THPZ7 544

To see the unseen: specific localization of proteins on two-dimensional electrophoresis gels using score imaging by FTICR MS/MS.

Application to the human umbilical vein endothelial cells (HUVECs) proteome.

Paulo Marcelo[1], Arnaud Bruneel[2], Iman Haddad[1], Pascal Pernet[2], Bruno Baudin[2,3], Jean Rossier[1], Joëlle Vinh[1]

[1] CNRS UMR 7637, Neurobiologie et Diversité Cellulaire, École Supérieure de Physique et Chimie Industrielles (ESPCI), 10 rue Vauquelin, 75231 Paris cedex 05, France
[2] AP-HP, Service de Biochimie A, Hôpital Saint-Antoine, (AP-HP), 184 rue du Faubourg Saint-Antoine, 75571 Paris cedex 12, France
[3] Laboratoire de Biochimie et Biologie cellulaire, UFR de Pharmacie, Université Paris 11- Orsay, 3 rue Jean-Baptiste Clément, 92290 Châtenay-Malabry, France

INTRODUCTION

We present an approach combining nano flow Liquid Chromatography Linear Ion Trap Fourier Transform Ion Cyclotron Resonance tandem Mass Spectrometry (nanoLC LIT-FTICR MS/MS) and two-dimensional gel electrophoresis (2-DGE) applied to the in-depth characterization of the human umbilical vein endothelial cells (HUVECs) proteome. In order to identify the undetected and comigrating minor proteins we decided to work with an unstained gel. The gel was blindly cut in 221 regular rectangles and analyzed using nanoLC LIT-FTICR MS/MS. Standard staining procedures were replaced by a MS score imaging of the gel for each protein. The proteins are visualized individually on the gel according to their associated identification score in the corresponding area. This score imaging enables specific protein localization on the gel with higher dynamic range and specificity than classical staining. It is in agreement with the localization of known proteins as described in public 2-DGE databases. 546 distinct proteins were unambiguously identified. Many of these proteins are known to be implicated in pathological situations such as cancer and cardiovascular diseases, situations where endothelial cells are particularly involved.

MATERIAL AND METHODS

Proteins extracted from primary cultures of HUVECs cells were obtained as previously described [1,2] and were then separated by 2-DGE according to previously described protocols [3,4]. Two gels were run in parallel; one was successively stained by colloidal blue (CCB) and silver nitrate while the other was not stained but only cut into 221 identical rectangles.

Unstained rectangles cut from 2-D gels and were treated as previously described [5]. The resulting peptide mixtures were extracted in 1% formic acid (FA) and analyzed by nanoscale capillary LC-MS/MS (Ultimate 3000 Dionex, LC-Packings) coupled online with a hybrid nanoESI linear trap and FT-ICR mass spectrometer (LTQ-FT, Thermo Fisher Scientific) using aqueous (buffer A: H20/ACN/FA, 98/2/0.1, v/v/v) and organic buffers (buffer B: H20/ACN/FA, 10/90/0.1, v/v/v). Chromatographic separations were conducted on an RP capillary column (Atlantis dC18, 75 µm id, 15 cm length, Waters) with a 220 nL/min flow. The gradient profile used consisted of two linear gradients from 0 to 20% B in 10 min and from 20% B to 60% B in 35 min.

Data were acquired in automatic mode using high dynamic range mode consisting of alternate acquisitions in FT-ICR MS full scan survey mode, FT-ICR MS SIM scan mode for accurate mass measurement and LIT MS/MS mode for sequencing as described [6].

Data were processed using Bioworks 3.1 cluster version software (ThermoElectron Fisher Scientific). Database search was run against Swissprot from UniProtKB release 5.5. Protein identification was validated according to the published standards[7]: an identification was validated only for human and if at least 2 different sequences were identified as first candidates in the protein with a minimal Xcorr and DeltaCn of 2.5 and 0.1 respectively. Mass accuracy tolerance was set to 0.01 Da in MS mode, to 0.5 Da in MS/MS mode.

For each protein identified according to the previous criteria, the values of the corresponding consensus score weighted by the Xcorr value were stored in a matrix representing the gel only for validated rectangles. The localization of the protein in the gel is visualized by a 3D representation of the matrix (x-axis for the pI, y-axis for molecular weight, z-axis for the consensus score). A linear scale enhances the major focalization spot(s), so called "hot spot(s)", for each protein in the gel.

RESULTS

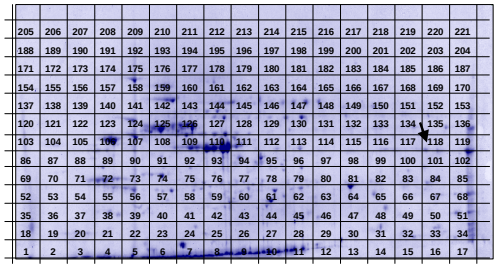

Figure 1 : Coomassie blue stained 2-D pattern of proteins (60 µg) from quiescent HUVECs in the pH range 4.0 to 7.0 (left to right) between the MW 10 Da and 120 kDa. First dimension: IPG 4–7,180 mm; second dimension: ExcelGel® 8-18%, 245 x 110 x 0.5 mm from Amersham Biosciences. Superimposed is the grid representing the pattern of the unstained twin gel cut in 221 regular rectangles.

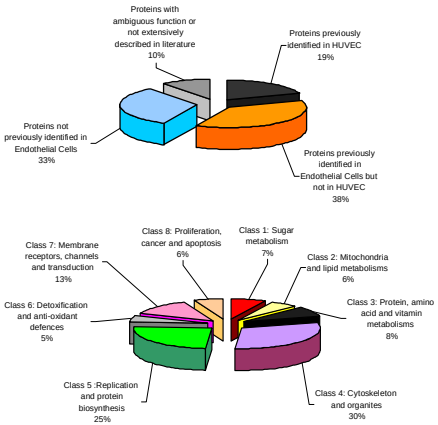

Figure 2 : Percentage distribution of identified proteins.

Considering the thickness of the 2-D gels, these results confirm the great capacity of the 2-DGE and MS techniques for the separation and subsequent identification of very low amounts of protein. As shown in Fig. 5A, 122 proteins (23%) were previously identified in HUVECs, 127 (24%) were identified only in other ECs, 206 proteins (40%) were not previously identified in HUVECs nor in other ECs, and 66 (13%) possess ambiguous or uncharacterized biological function. Not surprisingly, HUVECs express proteins from cytoskeleton and organelles, so called housekeeping proteins, e.g., from actin filaments, microtubules, intermediary filaments, endoplasmic reticulum and Golgi apparatus, and an almost complete proteasome. More surprisingly, they also express proteins from a functional actin/myosin complex, such as in motile cells, and, more specifically we found a fragment of von Willebrand Factor, a usual marker of ECs, that was never detected in previous studies on HUVECs.

Entry Name	Accession number	Protein Name	Mw	pI
AK1A1_HUMAN	P62736	Alcohol dehydrogenase [NADP+]	36441	6.35
ANXA1_HUMAN	P04083	Annexin A1	38583	6.64
ANXA2_HUMAN	P07355	Annexin A2	38472	7.56
BACH_HUMAN	O00154	Cytosolic acyl coenzyme A thioester hydrolase	41796	8.85
BUB3_HUMAN	O43684	Mitotic checkpoint protein BUB3	37154	6.36
EFTU_HUMAN	P49411	Elongation factor Tu, mitochondrial [Precursor]	49541	7.26
ENOA_HUMAN	P06733	Alpha-enolase	47037	6.99
G3P_HUMAN	P04406	Glyceraldehyde-3-phosphate dehydrogenase	35922	8.58
HNRPD_HUMAN	Q14103	Heterogeneous nuclear ribonucleoprotein D0	38434	7.61
HSP7C_HUMAN	P11142	Heat shock cognate 71 kDa protein	70020	5.56
IDHC_HUMAN	O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	46659	6.53
MK01_HUMAN	P28482	Mitogen-activated protein kinase 1	41389	6.5
PAI1_HUMAN	P05121	Plasminogen activator inhibitor 1 [Precursor]	45059	6.68
PCBP1_HUMAN	Q15365	Poly(rC)-binding protein 1	37497	6.66
PCBP2_HUMAN	Q15366	Poly(rC)-binding protein 2	38580	6.66
PDLI1_HUMAN	O00151	PDZ and LIM domain protein 1	35940	6.55
PGK1_HUMAN	P00558	Phosphoglycerate kinase 1	44483	8.3
PRS10_HUMAN	P62333	26S protease regulatory subunit S10B	44172	7.09
SIAS_HUMAN	Q9NR45	Sialic acid synthase	40307	6.29
TALDO_HUMAN	P37837	Transaldolase	37540	6.36

The results underline the higher sensitivity of the method presented here as compared with our previous studies[3,4]. As an example, rectangle 118 contains one spot visible in CCB identified as annexin A1 (P04083) (Fig. 1, black arrow) and also 4 visible spots in silver staining (not shown). After analysis of the whole rectangle, 20 proteins were unambiguously identified with at least 2 peptides in first hit for each protein according to our criteria (see Material and Methods). Except for HSP71, all proteins range between 35 kDa and 49 kDa for a basic pI range of 6.3-8.8. This is consistent with the localization of rectangle 118 in the gel.

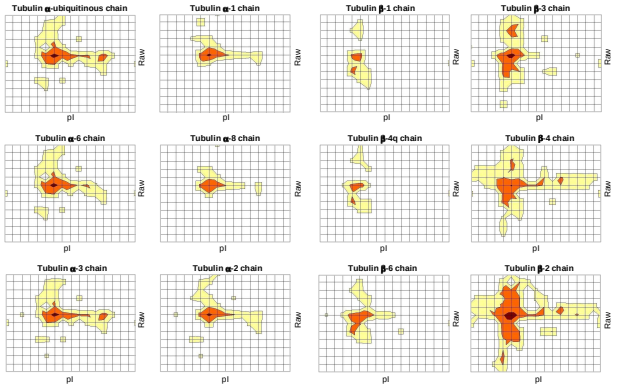

Figure 3 : Distribution in the 2-D gel of 12 isoforms of a major constitutive protein tubulin α and β. This representation was obtained using a logarithmic Score scale (see Material and Methods).

Studying isoforms of the same family of protein is always more complicated, because they exhibit large common domains so a major part of the detected tryptic peptides are the same of any of these isoforms. As we previously described [8] it is very challenging to uniquely identify some isoforms when they only differ from one another by a unique peptide. It means that it is difficult to image the position of specific isoforms except if specific peptides are unambiguously identified. This is the case for constitutive proteins such as tubulin or actin families, whose isoforms are hardly resolved in both first and second dimensions due to the combination of a high level of sequence homology between isoforms and massive heterogeneous post-translational modifications. Fig. 4 describes the distribution of 12 major tubulin isoforms on the gel. The logarithmic scale is used to enhance the lower scores. Two major behaviors could be distinguished in Fig. 4: the α isoforms spread in the gel following a coma-like profile, the β isoforms spread in the gel following a profile similar to a lying cross. This could be related to numerous and heterogeneous post-translational modifications such as polyglycylation and polyglutamylation, acetylation and phosphorylation [9], mainly inducing a pI shift towards acidic values.

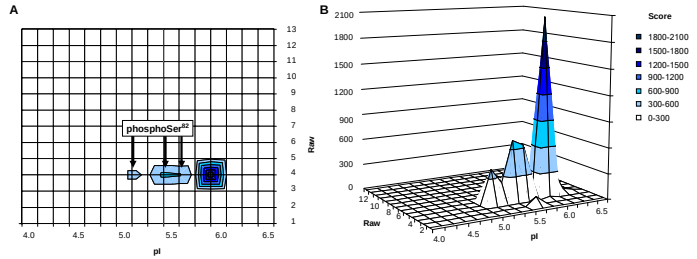

Figure 4 : Distribution of Heat-shock protein beta-1 (HSP27) in a virtual 2-D gel. For each rectangle where the identification was validated, this protein was represented as described in Material and Methods. This representation was obtained using a linear Score scale (A) with a 3D sight (B). The scale on this representation is given by the pI value in the center of the rectangle (x-axis) and the raw number along the Mr scale in 2nd dimension (y-axis).

Post-translational modifications: Three major spots are visible at around pI 5.2, 5.6 and 6 for a Mr of 28 kDa (Fig. 4A & B). This score imaging agrees with other publications focusing on HSP isoforms. An identical pattern was already described [10] They reported four isoforms : two isoforms PP1 (pI 5.7) and PP2 (pI 5.4) were shown to be phosphorylated and two others NP1 (pI 6.4) and NP2 (pI 5.8) were non-phosphoproteins. Looking further in our data the database search was repeated including serine phosphorylation as a variable modification, enabling us to localize the phosphorylation site in the sequence [80]QLpSSGVSEIR[90], in the rectangles 59, 61 and 62 corresponding to the apex of the 2 acidic isoforms.

CONCLUSION

We present an original approach combining nanoLC LIT-FTICR MS/MS and 2-DGE to further characterize the HUVECs proteome. The low dynamic range of 2-D gel staining is replaced by the nanoLC LIT-FTICR MS detection combining high resolution and high dynamic range SIM FT-MS mode and high sensitivity LIT MS/MS. This specific imaging is particularly interesting when looking for the less abundant proteins because they are often lost in the background using classical 2-DGE based techniques. We unambiguously identify more than 540 proteins and are able to localize them on the 2-D gel. Beside proteins involved in the general cellular machinery, HUVECs express a number of more specialized proteins, some of them being not yet identified in HUVECs or in other ECs. Many of these proteins could be implicated in various pathologies such as cancer and cardiovascular diseases, where EC are particularly involved. The localization of known proteins described in public 2-DGE databases is in agreement with the score imaging that we propose in this study. In summary, this new approach appears as a powerful and attractive alternative to the 2-DGE separations imaging for studying complex proteomes and more particularly the less abundant proteins.

REFERENCES

[1] Baudin, B., Bruneel, A., Bosselut, N., and Vaubourdolle, M. (2007). *Nat Protoc* **2**(3), 481-485.
[2] Mailloux, A., Grenet, K., Bruneel, A., Beneteau-Burnat, B., Vaubourdolle, M., and Baudin, B. (2001). *Eur J Cell Biol* **80**(6), 442-449.
[3] Bruneel, A., Labas, V., Mailloux, A., Sharma, S., Royer, N., Vinh, J., Pernet, P., Vaubourdolle, M., and Baudin, B. (2005). *Proteomics* **5**(15), 3876-3884.
[4] Bruneel, A., Labas, V., Mailloux, A., Sharma, S., Vinh, J., Vaubourdolle, M., and Baudin, B. (2003). *Proteomics* **3**(5), 714-723.
[5] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. (1996). *Proc Natl Acad Sci U S A* **93**(25), 14440-14445.
[6] Olsen, J. V., Ong, S. E., and Mann, M. (2004). *Mol Cell Proteomics* **3**(6), 608-614.
[7] Kapp, E. A., Schutz, F., Connolly, L. M., Chakel, J. A., Meza, J. E., Miller, C. A., Fenyo, D., Eng, J. K., Adkins, J. N., Omenn, G. S., and Simpson, R. J. (2005). *Proteomics* **5**(13), 3475-3490.
[8] Pflieger, D., Chabane, S., Gaillard, O., Bernard, B. A., Ducoroy, P., Rossier, J., and Vinh, J. (2006). *Proteomics* **6**(21), 5868-5879.
[9] McKean, P. G., Vaughan, S., and Gull, K. (2001) The extended tubulin superfamily. *J Cell Sci* **114**(Pt 15), 2723-2733.
[10] Meier, M., King, G. L., Clermont, A., Perez, A., Hayashi, M., and Feener, E. P. (2001). *Hypertension* **38**(6), 1260-1265.