DIFFERENTIAL PROTEOMIC ANALYSIS OF *Bacillus subtilis* BY 2D GELS, MALDI-TOF MS AND LC-MSMS

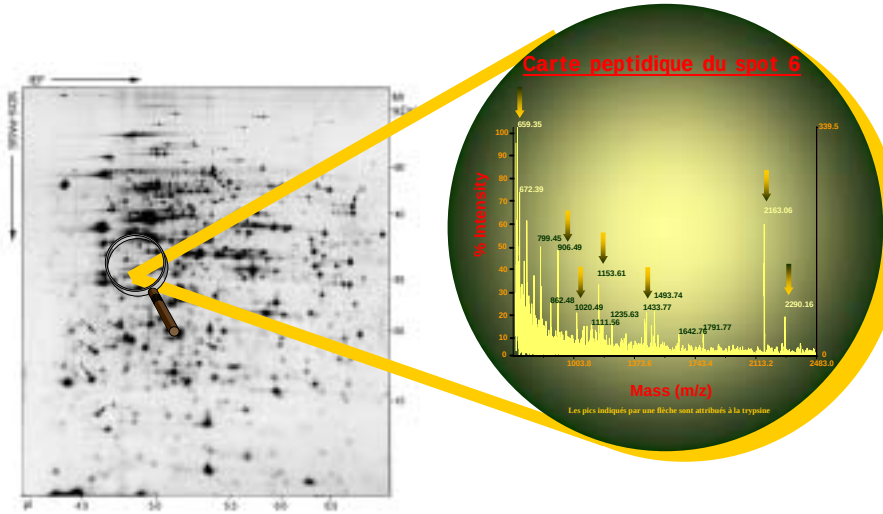
V. LABAS, J. VINH.
Neurobiologie et Diversité Cellulaire, ESPCI, CNRS UMR 7637, F75231 Paris Cedex 05
E. TURLIN, A. SEKOWSKA*
Institut Pasteur. Unité de Génétique des Génomes Bactériens, F75724 Paris cedex 15
** HKU Pasteur Research Center, Dexter HC Man Bldg, Pokfulam, Hong Kong.*

Introduction

The importance taken last years by the genomic analysis generates an increasing amount of information on the coding sequences and on the expression of genes. In this respect, the proteome reflects the consequence of cellular events at the translational or the post-translational level. A direct proteomic analysis can give a global picture of the macromolecular systems in their complexity. Proteome studies allow scientists to perform differential analyses. The protein identification by peptide fingerprinting was used to carry out a comparative proteomic analysis of *B.subtilis* [1] in response to different growth conditions. This bacterium pledged to plants is used as a model system for many Gram-positive bacteria. We compared the proteome of *B.subtilis* in two physiological states : growth on solid or in liquid medium.

Mass spectrometry MALDI-TOF is a powerful technology, allowing identification of a great number of proteins by peptide fingerprinting. However, this approach presents some limitations: small quantity of some proteins (slightly stained spots with silver nitrate), problem of detection of proteins of low (< with 15 kDa) or high (> with 100 kDa) molecular mass, co-purification of several polypeptides or contaminations (keratins or others), spectral extinction, post-translational modifications not indexed in the database, etc. We have shown that mass spectrometry coupled with reverse phase liquid chromatography (LC-MSMS) allowed identification of proteins previously not found using MALDI-TOF MS and search in the non-redundant database. This technique improves the limit of sensitivity of the analyses by mass spectrometry and allows one to work with protein mixtures[2].

Finally the use of an internal calibration and the manual data processing resulting from LC-MSMS allowed to improve the accuracy of mass measurements and leads to the identification of proteins without ambiguity.



Cell were harvested at the mid-exponential growth phase. The cell pellets were washed with the same culture medium and resuspended in 1mL Tris10mM pH 7.5, EDTA 1 mM, Urea 8 M, DTT 100 mM, Triton X100 1%, CHAPS 4%, PMSF 2mM, DNase (10 mg/mL) / RNase (5 mg/mL). Cells were subsequently disrupted with a "FP120 fastPrep Cell desruptor". Cell debris was removed by ultracentrifugation for 60 min at 90,000 x *g*. Approximately 120 mg of proteins were solubilized in 400 mL of rehydratation solution and loaded onto an 18 cm pH 4-7 immobilized pH gradient strip (IPG). Gel focusing was performed for 3 h at 300V, 1 h at 750 V, 30 min at 1500 V, 16 h at 2,500 V and 2 h at 3,500 V. The total focusing was 50 KVh. The equilibration of IPG strips was done as previously described (Görg *et al.*, 1987). The second dimension was performed with 11.5% SDS-PAGE gels using the Protean II xi 2D Multicell system (Bio-rad, Ivry, France). Proteins were stained with silver nitrate and gels were digitized using a JX-330 scanner (Sharp, Hamburg, Germany). After spot detection and quantification, 2-D gel patterns were edited and matched using the PDQUEST software package (PDI, Humington Station, NY).

In-gel digestion and MALDI-TOF analysis

Protein spots were cut off and in-gel digested with trypsin (Roche, EC 3.421.4) as described by Shevchenko *et al.* (1996). Digests were resuspended by sonication in 20 µL aq. formic acid 1%, desalted using C18 ZipTips™ (Millipore) and eluted with 50% and 80% acetonitrile. The desalted peptide mixture was dried and dissolved in 3µL aq. formic acid 1%. A saturated solution of 2,5-dihydroxybenzoic acid (DHB) in aq. TFA 0.1% was used as a matrix. Sample and matrix were loaded on the target plate using the dried droplet method (1:1, v/v). MALDI-TOF MS was performed with a Voyager-DE STR Biospectrometry Workstation mass spectrometer (PE Biosystems Inc., Framingham, MA), in positive reflector mode, with an accelerating voltage of 20 kV, a delayed extraction time of 200 ns and around 250 scans averaged for each MS. Data processing was performed using Data Explorer software (PE Biosystems). External calibration was realized using the $[M+H]^+$ ions from des-Arg Bradykinin (904.47 Da) and ACTH (clip 18-39) (2465.20 Da). The trypsin autoproteolysis products (clip 132-142, MW 1153.57 Da and clip 56-75, MW 2163.06 Da) were used as the internal calibrants.

Identified proteins

Spot	Identification	Analyse	ACC	pI Théo	M Théo	pI Obs	M Obs	Nbre peptides théo 650-4000 Da	Nbre peptides obs 650-4000 Da	% de couverture	Nbre peptides fragmentés	Nbre charges
1	Lysozyme C precursor(1)	LC-MSMS	P00695	-	16,5	4,7	16,5	10	0	8	1	2
2	Adenylate kinase(2)	LC-MSMS	P16304	4,6	24	4,8	19	16	2	9	2	2
3	Adenylate kinase(4)	LC-MSMS	P16304	4,6	24	4,8	21,5	16	3	4	1	2
4	Transcription elongation factor(5)	LC-MSMS	P80240	4,7	17	4,8	14,5	10	0	15	3	2
5	Flagellar hook protein	MALDI-TOF	NP_389511	4,9	27,45	5	23,5	15		27	-	-
6	Flagellin (19)	LC-MSMS	P02968	5	32	5	34,2	19	3	8	2	2
7	Flagellar M-ring protein	MALDI-TOF	P23447	4,6	59,28	4,9	43,1	32		21	-	-
8	Dihydrolipoamide dehydrogenase	MALDI-TOF	P21880	5	49,71	5,1	50	24		22	-	-
9	Glutamine synthase (30)	LC-MSMS	P12425	5,1	50	5,3	45,5	25	2	4	2	2
10	Homoserine dehydrogenase(31)	LC-MSMS	P19582	5,1	47,4	5,3	43	25	0	2	1	2
11	Lipoprotein lplA precursor	MALDI-TOF	P37966	5,7	56,21	5,3	50,5	37		26	-	-
12	Septum site-determining protein minD	MALDI-TOF	Q81464	5,2	29,39	5,7	29,8	18		23	-	-
13	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase	MALDI-TOF	P54482	6,8	48,57	6,7	44,5	24		27	-	-

Mass fingerprint searches

Data mining was performed using ProFound software (http://prowl.rockefeller.edu) . A mass accuracy of 25 ppm was selected for the database searches.

Data were first submitted to the SwissProt databank while restraining the search to the mass range 20-80 kDa and to mammalian taxonomy in order to identify a potential keratin contamination and trypsin autolysis products. All corresponding peptides were removed from the monoisotopic mass list.

Data were finally submitted to the *Bacillus subtilis* databank, while focusing the search on the experimental mass range and pI as measured by 2DE (tolerances M ± 50% and pI ± 2 pH unit).

In the example shown here (spot 6), we detected type I human keratins which most probably came from a contamination in the laboratory during digestion. Here no reliable identification of the protein of interest was possible. In this case the protein of interest (Flagellin) came out with a very low identification score.

LC-MS/MS analysis

Analyses by nanoscale capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS) of the digested proteins were performed using an Ultimate capillary LC system (LC Packings, Amsterdam, Netherlands) coupled to a quadrupole time-of-flight (QTOF2) mass spectrometer (Micromass, Manchester, UK) fitted with a Z-spray™ ion source. Samples were on line desalted and concentrated using a precolumn (C18, 0.3 mm I.D. mm length). Peptide separations were conducted on a reversed-phase capillary column (Pepmap C18, 75 µm I.D., 15 cm length, LC Packings) running with a 220 nL/min flow. The gradient profile used consisted in a linear gradient from 100% A (0.1% formic acid / 4% acetonitrile / 96% H_2O, v/v) to 45% B (0.09% formic acid / 10% H_2O / 90% acetonitrile, v/v) in 50 min, followed by a flush with 100% B in 10 min. Mass data were acquired using automatic switching between MS and MS/MS modes: one MS survey scan was followed by four MS/MS scans on the 3 most intense peptide ions detected. Acquisitions were performed with the dynamic exclusion of m/z ratios of already fragmented ions. Fragmentation was performed using Argon as the collision gas. The collected mass data were processed and converted into a PKL and submitted to a non-redondant database.

MS/MS data were searched against SwissProt database without restriction on taxonomy using the MS/MS ions search software available on the Matrix Science site (www.matrixscience.com).

Search result with default calibration

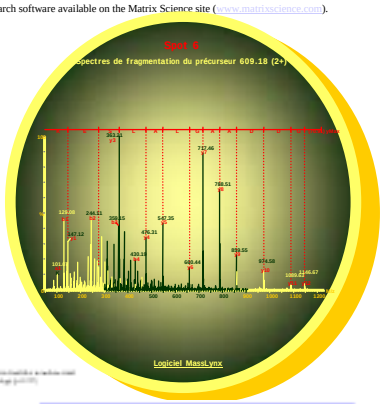
Default calibration gives a mass accuracy < 150 ppm on precursor ions and < 300 ppm on fragment ions. Protein identifications were obtained by comparison of experimental data to the SwissProt database. No taxonomic restriction was used to identify simultaneously proteins of interest and potential mammalian contaminants. We confirmed the presence of bovin trypsin and type I human keratins. Type II human keratins was then identified and two peptides were finally matched to flagellin from *Bacillus subtilis*. This result shows the possibility to identify several proteins in a mixture by LC-MS/MS, whereas MFP is limited in this case.

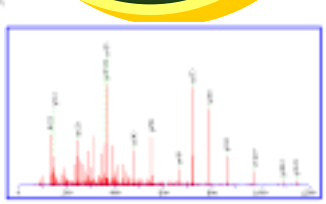
The identifications provided automatically by Mascot required sometimes manual validation. A protein identification was regarded as reliable in automatic mode when :

- the identification was based on the fragmentation of more than 3 peptides

-at least one of these peptides was associated with an MS/MS spectrum with a Mascot score superior to 50. MS/MS spectra with scores above 50 could usually be interpreted in a sequence of 5 amino acids or more (with informative fragments with S/N>5).

- the calibration was optimal.





Search result with internal calibration

To optimize the database search, it is possible to apply an internal calibration on nanoLC-QTOF MS/MS data using autolysis peaks of trypsin. For this, MS survey scans were combined and the resulting spectrum was smoothed, centroided and internally calibrated with 3 tryptic peptides displayed on the whole mass range (m/z : 577.29; 717.36; 1082.03) .

An internal calibration on combined survey scans allowed to obtain a mass measurement accuracy < 5 ppm on precursor ions and < 100 ppm on fragment ions.

When the search was performed in *Bacillus subtilis* taxonomy with the MS/MS spectrum of the doubly charged species at 609.18, a unique candidate was obtained. This improves significantly the identification scores for unambiguous identifications.

Sequence homology and sequence Tag

All data provided by LC-MS/MS analysis can be manually treated to identify proteins by sequence homology or by sequence tag search. In this example, the MS/MS spectrum of doubly charged peptide at m/z 609.18 was smoothed and centroided (converted into a DAT file) to be interpreted by PepSeq software (MassLynx software). The sequence information (AGDDAAGLAISER) allowed us to perform the unambiguous identification of flagellin using Fasta (http://www.ebi.ac.uk/fasta33). This result was also confirmed by the sequence tag search (www.narrador.embl-heidelberg.de).

Conclusion

MALDI-TOF mass finger printing (MFP) combined with 2D electrophoresis has been used to realize a differential proteomic analysis of *B.subtilis*. The results obtained by this method demonstrated the power of this tool but also its limitations.

In the present work, 6 proteins of interest have been identified out of 21 silver stained spots on 2D gels (pH 4-7), with a mass tolerance below 25 ppm. The high mass accuracy and probability, the percentage of coverage of the theoritical protein sequence (above 20%) provided by ProFound have allowed these unambiguous identifications. These represent a 30% success rate. The identification failures can be explained by a weak material quantity (very weakly silver stained spots) for proteins in the mass range 14-45 kDa. Indeed these molecular weight haven't generated many tryptic peptides and the signal was hidden by keratins contamination.

In this case, tandem mass spectrometry coupled to liquid chromatography (LC-MS/MS) allowed to identify proteins (7/15) which were unidentified by MALDI-TOF MFP in data bank (62% success rate). LC-MS/MS allowed to analyze samples with a better sensivity even for complex protein mixture. This study also showed that it was possible to validate manually the nanoHPLC-Q-TOF data using differents search modes (homology or sequence tag). To improve interrogation parameters, we manually run internal calibration on LC-MS/MS data using autolysis peaks of trypsin. The internal mass calibration allows to obtain a mass measurement accuracy < 5 ppm on precursor ions and < 100 ppm on fragment ions. This improves significantly the identification

References
1 - Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Danchin A, et al. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature. 1997 Nov 20;390(6657):249-56.

2 - Pflieger D., Le Caer JP., Lemaire C., Bernard BA., Dujardin G., Rossier J. Anal. Chem. 74 : 2400-2406

3 - Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal.Chem. 68 : 850-858.