IULIA M. LAZAR
Department of Biological Sciences, Virginia Polytechnic Institute and State University, 1981 Kraft Drive, Blacksburg, VA 24061, USA


Iulia M. Lazar

PEER REVIEWED

# LC-MS/Proteomics: a platform for the identification and selection of drug targets

**KEYWORDS: LC-MS, proteomics, drug targets.**

Abstract Conventional, hypothesis-driven biological research is conducted by studying the effect of a perturbation on a targeted cellular process. By using focused experimental analysis strategies, only a subset of the effects of such a perturbation can be explored, the large-scale implications having to be ignored. Proteomic technologies bring unsurpassed capabilities for generating a holistic view of the complex processes that control the growth and proliferation of a cell, and represent, therefore, a powerful approach for elucidating the molecular mechanisms of disease progression and response to therapeutic drugs. In this work, the potential of proteomic technologies, as completed on liquid chromatography (LC)-mass spectrometry detection (MS) platforms, for the discovery of networked drug-target clusters, as well as the implications for future drug discovery efforts, are discussed.

## INTRODUCTION

During the past two decades, the discovery of novel drugs and drug targets has become the lead objective of many academic, government, corporate and clinical laboratories. Drug development is a lengthy (8-15 years), costly (100 million to 5 billion USD per drug) and tedious process that consists of several major steps (1-3): (a) target identification and validation, (b) drug discovery through high-throughput screening of compound libraries and hit generation, (c) hit-to-lead, lead optimization and pre-clinical development, (d) clinical trials, and (e) FDA approval, launch and post-launch evaluations (Figure 1). The process is triggered by basic biomedical research, often occurring in academic laboratories, that results in the identification of a gene or gene product (mRNA or protein) with an essential biochemical and/or physiological role of relevance to a disease. Such molecular entities become drug targets after their presence, function and role are confirmed by validation assays.

The advent of "omic" technologies (genomics, transcriptomics, proteomics, metabolomics) has entertained hopes for a speedy and cost-effective discovery process of a multitude of drug targets. While much was initially expected, the information encoded in the large data sets quickly revealed a number of limitations. Gene-level information is not informative enough to assess the suitability of a target for pharmacological intervention, and, due to lack of specificity, many small drugs that target DNA have various toxic effects (4). mRNA translation into a functional protein can be effectively inhibited by antisense, ribozyme and RNAi-based therapies, but due to the hurdles associate with the stability of RNA-drug molecules, selectivity, imperfect base pairing, off-target effects, dosage and delivery, such treatments do not always result in promising clinical outcomes (4-6). Nevertheless, the potential of RNA-based

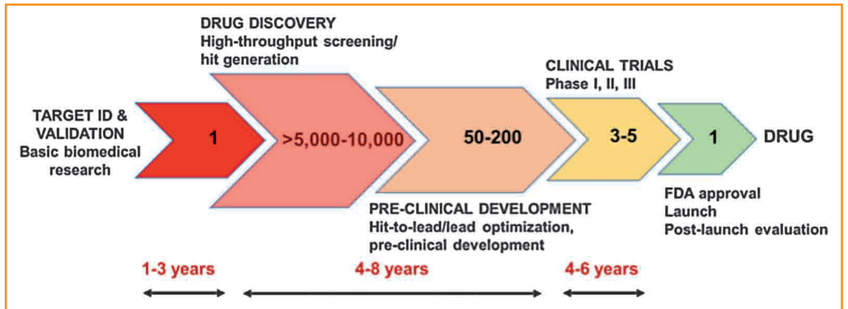


**Figure 1.** Drug discovery process.

gene therapies continues to spur interest and progress. Proteins are the workhorse of the cell that carry out all essential biological functions, their functional versatility being conferred by a broad range of posttranslational modifications (PTMs) that determine not only tertiary structure, but also location in the cell, as well as activity within a complex biomolecular network. While many protein structure databases (as determined by NMR and X-ray crystallography studies), as well as computational 3D structure prediction tools have been developed, a full understanding of a protein's function cannot be achieved outside its network of interactions with other proteins in the cell, and without characterizing the dynamic nature of its PTMs. The manually annotated SwissProt/UniProt database includes 541954 sequence entries representing 13046 species, of which 20274 belong to *Homo sapiens* (7). Moreover, the PhosphoSite Plus database identifies phosphorylation, ubiquitination and acetylation as the most abundant protein PTMs, describing 209508 phosphorylation, 51442 ubiquitination and 24380 acetylation non-redundant sites on a total of 19697 proteins (8). Earlier studies have estimated the number of disease modifying genes at ~3000, of druggable proteins at ~3000, and the outcome of overlapping these two categories at 600-1500 drug targets (2). Notably, the data pinpoint the central role that enzyme and receptor proteins had in stimulating the development of new drugs, roughly 50 % of the drug targets representing Ser/Thr/Tyr kinases (~22 %), G protein coupled receptors-GPCRs (~15 %), phosphatases (~4 %), Ser proteases (~4 %) and nuclear hormone receptors (~3 %). Despite continued challenges associated with determining *in vivo* protein structure and activity, due to the wealth of information that can be generated or predicted for each protein through a combination of biomedical/biophysical and computational approaches, it is anticipated that proteins will continue to prevail as drug targets in the foreseeable future (2).

## LC-MS IN PROTEOMICS RESEARCH

A quantitative correlation between mRNA and protein expression levels does not always exist. As a result, our ability to understand, intervene and control cellular events is severely hampered, unless the identity and abundance of the existing proteins, the nature, site and number of PTMs, and the specific functions associated with these proteins are first revealed. As many events under translational or post-translational control interfere in overall protein expression, the characterization of protein PTMs adds a new dimension to the already difficult task of protein profiling efforts (9). The advance of mass spectrometry instrumentation has brought superior capabilities to deciphering the information encoded by the protein complement of a cell, and has led proteomics to unexpected heights in the past decade (10, 11). The interfacing of 1D/2D separation techniques to MS detection, in particular nano-LC as a frontend to MS, has enabled the identification of as many as thousands of proteins in complex cellular extracts. As a result, aside the challenges associated with the complexity, dynamic range and dynamic composition of proteomic samples, LC-MS technologies are spearheading today the proteomic analysis efforts directed toward biomarker and drug target discovery. In this work, we explore the potential of LC-MS for enabling the discovery of novel, putative drug targets that can be evaluated in the broader context of biological pathways, and discuss the benefits of this approach to the drug discovery process. In our laboratory, the MCF-7 breast cancer cell line has been established as a model system for cell cycle studies and MS technology development. The cells are typically grown in EMEM with 10% fetal bovine serum and 10 µg/mL bovine insulin, arrested in the G1-phase by serum-deprivation for 48 h, released in the S-phase with hormones or growth factors, harvested, separated into nuclear and cytoplasmic fractions, digested with trypsin, and analyzed by nano-LC-MS/MS with a linear ion-trap quadrupole LTQ/Thermo Electron mass spectrometer. The LC-MS/MS mass spectra, acquired in data-dependent mode, are searched against a *Homo sapiens* SwissProt protein database with the Thermo Bioworks software. False discovery rates (FDR) are determined by searching the raw data against a forward-reversed protein sequence database, the upper limits being set at 3% and 1% at the protein and peptide levels, respectively. Such protocols typically result in the detection of ~3000 proteins in any particular cell type.

## MS DATA INTERPRETATION IN BIOLOGICAL CONTEXT

Once the peptide/protein amino acid sequence is extracted from the raw tandem MS data at the pre-established FDR values, the information must be placed in biological context. From a qualitative perspective, the list of protein IDs is queried for biological function, process and cellular location using tools such as enabled by the GoMiner/Gene Ontology (12) or DAVID (13) bioinformatics resources. To identify pertinent biological pathways, the data are mapped to public pathway databases enabled by KEGG (Kyoto Encyclopedia of Genes and Genomes) (14) or Pathway Commons (15). Protein-protein interaction diagrams can be produced with software packages such as STRING (16) or CYTOSCAPE (17) that build the diagrams from known experimental and/or computational data. The matching interactions are scored based on the reliability of the interaction. Alternatively, commercial, integrated software suites such as Metacore (18) or Ingenuity Pathway Analysis (19) can be used. The output of such an analysis pipeline provides ample information regarding the identifiable protein-protein interactions and pathways that are representative of the biological processes that are detectable in a particular cell state. For example, the analysis of the G1 and S stages of MCF-7 cells returned a total of 2725 proteins (20), among which, proteins representative of all hallmarks of cancer could be identified: self-sufficiency in proliferative capacity, evasion of apoptosis, insensitivity to antigrowth factors, limitless replicative potential, sustained angiogenesis, metastasis, deregulated metabolism and evasion of immune destruction (21). A STRING protein interaction diagram, as shown in Figure 2, highlights a set of 163 proteins as part of major interacting clusters that are implicated in the proliferative capacity of MCF-7 cells: cell cycle regulation/transcription, signaling/transcription, chromatin maintenance, ubiquitination, mitosis/cytokinesis, redox

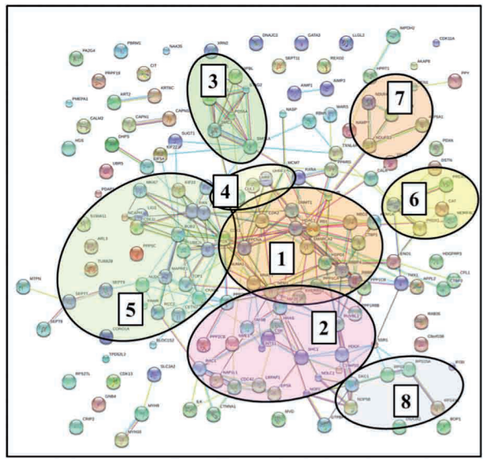


**Figure 2.** Protein clusters involved in MCF-7 cell proliferation: [1] cell cycle regulation/transcription, [2] signaling/transcription, [3] chromatin maintenance, [4] ubiquitination, [5] mitosis/ cytokinesis, [6] redox regulation, [7] mitochondrial respiration and [8] ribosome biogenesis.

regulation, mitochondrial respiration and ribosome biogenesis. From a quantitative perspective, when proteomic profiles of different cell states are compared, the use of computational methods can determine whether statistically significant differences between two, or more, biological states exist (22). The up- or down-regulated clusters and pathways can further provide cues into the molecular mechanisms that are affected by the experimental perturbation under study. Comparisons between the G1 and S stages of MCF-7 cell cycle, nuclear fractions, revealed that the major functional clusters that displayed a change in protein abundance or PTM status were implicated in signaling, transcription/translation, DNA damage response and chromatin remodeling (20). Typically, with the methodology described above, 60-150 proteins could be identified as exhibiting a change.

## PROTEIN INTERACTION NETWORKS AND DRUG TARGETS

The wealth of proteomic data has enabled the development of a number of protein, peptide and PTM databases that empower researchers, today, with a broad range of bioanalysis tools. Most importantly, however, the knowledge derived from proteomic experiments supports the development of system-level models that can provide insights into the molecular mechanisms that control cell functionality and progression of a disease. As derived from the intricate protein interaction networks and pathways, the differentially expressed or altered proteins represent a rich source of putative disease markers or drug targets. A close examination of such data in our MCF-7 studies revealed that these proteins were not random components of various unrelated biological processes, but rather grouped into functionally related clusters that matched pathways of relevance to the disease, in this case to the hallmarks of cancer. The enhanced proliferative capacity of MCF-7 breast cancer cells emerged as a result of transcriptional activities associated with the repression of CDK inhibitors, acquisition of insensitivity to growth inhibition, and DNA repair. The cell cycle signaling and transcriptional/ translational processes encompassed a plethora of potential therapeutic targets. Comparisons with non-tumorigenic MCF-10 cells revealed that ~20% of the proteins that displayed a change in abundance or PTM status were proteins that have been already targeted with drugs (according to information compiled from the Drug Bank and Therapeutic Target Database), and that about half of these targets were interacting partners within pathways linked to cancer. The abundance of known and identifiable targets suggests that additional drug targets that are part of the same interacting protein networks are likely to be present. Among the rest of 80% of proteins, several dozens of various enzymes, kinases/phosphatases and receptors emerged as putative candidates. As pointed out earlier, kinases and phosphatases represent the largest category of targets for small molecule drugs (2). The human genome encodes for 518 kinases and ~150 phosphatases (23), of which, a total of 73 kinases and 31 phosphatases were identifiable in the MCF-7 cells. A STRING diagram revealed, again, the complex set of interactions between this category of enzymes (Figure 3), uncovering two major functional clusters, i.e., signaling/cell cycle and glycolysis/ energy metabolism. The signaling cluster encompassed kinases and phosphatases that are essential to all major pathways that are involved in the progression of cancerous states: MAPK, insulin, ErbB, TGF-beta, mTOR, Wnt and VEGF. Out of roughly two dozen kinases/phosphatases for which differences were identified between MCF-7 and MCF-10, to date, less than half have been targeted. Abnormalities in kinase function due to mutations or up/down regulation were linked, so far, to over 400 diseases, and, at present, pharmaceutical companies spend a substantial percentage of their R&D funding on their targeting. With only 17 kinase inhibitors on the market, and roughly 150 in various phases of pre-clinical development and clinical trials (24), as well as dozens of phosphatase inhibitors in development (25), the stage is set, and the likelihood for further advancement of kinase/phosphatase-based therapies is high.

The ability of proteomic technologies to uncover a map of

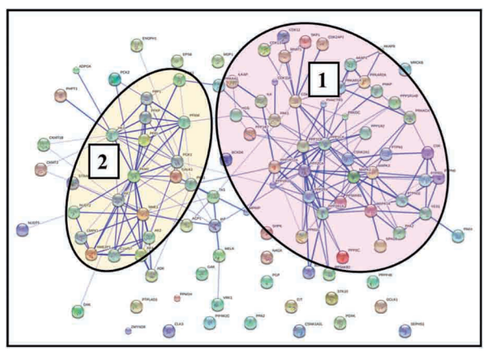


**Figure 3.** STRING diagram of kinases and phosphatases identified in MCF-7 cells: [1] signaling/cell cycle and [2] glycolysis/energy metabolism.

functionally related proteins that provides a systems-view over the status of a biological state is invaluable to understanding the mechanistic details of a disease. Such approaches can enable the selection of drug targets that act not as isolated, individual entities responsible for disease progression, but as aberrantly behaving components of a network with known function. Moreover, entire clusters of functionally related proteins could be selected as network signatures of a disease, to fuel the development of multi-target or combinatorial therapeutic strategies (26). Alternatively, in the case of networked proteins with multiple functions within multiple pathways, the development of specific drugs for each function of the same protein may also prove effective. Large-scale data integration and computational modeling studies have already investigated the potential of networked drug targets, and suggest that therapies that pursue the inhibition of a number of targets could be more effective than therapies aimed at the complete inhibition of a single target, even if the effect on each separate target is partial (26-28). The result could be a synergistic outcome with less toxicity and fewer side effects that suppresses cross-talk between signaling pathways, and, potentially, the development of resistance to therapeutics. Nevertheless, much remains yet to be investigated on the impact of network targeting on essentially every step of the drug discovery process, i.e., selection strategy of targets with druggable binding sites (active or allosteric), choice of drug cocktails or multi-target ligands (3), high-throughput screening for identifying the effective combinations of therapeutic compounds, lead optimization, selection of controls, and ultimately design of clinical trials.

## CONCLUSIONS

Advanced LC-MS proteomic technologies provide capabilities for an in-depth exploration of the cellular proteome, the identification of thousands of proteins, of protein structures, protein modifications, and protein-protein interactions, to ultimately enable the reconstruction of an integrated view of the dynamic molecular cell profile. The knowledge gained from the interpretation of such rich data will help delineate the differences between healthy and diseased biological regulatory networks, to undoubtfully accelerate the identification of novel therapeutic targets. With progressive improvements in the power of computational modeling and prediction tools, it is anticipated that the development of effective multi-target therapeutic strategies will take the lead in the drug discovery world.

## REFERENCES


1. Forbes, The Cost Of Creating A New Drug Now $5 Billion, Pushing Big Pharma To Change, August 11, 2013: http://www.forbes.com/sites/matthewherper/2013/08/11/how-the-staggering-cost-of-inventing-new-drugs-is-shaping-the-future-of-medicine/ (last checked on Jan. 21th 2014).
2. Hopkins A.L., Groom C.R., The druggable genome, *Nature Reviews, Drug Discovery* ***1***, 727-730 (2002).
3. Proschak E., Reconsidering the drug discovery pipeline for designed multitarget drugs, *Drug Discovery Today*, ***18(23/24)***, 1129-1130 (2013).
4. Hughes J.P., Rees S., Kalindjian S.B., Philpott K.L., Principles of early drug discovery, *BJP British Journal of Pharmacology*, ***162***, 1239–1249 (2011).
5. Castanotto D., Rossi J.J., The promises and pitfalls of RNA-interference-based therapeutics, *Nature*, ***457*** (7228), 426–433 (2009).
6. Bader AG, Brown D, Stoudemire J, Lammers P, Developing therapeutic microRNAs for cancer, *Gene Therapy*, ***18***, 1121–1126 (2011).
7. UniProtKB/Swiss-Prot protein knowledgebase release 2013_12 statistics: http://web.expasy.org/docs/relnotes/relstat.html (last checked on Jan. 21th 2014)
8. PhosphoSitePlus®: http://www.phosphosite.org/homeAction.do (last checked on Jan. 21th 2014)
9. Olsen J.V., Vermeulen M., Santamaria A., Kumar C., Miller M.L., Jensen L.J., Gnad F., Cox J., Jensen J.S., Nigg E.A., Brunak S., Mann M., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis, *Science Signaling* ***3(104)***, 1-15 (2010).
10. Yates J.R., Ruse C.I., Nakorchevsky A., Proteomics by mass spectrometry: approaches, advances, and applications., *Annu. Rev. Biomed.* Eng. ***11***, 49-79 (2009).
11. Savory J.J., Kaiser N.K., McKenna, A. M., Xian, F., Blakney, G.T., Rodgers, R.P., Hendrickson C.L., Marshall, A.G., Parts-Per-Billion Fourier Transform Ion Cyclotron Resonance Mass Measurement Accuracy with a "Walking" Calibration Equation, *Anal. Chem.* ***83(5)***, 1732–1736 (2011).
12. Zeeberg B.R., Feng, W., Wang, G., Wang, M.D. et al., GoMiner: a resource for biological interpretation of genomic and proteomic data, *Genome Biol.,* ***4***, R28 (2003).
13. Huang D.W., Sherman B.T., Lempicki R.A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nature Protocols,* ***4***, 44 - 57 (2008)
14. Kanehisa, M., A database for post-genome analysis. *Trends Genet.*, ***13***, 375-376 (1997).
15. Pathway Commons: http://www.pathwaycommons.org (last checked on Jan. 22nd 2014)
16. Franceschini, A., Szklarczyk D., Frankild S., Kuhn M., Simonovic M., Roth A., Lin J., Minguez P., Bork P., von Mering C., Jensen L.J., STRING v9.1: protein-protein interaction networks, with increased coverage and integration, *Nucleic Acids Res*, ***41***, D808-D815 (2013)Cytoscape: http://www.cytoscape.org/ (last checked on Jan. 22nd 2014).
17. Metacore: http://thomsonreuters.com/metacore/ (last checked on Jan. 22nd 2014).
18. Ingenuity Pathways Analysis: http://www.ingenuity.com/ (last checked on Jan. 22nd 2014).
19. Tenga M.J., Lazar, I.M., Proteomic snapshot of breast cancer cell cycle: G1/S transition point, *Proteomics*, ***13(1)***, 48-60 (2013).
20. Hanahan D., Weinberg R.A., The hallmarks of cancer: the next generation. *Cell*, ***144***, 646-674 (2011).
21. Gene Set Enrichment Analysis: http://www.broadinstitute.org/gsea/index.jsp (last checked on Jan. 22nd 2014).
22. Manning G. *et al.*, The protein kinase complement of the human genome, *Science,* ***298***, 1912-1934 (2002).
23. Cohen P., Alessi, D. R., Kinase drug discovery-what's next in the field? *Chem. Biol.*, ***8***, 96-104 (2012).
24. De Munter S., Kohn M., Bollen M., Challenges and opportunities of protein phosphatase-directed therapeutics, *Chem. Biol*. ***8***, 36-45 (2013).
25. Csermely P., Ágoston V., Pongor S.,The efficiency of multi-target drugs: the network approach might help drug design, *TRENDS in Pharmacol. Sci*., ***26***, 178-182 (2005).
26. Erler, J. T., Linding, R., Network-based drugs and biomarkers, *J. Pathol*., ***220***, 290-296 (2010).
27. Liebovitch L.S., Tsinoremas N., Pandya A., Developing combinatorial multi-component therapies (CMCT) of drugs that are more specific and have fewer side effects than traditional one drug therapies, *Nonlinear Biomedical Physics*, ***1(11)***, 1-5 (2007).