

V. LABAS, J.VINH.

Neurobiologie et Diversité Cellulaire, ESPCI, CNRS UMR 7637, F75231 Paris Cedex 05

E. TURLIN, P. BERTIN, L. du MERLE & C. LE BOUGUENEC.

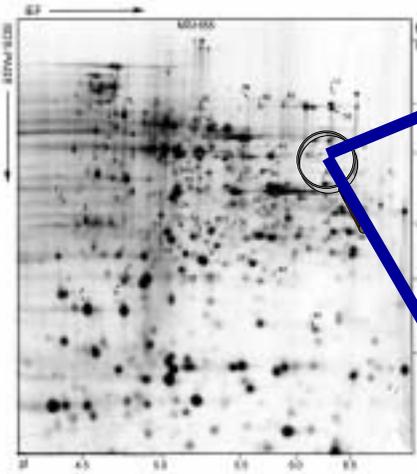
Institut Pasteur. Unité de Génétique des Génomes Bactériens. Unité de Pathogénie Bactérienne des Muqueuses. Paris 15.

Introduction

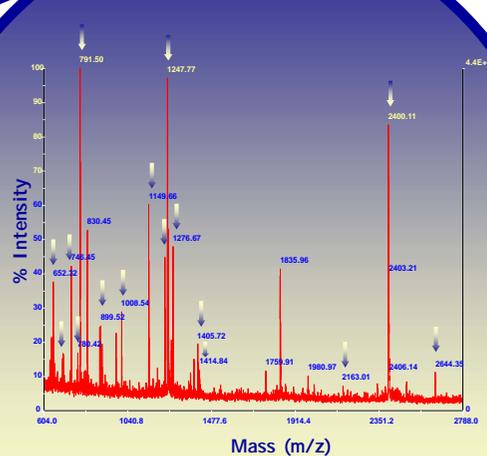
E. coli est une espèce hétérogène comprenant à la fois des souches commensales et des souches pathogènes. Différents types de *E. coli* causent différentes maladies, parmi lesquelles des infections de l'appareil urinaire (UTI), des méningites/septicémies et des infections entériques chez des hôtes variés. La plasticité du génome de *E. coli* est très importante. Durant les 5 dernières années, plusieurs études ont été publiées sur le chromosome de *E. coli*. La séquence du génome de la souche de laboratoire K-12 MG1655 a été publiée en 1997 (Blattner *et al.*, 1997) et il a été démontré que la taille du chromosome de *E. coli* varie de 4.5 à 5.5 mégabases (Mb) (Rode *et al.*, 1999). La séquence du génome de la souche O157:H7 (souche entérohémorragique EDL933) a été récemment déterminée (Perna *et al.*, 2001). Sa comparaison avec celle de la souche K-12 de référence a confirmé l'existence d'un squelette commun au chromosome de ces deux organismes, d'environ 4.1 Mb, et de segments spécifiques (îlots spécifiques) répartis sur les deux génomes en régions s'étendant jusqu'à 88 kilobases. Environ 26 % du génome de la souche EDL933 sont constitués d'îlots spécifiques comprenant 33 % de gènes de fonction inconnue. Le complément fonctionnel de la détermination de la séquence du génome est le protéome qui peut être analysé par la combinaison des techniques d'électrophorèse en gel bidimensionnel et de spectrométrie de masse. L'analyse comparative du protéome de plusieurs souches de *E. coli* semble prometteuse pour identifier des gènes dont l'expression est modifiée dans certaines souches. Il est montré ici que la spectrométrie de masse MALDI-TOF est une méthode simple, rapide et extrêmement sûre pour ce type d'étude différentielle.

Gel d'électrophorèse bi-dimensionnelle

Les cellules sont prélevées au cours de la phase exponentielle. Après centrifugation et lavages, le culot bactérien est resuspendu dans 1ml d'une solution Dnase(10 µg ml⁻¹) / Rnase(5 µg ml⁻¹). Les cellules sont lysées à l'aide de microbilles et les débris cellulaires sont éliminés par ultracentrifugation. Environ 120 µg de protéines sont solubilisés dans 400 µL de solution de réhydratation et chargés sur des gels d'isoelectrofocalisation de gradient de pH, immobilisé 4-7 de 18 cm. Pour cette étude, la première dimension est réalisée à l'aide d'un système horizontal Multiphor II (Pharmacia). Après SDS-PAGE 11.5% à l'aide du système Proteom II xi 2D Multicell (Biorad), les protéines sont colorées au nitrate d'argent. Les spots sont alors détectés et quantifiés à l'aide du logiciel PDQUEST (PDI, Huntington Station, NY).



Carte peptidique de MDH (spot 42)



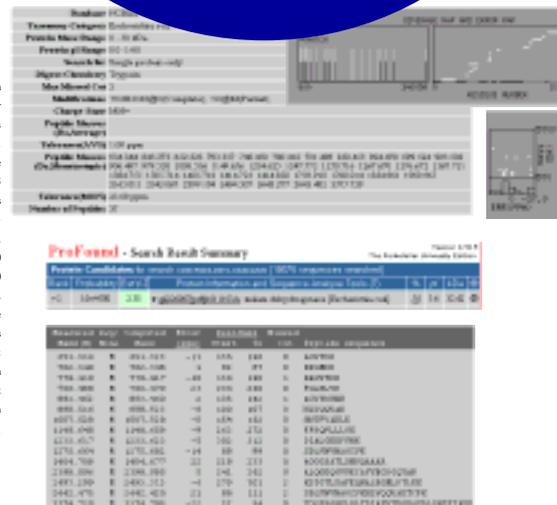
Les pics indiqués par une flèche sont attribués à la protéine MDH avec une erreur de masse inférieure à 1 ppm.

Protéines identifiées par MALDI-TOF

N°Spot	Identification	ACC	pI Théo	pI Obs	M Théo	M Obs	Couverture de séquence	
1	401	Palatine Phosphatase	4.9	33.6	4.3	31	20%	
2	1104	Glycose-specific HA component	P08837	4.79	18.1	4.6	17.1	50%
3	1201	Phagein Lipoprotein (X2)	AA054499	5.2	29.3	4.6	30.3	50%
4	1501	ATP Synthase beta chain	P00824	4.9	50.2	4.6	47.4	43%
5	1704	NISA Protein	P03003	4.5	55.0	4.6	58.1	23%
6	1708	Usher Factor	P22257	4.8	48.2	4.7	58.4	38%
7	2102	4,5-Dihydroxyheptadecanoic acid	P26427	5	20.9	4.8	20.1	49%
8	2301	Isocitrate lyase	P05313	5	47.8	4.8	35.7	28%
9	2802	Malate Dehydrogenase(X2)	P00681	4.8	38.2	4.7	41.4	50%
10	2904	Hexokinase II(Dehydrogenase(X2))	P32665	4.8	39.1	4.7	39.1	50%
11	3205	Thiol Peroxidase	P37901	4.7	17.3	4.8	19.8	28%
12	3405	Malate Dehydrogenase(X2) (malate dehydrogenase)	P28715	4.9	35.8	4.8	34.1	50%
13	3601	ATP Synthase Beta Chain	P00824	4.9	50.2	4.7	54.2	17%
14	4204	Acetoacetate CoA Transferase Alpha Subunit	Z23599	5.1	23.6	5.3	28.5	18%
15	4302	Phosphoenolpyruvate Carboxylase	P02900	4.6	22.8	5	29.7	28%
16	4306	Phosphoenolpyruvate Carboxylase	P02900	4.6	22.8	4.9	28.7	50%
17	4401	Hypothetical protein YJ21	P07739	5.3	35.35	4.9	35.5	43%
18	4504	Leucyl-tRNA Synthetase	P02917	5.5	39.1	5	42.2	50%
19	4506	Ulkation protein from 3D-page (YjZ)	P39179	5.3	36.2	5.1	39.7	24%
20	4704	Malate dehydrogenase	P05793	5.2	42.39	5.3	41.1	50%
21	5202	Hexokinase I(Dehydrogenase(X2))	P01256	5.2	32.9	5.2	30.4	50%
22	5305	Phosphoenolpyruvate Carboxylase	P02900	5.3	43.5	5.4	35.6	34%
23	5308	Hypothetical protein YJ21	P07739	5.3	35.35	5.2	39.7	46%
24	5401	Phosphoenolpyruvate Carboxylase	P02927	5.3	33.2	5.1	32.8	51%
25	5403	Phosphoenolpyruvate Carboxylase	P02997	5.2	30.3	5.2	33.8	46%
26	5407	Phosphoenolpyruvate Carboxylase	P11133	5.5	38.4	5.5	39.5	18%
27	5501	Hypothetical 36.1 kDa protein(X2)	P32697	5.1	36.6	5.1	38.8	21%
28	5511	Phosphoenolpyruvate Carboxylase	P02990	5.3	43.5	5.4	46.7	50%
29	6205	Phosphoenolpyruvate Carboxylase	P04049	5.6	20.1	5.6	20.8	40%
30	6303	Hypothetical protein YJ21	P07739	5.3	28.9	5.5	29.8	14%
31	6305	Dihydroxyacetone Phosphate Synthase	P00228	5.3	28.9	5.4	29.7	43%
32	6306	Hypothetical protein YJ21	P07739	5.3	28.9	5.4	29.7	43%
33	6315	2,3,4,5-Tetrahydroxyheptanoate-2-Carboxylate N-Succinylamide	P03948	5.5	30.5	5.7	33	99%
34	6401	Dihydroxyacetone Phosphate Synthase	P09625	5.3	32.4	5.6	36.1	47%
35	6413	Phosphoenolpyruvate Carboxylase	P11133	5.5	38.4	5.6	39.1	9%
36	6514	Beta-Ketothiolase(X2)	P14926	5.3	43.0	5.5	44.1	33%
37	6601	Dihydroxyacetone Phosphate Synthase	P12383	5.2	37.5	5.3	39.5	39%
38	6606	Phosphoenolpyruvate Carboxylase	P33550	5.7	58.3	5.5	58.8	17%
39	6705	RNA Polymerase	P10616	6.3	156.3	5.7	62	19%
40	7105	Hexokinase I(Dehydrogenase(X2))	P28715	5.8	39.1	5.9	39.7	51%
41	7206	D-Ribose-Binding Protein Monomer	P02525	5.7	28.51	5.9	29.7	31%
42	7406	Malate Dehydrogenase(X2)	P00694	5.6	32.49	6.3	39	54%
43	7408	Malate Dehydrogenase(X2)	P00694	5.6	32.49	5.8	34.4	43%
44	7501	Phosphoenolpyruvate Carboxylase	P11044	5.5	39.38	5.7	39.1	43%
45	8002	Phosphoenolpyruvate Carboxylase	P33550	5.7	58.36	5.8	58.6	50%
46	8004	DNA Polymerase Delta Subunit Protein	P27430	5.7	187.5	6.4	185.5	11%
47	8302	Uridine Phosphorylase	P12758	5.7	27.3	6.5	31.4	23%
48	8506	Dihydroxyacetone Phosphate Synthase(X2)	P09625	6.6	36.9	6.7	39.8	46%
49	8508	Malate Dehydrogenase(X2)	AA054235	6.9	28.6	6.4	33.1	41%
50	8509	Acetate Kinase	P15646	5.8	43.61	6.3	44.9	50%
51	8407	Malate Dehydrogenase(X2)	AA054235	6.9	28.6	6.4	33.1	41%
52	8702	Beta-Ketothiolase(X2)	P46923	6.8	99.25	6.5	111.4	0%
53	8705	Phosphoenolpyruvate Carboxylase	P23857	6.2	60.2	6.5	60.5	20%

Digestion dans le gel et analyses MALDI-TOF

Les spots d'intérêt sont découpés et digérés dans le gel par la trypsin (Roche Diagnostics) selon la méthode décrite par Shevchenko *et al.*, 1996. Après digestion des protéines, les extraits peptidiques sont dessalés par Zip Tip C18 (Millipore). L'échantillon est déposé sur la cible selon la méthode de la goutte séchée (w/v : 1 : 1) utilisant une solution saturée d'acide 2-5 dihydroxybenzoïque (DHB) en TFA 0.1%. Les cartes peptidiques sont alors obtenues à l'aide d'un spectromètre de masse MALDI-TOF Voyager DE-STR (Applied Biosystems, Framingham, MA), en mode réflexion positif avec une tension d'accélération de 20 000 V, une extraction retardée de 200 ns et environ 250 impulsions laser sont requis pour une acquisition spectrale. L'analyse et le traitement des données sont effectués par le logiciel Data Explorer (Applied Biosystems). Tous les spectres obtenus sont étalonnés, d'une part automatiquement en externe et d'autre part, en interne sur les produits d'autodigestion de la trypsin. Une erreur de précision de masse inférieure à 25 ppm est tolérée, pour les recherches dans les banques de données non redondantes, par l'intermédiaire des logiciels Protein and MS-FIT.



Résultats et conclusion

Dans le cadre de cette étude, l'association de l'électrophorèse 2D et de la spectrométrie de masse MALDI-TOF a permis d'identifier par MFP (Mass Finger Print) 93% des protéines analysées. D'une part, sur l'ensemble des 80 spots, faiblement à fortement colorés au nitrate d'argent, présentant des masses apparentes allant de 18 à 150 kDa avec des pI de 4 à 7, il est possible d'obtenir des résultats extrêmement reproductibles tout en gardant les mêmes conditions de traitement. Pour les spots de haut poids moléculaire contenant moins de 90 ng de matériel biologique dans le gel, la sensibilité de l'appareil MALDI-TOF permet ainsi d'obtenir des cartes peptidiques avec des quantités largement inférieures à 1 picomole.

D'autre part, la bonne précision de mesure (tolérance inférieure à 25 ppm), les probabilités très élevées données par les logiciels de recherche et le % de couverture de séquence (supérieur à 30% dans la majorité des cas) ont permis d'identifier 53 protéines différentes sans aucune ambiguïté, comme par exemple l'identification de la protéine MDH (malate dehydrogenase) présentée ci-contre.

La MDH est une enzyme métabolique dont le polymorphisme sert de marqueur dans des études de phylogénie en utilisant en particulier la technique du Multilocus Enzyme Electrophoresis (MLEE). L'étude de la variabilité génétique et moléculaire de la MDH a été utilisée pour comparer des souches commensales et des souches pathogènes de *E. coli* (Boyd *et al.*, 1994).

Cette étude montre qu'il est devenu simple et rapide d'étudier des organismes variants dont la séquence du génome est en partie connue. En effet, les variations du génome et la grande homologie de séquence des protéines pour ce groupe des souches étudiées ici, n'ont pas limité l'identification par la méthode de l'empreinte peptidique.

Références

- Blattner, F.R., G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, *et al.* 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277**:1453-1462.
- Rode, C.K., L.J. Munkers-Watson, A.T. Johnson, and C.A. Bloch. 1999. Type-specific contributions to chromosome size differences in *Escherichia coli*. Infect. Immun. **19**:230-236.
- Perna, N.T., G. Plunkett III, V. Burland, B. Mau, J.D. Glasner, D.J. Rose, *et al.* 2001. Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7. Nature **409**:529-533.
- Boyd, E.F., K. Nelson, F.S. Wang, T.S. Wittman, and R.K. Selander. 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **91**.
- 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.