribosyl)ation. Until now, most methods used to study poly(ADP-ribosyl)ation sites did not generate the same level of direct evidence of covalent modification that can be provided with mass spectrometry. The present approach to mapping pADPr sites is based on established phosphopeptide sample preparation and tandem mass spectrometric analysis methods making it readily accessible to the scientific community. Notably, the simplification of pADPr to ribose-5'-phosphate results in tandem mass spectra consisting mainly of *b*- and *y*-ion signals. This allows researchers to use widely available peptide search analysis techniques to localize modification sites and also makes this workflow suitable to substrate-specific or high-throughput proteomics analysis. Additionally, the unique mass of the ribose-5'-phosphate modification to those contained in the ABRF protein modification database (http://www.abrf.org) and the use of sufficient control samples allow for confident differentiation between ribose-5'-phosphate and other possible protein modifications. Finally, it may be noted that while only PARP has been studied here, neither the PDE cleavage step nor the TiO_2 enrichment step is protein dependent. Therefore, the method should work on any number of proteins or protein mixtures such as a whole cell lysate. Currently, parylation patterns of a renal cell line, HEK293, following exposure to methylnitronitrosoguanidine (MNNG), a strong DNA methylation agent. While only preliminary studies have been completed, the application of this method to more complex protein mixtures is resulting in promising discoveries.

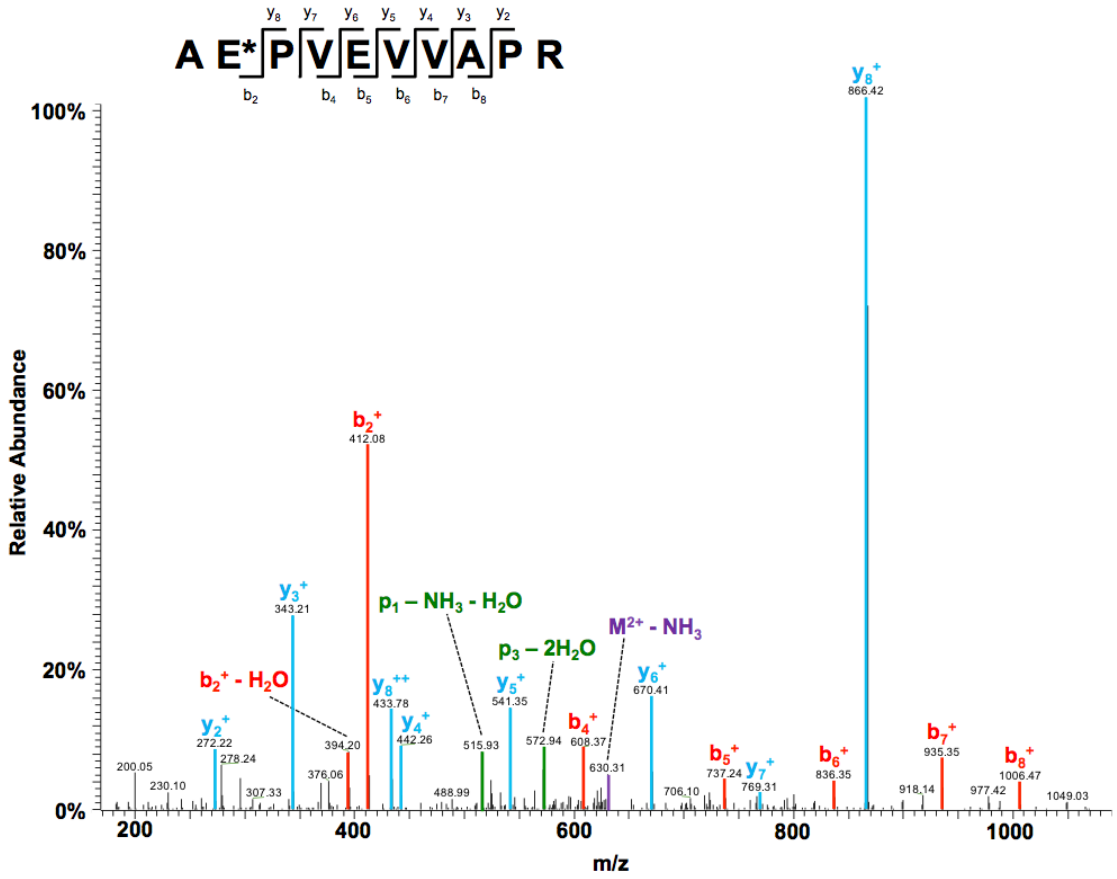
4.6 Notes to Chapter 4

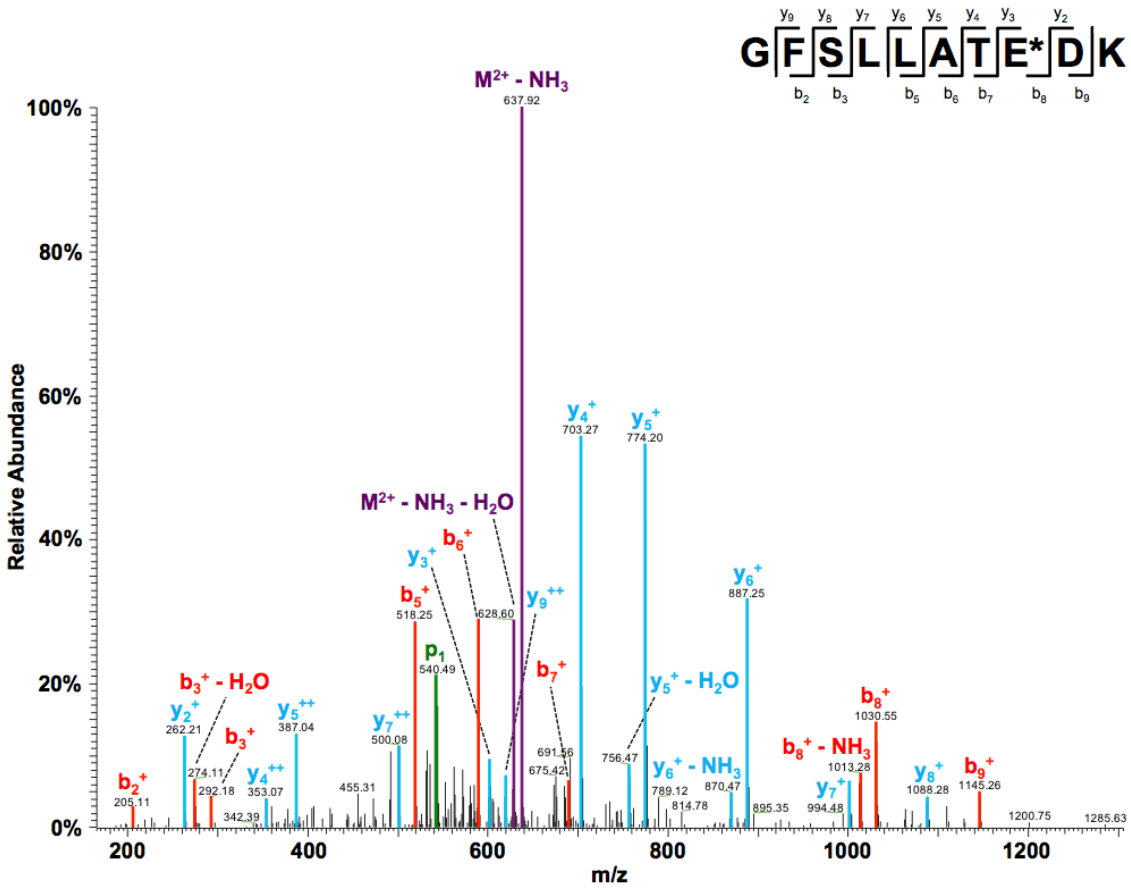
Acknowledgments

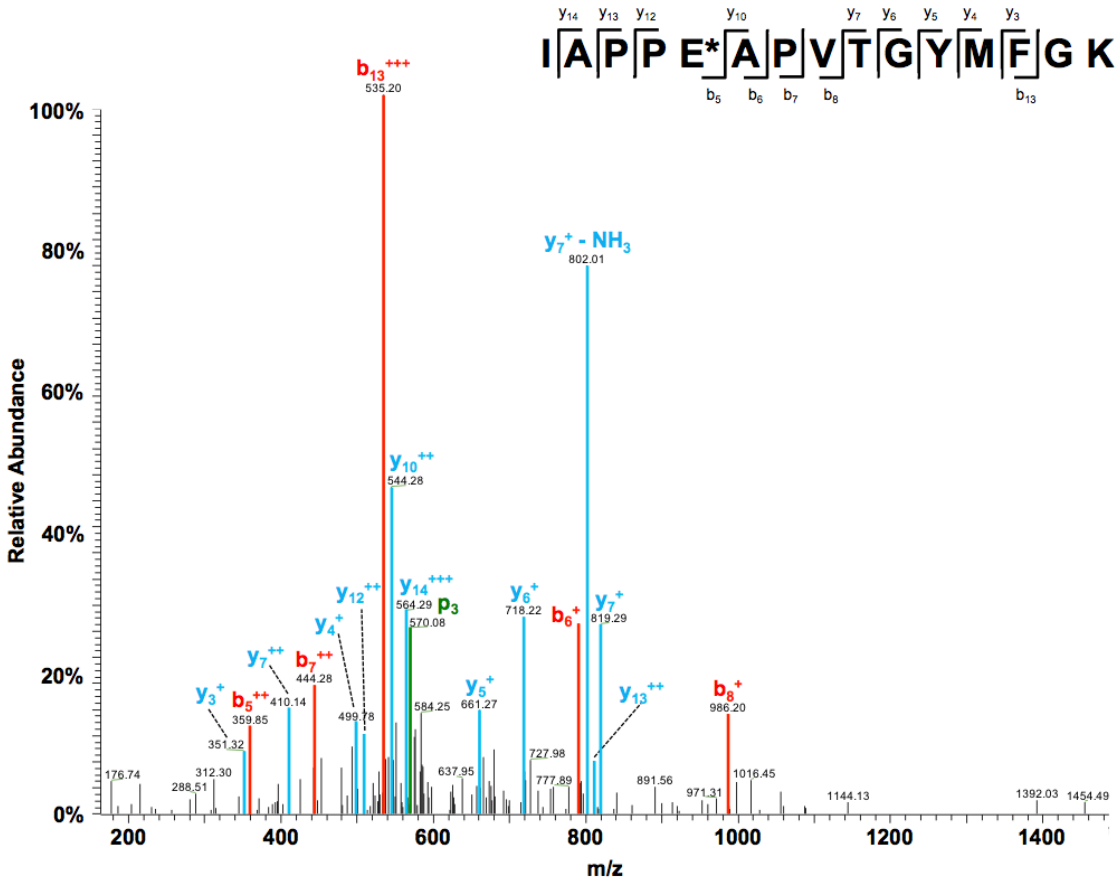
This work was supported by the Canadian Institutes of Health Research [grant number MO-178013 to G.G.P.]. This work is supported in part by the University of Washington's Proteomics Resource (UWPR95794) and the NW Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Mass Spectrometry Core (5 U54 AI057141). G.G.P. holds a Canada Research Chair in Proteomics.

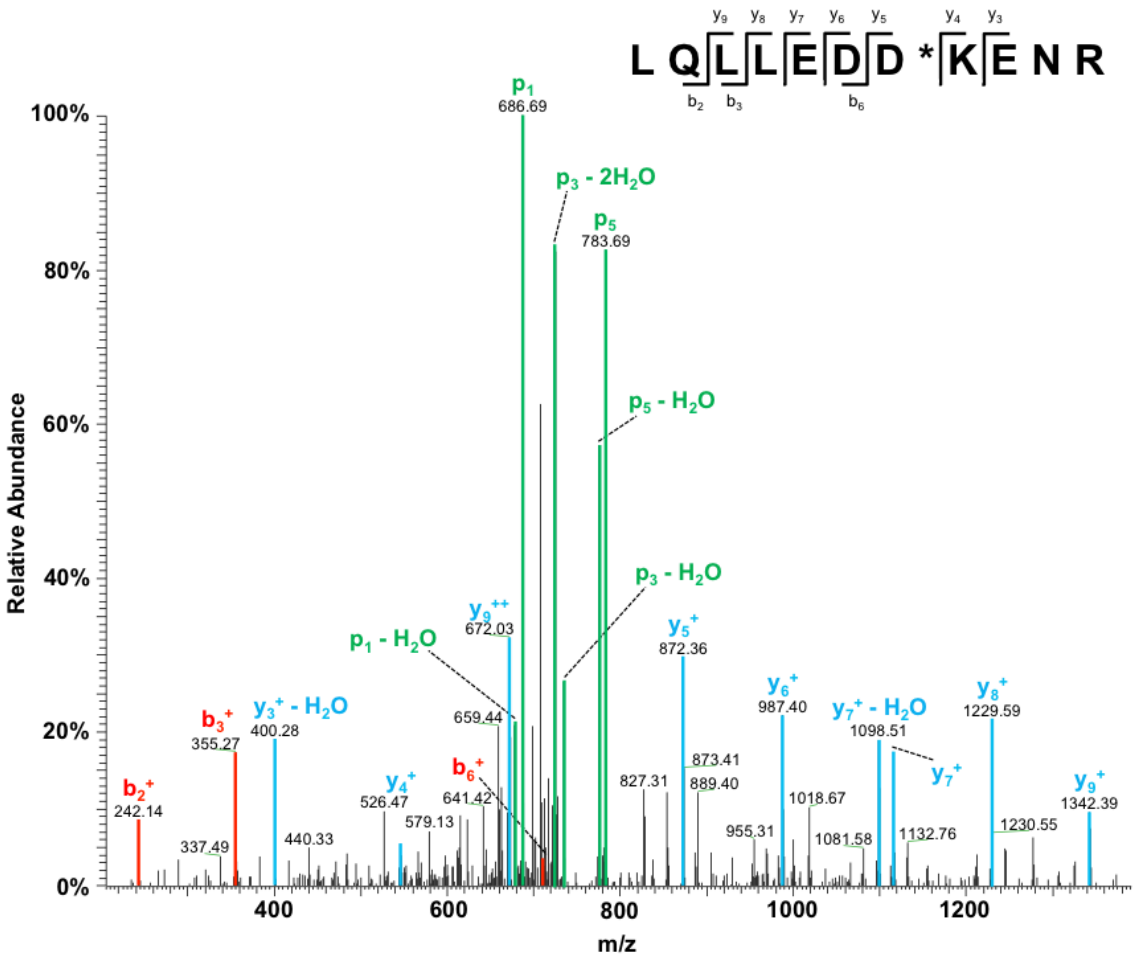
Supplemental Figures

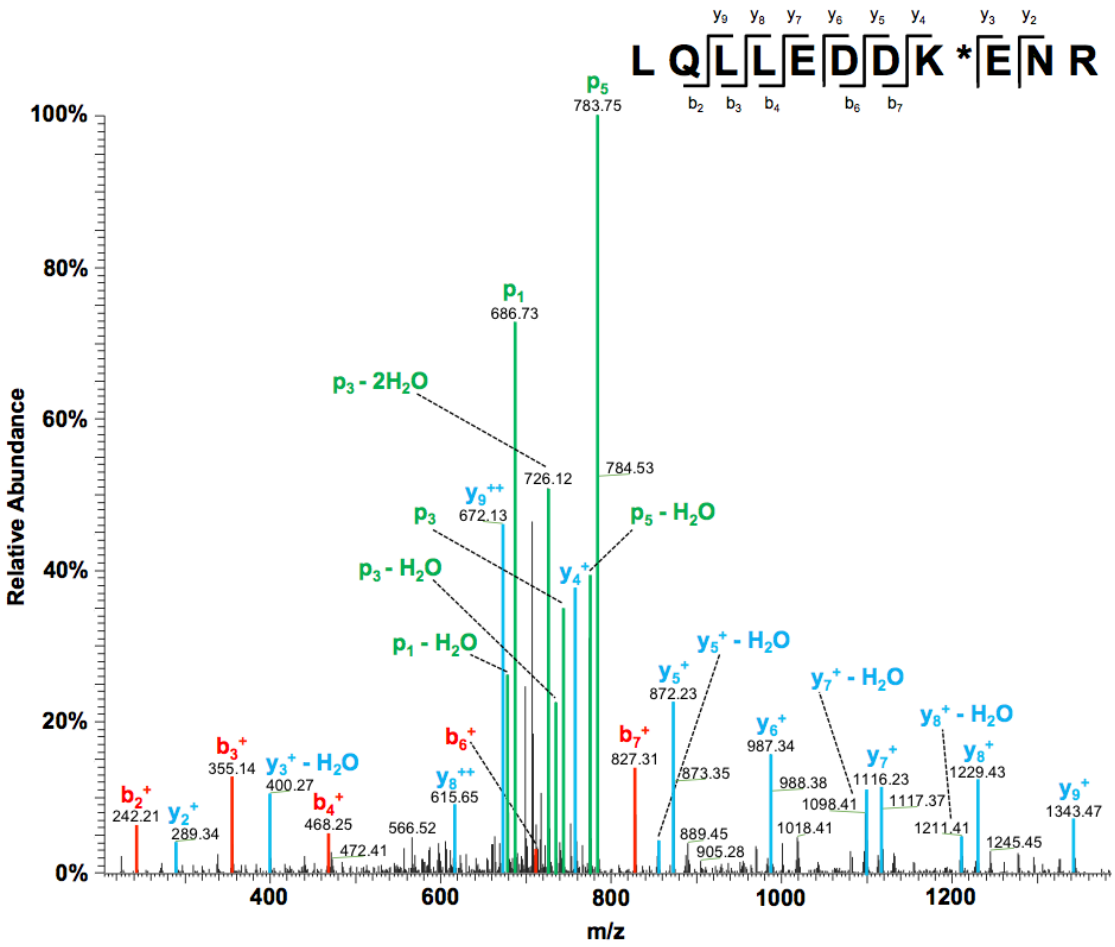


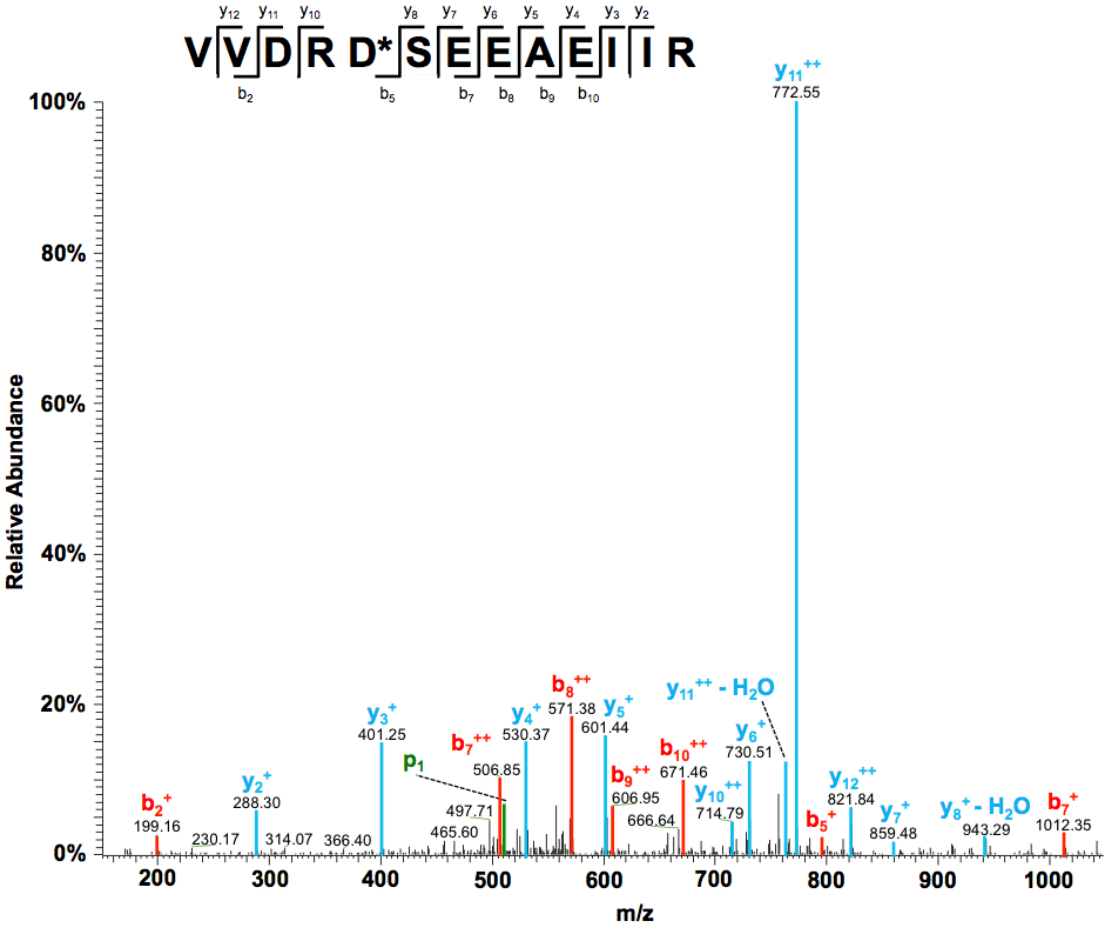

A E* P V E V V A P R
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
200
400
600
800
1000
y8+ 866.42
b2+ 412.08
y3+ 343.21
p1 – NH3 - H2O
M2+ - NH3
p3 – 2H2O
b2+ - H2O
y8++ 433.78
y5+ 541.35
y6+ 670.41
y2+ 272.22
y4+ 442.26
b4+ 608.37
b5+ 737.24
y7+ 769.31
b6+ 836.35
b7+ 935.35
b8+ 1006.47
394.20
515.93
572.94
630.31
278.24
376.06
200.05
230.10
307.33
488.99
706.10
918.14
977.42
1049.03

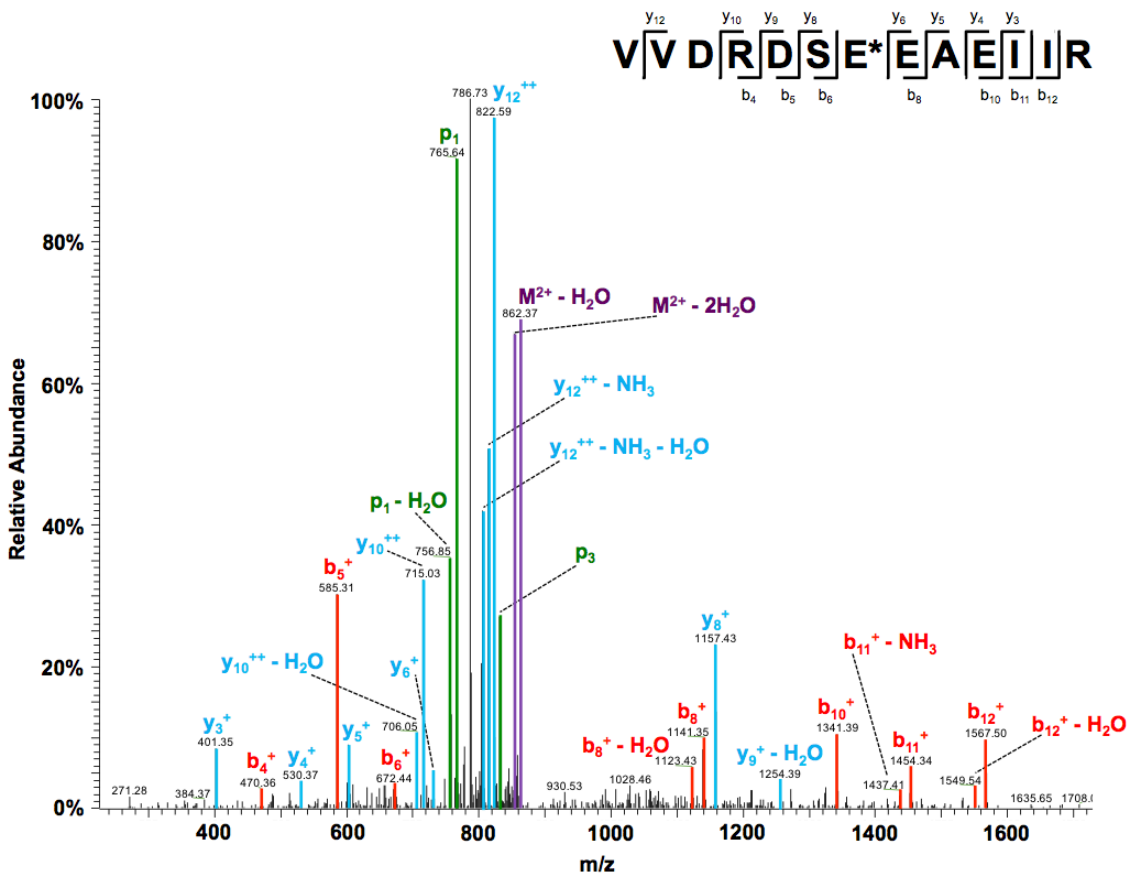

GFSLLATE*DK
y9 y8 y7 y6 y5 y4 y3 y2
b2 b3 b5 b6 b7 b8 b9
M2+ - NH3
637.92
M2+ - NH3 - H2O
y4+
703.27
y5+
774.20
y6+
887.25
y3+
b6+
b5+
518.25
628.60
y9++
p1
540.49
b7+
b3+ - H2O
y2+
262.21
y5++
387.04
b3+
274.11
y7++
500.08
y4++
353.07
292.18
342.39
455.31
691.56
675.42
y5+ - H2O
y6+ - NH3
789.12
870.47
814.78
756.47
b8+
1030.55
b8+ - NH3
1013.28
y7+
994.48
y8+
1088.28
b9+
1145.26
b2+
205.11
895.35
1200.75
1285.63
Relative Abundance
100%
80%
60%
40%
20%
0%
200
400
600
800
1000
1200
m/z

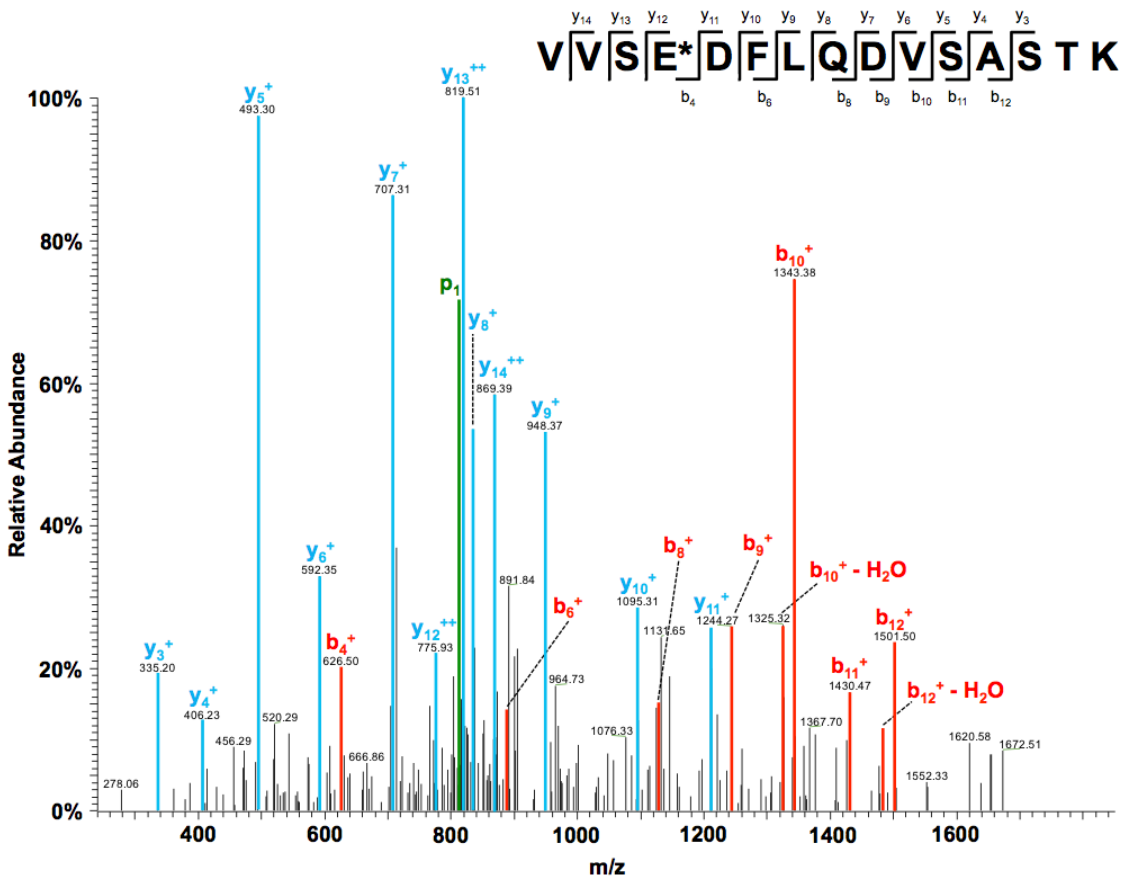

I A P P E* A P V T G Y M F G K
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
200
400
600
800
1000
1200
1400
b13+++ 535.20
y7+ - NH3 802.01
y10++ 544.28
y14+++ 564.29
p3 570.08
y12++
y4+
y7++
b7++ 444.28
y3+
b5++ 410.14
359.85
351.32
499.78
584.25
y5+ 661.27
y6+ 718.22
b6+
y7+ 819.29
y13++
b8+ 986.20
727.98
777.89
637.95
891.56
971.31
1016.45
1144.13
1392.03
1454.49
176.74
312.30
288.51

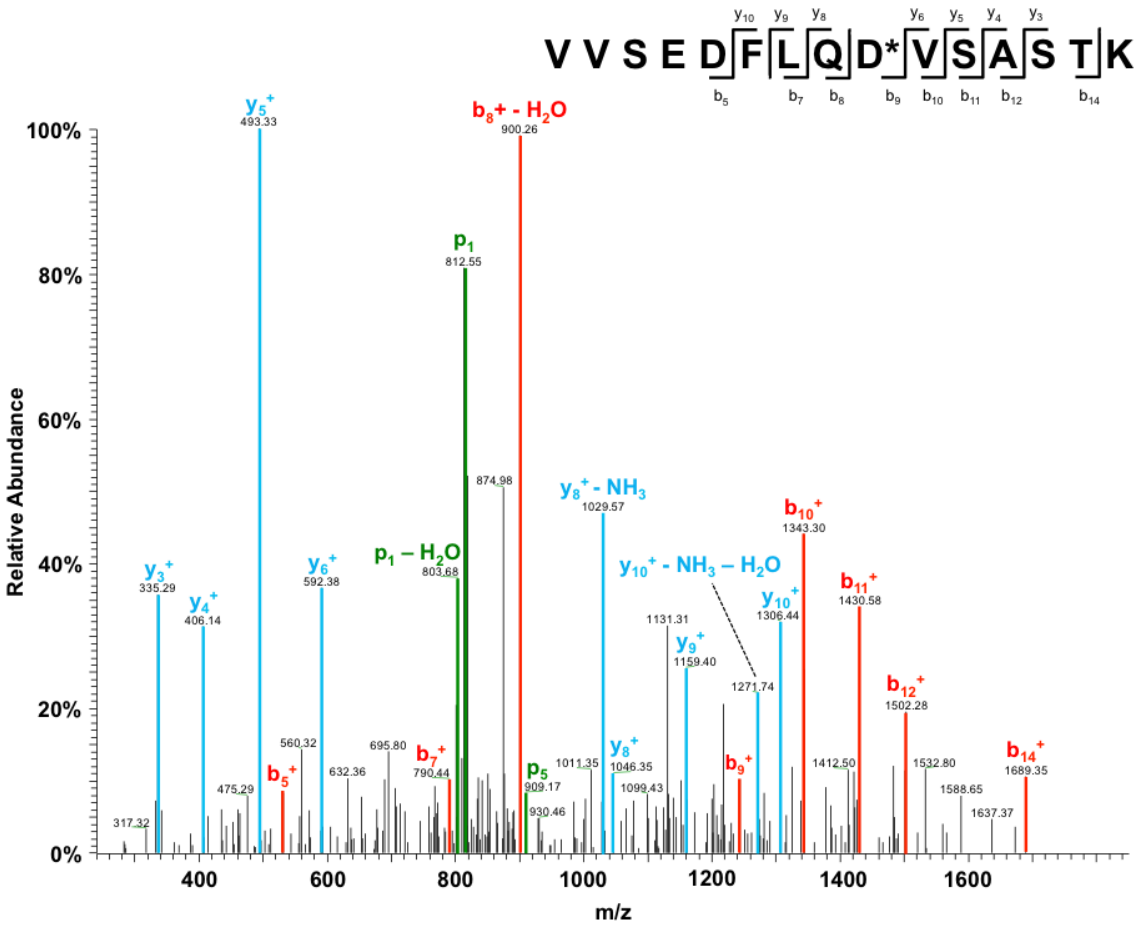

L Q L L E D D * K E N R
y9 y8 y7 y6 y5 y4 y3
b2 b3 b6
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
400
600
800
1000
1200
p1 686.69
p3 - 2H2O
p5 783.69
p5 - H2O
p3 - H2O
p1 - H2O
y9++ 672.03
y5+ 872.36
y6+ 987.40
y7+ - H2O 1098.51
y8+ 1229.59
y3+ - H2O 400.28
b3+ 355.27
b2+ 242.14
y4+ 526.47
b6+ 641.42
659.44
579.13
337.49
440.33
827.31
873.41
889.40
955.31
1018.67
1081.58
y7+
1132.76
1230.55
y9+ 1342.39

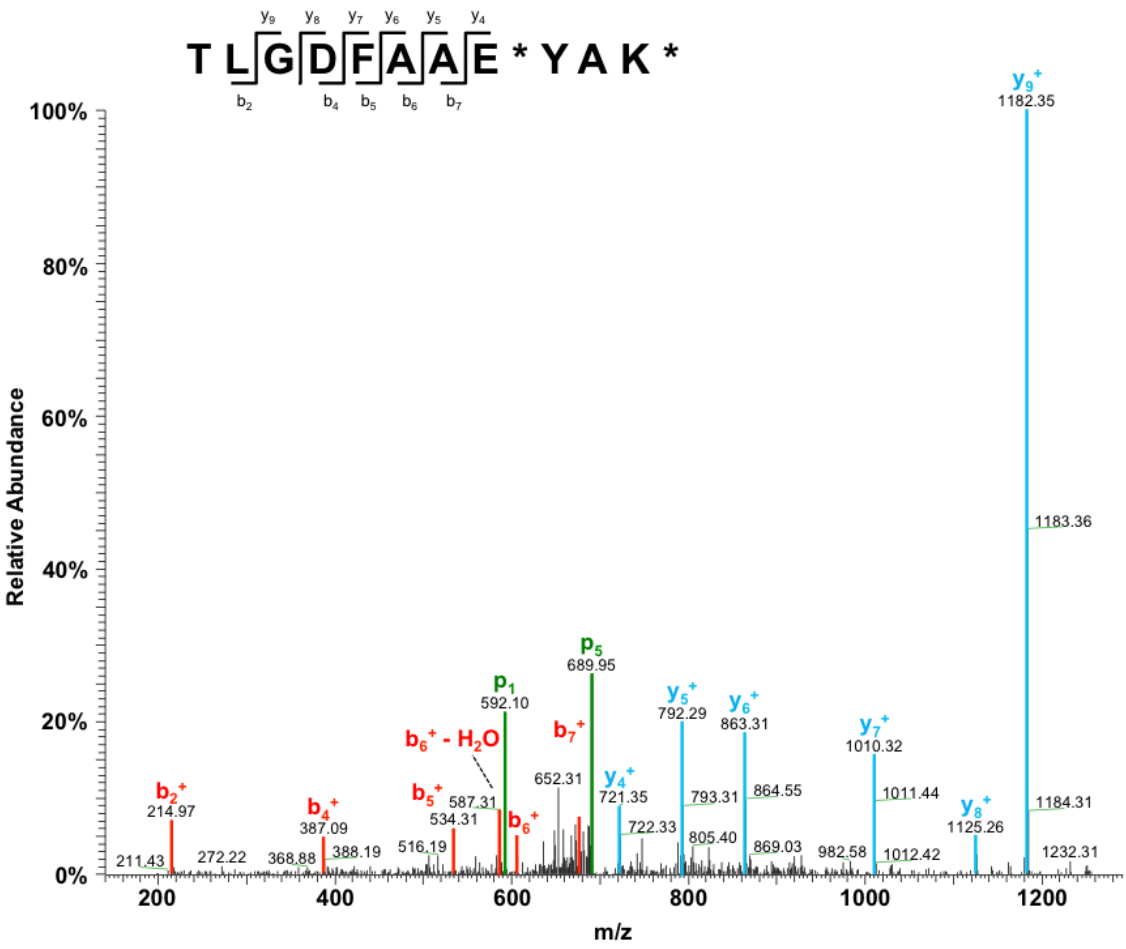

L Q L L E D D K *E N R
y9 y8 y7 y6 y5 y4 y3 y2
b2 b3 b4 b6 b7
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
400
600
800
1000
1200
p5
783.75
p1
686.73
p3 - 2H2O
726.12
784.53
y9++
672.13
p3
p5 - H2O
p3 - H2O
y4+
p1 - H2O
y5+ - H2O
y5+
872.23
y7+ - H2O
y8+ - H2O
y6+
987.34
y7+
1116.23
y8+
1229.43
y9+
1343.47
b6+
b7+
827.31
b3+
355.14
y3+ - H2O
400.27
y8++
615.65
b2+
242.21
y2+
289.34
b4+
468.25
566.52
472.41
873.35
889.45
905.28
988.38
1098.41
1018.41
1117.37
1211.41
1245.45

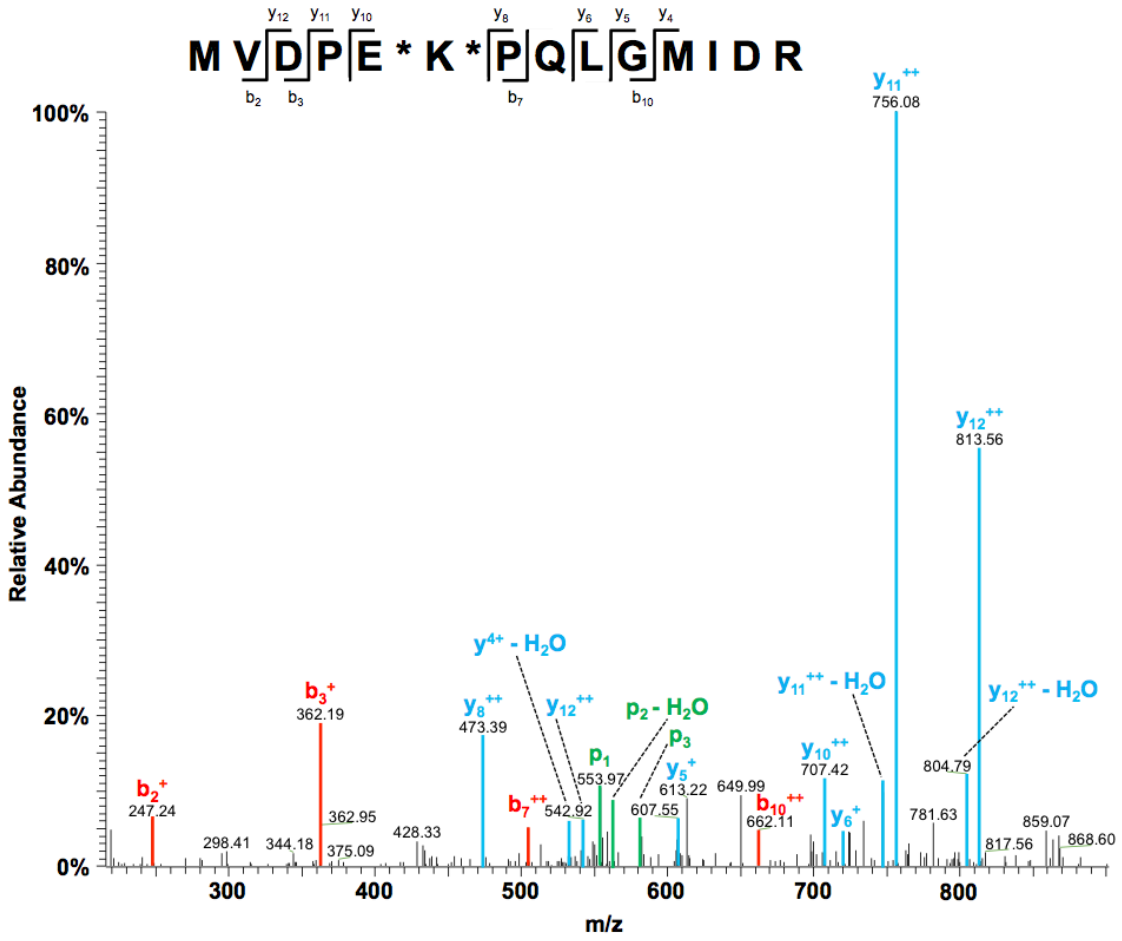

V V D R D* S E E A E I I R
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
200
300
400
500
600
700
800
900
1000
y11++ 772.55
b8++ 571.38
y5+ 601.44
y4+ 530.37
y3+ 401.25
p1
b7++ 506.85
y11++ - H2O
y6+ 730.51
b10++ 671.46
b9++ 606.95
y10++ 714.79
y12++ 821.84
y2+ 288.30
b2+ 199.16
497.71
465.60
666.64
b5+
y7+ 859.48
y8+ - H2O 943.29
b7+ 1012.35
230.17
314.07
366.40


VVDRDSE*EAEIIR
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
400
600
800
1000
1200
1400
1600
p1
765.64
y12++
822.59
786.73
M2+ - H2O
862.37
M2+ - 2H2O
y12++ - NH3
y12++ - NH3 - H2O
p1 - H2O
y10++
756.85
715.03
p3
b5+
585.31
y10++ - H2O
y6+
y8+
1157.43
b11+ - NH3
y3+
401.35
y5+
706.05
b4+
470.36
y4+
530.37
b6+
672.44
b8+
1141.35
b8+ - H2O
1123.43
y9+ - H2O
1254.39
b10+
1341.39
b11+
1454.34
b12+
1567.50
b12+ - H2O
271.28
384.37
930.53
1028.46
1437.41
1549.54
1635.65


y14 y13 y12 y11 y10 y9 y8 y7 y6 y5 y4 y3
V V S E* D F L Q D V S A S T K
b4 b6 b8 b9 b10 b11 b12
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
400
600
800
1000
1200
1400
1600
y5+ 493.30
y13++ 819.51
y7+ 707.31
p1
y8+
y14++ 869.39
y9+ 948.37
b10+ 1343.38
y6+ 592.35
b8+
b9+
b10+ - H2O
891.84
y10+ 1095.31
b6+
y11+ 1244.27
1325.32
b12+ 1501.50
y12++ 775.93
b4+ 626.50
y3+ 335.20
b11+ 1430.47
y4+ 406.23
b12+ - H2O
964.73
520.29
456.29
1076.33
1367.70
1620.58
1672.51
666.86
278.06
1552.33


V V S E D F L Q D* V S A S T K
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
400
600
800
1000
1200
1400
1600
y5+ 493.33
b8+ - H2O 900.26
p1 812.55
874.98
y8+ - NH3 1029.57
b10+ 1343.30
y3+ 335.29
y6+ 592.38
p1 – H2O 803.68
y10+ - NH3 – H2O
y4+ 406.14
y10+ 1306.44
b11+ 1430.58
1131.31
y9+ 1159.40
1271.74
b12+ 1502.28
560.32
695.80
b7+ 790.44
b5+
632.36
p5 909.17
1011.35
y8+ 1046.35
b9+
1412.50
1532.80
b14+ 1689.35
475.29
1099.43
1588.65
317.32
930.46
1637.37


T L G D F A A E * Y A K *
y9 y8 y7 y6 y5 y4
b2 b4 b5 b6 b7
Relative Abundance
m/z
100%
80%
60%
40%
20%
0%
200
400
600
800
1000
1200
y9+ 1182.35
1183.36
p5 689.95
p1 592.10
b6+ - H2O
b7+
y5+ 792.29
y6+ 863.31
y7+ 1010.32
b2+ 214.97
b4+ 387.09
b5+ 534.31
587.31
b6+
652.31
y4+ 721.35
793.31
864.55
1011.44
y8+ 1125.26
1184.31
722.33
805.40
869.03
982.58
1012.42
1232.31
211.43
272.22
368.88
388.19
516.19


M V D P E * K * P Q L G M I D R
y12 y11 y10 y8 y6 y5 y4
b2 b3 b7 b10
Relative Abundance
100%
80%
60%
40%
20%
0%
m/z
300
400
500
600
700
800
y11++ 756.08
y12++ 813.56
y12++ - H2O
y11++ - H2O
y4+ - H2O
b3+ 362.19
y8++ 473.39
y12++
p2 - H2O
p3
p1 553.97
y5+ 613.22
b2+ 247.24
b7++ 542.92
607.55
649.99
b10++ 662.11
y10++ 707.42
y6+
804.79
781.63
859.07
868.60
817.56
362.95
375.09
298.41
344.18
428.33

A.1 List of Abbreviations

ACN	-	Acetonitrile
ADP	-	Adenosine diphosphate
AIF	-	All-ion fragmentation
Akt	-	Protein Kinase B
AQUA	-	Absolute quantification
ARH3	-	ADP-ribosylhydrolase 3
ART	-	ADP-ribosyl transferase
atRA	-	All-*trans* retinoic acid
ADH	-	Alcohol dehydrogenases
BCA	-	Bicinchoninic acid
BRCT	-	BRCA1 C terminus
CID	-	Collision induced dissociation
CRABP-II	-	Cellular retinoic acid binding protein 2
CRBP	-	Cellular retinol binding protein
CSI	-	Captive spray ionization
BAD	-	Bcl-2-associated death promoter
DDA	-	Data-dependent acquisition
DE	-	Dynamic exclusion
DIA	-	Data-independent acquisition
DMEM	-	Dulbecco's Modified Eagle Medium
DMSO	-	Dimethyl sulfoxide
DNA	-	Deoxyribonucleic acid
DR	-	Direct repeats
ETD	-	Electron transfer dissociation
ESI	-	Electrospray ionization
FA	-	Formic acid
FABP5	-	Fatty acid binding protein
FBS	-	Fetal bovine serum
FDR	-	False discovery rate
FT-ARM	-	Fourier transform-all reaction monitoring
GPF	-	Gas phase fractionation
HCD	-	Higher-energy collisional dissociation
HepG2	-	Hepatocellular carcinoma cell line
HPLC	-	High-performance liquid chromatography
ICAT	-	Isotope coded affinity tags
ID	-	Inner diameter
IMAC	-	Immobilized metal ion affinity chromatography
IRMPD	-	Infrared multiphoton dissociation
iTRAQ	-	Isobaric tags for relative and absolute quantitation
LC	-	Liquid chromatography
LC-MS/MS	-	Liquid chromatography-tandem mass spectrometric
MCF-7	-	Michigan Cancer Foundation-7 (breast carcinoma cell line)
MeOH	-	Methanol
MNNG	-	Methylnitronitrosoguanidine

MRM	-	Multiple reaction monitoring
mRNA	-	Messenger ribonucleic acid
MS	-	Mass spectrometry
mTOR	-	Mammalian target of rapamycin
NAD	-	Nicotinamide adenine dinucleotide
nESI	-	Nano-Elctrospray ionization
PAcIFIC	-	Precursor acquisition independent from ion count
pADPr	-	Poly(ADP-ribose)
PARG	-	Poly(ADP-ribose) glycohydrolase
PARP-1	-	Poly(ADP-ribose) polymerase-1
PBS	-	Phosphate buffered saline
PDE	-	Phosphodiesterase
PI3K	-	Phosphatidylinositol 3-kinase
PKB	-	Protein Kinase B
PPAR-β/δ	-	Peroxisome proliferator-activated receptor beta/delta
PRAMP	-	Phosphoribosyl-AMP
PR_2AMP	-	Di(phosphoribosyl)-AMP
PSI	-	Pounds per square inch
qPAcIFIC	-	Quantitative PAcIFIC
RALDH	-	Retinaldehyde dehydrogenase
RAR	-	Retinoic acid receptor
RARE	-	Retinoic acid response element
RBP	-	Retinol binding protein
RDH	-	Retinol dehydrogenases
RNA	-	Ribonucleic acid
RPM	-	Revolutions per minute
RXR	-	Retinoid X receptor
SILAC	-	Stable isotope labeling by amino acids in cell culture
S/N	-	Signal to noise ratio
SRM	-	Selected reaction monitoring
STRA6	-	Stimulated by retinoic acid gene 6 protein
TCEP	-	Tris(2-carboxyethyl)phosphine
TFA	-	Trifluoroacetic acid
TLC	-	Thin layer chromatography
TMT	-	Tandem mass tags
TOF	-	Time of flight
XDIA	-	Extended data-independent acquisition
YEPD	-	Yeast extract peptone dextrose
2-DE	-	Two-dimensional gel electrophoresis
2D-DIGE	-	Two-dimensional difference gel electrophoresis
5'-AMP	-	5'-adenosine monophosphate

*Protein and gene names discussed in Chapter 3 in relation to Table 1 are not listed in this list. Please refer to Chapter 3, Table 1, page 89-91.

A.2 Bibliography

References:

1. Valcarcel, R., Holz, H., Jimenez, C. G., Barettino, D., and Stunnenberg, H. G. (1994) Retinoid-dependent in vitro transcription mediated by the RXR/RAR heterodimer, *Genes Dev 8*, 3068-3079.

2. Maiani, G., Caston, M. J., Catasta, G., Toti, E., Cambrodon, I. G., Bysted, A., Granado-Lorencio, F., Olmedilla-Alonso, B., Knuthsen, P., Valoti, M., Bohm, V., Mayer-Miebach, E., Behsnilian, D., and Schlemmer, U. (2009) Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans, *Mol Nutr Food Res 53 Suppl 2*, S194-218.

3. Soprano, D. R., Qin, P., and Soprano, K. J. (2004) Retinoic acid receptors and cancers, *Annu Rev Nutr 24*, 201-221.

4. Zancai, P., Dal Col, J., Piccinin, S., Guidoboni, M., Cariati, R., Rizzo, S., Boiocchi, M., Maestro, R., and Dolcetti, R. (2005) Retinoic acid stabilizes p27Kip1 in EBV-immortalized lymphoblastoid B cell lines through enhanced proteasome-dependent degradation of the p45Skp2 and Cks1 proteins, *Oncogene 24*, 2483-2494.

5. Olson, J. A. (1994) Absorption, transport and metabolism of carotenoids in humans, *Pure Appl Chem 66*, 1011-1016.

6. Harrison, E. H. (2005) Mechanisms of digestion and absorption of dietary vitamin A, *Annu Rev Nutr 25*, 87-103.

7. Hodam, J. R., and Creek, K. E. (1998) Comparison of the metabolism of retinol delivered to human keratinocytes either bound to serum retinol-binding protein or added directly to the culture medium, *Exp Cell Res 238*, 257-264.

8. Kawaguchi, R., Zhong, M., Kassai, M., Ter-Stepanian, M., and Sun, H. (2012) STRA6-catalyzed vitamin A influx, efflux, and exchange, *J Membr Biol 245*, 731-745.

9. Napoli, J. L. (2000) Retinoic acid: Its biosynthesis and metabolism, *Prog Nucleic Acid Re 63*, 139-188.

10. Duester, G. (1996) Involvement of alcohol dehydrogenase, short-chain dehydrogenase/reductase, aldehyde dehydrogenase, and cytochrome P450 in the control of retinoid signaling by activation of retinoic acid synthesis, *Biochemistry 35*, 12221-12227.

11. Chen, H., Howald, W. N., and Juchau, M. R. (2000) Biosynthesis of all-trans-retinoic acid from all-trans-retinol: catalysis of all-trans-retinol oxidation by human P-450 cytochromes, *Drug Metab Dispos 28*, 315-322.

12. Budhu, A. S., and Noy, N. (2002) Direct channeling of retinoic acid between cellular retinoic acid-binding protein II and retinoic acid receptor sensitizes mammary carcinoma cells to retinoic acid-induced growth arrest, *Mol Cell Biol 22*, 2632-2641.

13. Tan, N. S., Shaw, N. S., Vinckenbosch, N., Liu, P., Yasmin, R., Desvergne, B., Wahli, W., and Noy, N. (2002) Selective Cooperation between Fatty Acid Binding Proteins and Peroxisome Proliferator-Activated Receptors in Regulating Transcription, *Mol Cell Biol 22*, 5114-5127.

14. Ray, W. J., Bain, G., Yao, M., and Gottlieb, D. I. (1997) CYP26, a novel mammalian cytochrome P450, is induced by retinoic acid and defines a new family, *J Biol Chem 272*, 18702-18708.

15. Hill, D. L., and Struck, R. F. (1983) Pharmacologic disposition of chemopreventive retinoids (review), *Anticancer Res 3*, 171-180.

16. Wolf, G. (1996) A history of vitamin A and retinoids, *FASEB J 10*, 1102-1107.

17. Wang, K. C., Cheng, A. L., Chuang, S. E., Hsu, H. C., and Su, I. J. (2000) Retinoic acid-induced apoptotic pathway in T-cell lymphoma: Identification of four groups of genes with differential biological functions, *Exp Hematol 28*, 1441-1450.

18. Orfanos, C. E., Zouboulis, C. C., AlmondRoesler, B., and Geilen, C. C. (1997) Current use and future potential role of retinoids in dermatology, *Drugs 53*, 358-388.

19. Goldfarb, M. T., Ellis, C. N., and Voorhees, J. J. (1987) Retinoids in dermatology, *Mayo Clin Proc 62*, 1161-1164.

20. Huang, M. E., Ye, Y. C., Chen, S. R., Chai, J. R., Lu, J. X., Zhoa, L., Gu, L. J., and Wang, Z. Y. (1988) Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia, *Blood 72*, 567-572.

21. Matthay, K. K., Villablanca, J. G., Seeger, R. C., Stram, D. O., Harris, R. E., Ramsay, N. K., Swift, P., Shimada, H., Black, C. T., Brodeur, G. M., Gerbing, R. B., and Reynolds, C. P. (1999) Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. Children's Cancer Group, *N Engl J Med 341*, 1165-1173.

22. Larsen, F. G., Jakobsen, P., Knudsen, J., Weismann, K., Kragballe, K., and Nielsenkudsk, F. (1993) Conversion of acitretin to etretinate in psoriatic patients in influenced by ethanol, *J Invest Dermatol 100*, 623-627.

23. Duvic, M., Hymes, K., Heald, P., Breneman, D., Martin, A. G., Myskowski, P., Crowley, C., Yocum, R. C., and Bexarotene Worldwide Study, G. (2001) Bexarotene is effective and safe for treatment of refractory advanced-stage cutaneous T-cell lymphoma: Multinational phase II-III trial results, *J Clin Oncol 19*, 2456-2471.

24. Fenaux, P. (1993) The role of all-trans-retinoic acid in the treatment of acute promyelocytic leukemia, *Acta Haematol 89 Suppl 1*, 22-27.

25. Cheng, A. L., Su, I. J., Chen, C. C., Tien, H. F., Lay, J. D., Chen, B. R., Pu, Y. S., Hang, R. L., Shen, M. C., Wang, C. H., and Chen, Y. C. (1994) Use of retinoic acids in the treatment of peripheral T-cell lymphoma - A pilot study, *J Clin Oncol 12*, 1185-1192.

26. Coffey, D. C., Kutko, M. C., Glick, R. D., Butler, L. M., Heller, G., Rifkind, R. A., Marks, P. A., Richon, V. M., and La Quaglia, M. P. (2001) The histone deacetylase inhibitor, CBHA, inhibits growth of human neuroblastoma xenografts in vivo, alone and synergistically with all-trans retinoic acid, *Cancer Res 61*, 3591-3594.

27. Culine, S., Kramar, A., Droz, J. P., and Theodore, C. (1999) Phase II study of all-trans retinoic acid administered intermittently for hormone refractory prostate cancer, *J Urol 161*, 173-175.

28. Petkovich, M., Brand, N. J., Krust, A., and Chambon, P. (1987) A human retinoic acid receptor which belongs to the family of nuclear receptors, *Nature 330*, 444-450.

29. Brand, N., Petkovich, M., Krust, A., Chambon, P., Dethe, H., Marchio, A., Tiollais, P., and Dejean, A. (1988) Identification of a 2nd human retinoic acid receptor, *Nature 332*, 850-853.

30. Krust, A. (1989) A third human retinoic acid receptor, hRAR-gamma, *PNAS 86*, 5310-5314.

31. Zhang, X. K., Hoffmann, B., Tran, P. B., Graupner, G., and Pfahl, M. (1992) Retinoid X receptor is an auxiliary protein for thyroid hormone and retinoic acid receptors, *Nature 355*, 441-446.

32. de The, H., Vivanco-Ruiz, M. M., Tiollais, P., Stunnenberg, H., and Dejean, A. (1990) Identification of a retinoic acid responsive element in the retinoic acid receptor beta gene, *Nature 343*, 177-180.

33. Donato, L. J., Suh, J. H., and Noy, N. (2007) Suppression of mammary carcinoma cell growth by retinoic acid: the cell cycle control gene Btg2 is a direct target for retinoic acid receptor signaling, *Cancer Res 67*, 609-615.

34. Gibbs, S., Backendorf, C., and Ponec, M. (1996) Regulation of keratinocyte proliferation and differentiation by all-trans-retinoic acid, 9-cis-retinoic acid and 1,25-dihydroxy vitamin D3, *Arch Derm Res 288*, 729-738.

35. Schug, T. T., Berry, D. C., Shaw, N. S., Travis, S. N., and Noy, N. (2007) Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors, *Cell 129*, 723-733.

36. Juge-Aubry, C. (1997) DNA binding properties of peroxisome proliferator-activated receptor subtypes on various natural peroxisome proliferator response elements, *J Biol Chem 272*, 25252-25259.

37. Tachibana, K., Yamasaki, D., Ishimoto, K., and Doi, T. (2008) The role of PPARs in cancer, *PPAR Res 2008*, 102737.

38. Wagner, K. D., and Wagner, N. (2010) Peroxisome proliferator-activated receptor beta/delta (PPARbeta/delta) acts as regulator of metabolism linked to multiple cellular functions, *Pharmacol Ther 125*, 423-435.

39. Yu, S., Levi, L., Siegel, R., and Noy, N. (2012) Retinoic acid induces neurogenesis by activating both retinoic acid receptors (RARs) and peroxisome proliferator-activated receptor beta/delta (PPARbeta/delta), *J Biol Chem 287*, 42195-42205.

40. Fresno Vara, J. A., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C., and Gonzalez-Baron, M. (2004) PI3K/Akt signalling pathway and cancer, *Cancer Treat Rev 30*, 193-204.

41. Carnero, A. (2010) The PKB/AKT pathway in cancer, *Curr Pharm Des 16*, 34-44.

42. Noy, N. (2010) Between death and survival: retinoic acid in regulation of apoptosis, *Annu Rev Nutr 30*, 201-217.

43. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics, *Nature 422*, 198-207.

44. Trojanowicz, B., Sekulla, C., Lorenz, K., Kohrle, J., Finke, R., Dralle, H., and Hoang-Vu, C. (2010) Proteomic approach reveals novel targets for retinoic acid-mediated therapy of thyroid carcinoma, *Mol Cell Endocrinol 325*, 110-117.

45. Harris, M. N., Ozpolat, B., Abdi, F., Gu, S., Legler, A., Mawuenyega, K. G., Tirado-Gomez, M., Lopez-Berestein, G., and Chen, X. (2004) Comparative proteomic analysis of all-trans-retinoic acid treatment reveals systematic posttranscriptional control mechanisms in acute promyelocytic leukemia, *Blood 104*, 1314-1323.

46. Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fey, S. J., Larsen, P. M., and Gorg, A. (2000) Towards higher resolution: Two-dimensional electrophoresis of Saccharomyces cerevisiae proteins using overlapping narrow immobilized pH gradients, *Electrophoresis 21*, 2610-2616.

47. Voss, T., and Haberl, P. (2000) Observations on the reproducibility and matching efficiency of two-dimensional electrophoresis gels: consequences for comprehensive data analysis, *Electrophoresis 21*, 3345-3350.

48. Marouga, R., David, S., and Hawkins, E. (2005) The development of the DIGE system: 2D fluorescence difference gel analysis technology, *Anal Bioanal Chem 382*, 669-678.

49. Unlu, M., Morgan, M. E., and Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts, *Electrophoresis 18*, 2071-2077.

50. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review, *Anal Bioanal Chem 389*, 1017-1031.

51. Nilsson, T., Mann, M., Aebersold, R., Yates, J. R., 3rd, Bairoch, A., and Bergeron, J. J. (2010) Mass spectrometry in high-throughput proteomics: ready for the big time, *Nat Methods 7*, 681-685.

52. Cravatt, B. F., Simon, G. M., and Yates, J. R., 3rd. (2007) The biological impact of mass-spectrometry-based proteomics, *Nature 450*, 991-1000.

53. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed, *Mol Cell Proteomics 8*, 2759-2769.

54. Belov, M. E., Gorshkov, M. V., Udseth, H. R., Anderson, G. A., Tolmachev, A. V., Prior, D. C., Harkewicz, R., and Smith, R. D. (2000) Initial implementation of an electrodynamic ion funnel with Fourier transform ion cyclotron resonance mass spectrometry, *J Am Soc Mass Spectrom 11*, 19-23.

55. Shen, Y., Tolic, N., Masselon, C., Pasa-Tolic, L., Camp, D. G., 2nd, Hixson, K. K., Zhao, R., Anderson, G. A., and Smith, R. D. (2004) Ultrasensitive proteomics using high-efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS, *Anal Chem 76*, 144-154.

56. Shen, Y., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., Zhao, R., Livesay, E. A., Udseth, H. R., and Smith, R. D. (2005) Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics, *Anal Chem 77*, 3090-3100.

57. Wilkins, M. (2009) Proteomics data mining, *Expert Rev Proteomics 6*, 599-603.

58. Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., and Humphery-Smith, I. (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium, *Electrophoresis 16*, 1090-1094.

59. Michalski, A., Cox, J., and Mann, M. (2011) 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 10*, 1785-1793.

60. Stasyk, T., and Huber, L. A. (2004) Zooming in: fractionation strategies in proteomics, *Proteomics 4*, 3704-3716.

61. Gilar, M., Olivova, P., Chakraborty, A. B., Jaworski, A., Geromanos, S. J., and Gebler, J. C. (2009) Comparison of 1-D and 2-D LC MS/MS methods for proteomic analysis of human serum, *Electrophoresis 30*, 1157-1167.

62. Courchesne, P. L., Jones, M. D., Robinson, J. H., Spahr, C. S., McCracken, S., Bentley, D. L., Luethy, R., and Patterson, S. D. (1998) Optimization of capillary chromatography ion trap-mass spectrometry for identification of gel-separated proteins, *Electrophoresis 19*, 956-967.

63. Wang, N., and Li, L. (2008) Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis, *Anal Chem 80*, 4696-4710.

64. Berg, M., Parbel, A., Pettersen, H., Fenyo, D., and Bjorkesten, L. (2006) Reproducibility of LC-MS-based protein identification, *J Exp Bot 57*, 1509-1514.

65. Stalder, D., Haeberli, A., and Heller, M. (2008) Evaluation of reproducibility of protein identification results after multidimensional human serum protein separation, *Proteomics 8*, 414-424.

66. Bern, M., Finney, G., Hoopmann, M. R., Merrihew, G., Toth, M. J., and MacCoss, M. J. (2010) Deconvolution of mixture spectra from ion-trap data-independent-acquisition tandem mass spectrometry, *Anal Chem 82*, 833-841.

67. Masselon, C., Anderson, G. A., Harkewicz, R., Bruce, J. E., Pasa-Tolic, L., and Smith, R. D. (2000) Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures, *Anal Chem 72*, 1918-1924.

68. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., and Yates, J. R. (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra, *Nat Methods 1*, 39-45.

69. Purvine, S., Eppel, J. T., Yi, E. C., and Goodlett, D. R. (2003) Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer, *Proteomics 3*, 847-850.

70. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database, *J Am Soc Mass Spec 5*, 976-989.

71. Conboy, J. J., and Henion, J. D. (1992) The determination of glycopeptides by liquid chromatography/mass spectrometry with collision-induced dissociation, *J Am Soc Mass Spec 3*, 804-814.

72. Huddleston, M. J., Bean, M. F., and Carr, S. A. (1993) Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests, *Anal Chem 65*, 877-884.

73. van Dongen, W. D., van Wijk, J. I. T., Green, B. N., Heerma, W., and Haverkamp, J. (1999) Comparison between collision induced dissociation of electrosprayed protonated peptides in the up‐front source region and in a low‐energy collision cell, *Rapid Commun Mass Spectrom 13*, 1712-1716.

74. Hakansson, K., Zubarev, R., and Hakansson, P. (1998) Combination of nozzle-skimmer fragmentation and partial acid hydrolysis in electrospray ionization time-of-flight mass spectrometry of synthetic peptides, *Rapid Commun Mass Spectrom 12*, 705-711.

75. Ramos, A. A., Yang, H., Rosen, L. E., and Yao, X. (2006) Tandem parallel fragmentation of peptides for mass spectrometry, *Anal Chem 78*, 6391-6397.

76. Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., and Geromanos, S. (2005) Quantitative proteomic analysis by accurate mass retention time pairs, *Anal Chem 77*, 2187-2200.

77. Hughes, M. A., Silva, J. C., Geromanos, S. J., and Townsend, C. A. (2006) Quantitative proteomic analysis of drug-induced changes in mycobacteria, *J Proteome Res 5*, 54-63.

78. Mbeunkui, F., Scholl, E. H., Opperman, C. H., Goshe, M. B., and Bird, D. M. (2010) Proteomic and bioinformatic analysis of the root-knot nematode Meloidogyne hapla: the basis for plant parasitism, *J Proteome Res 9*, 5370-5381.

79. Li, G. Z., Vissers, J. P., Silva, J. C., Golick, D., Gorenstein, M. V., and Geromanos, S. J. (2009) Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures, *Proteomics 9*, 1696-1719.

80. Plumb, R. S., Rainville, P. D., Potts, W. B., 3rd, Johnson, K. A., Gika, E., and Wilson, I. D. (2009) Application of ultra performance liquid chromatography-mass spectrometry to profiling rat and dog bile, *J Proteome Res 8*, 2495-2500.

81. Cortezzi, S. S., Garcia, J. S., Ferreira, C. R., Braga, D. P., Figueira, R. C., Iaconelli, A., Jr., Souza, G. H., Borges, E., Jr., and Eberlin, M. N. (2011) Secretome of the preimplantation human embryo by bottom-up label-free proteomics, *Anal Bioanal Chem 401*, 1331-1339.

82. D'Aguanno, S., D'Alessandro, A., Pieroni, L., Roveri, A., Zaccarin, M., Marzano, V., De Canio, M., Bernardini, S., Federici, G., and Urbani, A. (2011) New insights into neuroblastoma cisplatin resistance: a comparative proteomic and meta-mining investigation, *J Proteome Res 10*, 416-428.

83. Chakraborty, A. B., Berger, S. J., and Gebler, J. C. (2007) Use of an integrated MS--multiplexed MS/MS data acquisition strategy for high-coverage peptide mapping studies, *Rapid Commun Mass Spectrom 21*, 730-744.

84. Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G. Z., Gorenstein, M. V., Bateman, R. H., and Langridge, J. I. (2009) The detection, correlation, and comparison of

peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS, *Proteomics 9*, 1683-1695.

85. Blackburn, K., Mbeunkui, F., Mitra, S. K., Mentzel, T., and Goshe, M. B. (2010) Improving protein and proteome coverage through data-independent multiplexed peptide fragmentation, *J Proteome Res 9*, 3621-3637.

86. Levin, Y., Hradetzky, E., and Bahn, S. (2011) Quantification of proteins using data-independent analysis (MSE) in simple andcomplex samples: a systematic evaluation, *Proteomics 11*, 3273-3287.

87. Hoaglund-Hyzer, C. S., Li, J., and Clemmer, D. E. (2000) Mobility labeling for parallel CID of ion mixtures, *Anal Chem 72*, 2737-2740.

88. Baker, E. S., Tang, K., Danielson, W. F., 3rd, Prior, D. C., and Smith, R. D. (2008) Simultaneous fragmentation of multiple ions using IMS drift time dependent collision energies, *J Am Soc Mass Spectrom 19*, 411-419.

89. Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006) Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer, *J Am Soc Mass Spectrom 17*, 977-982.

90. Geiger, T., Cox, J., and Mann, M. (2010) Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation, *Mol Cell Proteomics 9*, 2252-2261.

91. Panchaud, A., Scherl, A., Shaffer, S. A., von Haller, P. D., Kulasekara, H. D., Miller, S. I., and Goodlett, D. R. (2009) Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean, *Anal Chem 81*, 6481-6488.

92. Carvalho, P. C., Han, X., Xu, T., Cociorva, D., Carvalho Mda, G., Barbosa, V. C., and Yates, J. R., 3rd. (2010) XDIA: improving on the label-free data-independent analysis, *Bioinformatics 26*, 847-848.

93. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Mol Cell Proteomics 11*, O111 016717.

94. Weisbrod, C. R., Eng, J. K., Hoopmann, M. R., Baker, T., and Bruce, J. E. (2012) Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification, *J Proteome Res 11*, 1621-1632.

95. Scherl, A., Shaffer, S. A., Taylor, G. K., Kulasekara, H. D., Miller, S. I., and Goodlett, D. R. (2008) Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides, *Anal Chem 80*, 1182-1191.

96. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K., and Hamon, C. (2003) Tandem mass tags: a novel

quantification strategy for comparative analysis of complex protein mixtures by MS/MS, *Anal Chem 75*, 1895-1904.

97. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents, *Mol Cell Proteomics 3*, 1154-1169.

98. Panchaud, A., Jung, S., Shaffer, S. A., Aitchison, J. D., and Goodlett, D. R. (2011) Faster, quantitative, and accurate precursor acquisition independent from ion count, *Anal Chem 83*, 2250-2257.

99. Scherl, A., Shaffer, S. A., Taylor, G. K., Hernandez, P., Appel, R. D., Binz, P. A., and Goodlett, D. R. (2008) On the benefits of acquiring peptide fragment ions at high measured mass accuracy, *J Am Soc Mass Spectrom 19*, 891-901.

100. Voss, J., Goo, Y. A., Cain, K., Woods, N., Jarrett, M., Smith, L., Shulman, R., and Heitkemper, M. (2011) Searching for the noninvasive biomarker holy grail: are urine proteomics the answer?, *Biol Res Nurs 13*, 235-242.

101. Pak, H., Pasquarello, C., and Scherl, A. (2011) Label-free protein quantification on tandem mass spectra acquired in a data-independent mode provides accurate measurements over five orders of concentration magnitude in complex matrices, pp 211-215, J Integrated Omics.

102. Goo, Y. A., Cain, K., Jarrett, M., Smith, L., Voss, J., Tolentino, E., Tsuji, J., Tsai, Y. S., Panchaud, A., Goodlett, D. R., Shulman, R. J., and Heitkemper, M. (2012) Urinary proteome analysis of irritable bowel syndrome (IBS) symptom subgroups, *J Proteome Res 11*, 5650-5662.

103. Hengel, S. M., Murray, E., Langdon, S., Hayward, L., O'Donoghue, J., Panchaud, A., Hupp, T., and Goodlett, D. R. (2011) Data-independent proteomic screen identifies novel tamoxifen agonist that mediates drug resistance, *J Proteome Res 10*, 4567-4578.

104. Smith, R. D., Pasa-Tolic, L., Lipton, M. S., Jensen, P. K., Anderson, G. A., Shen, Y., Conrads, T. P., Udseth, H. R., Harkewicz, R., Belov, M. E., Masselon, C., and Veenstra, T. D. (2001) Rapid quantitative measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry, *Electrophoresis 22*, 1652-1668.

105. Kasicka, V. (2001) Recent advances in capillary electrophoresis of peptides, *Electrophoresis 22*, 4139-4162.

106. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation, *Proc Natl Acad Sci U S A 96*, 6591-6596.

107. Ong, S. E. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics, *Mol Cell Proteomics 1*, 376-386.

108. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat Biotechnol 17*, 994-999.

109. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS, *Proc Natl Acad Sci U S A 100*, 6940-6945.

110. Kirkpatrick, D. S., Gerber, S. A., and Gygi, S. P. (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications, *Methods 35*, 265-273.

111. Desiderio, D. M., and Kai, M. (1983) Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue, *Biomed Mass Spectrom 10*, 471-479.

112. Liu, H., Sadygov, R. G., and Yates, J. R., 3rd. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics, *Anal Chem 76*, 4193-4201.

113. Tsai, Y. S., Scherl, A., Shaw, J. L., MacKay, C. L., Shaffer, S. A., Langridge-Smith, P. R., and Goodlett, D. R. (2009) Precursor ion independent algorithm for top-down shotgun proteomics, *J Am Soc Mass Spectrom 20*, 2154-2166.

114. Rowen, L., Mahairas, G., and Hood, L. (1997) Sequencing the human genome, *Science 278*, 605-607.

115. Macilwain, C. (2000) World leaders heap praise on human genome landmark, *Nature 405*, 983-984.

116. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei,

M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001) The sequence of the human genome, *Science 291*, 1304-1351.

117. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C.,

Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., and International Human Genome Sequencing, C. (2001) Initial sequencing and analysis of the human genome, *Nature 409*, 860-921.

118. Gregory, S. G., Barlow, K. F., McLay, K. E., Kaul, R., Swarbreck, D., Dunham, A., Scott, C. E., Howe, K. L., Woodfine, K., Spencer, C. C., Jones, M. C., Gillson, C., Searle, S., Zhou, Y., Kokocinski, F., McDonald, L., Evans, R., Phillips, K., Atkinson, A., Cooper, R., Jones, C., Hall, R. E., Andrews, T. D., Lloyd, C., Ainscough, R., Almeida, J. P., Ambrose, K. D., Anderson, F., Andrew, R. W., Ashwell, R. I., Aubin, K., Babbage, A. K., Bagguley, C. L., Bailey, J., Beasley, H., Bethel, G., Bird, C. P., Bray-Allen, S., Brown, J. Y., Brown, A. J., Buckley, D., Burton, J., Bye, J., Carder, C., Chapman, J. C., Clark, S. Y., Clarke, G., Clee, C., Cobley, V., Collier, R. E., Corby, N., Coville, G. J., Davies, J., Deadman, R., Dunn, M., Earthrowl, M., Ellington, A. G., Errington, H., Frankish, A., Frankland, J., French, L., Garner, P., Garnett, J., Gay, L., Ghori, M. R., Gibson, R., Gilby, L. M., Gillett, W., Glithero, R. J., Grafham, D. V., Griffiths, C., Griffiths-Jones, S., Grocock, R., Hammond, S., Harrison, E. S., Hart, E., Haugen, E., Heath, P. D., Holmes, S., Holt, K., Howden, P. J., Hunt, A. R., Hunt, S. E., Hunter, G., Isherwood, J., James, R., Johnson, C., Johnson, D., Joy, A., Kay, M., Kershaw, J. K., Kibukawa, M., Kimberley, A. M., King, A., Knights, A. J., Lad, H., Laird, G., Lawlor, S., Leongamornlert, D. A., Lloyd, D. M., Loveland, J., Lovell, J., Lush, M. J., Lyne, R., Martin, S., Mashreghi-Mohammadi, M., Matthews, L., Matthews, N. S., McLaren, S., Milne, S., Mistry, S., Moore, M. J., Nickerson, T., O'Dell, C. N., Oliver, K., Palmeiri, A., Palmer, S. A., Parker, A., Patel, D., Pearce, A. V., Peck, A. I., Pelan, S., Phelps, K., Phillimore, B. J., Plumb, R., Rajan, J., Raymond, C., Rouse, G., Saenphimmachak, C., Sehra, H. K., Sheridan, E., Shownkeen, R., Sims, S., Skuce, C. D., Smith, M., Steward, C., Subramanian, S., Sycamore, N., Tracey, A., Tromans, A., Van Helmond, Z., Wall, M., Wallis, J. M., White, S., Whitehead, S. L., Wilkinson, J. E., Willey, D. L., Williams, H., Wilming, L., Wray, P. W., Wu, Z., Coulson, A., Vaudin, M., Sulston, J. E., Durbin, R., Hubbard, T., Wooster, R., Dunham, I., Carter, N. P., McVean, G., Ross, M. T., Harrow, J., Olson, M. V., Beck,

S., Rogers, J., Bentley, D. R., Banerjee, R., Bryant, S. P., Burford, D. C., Burrill, W. D., Clegg, S. M., Dhami, P., Dovey, O., Faulkner, L. M., Gribble, S. M., Langford, C. F., Pandian, R. D., Porter, K. M., and Prigmore, E. (2006) The DNA sequence and biological annotation of human chromosome 1, *Nature 441*, 315-321.

119. Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., and Hochstrasser, D. F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis, *Biotechnology (N Y) 14*, 61-65.

120. Wilkins, M. R., Appel, R. D., Van Eyk, J. E., Chung, M. C., Gorg, A., Hecker, M., Huber, L. A., Langen, H., Link, A. J., Paik, Y. K., Patterson, S. D., Pennington, S. R., Rabilloud, T., Simpson, R. J., Weiss, W., and Dunn, M. J. (2006) Guidelines for the next 10 years of proteomics, *Proteomics 6*, 4-8.

121. Domon, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis, *Science 312*, 212-217.

122. Acosta-Martin, A. E., Panchaud, A., Chwastyniak, M., Dupont, A., Juthier, F., Gautier, C., Jude, B., Amouyel, P., Goodlett, D. R., and Pinet, F. (2011) Quantitative mass spectrometry analysis using PAcIFIC for the identification of plasma diagnostic biomarkers for abdominal aortic aneurysm, *PLoS One 6*, e28698.

123. Ramanathan, R., Raghavan, N., Comezoglu, S. N., and Humphreys, W. G. (2011) A low flow ionization technique to integrate quantitative and qualitative small molecule bioanalysis, *Int J Mass Spectrom 301*, 127-135.

124. Dahal, U. P., Jones, J. P., Davis, J. A., and Rock, D. A. (2011) Small molecule quantification by liquid chromatography-mass spectrometry for metabolites of drugs and drug candidates, *Drug Metab Dispos 39*, 2355-2360.

125. Ghaemmaghami, S., Huh, W., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast, *Nature 425*, 737-741.

126. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry, *Anal Chem 75*, 4646-4658.

127. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search, *Anal Chem 74*, 5383-5392.

128. Hoopmann, M. R., Finney, G. L., and MacCoss, M. J. (2007) High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry, *Anal Chem 79*, 5620-5632.

129. Katz, E., Fon, M., Eigenheer, R. A., Phinney, B. S., Fass, J. N., Lin, D., Sadka, A., and Blumwald, E. (2010) A label-free differential quantitative mass spectrometry method for

the characterization and identification of protein changes during citrus fruit development, *Proteome Sci 8*, 68.

130. Wang, X., Stoll, D. R., Schellinger, A. P., and Carr, P. W. (2006) Peak capacity optimization of peptide separations in reversed-phase gradient elution chromatography: fixed column format, *Anal Chem 78*, 3406-3416.

131. Houel, S., Abernathy, R., Renganathan, K., Meyer-Arendt, K., Ahn, N. G., and Old, W. M. (2010) Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies, *J Proteome Res 9*, 4152-4160.

132. Berry, D. C., and Noy, N. (2009) All-trans-retinoic acid represses obesity and insulin resistance by activating both peroxisome proliferation-activated receptor beta/delta and retinoic acid receptor, *Mol Cell Biol 29*, 3286-3296.

133. Mangiarotti, R., Danova, M., Alberici, R., and Pellicciari, C. (1998) All-trans retinoic acid (ATRA)-induced apoptosis is preceded by G1 arrest in human MCF-7 breast cancer cells, *Brit J Cancer 77*, 186-191.

134. Ryu, S., Gallis, B., Goo, Y. A., Shaffer, S. A., Radulovic, D., and Goodlett, D. R. (2008) Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method, *Cancer Inform 6*, 243-255.

135. Laperriere, D., Wang, T. T., White, J. H., and Mader, S. (2007) Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution, *BMC Genomics 8*, 23.

136. Topletz, A. R., Le, H. N., Lee, N., Chapman, J. D., Kelly, E. J., Wang, J., and Isoherranen, N. (2013) Hepatic Cyp2d and Cyp26a1 mRNAs and activities are increased during mouse pregnancy, *Drug Metab Dispos 41*, 312-319.

137. Mosmann, T. (1983) Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays, *J Immunol Methods 65*, 55-63.

138. Hu, J. Z., He, Y., Yan, M., Zhu, C., Ye, W. M., Zhu, H. G., Chen, W. T., Zhang, C. P., and Zhang, Z. Y. (2013) Dose Dependent Activation of Retinoic Acid-Inducible Gene-I Promotes Both Proliferation and Apoptosis Signals in Human Head and Neck Squamous Cell Carcinoma, *Plos One 8*.

139. Adamson, P. C., Balis, F. M., Smith, M. A., Murphy, R. F., Godwin, K. A., and Poplack, D. G. (1992) Dose-dependent pharmacokinetics of all-trans-retinoic acid, *J Natl Cancer Inst 84*, 1332-1335.

140. Serfling, E., Jasin, M., and Schaffner, W. (1985) Enhancers and eukaryotic gene transcription, *Trends Genet 1*, 224-230.

141. Farr, S., and Dunn, R. T., 2nd. (1999) Concise review: gene expression applied to toxicology, *Toxicol Sci 50*, 1-9.

142. Goldberg, A. L. (2003) Protein degradation and protection against misfolded or damaged proteins, *Nature 426*, 895-899.

143. Fanis, P., Gillemans, N., Aghajanirefah, A., Pourfarzad, F., Demmers, J., Esteghamat, F., Vadlamudi, R. K., Grosveld, F., Philipsen, S., and van Dijk, T. B. (2012) Five friends of methylated chromatin target of protein-arginine-methyltransferase[prmt]-1 (chtop), a complex linking arginine methylation to desumoylation, *Mol Cell Proteomics 11*, 1263-1273.

144. Andersson, M. K., Stahlberg, A., Arvidsson, Y., Olofsson, A., Semb, H., Stenman, G., Nilsson, O., and Aman, P. (2008) The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response, *BMC Cell Biol 9*, 37.

145. Kelman, Z. (1997) PCNA: Structure, functions and interactions, *Oncogene 14*, 629-640.

146. Noskov, V. N., Araki, H., and Sugino, A. (1998) The RFC2 gene, encoding the third-largest subunit of the replication factor C complex, is required for an S-phase checkpoint in Saccharomyces cerevisiae, *Mol Cell Bio 18*, 4914-4923.

147. Smallwood, A., Hon, G. C., Jin, F., Henry, R. E., Espinosa, J. M., and Ren, B. (2012) CBX3 regulates efficient RNA processing genome-wide, *Genome Res 22*, 1426-1436.

148. Wang, Y., He, Q. Y., Chen, H., and Chiu, J. F. (2007) Synergistic effects of retinoic acid and tamoxifen on human breast cancer cells: proteomic characterization, *Exp Cell Res 313*, 357-368.

149. Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation, *Cell 108*, 557-572.

150. Helmbrecht, K., Zeise, E., and Rensing, L. (2000) Chaperones in cell cycle regulation and mitogenic signal transduction: a review, *Cell Prolif 33*, 341-365.

151. Voges, D., Zwickl, P., and Baumeister, W. (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis, *Annu Rev Biochem 68*, 1015-1068.

152. Miyauchi, Y., Michigami, T., Sakaguchi, N., Sekimoto, T., Yoneda, Y., Pike, J. W., Yamagata, M., and Ozono, K. (2005) Importin 4 is responsible for ligand-independent nuclear translocation of vitamin D receptor, *J Biol Chem 280*, 40901-40908.

153. Mahajan, R., Delphin, C., Guan, T. L., Gerace, L., and Melchior, F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2, *Cell 88*, 97-107.

154. Pichler, A., Gast, A., Seeler, J. S., Dejean, A., and Melchior, F. (2002) The nucleoporin RanBP2 has SUMO1 E3 ligase activity, *Cell 108*, 109-120.

155. Schoonjans, K., Staels, B., and Auwerx, J. (1996) The peroxisome proliferator activated receptors (PPARs) and their effects on lipid metabolism and adipocyte differentiation, *Biochimica Et Biophysica Acta-Lipids and Lipid Metabolism 1302*, 93-109.

156. Ricote, M., Valledor, A. F., and Glass, C. K. (2004) Decoding transcriptional programs regulated by PPARs and LXRs in the macrophage: effects on lipid homeostasis, inflammation, and atherosclerosis, *Arterioscler Thromb Vasc Biol 24*, 230-239.

157. Makde, R. D., England, J. R., Yennawar, H. P., and Tan, S. (2010) Structure of RCC1 chromatin factor bound to the nucleosome core particle, *Nature 467*, 562-566.

158. Zhu, W. Y., Jones, C. S., Kiss, A., Matsukuma, K., Amin, S., and De Luca, L. M. (1997) Retinoic acid inhibition of cell cycle progression in MCF-7 human breast cancer cells, *Experimental Cell Research 234*, 293-299.

159. Dorner, D., Vlcek, S., Foeger, N., Gajewski, A., Makolm, C., Gotzmann, J., Hutchison, C. J., and Foisner, R. (2006) Lamina-associated polypeptide 2alpha regulates cell cycle progression and differentiation via the retinoblastoma-E2F pathway, *J Cell Biol 173*, 83-93.

160. Dechat, T., Vlcek, S., and Foisner, R. (2000) Review: lamina-associated polypeptide 2 isoforms and related proteins in cell cycle-dependent nuclear structure dynamics, *J Struct Biol 129*, 335-345.

161. Weber, P. J., Eckhard, C. P., Gonser, S., Otto, H., Folkers, G., and Beck-Sickinger, A. G. (1999) On the role of thymopoietins in cell proliferation. Immunochemical evidence for new members of the human thymopoietin family, *Biol Chem 380*, 653-660.

162. Boutros, R., and Byrne, J. A. (2005) D53 (TPD52L1) is a cell cycle-regulated protein maximally expressed at the G2-M transition in breast cancer cells, *Exp Cell Res 310*, 152-165.

163. Mendez, J., and Stillman, B. (2000) Chromatin association of human origin recognition complex, Cdc6, and minichromosome maintenance proteins during the cell cycle: Assembly of prereplication complexes in late mitosis, *Mol Cell Biol 20*, 8602-8612.

164. Varisli, L., Gonen-Korkmaz, C., Debelec-Butuner, B., Erbaykent-Tepedelen, B., Muhammed, H. S., Bogurcu, N., Saatcioglu, F., and Korkmaz, K. S. (2011) Ubiquitously expressed hematological and neurological expressed 1 downregulates Akt-mediated GSK3beta signaling, and its knockdown results in deregulated G2/M transition in prostate cells, *DNA Cell Biol 30*, 419-429.

165. Laughlin, K. M., Luo, D. F., Liu, C., Shaw, G., Warrington, K. H., Law, B. K., and Harrison, J. K. (2009) Hematopoietic- and neurologic-expressed sequence 1 (Hn1) depletion in B16.F10 melanoma cells promotes a differentiated phenotype that includes increased melanogenesis and cell cycle arrest, *Differentiation 78*, 35-44.

166. Montagnoli, A., Valsasina, B., Brotherton, D., Troiani, S., Rainoldi, S., Tenca, P., Molinari, A., and Santocanale, C. (2006) Identification of Mcm2 phosphorylation sites by S-phase-regulating kinases, *J Biol Chem 281*, 10281-10290.

167. Azuma, Y., Tan, S. H., Cavenagh, M. M., Ainsztein, A. M., Saitoh, H., and Dasso, M. (2001) Expression and regulation of the mammalian SUMO-1 E1 enzyme, *FASEB J 15*, 1825-1827.

168. Moore, J. D. (2001) The Ran-GTPase and cell-cycle control, *Bioessays 23*, 77-85.

169. Maddika, S., Ande, S. R., Panigrahi, S., Paranjothy, T., Weglarczyk, K., Zuse, A., Eshraghi, M., Manda, K. D., Wiechec, E., and Los, M. (2007) Cell survival, cell death and cell cycle pathways are interconnected: implications for cancer therapy, *Drug Resist Updat 10*, 13-29.

170. Barnes, C. J., Li, F., Mandal, M., Yang, Z., Sahin, A. A., and Kumar, R. (2002) Heregulin induces expression, ATPase activity, and nuclear localization of G3BP, a Ras signaling component, in human breast tumors, *Cancer Res 62*, 1251-1255.

171. Qiu, X. B., Lin, Y. L., Thome, K. C., Pian, P., Schlegel, B. P., Weremowicz, S., Parvin, J. D., and Dutta, A. (1998) An eukaryotic RuvB-like protein (RUVBL1) essential for growth, *J Biol Chem 273*, 27786-27793.

172. Lee, S. M., Koh, H. J., Park, D. C., Song, B. J., Huh, T. L., and Park, J. W. (2002) Cytosolic NADP(+)-dependent isocitrate dehydrogenase status modulates oxidative damage to cells, *Free Radic Biol Med 32*, 1185-1196.

173. Xie, H., Hu, Z., Chyna, B., Horrigan, S. K., and Westbrook, C. A. (2000) Human mortalin (HSPA9): a candidate for the myeloid leukemia tumor suppressor gene on 5q31, *Leukemia 14*, 2128-2134.

174. Desagher, S. (1999) Bid-induced Conformational Change of Bax Is Responsible for Mitochondrial Cytochrome c Release during Apoptosis, *J Cell Biol 144*, 891-901.

175. Satelli, A., and Rao, U. S. (2011) Galectin-1 is silenced by promoter hypermethylation and its re-expression induces apoptosis in human colorectal cancer cells, *Cancer Lett 301*, 38-46.

176. Stanelle, J., and Putzer, B. M. (2006) E2F1-induced apoptosis: turning killers into therapeutics, *Trends Mol Med 12*, 177-185.

177. Cao, W., Liu, N., Tang, S., Bao, L., Shen, L., Yuan, H., Zhao, X., and Lu, H. (2008) Acetyl-coenzyme A acyltransferase 2 attenuates the apoptotic effects of BNIP3 in two human cell lines, *Biochimica Et Biophysica Acta-General Subjects 1780*, 873-880.

178. Raines, E., Dower, S., and Ross, R. (1989) Interleukin-1 mitogenic activity for fibroblasts and smooth muscle cells is due to PDGF-AA, *Science 243*, 393-396.

179. Reott, M. A., Parker, A. C., Rocha, E. R., and Smith, C. J. (2009) Thioredoxins in redox maintenance and survival during oxidative stress of Bacteroides fragilis, *J Bacteriol 191*, 3384-3391.

180. Wei, J. G., Ji, H. S., Guo, M. L., and Qin, Q. W. (2012) Isolation and characterization of a thioredoxin domain-containing protein 12 from orange-spotted grouper, Epinephelus coioides, *Fish & Shellfish Immunology 33*, 667-673.

181. de Groot, J. W., Links, T. P., Plukker, J. T., Lips, C. J., and Hofstra, R. M. (2006) RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors, *Endocr Rev 27*, 535-560.

182. Saitoh, M., Nishitoh, H., Fujii, M., Takeda, K., Tobiume, K., Sawada, Y., Kawabata, M., Miyazono, K., and Ichijo, H. (1998) Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1, *EMBO J 17*, 2596-2606.

183. Wang, Z., Hao, Y., and Lowe, A. W. (2008) The adenocarcinoma-associated antigen, AGR2, promotes tumor growth, cell migration, and cellular transformation, *Cancer Res 68*, 492-497.

184. Hu, Z. Y., Gu, Y. Y., Han, B., Zhang, J. S., Li, Z. L., Tian, K. L., Young, C. Y. F., and Yuan, H. Q. (2012) Knockdown of AGR2 induces cellular senescence in prostate cancer cells, *Carcinogenesis 33*, 1178-1186.

185. Zechel, C., Shen, X. Q., Chen, J. Y., Chen, Z. P., Chambon, P., and Gronemeyer, H. (1994) The dimerization interfaces formed between the DNA-binding domains of RXR, RAR, and TR determine the binding-specificity and polarity of the full-length receptors to direct repeats *EMBO J 13*, 1425-1433.

186. Kliewer, S. A., Umesono, K., Noonan, D. J., Heyman, R. A., and Evans, R. M. (1992) Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors, *Nature 358*, 771-774.

187. Rastinejad, F., Wagner, T., Zhao, Q., and Khorasanizadeh, S. (2000) Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1, *Embo Journal 19*, 1045-1054.

188. Burkle, A. (2005) Poly(ADP-ribose). The most elaborate metabolite of NAD+, *FEBS J 272*, 4576-4589.

189. D'Amours, D., Desnoyers, S., D'Silva, I., and Poirier, G. G. (1999) Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions, *Biochem J 342 (Pt 2)*, 249-268.

190. Ali, A. A., Timinszky, G., Arribas-Bosacoma, R., Kozlowski, M., Hassa, P. O., Hassler, M., Ladurner, A. G., Pearl, L. H., and Oliver, A. W. (2012) The zinc-finger domains of PARP1 cooperate to recognize DNA strand breaks, *Nat Struct Mol Biol 19*, 685-692.

191. Altmeyer, M., Messner, S., Hassa, P. O., Fey, M., and Hottiger, M. O. (2009) Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites, *Nucleic Acids Res 37*, 3723-3738.

192. Langelier, M. F., Planck, J. L., Roy, S., and Pascal, J. M. (2012) Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1, *Science 336*, 728-732.

193. de Murcia, G., Huletsky, A., Lamarre, D., Gaudreau, A., Pouyet, J., Daune, M., and Poirier, G. G. (1986) Modulation of chromatin superstructure induced by poly(ADP-ribose) synthesis and degradation, *J Biol Chem 261*, 7011-7017.

194. Poirier, G. G., de Murcia, G., Jongstra-Bilen, J., Niedergang, C., and Mandel, P. (1982) Poly(ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure, *Proc Natl Acad Sci U S A 79*, 3423-3427.

195. Rouleau, M., Aubin, R. A., and Poirier, G. G. (2004) Poly(ADP-ribosyl)ated chromatin domains: access granted, *J Cell Sci 117*, 815-825.

196. Malanga, M., and Althaus, F. R. (2005) The role of poly(ADP-ribose) in the DNA damage signaling network, *Biochem Cell Biol 83*, 354-364.

197. Hilz, H. (1981) ADP-ribosylation of proteins--a multifunctional process, *Hoppe Seylers Z Physiol Chem 362*, 1415-1425.

198. Riquelme, P. T., Burzio, L. O., and Koide, S. S. (1979) ADP ribosylation of rat liver lysine-rich histone in vitro, *J Biol Chem 254*, 3018-3028.

199. Burzio, L. O., Riquelme, P. T., and Koide, S. S. (1979) ADP ribosylation of rat liver nucleosomal core histones, *J Biol Chem 254*, 3029-3037.

200. Ogata, N., Ueda, K., and Hayaishi, O. (1980) ADP-ribosylation of histone H2B. Identification of glutamic acid residue 2 as the modification site, *J Biol Chem 255*, 7610-7615.

201. Ogata, N., Ueda, K., Kagamiyama, H., and Hayaishi, O. (1980) ADP-ribosylation of histone H1. Identification of glutamic acid residues 2, 14, and the COOH-terminal lysine residue as modification sites, *J Biol Chem 255*, 7616-7620.

202. Kawaichi, M., Ueda, K., and Hayaishi, O. (1981) Multiple autopoly(ADP-ribosyl)ation of rat liver poly(ADP-ribose) synthetase. Mode of modification and properties of automodified synthetase, *J Biol Chem 256*, 9483-9489.

203. Haenni, S. S., Hassa, P. O., Altmeyer, M., Fey, M., Imhof, R., and Hottiger, M. O. (2008) Identification of lysines 36 and 37 of PARP-2 as targets for acetylation and auto-ADP-ribosylation, *Int J Biochem Cell Biol 40*, 2274-2283.

204. Messner, S., Altmeyer, M., Zhao, H., Pozivil, A., Roschitzki, B., Gehrig, P., Rutishauser, D., Huang, D., Caflisch, A., and Hottiger, M. O. (2010) PARP1 ADP-ribosylates lysine residues of the core histone tails, *Nucleic Acids Res 38*, 6350-6362.

205. Adamietz, P., and Hilz, H. (1976) Poly(adenosine diphosphate ribose) is covalently linked to nuclear proteins by two types of bonds, *Hoppe Seylers Z Physiol Chem 357*, 527-534.

206. Hottiger, M. O. (2011) ADP-ribosylation of histones by ARTD1: an additional module of the histone code?, *FEBS Lett 585*, 1595-1599.

207. Pic, E., Gagne, J. P., and Poirier, G. G. (2011) Mass spectrometry-based functional proteomics of poly(ADP-ribose) polymerase-1, *Expert Rev Proteomics 8*, 759-774.

208. Hengel, S. M., Shaffer, S. A., Nunn, B. L., and Goodlett, D. R. (2009) Tandem mass spectrometry investigation of ADP-ribosylated kemptide, *J Am Soc Mass Spectrom 20*, 477-483.

209. Margarit, S. M., Davidson, W., Frego, L., and Stebbins, C. E. (2006) A steric antagonism of actin polymerization by a salmonella virulence protein, *Structure 14*, 1219-1229.

210. Osago, H., Yamada, K., Shibata, T., Yoshino, K., Hara, N., and Tsuchiya, M. (2009) Precursor ion scanning and sequencing of arginine-ADP-ribosylated peptide by mass spectrometry, *Anal Biochem 393*, 248-254.

211. Zee, B. M., and Garcia, B. A. (2010) Electron transfer dissociation facilitates sequencing of adenosine diphosphate-ribosylated peptides, *Anal Chem 82*, 28-31.

212. Oetjen, J., Rexroth, S., and Reinhold-Hurek, B. (2009) Mass spectrometric characterization of the covalent modification of the nitrogenase Fe-protein in Azoarcus sp. BH72, *FEBS J 276*, 3618-3627.

213. Marsischky, G. T., Wilson, B. A., and Collier, R. J. (1995) Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation. Evidence for active site similarities to the ADP-ribosylating toxins, *J Biol Chem 270*, 3247-3254.

214. Rolli, V., O'Farrell, M., Menissier-de Murcia, J., and de Murcia, G. (1997) Random mutagenesis of the poly(ADP-ribose) polymerase catalytic domain reveals amino acids involved in polymer branching, *Biochemistry 36*, 12147-12154.

215. Tao, Z., Gao, P., and Liu, H. W. (2009) Identification of the ADP-ribosylation sites in the PARP-1 automodification domain: analysis and implications, *J Am Chem Soc 131*, 14258-14260.

216. Eng, J. K., McCormack, A. L., and Yates Iii, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *Journal of the American Society for Mass Spectrometry 5*, 976-989.

217. Gagne, J. P., Isabelle, M., Lo, K. S., Bourassa, S., Hendzel, M. J., Dawson, V. L., Dawson, T. M., and Poirier, G. G. (2008) Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes, *Nucleic Acids Res 36*, 6959-6976.

218. Oka, J., Ueda, K., and Hayaishi, O. (1978) Snake venom phosphodiesterase: simple purification with Blue Sepharose and its application to poly(ADP-ribose) study, *Biochem Biophys Res Commun 80*, 841-848.

219. Menard, L., and Poirier, G. G. (1987) Rapid assay of poly(ADP-ribose) glycohydrolase, *Biochem Cell Biol 65*, 668-673.

220. Gagne, J. P., Pic, E., Isabelle, M., Krietsch, J., Ethier, C., Paquet, E., Kelly, I., Boutin, M., Moon, K. M., Foster, L. J., and Poirier, G. G. (2012) Quantitative proteomics profiling of the poly(ADP-ribose)-related response to genotoxic stress, *Nucleic Acids Res 40*, 7788-7805.

221. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, *Anal Chem 74*, 5383-5392.

222. Thiele, A., Zerweck, J., and Schutkowski, M. (2009) Peptide arrays for enzyme profiling, *Methods Mol Biol 570*, 19-65.

223. Hengel, S. M., Icenogle, L., Collins, C., and Goodlett, D. R. (2010) Sequence assignment of ADP-ribosylated peptides is facilitated as peptide length increases, *Rapid Commun Mass Spectrom 24*, 2312-2316.

224. Dai, Z., Wang, B., Sun, G., Fan, X., Anderson, V. E., and Monnier, V. M. (2008) Identification of glucose-derived cross-linking sites in ribonuclease A, *J Proteome Res 7*, 2756-2768.

225. Molina, H., Horn, D. M., Tang, N., Mathivanan, S., and Pandey, A. (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry, *Proc Natl Acad Sci U S A 104*, 2199-2204.

226. Gallis, B., Corthals, G. L., Goodlett, D. R., Ueba, H., Kim, F., Presnell, S. R., Figeys, D., Harrison, D. G., Berk, B. C., Aebersold, R., and Corson, M. A. (1999) Identification of flow-dependent endothelial nitric-oxide synthase phosphorylation sites by mass spectrometry and regulation of phosphorylation and nitric oxide production by the phosphatidylinositol 3-kinase inhibitor LY294002, *J Biol Chem 274*, 30101-30108.

227. Kyono, Y., Sugiyama, N., Imami, K., Tomita, M., and Ishihama, Y. (2008) Successive and selective release of phosphorylated peptides captured by hydroxy acid-modified metal oxide chromatography, *J Proteome Res 7*, 4585-4593.

228. Hayaishi, O., and Ueda, K. (1977) Poly(ADP-ribose) and ADP-ribosylation of proteins, *Annu Rev Biochem 46*, 95-116.

229. Gibson, B. A., and Kraus, W. L. (2012) New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs, *Nat Rev Mol Cell Biol 13*, 411-424.

230. Kalisch, T., Ame, J. C., Dantzer, F., and Schreiber, V. (2012) New readers and interpretations of poly(ADP-ribosyl)ation, *Trends Biochem Sci 37*, 381-390.

231. Ogata, N., Ueda, K., Kawaichi, M., and Hayaishi, O. (1981) Poly(ADP-ribose) synthetase, a main acceptor of poly(ADP-ribose) in isolated nuclei, *J Biol Chem 256*, 4135-4137.

232. Cervantes-Laurean, D., Minter, D. E., Jacobson, E. L., and Jacobson, M. K. (1993) Protein glycation by ADP-ribose: studies of model conjugates, *Biochemistry 32*, 1528-1534.

233. Jacobson, E. L., Cervantes-Laurean, D., and Jacobson, M. K. (1994) Glycation of proteins by ADP-ribose, *Mol Cell Biochem 138*, 207-212.

234. Coquelle, N., and Glover, J. N. (2012) PARP pairs up to PARsylate, *Nat Struct Mol Biol 19*, 660-661.