

Mudpit vs. GeLC-MS. effective strategies to enhance the proteome coverage

Abstract

Characterization of the proteome component in complex sample such cell lysates and biological fluids can be challenging, mainly due to the wide dynamic concentration range of the proteins in the sample. Many techniques have been developed to yield better proteome coverage and analyze low abundant species; this mini-review will focus on two prefractionation techniques, namely MudPIT and GeLCMS, offering a general view of their use, stressing on their strengths and weaknesses and providing practical information to get the best results.

Keywords: prefractionation, multidimensional chromatography, proteomics, MudPIT, GeLCMS, protein identification and characterization

Volume 4 Issue 2 - 2016

Filippo Genovese

Centro Interdipartimentale Grandi Strumenti, University of Modena and Reggio Emilia, Italy

Correspondence: Filippo Genovese, C.I.G.S., University of Modena and Reggio Emilia, Via G. Campi 213/A, Modena, Italy, Tel 00390592055730, Fax 00390592055600, Email filippo.genovese@unimore.it

Received: September 06, 2016 | **Published:** October 04, 2016

Abbreviations: 2D, two-dimensions; DIA, data independent acquisition; GeLC-MS, gel-enhanced liquid chromatography-mass spectrometry; MudPIT, multidimensional protein identification technology; RP, reverse phase; SCX, strong cation exchange; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

Introduction

MudPIT, acronym for Multidimensional Protein Identification Technology, and GeLC-MS, namely Gel-enhanced Liquid Chromatography-Mass Spectrometry, are among the most common strategies to deepen the proteome coverage. Such approaches use two or more separation techniques based on different physicochemical interactions to “dilute” in terms of space and time (*i.e.*, fractionate) a complex peptide mixture before it reaches the mass spectrometer, in order to gain a deeper insight into its protein composition.

Although for an analyst the temptation of injecting the sample into the instrument “as it is” on a dilute-and-shoot basis is always around the corner, complex peptide mixtures such as, for instance, serum or whole cell lysates yield poor results with such approach, as proteins of complex matrices are present at very different concentration levels, with the most abundant peptides tending to cover and suppress the ionization of the poorly represented ones; furthermore, chromatographic separation performance tends to decay as long as the column interaction sites are saturated with too many analytes. The so-called “Shotgun” approach has, indeed, its limits: although mass spectrometry has made huge progress in terms of speed of analysis and sensitivity, flooding an instrument with ions will not leave it the time to gather the required information for all of them.

A common strategy which scientists have adopted to overcome these issues consists in using two (or more) fractionation steps to simplify the composition of the single fractions, thereby easing the job of the mass spectrometer. Chosen fractionation protocols must be complementary, *i.e.* exploit different physicochemical properties of the components of the sample, in order to grant the best results in terms of protein identification and characterization.

Although time-consuming, as every single fraction must be analyzed separately, a multidimensional analytical approach provides a definitely deeper degree of characterization of “difficult” biological

samples. Software that can handle multiple fractions protein identifications is necessary to integrate and elaborate properly the data arising from the different fractions. Indeed, peptides originating from the same protein can fall in different fractions; thus, it makes more sense, even from a statistical point of view, to post-process all the data as a unique sample.

This article will focus on two of the most commonly used techniques of peptide/protein prefractionation, MudPIT and GeLC-MS, comparing their strengths and weaknesses and providing some examples of their application in scientific literature.

Discussion

The core of the MudPIT, in its original version, is a chromatographic column packed with two different stationary phases: the first element, a strong cationic exchange (SCX) layer, separates peptides based on their affinity towards a negatively charged resin, while the second part of the column is made of silica modified with hydrophobic chains containing eighteen carbon atoms, a classical reverse phase (RP) element which separates peptides based on their polarity. In a typical MudPIT experiment, as the whole digestion mixture is loaded on the chromatograph, peptides are trapped in the SCX region of the column due to the interactions between the negatively charged resin and the positive charge on the peptides’ N-terminus and on basic amino acids. By injecting into the chromatographic system a set of solutions containing increasing concentrations of a chemical, typically potassium chloride, the positive ions produced by the salt progressively displace the peptides from the SCX trap column, starting from the more weakly bound, so they can reach the RP element of the column, being further separated before entering the mass spectrometer for the analysis. The technique can be easily automated by programming the adequate injections sequence in the HPLC auto-sampler.

Tip: Divert the column flow to waste in the very first minutes of the run as the high salt content used to elude the peptides from the SCX could rapidly contaminate the MS ion source, progressively compromising the instrument sensitivity. Other fractionation techniques have been adopted for MudPIT, a detailed review has been published by Kislinger & Emili;¹ promising results have arisen from using RP chromatography even in the first step of the separation, but with a mobile phase buffered at an alkaline pH;² the authors showed

that this approach outperforms the classic MudPIT, due to the better chromatographic behavior of RP compared to SCX, ending up with a higher number of identified proteins.

As a further step forward, an “advanced MudPIT”, exploiting a heated column block to raise the elution temperature, coupled with a high performance mass spectrometer provided a significant coverage of two model bacterial proteomes, identifying thousands of proteins in a significantly lower amount of time.³

In GeLCMS, intact proteins are initially separated according to their molecular weight exploiting Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE). With this technique, the proteins surrounded by a detergent (SDS) lose their native charge; thus, their extent of migration in an electric field depends only on their molecular size: the smaller the protein the faster it moves into the narrow grid of the polyacrylamide gel. A gradient of cross-linked polyacrylamide along the gel length together with fine-tuning of the elution buffer composition will yield better resolution of proteins.

Tip: In order to get reproducible migration pattern, especially for gradient gels, and reduce keratin contamination from the operator skin/hair which is inevitably incorporated in “home-made” gels and could suppress the signal of valuable peptides, it is recommended to use pre-cast gels. At the end of the electrophoretic run, the whole lane is cut into smaller bands, usually ranging from 10 to 40; each band then undergoes tryptic digestion (for a general protocol, refer to).⁴ The length of the fractions can be tailored according to the lane migration pattern: pale colored zones can be kept together while densely populated regions should be kept apart. After extracting the peptides from the gel, each fraction is further separated by RP chromatography before being analyzed by MS. A simple protocol of this multidimensional separation technique has been published by Lundby & Olsen.⁵

GeLCMS has been recently adopted for an in-depth characterization of the proteome of the chicken egg white; this substrate has an important role in nutrition science and industry; thanks to GeLCMS, the authors were able to detect proteins which had never been reported so far. Furthermore, this study proved that modern MS instrumentation coupled with proper fractionation strategy can analyze in depth proteins even in a complex matrix without the use of, for instance, enrichment tools which could bias the quantification process.⁶

If the reader feels puzzled on which technique to use, many comparative studies have been published so far; e.g.,⁷ a recent study examined both the qualitative and the quantitative performances of these two techniques;⁸ as a general trend, GeLC-MS seems to outperform MudPIT in many cases.

Pros for GeLC-MS:

- Less protein material required;
- Abundant proteins concentrate in distinct gel slices; thus, the correspondent peptides are not smeared over different LC-MS/MS runs as it happens for MudPIT, thereby increasing the chance of identifying less copious proteins;
- Gel-determined approximate molecular weight can be used (with parsimony, do not forget about truncated isoforms) as an additional evidence of the presence of the identified proteins;
- Low molecular weight MS-annoying species (e.g., detergents, chaotropic agents) are washed away before the digestion step;

- The intensity of the colored bands on the gel provides an idea of the proteins concentration. Cons for GeLC-MS:
- More time-consuming compared to MudPIT;
- Automation unfeasible;
- Bias in the identified protein population due to poor solubility of some components of the sample in the gel loading buffer (e.g., membrane proteins);
- MudPIT and other 2D separation approaches are more compatible with quantitative biomarkers validation studies, as the fractionation steps occur at a peptide level.⁹

Moreover, horizontal carryover (i.e., cross-contamination between proteins in different lanes of the same gel) must be taken into account;¹⁰ the authors recommend to always run control lanes on different gels to exclude the occurrence of this unwanted effect.

Many efforts have been put in pushing instrumental performances to further dilute upcoming ions in time and optimize the MS acquisition process. An interesting article examines the effect of the chromatographic gradient length on the number of identified peptides.¹¹ On the precursor selection process side, where usually only the most abundant peptides are selected for MS/MS sequencing following a data-dependent paradigm, compiling a recursive exclusion lists while performing technical replicates has shown to improve proteome coverage.¹² Smart database real-time peptide identification and ion selection¹³ has proven to be effective in optimizing acquisition times, although it cannot cope with ion suppression.

An interesting approach to overcome the bias towards most intense peptide signals during ion selection for MS/MS sequencing is the so-called DIA, Data Independent Acquisition,¹⁴ where the first element of a mass spectrometer, usually a quadrupole, filters a wide range of ions which are fragmented all together in the collision cell; the originating MS/MS spectra are assigned to the corresponding precursor ions by spectra deconvolution and chromatographic behavior analysis. Many mass spectrometer vendors have developed a proprietary approach to this acquisition paradigm.^{15–17}

Conclusion

Although nowadays an almost full characterization of the proteome of a simple organism like yeast can be achieved in a few hours,¹⁸ complex samples from higher organisms usually require a prefractionation step for better outcomes. Both MudPIT and GeLCMS have proved to be effective strategies to enhance the proteome coverage and characterize low abundant species; a trade-off must be made in choosing the number of fractions and instrumental analyses duration for both techniques, as this decision will significantly limit the analyses throughput.

If the reader is still confused about which technique to adopt for his/her precious sample, I would recommend trying both of them with real samples, as the performances of these proteome coverage enhancing protocols can dramatically change from setting to setting and yield results which can be complementary in terms of identified proteins.

Acknowledgements

None.

Conflict of interest

The author declares no conflict of interest.

References

1. Kislinger T, Emili A. Multidimensional protein identification technology: current status and future prospects. *Expert Rev Proteomics*. 2005;2(1):27–39.
2. Zhou F, Cardoza JD, Ficarro SB, et al. Online Nanoflow RP–RP–MS Reveals Dynamics of Multicomponent Ku Complex in Response to DNA Damage. *J Proteome Res*. 2010;9(12):6242–6255.
3. Franzel B, Wolters DA. Advanced MudPIT as a next step toward high proteome coverage. *Proteomics*. 2011;11(18):3651–3656.
4. Shevchenko A, Tomas H, Havlis J, et al. In–gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. 2007;1(6):2856–2860.
5. Lundby A, Olsen JV. GeLCMS for In–Depth Protein Characterization and Advanced Analysis of Proteomes. *Methods Mol Biol*. 2011;753:143–155.
6. Mann K, Mann M. In–depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. *Proteome Sci*. 2011;9(1):7.
7. de Godoy LM, Olsen JV, de Souza GA, et al. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol*. 2006;7(6):R50.
8. Fang Y, Robinson DP, Foster LJ. Quantitative Analysis of Proteome Coverage and Recovery Rates for Upstream Fractionation Methods in Proteomics. *J Proteome Res*. 2010;9(4):1902–1912.
9. Cao Z, Tang HY, Wang H, et al. Systematic comparison of fractionation methods for in–depth analysis of plasma proteomes. *J Proteome Res*. 2012;11(6):3090–3100.
10. Knaust A, Shevchenko A, Shevchenko A. Horizontal carryover of proteins on one–dimensional polyacrylamide gels may jeopardize gel–enhanced liquid chromatography mass spectrometry proteomic interpretations. *Anal Biochem*. 2012;421(2):779–781.
11. Hsieh EJ, Bereman MS, Durand S, et al. Effects of Column and Gradient Lengths on Peak Capacity and Peptide Identification in nanoflow LC–MS/MS of Complex Proteomic Samples. *J Am Soc Mass Spectrom*. 2013;24(1):148–153.
12. Muntel J, Hecker M, Becher D. An exclusion list based label–free proteome quantification approach using an LTQ Orbitrap. *Rapid Commun Mass Spectrom*. 2012;26(6):701–709.
13. Graumann J, Scheltema RA, Zhang Y, et al. A Framework for Intelligent Data Acquisition and Real–Time Database Searching for Shotgun Proteomics. *Mol Cell Proteomics*. 2012;11(3):1–11.
14. Doerr A. DIA mass spectrometry. *Nat Methods*. 2015;12:35.
15. Gillet LC, Navarro P, Tate S, et al. 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*. 2011;11(6):O111.016717.
16. Peterson AC, Russell JD, Bailey DJ, et al. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics*. 2012;11(11):1475–1488.
17. Distler U, Kuharev J, Navarro P, et al. Drift time–specific collision energies enable deep–coverage data–independent acquisition proteomics. *Nat Methods*. 2014;11:167–170.
18. Richards AL, Hebert AS, Ulbrich A, et al. One–hour proteome analysis in yeast. *Nat Protoc*. 2015;10:701–714.