

**Qualitative and Quantitative Mass Spectrometry in
Proteomics**

Ph.D. thesis

Szájli Emília

Supervisor: Dr. Medzihradszky-Fölkl Katalin

Chemistry Graduate School

University of Szeged

**Biological Research Center of the Hungarian Academy of Sciences
Proteomics Research Group**

Szeged

2009

Introduction

In the Proteomics Research Group, biologically relevant samples are analyzed by mass spectrometry. My Ph.D. work can be divided into three parts:

- I. protein identification
 - II. protein quantitation
 - III. analysis of post-translational modifications
-
- I. ***Protein identification:*** Protein analysis by mass spectrometry is based on the determination of the molecular weight and sequence of their peptides obtained by digestion, generally using trypsin. MALDI-TOF MS is a widely used analytical technique because of its excellent sensitivity, relatively high speed, simplicity and low cost. For the study of complex samples, LC-MS/MS analysis is more suitable than MALDI-TOF MS due to the large number of fragmentation spectra generated by LC-MS/MS analyses. As a part of my PhD work, I participated in the identification of the interaction partners of Tubulin Polymerization Promoting Protein (TPPP/p25, $M_w=25$ kDa) from bovine brain tissue. Furthermore, I identified several proteins differentially expressed in *Streptomyces griseus* NRRL B-2682, its A-factor-non-producing mutant and its mutant transformed with plasmid AFN/pSGF4, using MALDI-TOF mass spectrometry. The incomplete or missing sequence-database concerning the *Streptomyces griseus* species posed a great difficulty in this project. Furthermore, the obtained MS/MS spectra were not suitable for de novo sequencing.
 - II. ***Protein quantitation:*** Quantitation is a crucial tool for better understanding of biological processes. Changes in the quantity of certain intra- or extracellular proteins can correlate with various physiological states, diseases, drug treatments etc. Since the amount of the analyte in the sample does not correlate directly with the ion-current intensity of its mass spectrometry signal, additional techniques have been and have to be implemented to enable differential quantitation of proteins with mass spectrometry. The relative intensity of a certain component depends on several factors, such as its structure, hydrophobicity, basicity, charge and the presence of other molecules. Lately,

MALDI-TOF MS has become a popular method for quantitative analysis of biomolecules (oligonucleotides, proteins, glycoproteins etc.) originating from various sample types or even for imaging tissue sections. At the same time, several studies discuss the non-quantitative nature of MALDI-TOF MS. In my thesis, I discuss the quantitation of unfractionated samples based on MALDI-TOF.

III. *The analysis of post-translational modifications (PTMs)*: About 500 post-translational modifications have been described in the literature, including chemical modifications. The disulfide-bond, oxidation, acetylation, methylation, phosphorylation, sulfation, glycosylation and ubiquitination are examples of possible PTMs. Both the biological characteristics and the function of proteins can be modified by PTMs. Since the physical properties, like the chromatographic character or the MS behavior (for example, fragmentation) of the peptides can also be affected by these changes, special tools are required for their analysis. Furthermore, several modifications are only present in a small fraction of the protein population (several %), so in almost all cases their enrichment is required. In this project, my task was to verify the ubiquitination of various proteins [human DNA polymerase ι , DNA polymerase ϵ , Werner helicase interacting protein (ATPase WRNIP1) and *Saccharomyces cerevisiae* Rad5 protein], and to identify the site(s) of this modification. The peptide containing ubiquitination did not show specific fragmentation and there were no procedures available for specific enrichment of the modified subpopulation, so specific MS methods were applied for their analysis.

The aims

I. *Protein identification:*

- The identification of interaction partners of Tubulin Polymerization Promoting Protein (TPPP/p25) from bovine brain tissue with MALDI-TOF mass spectrometry.
- The identification of proteins which are differentially expressed in *Streptomyces griseus* NRRL B-2682, its A-factor-non-producing mutant and its mutant transformed with plasmid AFN/pSGF4.

II. *The quantitation of proteins:*

The investigation of using Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for quantitation. This means studying the reproducibility and reliability of quantitative measurements, developing novel methods to describe the reliability of calculated data and recommending changes in the methodology to improve these measures.

III. *The analysis of post-translational modifications*

Confirming the ubiquitination of the following proteins, and identifying its exact site: *human* DNA polymerase iota, DNA polymerase eta, Werner helicase interacting protein and *Saccharomyces cerevisiae* Rad5 protein.

Applied methods

The proteins were digested with trypsin in almost all cases, but in case of some samples cyanobromide was used.

The digestions were purified with C₁₈ ZIP-TIP, when necessary.

The analyses were performed using MALDI-TOF MS and/or LC-MS/MS.

The samples were loaded on the target using the “dried-droplet” method and spectra were acquired with delayed extraction, in reflectron, positive ion mode.

Reversed-phase HPLC was performed in LC-MS/MS analysis.

Instruments:

Reflex III MALDI-TOF mass spectrometer

Agilent 1100 nanoHPLC system on-line coupled Agilent XCT Plus Ion Trap

For investigating the quantitative nature of MALDI-TOF MS, statistical methods were applied on the data: variance analysis, regression analysis, F-probe, t-test, confidence interval and inverse confidence limit calculations.

The most relevant results

I. *Protein identification:*

1. The two most significant interacting partners of TPPP/p25 identified were glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and elongation factor 1 alpha.
2. The expression of *facC* in an A-factor-non-producing mutant of *S. griseus* results in significant changes in the expression profiles of at least 50 proteins of the extracellular proteome, restoring their expression to levels similar to that in the parent B2682. Although the genome sequence of *S. griseus* was not available, we positively identified an initial set of nine proteins. Four of these are known members of the A-factor regulon, and all of the five others are probably regulated by A-factor since their gene's promoter regions displayed candidate AdpA binding sites.
3. Since the *Streptomyces griseus* protein database has become accessible, further 20 proteins were identified (data unpublished, manuscript in progress).

II. *Protein quantitation:*

I made quantitative analysis of peptides from complex samples using known amounts of reference molecules. Statistical analyses were performed on the data to determine the reliability of the measurements.

1. The inverse confidence limit was applied as a new parameter to determine the reliability of the calculated concentration ratios using calibration lines.
2. The accuracy and precision of the calculated concentrations were unacceptably low in the systems examined.
3. Using a stable isotope-labeled analogue of the analyte as an internal standard resulted in a modest improvement of the accuracy but a major development in the precision of the calculated concentration ratios. However, this improvement was only seen if the concentration ratio of the investigated and the labeled peptides fell in the range of 0.5 to 5.

4. Thus, preliminary measurements (min. 2) are needed to approximate the order of magnitude of the concentration of the compounds investigated, prior to generation of the calibration line used for exact quantitation.
5. I demonstrated that even a minor change in the components of the sample may significantly alter the slope of the calibration curve, indicating the inaccuracy of external calibration.
6. Accordingly, quantitation should be performed ideally within the sample with two isotopically labeled versions of the molecule(s) of interest.
7. The accuracy of the calculated concentration can be improved by increasing the number of data points used for the calibration line ($n_{\min} = 8-10$ in every concentration ratio (k); by increasing the total number of measurements ($N=n*k$) used for calibration; by averaging of many laser shots over a large sample area (min. 1000/spot); and by increasing the number of measurements (N') corresponding to the unknown concentration ratio (X').
8. The calibration line should be generated using every data point, not just the averages of measurements made at equal concentration ratios. This way, we get real information on the reliability of the calibration line, indicated by the R^2 value and the confidence limits.
9. The calculation and publishing of the 95 % inverse confidence limits of the estimated concentration ratios is recommended in every case.

III. *The analysis of post-translational modifications:*

In every case, the detection of tryptic peptide fragments of the ubiquitin in the mass spectra proved the post-translational modification. The determination of the exact ubiquitination site for each of the four proteins has not been successful so far by mass spectrometry, but the number of potential Lys amino acids, where the post-translation modification could take place, was significantly reduced.

Acknowledgment

I thank my supervisor, Dr. Medzihradzky F. Katalin for her useful advice and constructive criticism regarding my project.

The technical help and expertise of Dr. Hunyadi-Gulyás Éva, Klement Éva and Dr. Darula Zsuzsa for useful discussions.

I thank Dr. Fehér Tamás for help in statistical analysis and Dr. Vizler Csaba for useful advice.

I thank Bio-science Inc, especially Andrea Süle for providing a Thermo Inc. AQUA Demo kit.

Finishing this work was supported by a fellowship from the G. Richter Centenarian Foundation (to E. S.).

Finally, I thank Dr. Kovács Lajos and Dr. Paragi Gábor for supervising my scientific development at the beginning of my Ph.D. work.

The collaborating partners, Dr. Ovádi Judit, Oláh Judit, Dr. Biró Sándor, Dr. Birkó Zsuzsanna, Dr. Unk Ildikó, Dr. Haracska Lajos and Burkovics Péter also contributed to his project.

I thank the Director General, Dr. Dudits Dénes and the Supervisor of administrators of the Director General's Group, Dr. Páy Anikó for supporting my work in BRC.

List of publications

Publications relating to the Ph.D. thesis

1. Oláh J, Tokési N, Vincze O, Horváth I, Lehotzky A, Erdei A, **Szájli E**, Medzihradzsky KF, Orosz F, Kovács GG, Ovádi J. *Interaction of TPPP/p25 protein with glyceraldehyde-3-phosphate dehydrogenase and their colocalization in Lewy bodies*, FEBS Lett. 2006 Oct 30; 580 (25):5807-14. **IF:** 3.37
2. Birkó Z, Bialek S, Buzás K, **Szájli E**, Traag BA, Medzihradzsky KF, Rigali S, Vijgenboom E, Penyige A, Kele Z, van Wezel GP, Biró S., *The secreted signaling protein factor C triggers the A-factor response regulon in Streptomyces griseus: overlapping signaling routes*, Mol Cell Proteomics. 2007 Jul; 6 (7):1248-56. **IF:** 9.425 (2007)
3. **Szájli E**, Fehér T, Medzihradzsky KF., *Investigating the quantitative nature of MALDI-TOF MS*. Mol Cell Proteomics 2008: 2410–2418 First Published on July 24, 2008 doi:10.1074/ mcp.M800108- MCP 200 **IF:** 9.425 (2007)

Publications not relating to the Ph.D. thesis

1. **Szajli E**, Paragi G, Kovacs L., *A study of H-bonding of 3- and 5-substituted 6-aminouracils in duplex and triplex structures*, Nucleosides Nucleotides Nucleic Acids. 2005; 24 (5-7): 907-10. **IF:** 0.565
2. Gábor Paragi, **Emília Szájli**, Ferenc Bogár, Lajos Kovács, Célia Fonseca Guerra and F. Matthias Bickelhaupt, *Hydrogen bonding of 3- and 5-methyl-6-aminouracil with natural DNA bases*, New Journal of Chemistry, 2008, **IF:** 2.651 (2007)

List of posters

- 26-28.06.2003.: **Emília Szájli**, Gyula Schneider, *The stereoselective synthesis of steroid carbamates*, Conference of chemist, Hajdúszoboszló, Hungary.
- 6-7.11.2003.: Ágota Szájli, **Emília Szájli**, János Wölfling, Gyula Schneider, *Synthesis of estrone derivatives with halogen content*, The Vth International Symposium "Young people and multidisciplinary research", Timisoara, Rumania.
- 12-16. 09.2004.: **Emília Szájli**, Gábor Paragi and Lajos Kovács, *A study of H-bonding of 3- and 5-substituted 6-aminouracils in duplex and triplex structure*, XVI International Roundtable, International Society of Nucleosides, Nucleotides and Nucleic Acids, Minneapolis, USA.
- 27-01. 08./09.2006.: **Emília Szájli** and Medzihradzky KF. *Quantitative measurements using MALDI-TOF mass spectrometry*, 17th International Mass Spectrometry Conference, Prága.
- 06-10.05.2007.: Z. Darula; E. Klement; E. Hunyadi-Gulyás; A. Csorba; K. Buzas; **E. Szájli** and K.F. Medzihradzky, *Unexpected guests in protein identification by mass spectrometry* 25th Informal Meeting on Mass Spectrometry (25th IMMS), Nyíregyháza-Sóstó, Hungary.

Lectures

- 28-30.10.2002.: **Emília Szájli**, Gyula Schneider, *The stereoselective synthesis of cyclic steroid carbonates*. XXV. Chemistry days of Szeged.
- 28.11.2002.: **Emília Szájli**, Gyula Schneider, *The stereoselective synthesis of cyclic steroid carbonates*. Scientific conference of students in Szeged: prize 3rd.
- 16.04.2003.: **Emília Szájli**, Gyula Schneider, *The stereoselective synthesis of cyclic steroid carbonates*. National scientific conference of students, prize 3rd, Budapest.
20. 11.2003.: Paragi, **E. Szájli**, B. Penke, Z. Timár: *Ab initio calculations on dimer and triplex DNA systems* Dep. of Theoret and Phys. Chemistry, Fac. of Sci. Masaryk Univ.Brno, G.

- 25-26.03.2004.: **G. Paragi, E. Szájli, B. Penke, Z. Timár:** *A study of H-bonding of DNA-bases in duplex and triplex structures using ab initio (HF and DFT) methods.* Regular Meeting of the QSAR and Modeling Group of the Medical Chemistry Section of the Hungarian Chemical Society and the Chemometry and Molecular Modeling Council of the Academic Committee of Szeged:
- 05.2004.: **Emília Szájli,** Dr. Lajos Kovács, Gábor Paragi, *A study of H-bonding of 3- and 5-substituted 6-aminouracils in duplex and triplex structures.* Meeting of Peptide Workgroups, Balatonszemes.
- 25-27.08.2007.: **Emília Szájli,** Dr. Katalin F. Medzihradzky, *MALDI-TOF MS Analysis of Complex Samples.* Meeting of Proteomics Company (II.), Debrecen,