ELSEVIER

Journal of Chromatography B, 753 (2001) 23–28

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

The BPP (protein biochemistry and proteomics) two-dimensional electrophoresis database

Florence Poirier[a], Naïma Imam[a], Michel Pontet[a], Raymonde Joubert-Caron[a], Michel Caron[a,b,*]

[a] *Biochimie des Protéines et Protéomique, Biochimie Cellulaire des Hémopathies Lymphoïdes (EA 1625), UFR SMBH Léonard de Vinci, Université Paris 13, 74 Rue Marcel Cachin, 93017 Bobigny Cedex, France*

[b] *Laboratoire Central de Biochimie Médicale, Hôpital Avicenne, 93009 Bobigny Cedex, France*

Abstract

The BPP (protein biochemistry and proteomics) two-dimensional electrophoresis (2-DE) database (http://www-smbh.univ-paris13.fr/lbtp/Biochemistry/Biochimie/bque.htm) was established in 1998. The current release contains 11 reference maps from human hematopoietic and lymphoid cell line samples. These reference maps have now 255 identified spots, corresponding to 84 protein entries. The World Wide Web (WWW) presentation is designed to allow public access to the available 2-DE data together with logical connections to databases providing complementary information.

Keywords: Proteome; Proteomics

1. Introduction

As we are entering the post-genomic era, greater opportunities are presented for proteome investigations [1]. Not all cellular processes are controlled at the level of gene expression, and protein profiling thus provides a complementary tool to the use of the genomic approach [2]. As a matter of fact, proteins are usually the functional molecules and, therefore, the most likely components to reflect qualitative and quantitative differences in gene expression [3]. Proteomics uses a combination of techniques to resolve (high-resolution two-dimensional polyacrylamide gel electrophoresis, 2-DE), and identify (mass spectrometry, western blotting, etc.) cell proteins, and to store and communicate (2-DE databases) the information on protein expression [4]. Two-dimensional electrophoresis is a highly resolving technique for arraying proteins by isoelectric point (p*I*) and molecular mass [5]. To date, the resolving ability of 2-DE for protein separation is unsurpassed, thus ensuring its use as the fundamental separation method for proteomics. Since the initial concept of 2-DE [6] proteomics has progressed substantially into a technology capable of investigating the protein content of a cell and its response to changing conditions. Global studies of protein expression rely on the construction of 2-DE protein reference maps [7–10]. These maps can be used to characterise a proteome for a biological sample, to track changes in this proteome over time, or to identify differences in protein expression between

*Corresponding author. Biochimie des Protéines et Protéomique, Biochimie Cellulaire des Hémopathies Lymphoïdes (EA 1625), UFR SMBH Léonard de Vinci, Université Paris 13, 74 Rue Marcel Cachin, 93017 Bobigny Cedex, France. Tel.: +33-1-4838-7777; fax: +33-1-4838-7754.

E-mail address: caron@smbh.univ-paris13.fr (M. Caron).

0378-4347/01/$ – see front matter
PII: S0378-4347(00)00459-X

samples or sample treatments. The presence or abundance of a peculiar protein found in one sample and not another can be the basis for diagnostic tests or lead to the identification of targets for drug development [11]. Further, where the expression of a protein (or its modification) can be correlated with progress of the disease, an avenue for the treatment of the disease may become evident.

We are using proteomics for a systematic analysis and identification of soluble proteins in human leukaemia and lymphoid cells, and to determine which proteins are involved in response to stimuli that trigger cell differentiation or after drug treatments. The 2-DE data collected are being made publicly available through a World Wide Web (WWW) site that links the protein data with databases containing protein sequence and biological information. This 2-DE database includes reference maps that constitute a necessary prerequisite for further studies. Further information can be obtained by directly examining the database on the WWW: http://www-smbh.univ-paris13.fr/lbtp/Biochemistry/Biochimie/bque.htm.

2. Experimental

Different leukaemia and lymphoid cell lines were used during the construction of the database: erythroleukemia cells (TF1, K562), promyelocytic cells (KG1a, HL60), lymphoma cells (DG75), and lymphoblastoid cells (PRI). The cells were collected during the exponential growth phase. Treatment of the erythroleukemia lines by erythropoietin (EPO) of aphidicolin (APH) was used to obtain an erythroid differentiation [12]. Soluble proteins were extracted using $3 \cdot 10^8$ cells/ml extraction buffer as previously described [13,14].

Proteins were separated using the IPG-DALT method [15,16]: first by electrofocusing in pH ranges of pH 3–10 or 4–7, and then by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) on horizontal 12–14% [13,17,18] or vertical 10–24% [14] gels. The analytical gels were stained with silver nitrate according to Ref. [19]. Preparative gels for mass spectrometry (MS) were stained with Colloidal G250 Coomassie Blue [20]. The stained gels were scanned using a GS-700 densitometer, and analysed using Melanie 2.3 software [21] (Biorad, Marnes La Coquette, France) as previously described [13].

The identification of proteins on 2-DE maps was obtained using a combination of techniques, including gel comparison and matching, immunoblotting [17], and peptide mass fingerprinting using MS [14]. For MS analysis, a matching was done between analytical silver stained gels and preparative gels in order to correlate the precise position of the spots to be excised. Proteins were in-gel digested [22], and desalting of peptides was performed using ZipTips_{C18} (Millipore, Saint Quentin en Yvelines, France). Mass spectra were recorded in the positive reflectron mode of a matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) MS system (Voyager, DE STR; PE Biosystems, Framingham, MA, USA) equipped with a delayed extraction device. 2,5-Dihydroxybenzoic acid was used as matrix. The peptide mass profiles produced by MALDI-TOF-MS were analysed using Peptident [23] and ProFound [24]. Peptide masses were compared with the theoretical masses derived from the sequences contained in the SWISS-PROT/TrEMBL database or deduced from the GenBank Data Bank [25].

3. Database structure

The BPP 2-DE database (http://www-smbh.univ-paris13.fr/lbtp/Biochemistry/Biochimie/bque.htm) contains data on soluble proteins identified on various 2-DE maps of hematopoietic cell lines. Eleven reference maps for the following cellular types are currently available: erythroleukemia cells (TF1, K562), promyelocytic cells (KG1a, HL60), lymphoma cells (DG75), and lymphoblastoid cells (PRI). The four maps of TF1 and the three maps of K562 allow one to compare the proteomes of these erythroleukemia lines during the exponential growth phase or after the induction of an erythroid differentiation by different drugs.

Table 1 gives detailed descriptions for each of the maps of the database, including the number of detected spots, number of identified spots, and number of distinct identified proteins.

A view of any map with labels of identified

Table 1
Content of the BPP database[a]

Map	p*I* range	Detected spots	Identified spots	Identified proteins
TF1 (growth phase)	3–10	410	18	12
APH	3–10	707	23	14
EPO	3–10	1259	26	16
EPO/GM	3–10	658	24	15
K562 (growth phase)	3–10	476	14	8
APH	3–10	696	35	21
EPO	3–10	745	32	18
KG1a (growth phase)	3–10	512	6	4
HL60 (growth phase)	3–10	377	6	5
DG75 (growth phase)	3–10	719	10	9
PRI (growth phase)	4–7	906	61	59

[a] Abbreviations are as follows: TF1 and K562, erythroleukemia derived cells; KG1a and HL60, promyelocytic leukemia derived cells; DG75, Burkitt lymphoma derived cells; PRI lymphoblastoid B cell line; APH, differentiated with aphidicolin; EPO, differentiated with erythropoietin; EPO/GM, differentiated with erythropoietin and GM-CSF.

proteins is obtained by clicking on the gel icon, while a view of the gel without labels is obtained by clicking on «without index» inside the window.

Fig. 1 shows the 2-DE image corresponding to the lymphoblastoid cell line PRI, during the exponential growth phase, depicted in the WWW (http://www-smbh.univ-paris13.fr/lbtp/Biochemistry/Biochimie/gelpri47.htm).

Spots flagged with a label correspond to identified polypeptides. The labels correspond to the accession number in the SWISS-PROT/TrEMBL protein sequence database [26,27]. Information obtained on any given identified polypeptide can be easily retrieved by clicking on a link (http://www-smbh.univ-paris13.fr/lbtp/Biochemistry/Biochimie/ident.htm). A file listing the information entered for the polypeptides identified on the gels is displayed in the format shown in Fig. 2 (only a fraction of the files is shown). For any given polypeptide found in the file is listed what types of maps the protein has been identified on (examples are K562 APH and K562 EPO), the mapping procedure (matching with another gel, MALDI-TOF, etc.) molecular mass and p*I* of the spot. In addition, links allow to open their pages in SWISS-PROT, and in SWISS-2DPAGE [28,29], HSC-2DPAGE, and SIENA-2DPAGE for the proteins indexed in these databases. It is therefore possible to navigate throughout various databases (EMBL, Genbank, PROSITE, OMIM, etc.) containing complementary information (description of the proteins identified and bibliographic references) using the different links of these pages. Therefore, in addition to the 2-DE images showing the protein location, the different links give cross-references to other databases and to the reference maps, which themselves provide experimental data (isoelectric point, molecular mass, etc.). In additions other links allow one to obtain information on the protocols used to construct the 2-DE (http://www-smbh.univ-paris13.fr/lbtp/Biochimie/techn.htm), and to contact other related databases (http://www-smbh.univ-paris13.fr/lbtp/Biochimie/autprot.htm). In the future, new links and entries will be created so as to gather information on chemical and biological characteristics of proteins.

4. Future trends and conclusion

It is apparent that the multi-level control of protein synthesis and degradation means that only the direct analysis of mature protein products can reveal their correct identities, their amount, and their relevant state of modification. At present, the most direct way to define the proteome of an organism is to use 2-DE in combination with identification methods such as MS. With the 2-DE technology available today, it is rather straightforward to reveal major protein expression changes associated with a given disease, or with the response to a given effector (differentiation or transcription factors, drugs, etc.). However, many limitations associated with building up 2-DE data-

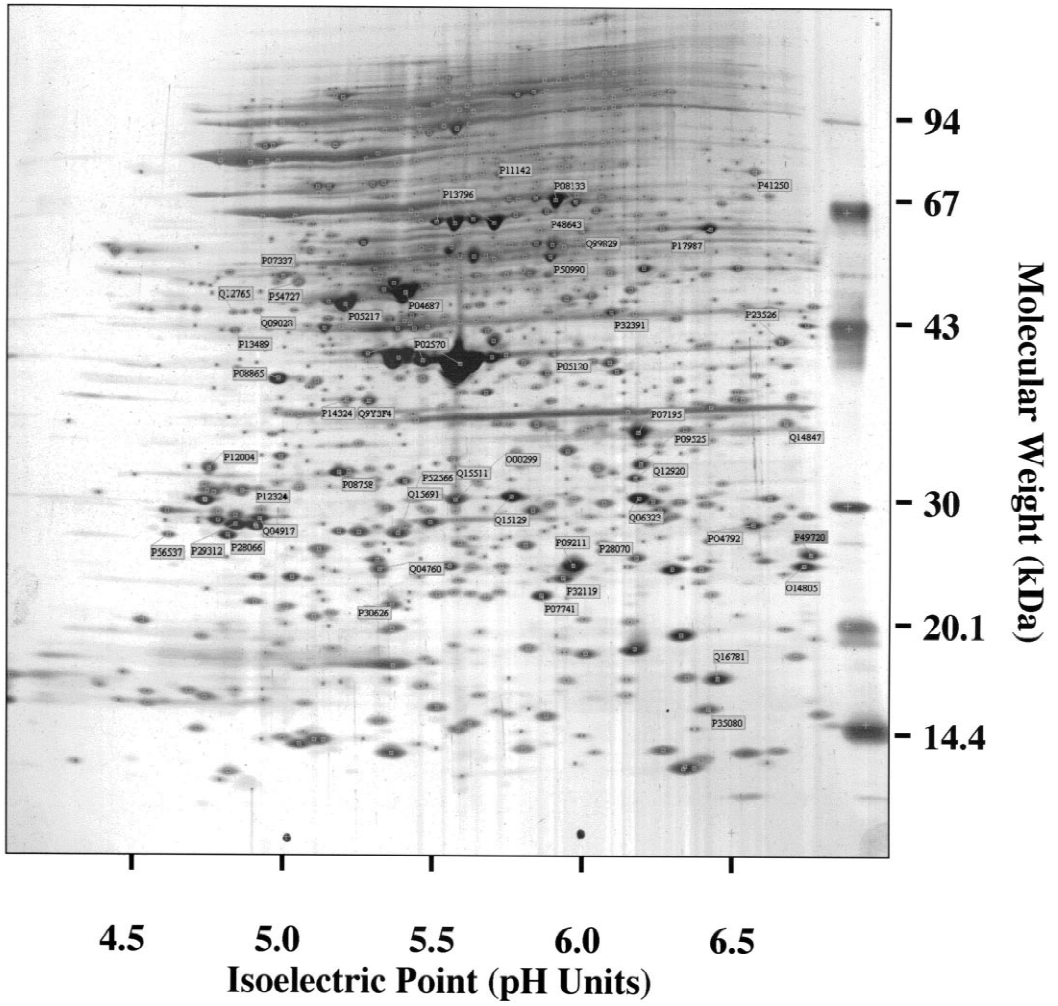
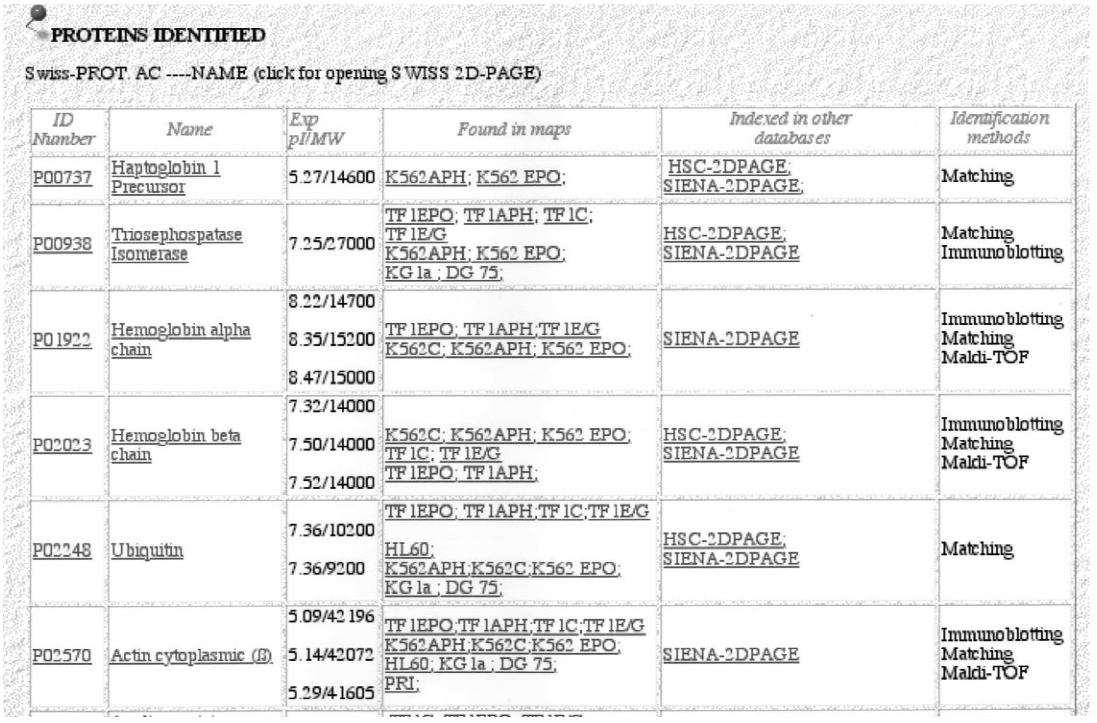


Fig. 1. 2-DE map of proteins from the lymphoblastoid cell line PRI during exponential growth phase.

bases remain, that include the detection of very low abundancy polypeptides as well as the lack of satisfactory quantitation procedures for the differential analysis of all the proteins resolved in several gels. Some of these shortcomings are being addressed by the analysis of fractionated protein samples as most of the high abundance proteins on 2-DE gels will be progressively identified [3,30]. A challenge to proteomics is therefore to refine current procedures as well as to use old or new separation and analysis technologies in combination with these procedures, for the purpose of large-scale proteome studies.

5. Nomenclature

APH	Aphidicolin
BPP	Protein biochemistry and proteomics
EPO	Erythropoietin
IPG-DALT	Bidimensional electrophoresis using immobilines in the first dimension



PROTEINS IDENTIFIED

Swiss-PROT AC ----NAME (click for opening SWISS 2D-PAGE)

ID Number	Name	Exp pI/MW	Found in maps	Indexed in other databases	Identification methods
P00737	Haptoglobin 1 Precursor	5.27/14600	K562APH; K562 EPO;	HSC-2DPAGE; SIENA-2DPAGE;	Matching
P00938	Triosephospatase Isomerase	7.25/27000	TF1EPO; TF1APH; TF1C; TF1E/G K562APH; K562 EPO; KG1a ; DG 75;	HSC-2DPAGE; SIENA-2DPAGE	Matching Immunoblotting
P01922	Hemoglobin alpha chain	8.22/14700 8.35/15200 8.47/15000	TF1EPO; TF1APH;TF1E/G K562C; K562APH; K562 EPO;	SIENA-2DPAGE	Immunoblotting Matching Maldi-TOF
P02023	Hemoglobin beta chain	7.32/14000 7.50/14000 7.52/14000	K562C; K562APH; K562 EPO; TF1C; TF1E/G TF1EPO; TF1APH;	HSC-2DPAGE; SIENA-2DPAGE	Immunoblotting Matching Maldi-TOF
P02248	Ubiquitin	7.36/10200 7.36/9200	TF1EPO; TF1APH;TF1C;TF1E/G HL60; K562APH;K562C;K562 EPO; KG1a ; DG 75;	HSC-2DPAGE; SIENA-2DPAGE	Matching
P02570	Actin cytoplasmic (β)	5.09/42196 5.14/42072 5.29/41605	TF1EPO;TF1APH;TF1C;TF1E/G K562APH;K562C;K562 EPO; HL60; KG1a ; DG 75; PRI;	SIENA-2DPAGE	Immunoblotting Matching Maldi-TOF

Fig. 2. A partial view of the data of the table listing identified proteins in the BPP 2-DE database. All the data can be seen by scrolling the table.

MALDI-TOF	Matrix-assisted laser desorption ionisation time-of-flight
MS	Mass spectrometry
SDS–PAGE	Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate
2-DE	Bidimensional electrophoresis
WWW	World Wide Web

Acknowledgements

This work was supported in part by Le Ministère de l'Éducation Nationale de la Recherche et de la Technologie (MENRT), and by grants from La Ligue Française contre le Cancer (Comité de Seine Saint-Denis) and from the Association pour la Recherche contre le Cancer (ARC, GRANT 9348). F.P. is a recipient of a fellowship from Fondation Midy.

We would like to recognise the following individuals who have contributed significantly to the various aspects of the work highlighted in this paper. We thank Noelle Leroux (Service de la Communication, University Paris 13, France) for her precious help during the construction of the website. The MS data mentioned in this paper were obtained by Jean Pierre Lecaer and Valerie Labas, in the laboratory of Jean Rossier (ESPCI, Paris, France).

References

[1] N.L. Anderson, N.G. Anderson, Electrophoresis 19 (1998) 1853.
[2] S.P. Gygi, Y. Rochon, B.R. Franza, R. Aebersold, Mol. Cell Biol. 19 (1999) 1720.
[3] M.F. Lopez, J. Chromatogr. B 722 (1999) 191.
[4] D.F. Hochstrasser, Clin. Chem. Lab. Med. 36 (1998) 825.

[5] A. Gorg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, W. Weiss, Electrophoresis 21 (2000) 1037.
[6] P. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[7] H. Boucherie, F. Sagliocco, R. Joubert, I. Maillet, J. Labarre, M. Perrot, Electrophoresis 17 (1996) 1683.
[8] C.S. Giometti, K. Williams, S.L. Tollaksen, Electrophoresis 18 (1997) 573.
[9] J.E. Celis, M. Ostergaard, N.A. Jensen, I. Gromova, H.H. Rasmussen, P. Gromov, FEBS Lett. 430 (1998) 64.
[10] M.C. Costanzo, J.D. Hogan, M.E. Cusick, B.P. Davis, A.M. Fancher, P.E. Hodges, P. Kondu, C. Lengieza, J.E. Lew-Smith, C. Lingner, K.J. Roberg-Perez, M. Tillberg, J.E. Brooks, J.I. Garrels, Nucleic Acids Res. 28 (2000) 73.
[11] W.F. Patton, J. Chromatogr. B 722 (1999) 203.
[12] D. Lutomski, P. Bourin, D. Bladier, M. Caron, R. Joubert-Caron, Eur. J. Cell Biol. (1997) 20.
[13] R. Joubert-Caron, J. Feuillard, S. Kohanna, F. Poirier, J.P. Le Caer, M. Schuhmacher, G.W. Bornkamm, A. Polack, M. Caron, D. Bladier, M. Raphael, Electrophoresis 20 (1999) 1017.
[14] R. Joubert-Caron, J.P. LeCaer, F. Montandon, F. Poirier, N. Imam, J. Feuillard, D. Bladier, J. Rossier, M. Caron, Electrophoresis 21 (2000) 2566.
[15] B. Bjellqvist, K. Ek, P.G. Righetti, E. Gianazza, A. Görg, R. Weistermeier, W. Postel, J. Biochem. Biophys. Methods 6 (1982) 317.
[16] P.G. Righetti, A. Bossi, Anal. Biochem. 247 (1997) 1.
[17] D. Lutomski, M. Caron, J.D. Cornillot, P. Bourin, C. Dupuy, M. Pontet, D. Bladier, R. Joubert-Caron, Electrophoresis 17 (1996) 600.
[18] D. Lutomski, M. Fouillit, P. Bourin, D. Mellottée, N. Denize, M. Pontet, D. Bladier, M. Caron, R. Joubert-Caron, Glycobiology 7 (1997) 1193.
[19] J. Heukeshoven, R. Dernick, Electrophoresis 6 (1985) 103.
[20] N.L. Anderson, D.C. Copple, R.A. Bendele, G.S. Probst, F.C. Richardson, Fund. Appl. Toxicol. 18 (1992) 570.
[21] R.D. Appel, P. Palagi, D. Walther, R. J Vargas, J.C. Sanchez, F. Ravier, C. Pasquali, D.F. Hochstrasser, Electrophoresis 18 (1997) 2724.
[22] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Anal. Chem. 68 (1996) 850.
[23] URL: http://www.expasy.ch/tools/peptidident.html
[24] URL: http://prowl.rockefeller.edu/cgi-bin/ProFound
[25] URL: http://www.ncbi.nlm.nih.gov/web/genbank
[26] A. Bairoch, R. Apweiler, Nucleic Acids Res. 27 (1999) 49.
[27] A. Bairoch, R. Apweiler, Nucleic Acids Res. 28 (2000) 45.
[28] R.D. Appel, J.C. Sanchez, A. Bairoch, O. Golaz, F. Ravier, C. Pasquali, G.J. Hughes, D.F. Hochstrasser, Nucleic Acids Res. 24 (1996) 180.
[29] C. Hoogland, J.C. Sanchez, L. Tonella, P.A. Binz, A. Bairoch, D.F. Hochstrasser, R.D. Appel, Nucleic Acids Res. 28 (2000) 286.
[30] M.P. Molloy, B.R. Herbert, B.J. Walsh, M.I. Tyler, M. Traini, J.C. Sanchez, D.F. Hochstrasser, K.L. Williams, A.A. Gooley, Electrophoresis 19 (1998) 837.