

세포생물학과 Proteomics 응용

김동욱*

예일대 의대 세포생물학과/분자 의과학 연구소

많은 생물체의 완전한 genome sequence가 속속 밝혀지면서 세포의 기능을 종합적으로 평가하려는 노력들이 이어져 왔다. DNA microarray는 세포 전체의 유전자 전사, 즉 mRNA 레벨을 측정해주므로 세포가 처해있는 서로 다른 환경 속에서 유전자 발현의 차이를 측정할 수 있다. 그러나 유전자 발현의 최종 산물은 mRNA를 통해 번역된 단백질에 해당되고, 많은 단백질이 번역 후 수식(post-translational modification) 과정을 거쳐 세포 내에서 기능을 발휘하므로 진정한 세포의 생리학적 상태를 평가하기 위해선 단백질 레벨의 분석이 필수적이다. Proteomics란 유전자 산물 즉 단백질의 기능을 large-scale로 분석하는 것으로 정의된다. 이것은 genome에 의해 만들어지는 모든 단백질(proteome)을 의미하기도 하고 좁은 의미에서는 세포내의 어떤 organelle (예: Golgi Complex)에 존재하는 단백질 혹은 어떤 protein complex를 지칭하기도 한다. Proteomics는 어떤 주어진 조건에서 특별한 세포 또는 organelle에서 발현되는 단백질들을 연구하고 이해하는데 강력한 수단이 되고 있다. 이런 proteomics는 genomics, bioinformatics 등과 유기적으로 연결되어 세포의 기능을 입체적으로 이해하는데 도움을 준다. 본고에서는 proteomic analysis 과정을 간단히 살피고 최근 세포 생물학에서 이루어지는 proteomics의 응용을 살펴본다.

Proteomic analysis 과정

1) Sample 준비

어떤 단백질들을 대량으로 분석하는 과정은 Fig. 1에 나와 있는 것처럼 몇 단계의 과정을 거치게 된다. 먼저 분석하고자 하는 sample을 준비해야 하는데 이는 total soluble protein fraction일 수도 있고 detergent 처리에 의한 membrane protein들일 수도 있다(1,2). 또는 특정 세포내 기관, 예컨대 Golgi Complex (3,4) 나 Mitochondria (5) 등이 될 수도 있다.

2) 2D-GE

분석하고자 하는 sample이 준비되면 two-dimensional gel electrophoresis (2D-GE)에 의해서 단백질을 분리하고 염색하게

된다. 2D-GE는 많은 단백질을 분리할 수 있으므로 proteomics의 핵심 기술에 속한다. 2D-GE의 원리는 두 가지 분리기술, 즉 단백질의 isoelectric point (pI)와 molecular weight (MW)을 이용하므로 분해력이 뛰어나다. 그러나 membrane protein들을 분리하는데 있어 한계를 지닌다(2). 이들은 대개 세포내 소량으로 존재하거나 pI 8.0 이상의 basic protein들이 많고, isoelectric focusing을 위해 사용하는 aqueous media에 solubility가 낮기 때문이다. 이를 위해 carrier ampholyte gel 이나 immobilized pH gradient가 사용되는 등 그 개선책이 시도되고 있다(2).

3) MS에 의한 분석

Mass spectrometry (MS)는 2D-GE과 더불어 proteomics의 중요한 기술이다. 단백질들이 2D-GE에 의해 분리되고 gel의 염색 및 탈색후 원하는 protein band가 추출된다. 이 gel 속의 단백질이 trypsin과 같은 그 specificity를 알고 있는 protease에 의해 분해된 후 MS에 의해 peptide들의 분석이 이루어진다. 모든 MS는

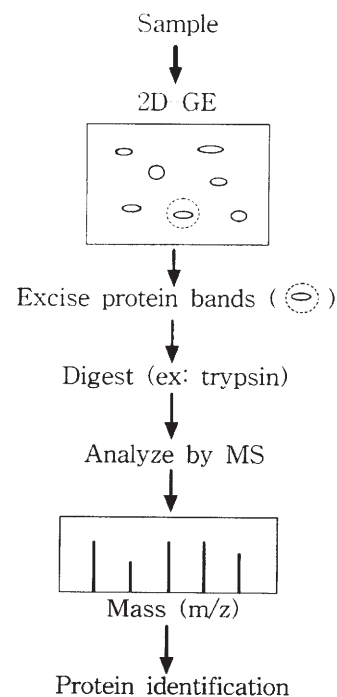


Fig. 1. A schematic diagram for proteomic analysis.

*To whom correspondence should be addressed.
Tel: 1-203-865-0100, Fax: 1-203-787-7876
E-mail: dong-wook.kim@yale.edu, dong-wook1@yahoo.com

기본적으로 ion source, mass analyzer, ion detector로 구성되어 있으며, 두 타입의 MS가 proteomics에서 흔히 사용되어 진다 (6,7). Time-of-flight (TOF) MS는 single stage MS로 proteolytic enzyme에 의해 만들어진 peptide들의 mass를 분석함으로써 mass map (mass finger print)을 얻고 이러한 mass map을 database 내에 있는 단백질들의 mass map과 비교함으로써 단백질을 동정할 수 있다. 또 다른 타입의 MS로는 tandem MS로 mass analyzer가 2개(MS/MS)로 구성되어 있다. 효소 처리된 단백질이 1차 mass analyzer 통과 후 특정 peptide가 분리되어진다. 이 peptide는 collision cell 내에서 다시 분해(fragmentation)되고 fragmentation들의 mass가 2차 mass analyzer에서 측정되어진다. 대개 fragmentation이 아미노산 사이에서 일어나므로 mass spectrum에서 peak 사이의 mass 차이를 이용해 아미노산 sequence를 추론할 수 있다. 이러한 tandem MS는 제한된 database의 극복이라는 장점이 있으나 노동이 많이 요구되는 단점이 있다. MS/MS 분석을 위한 peptide sample들을 깨끗이 하기 위해 HPLC (high performance liquid chromatography)를 사용하는데(LC-MS/MS), 이는 sample 내의 염들을 제거하거나 peptide를 농축하는 역할을 한다.

Proteomic Applications

현재 cell biology의 많은 분야에서 proteomics의 응용이 이루어지고 있다. 여기에서는 최근의 연구 결과들을 중심으로 몇가지 예를 설명함으로써 이 분야에서 proteomic application의 이해를 돕고자 한다.

1) 단백질 발현의 정확한 정량

이것은 두 가지 서로 다른 cell type, 예를 들면 wild type과 mutant에서, 또는 같은 cell의 서로 다른 환경 속에서 특정 단백질의 발현 정도를 정확히 비교하기 위한 것으로 배양하는 cell을 ^{15}N 으로 labeling하거나(8) 배양후 단백질의 cysteine을 labeling하는 방법(9)을 사용한다. ^{15}N labeling을 예를 들어 설명하면 (Fig. 2) #1 cell pool은 unlabeled (^{14}N) 되고 #2 cell pool은 질소원으로 ^{15}N 을 사용해 labeling된다(8). 적당히 배양된 후 각각의 cell pool은 같이 모아지고 2D-GE와 MS에 의해 특정 단백질이 분석되어진다. ^{15}N 의 peptide 속으로의 incorporation은 peptide의 mass를 약간 변형시키므로 결국은 두 종에서 유래한 ^{14}N -peptide와 ^{15}N -peptide가 peak pair를 만든다. 이러한 peak pair의 intensity를 측정(Fig. 2에서는 ^{15}N -peptide의 intensity가 ^{14}N -peptide의 intensity의 약 50%에 해당)함으로써 peptide (protein)의 상대적 발현 정도를 정확히 측정할 수 있다.

2) Post-translational modification의 연구

Proteomics의 가장 큰 잠재력을 가진 응용 분야중 하나가 단백질 전사후 수식(post-translational modification)에 대한 연구이다. 세포에는 수많은 단백질의 전사후 수식이 일어나며(8,10,11) 이것은 단백질의 기능에 지대한 영향을 미치므로 이런 변화를 체크할 수 있는 능력은 대단히 중요한 생물학적 의미를 갖는다. 대표

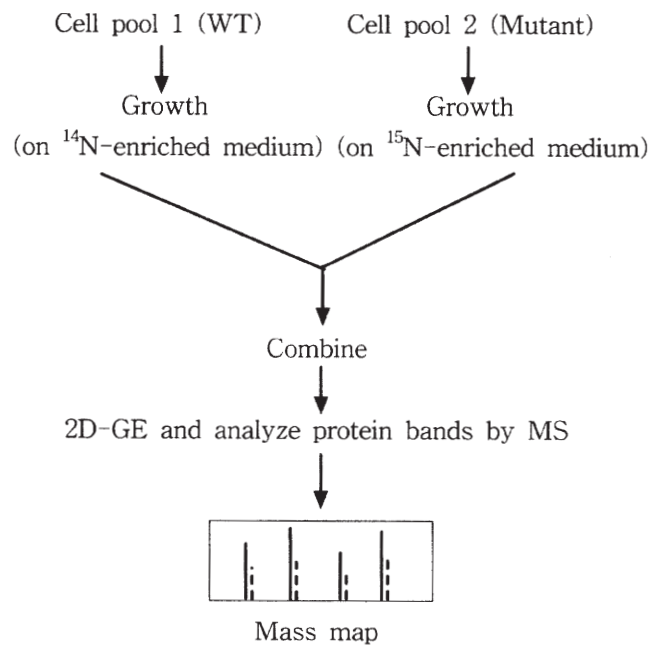


Fig. 2. Method for quantifying the expression of protein from two different cell types. In a mass map, solid lines indicate ^{14}N -peptide peaks of a protein from wild type (cell pool 1) and dashed lines denote ^{15}N -peptide peaks of that from a mutant (cell pool 2).

적인 단백질의 전사후 수식중 하나가 phosphorylation인데 MS에 의해 그 수식의 정량이 가능하다. 예를 들어 Fig. 2와 같은 배양 조건 아래에서 두 가지 다른 cell pool에서 유래된 같은 단백질에 대한 phosphorylation의 정도를 MS에 의해 측정한다면 Fig. 3과 같은 패턴의 mass spectrum을 얻게된다. 여기서 두 cell pool 사이의 한 단백질에 있어서 phosphorylation이 되지 않은 peptide들은 같은 양(같은 peak intensity)으로 존재하게 된다. 그리고 phosphorylation된 peptide (Xp)의 경우 그 peptide의 unphosphorylation form (Xu)으로부터 peak가 이동된다. 여기서 두 cell pool 사이의 peak의 intensity를 비교하면 site-specific phosphorylation의 정도를 측정할 수 있다(8).

3) Protein/Lipid Secretion에서의 응용

최근에 몇몇 논문이 단백질 또는 지질 분비와 관련되어 보고되었다(3,12,13). Wu 등(3)은 mammary epithelial cell의 두 functional state 즉 basal secretory state와 maximal secretory state로부터 Golgi를 분리해 Golgi proteome을 분석 비교하였다. 이러한 비교로부터 그들은 이 두 상태의 transition으로부터 membrane fusion 및 secretion의 조절에 관련된 약 30개의 단백질들이 Golgi에서 up-regulation 되고 있다는 것을 발견했다. 이러한 Golgi Complex의 functional proteomic analysis는 organelle function의 연구를 위한 proteomics 응용 가능성을 시사해 준다. 또한 그들은 mammary epithelial cell의 지질 분비에 관련된 두 subcellular fraction의 proteome을 분석하였다(12). 그리하여 그들은 지질 분비 기작과 endoplasmic reticulum (ER)이 기능적으로 연관돼 있음을 보여주고 있다.

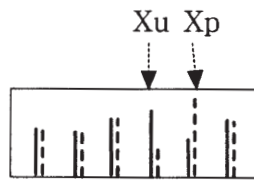


Fig. 3. Mass map of site-specific phosphorylation on protein. Solid lines indicate peptide peaks of a protein from wild type and dashed lines show peptide peaks of that from a mutant. Xu, unphosphorylated peptide X; Xp, phosphorylated peptide X.

4) Pathogen의 proteomics

현재 대략 30 여종의 bacteria genome의 완전한 sequence가 얻어졌고(1), 이에 대한 proteome들의 분석도 활발히 전개되고 있다. 이중 특히 각종 병을 일으키는데 관여하는 병원균들의 proteome에 대한 분석이 상당 부분을 차지한다(1,14-16). 이는 이런 병원균들의 proteome을 분석해 pathogen-host interaction, cell-cell signalling, antibiotic resistance 및 pathogenicity에 관여하는 인자들을 찾아내기 위한 기초로 사용하기 위함이다. 각종 분리 기술을 사용해 Regula 등은 human pathogen인 *Mycoplasma pneumoniae*로부터 688개의 predicted protein중 224개를 동정하였다(1). 또한 *Pseudomonas aeruginosa*에서는 membrane subproteome map이 만들어졌고(14), *Haemophilus influenzae*에서는 1,742개의 단백질중 502개가 동정되었다(15).

5) Subcellular organelle의 proteomic analysis

세포 레벨의 proteomic analysis가 때때로 단백질 분리나 확인 과정에서 어느 정도 한계를 갖고 있는 것이 사실이다. 어느 특정한 organelle의 proteome 연구는 분석되는 단백질의 수가 적고 적게 발현되는 단백질이 농축될 수 있다는 이점을 갖고 있다. 그리고 그 organelle 내에 존재하는 단백질의 기능을 연구하는데 매우 유용하다. Organelle proteomics의 몇몇 예를 들면 yeast의 spindle pole body (17)와 nuclear pore complex (18), human placental mitochondria (5), rat liver Golgi complex (4)의 proteome 분석 등이 해당된다.

6) Protein mixture의 shotgun identification

Multiprotein complex의 단백질을 1D- 또는 2D-GE를 거치지 않고 LC/MS/MS에 의해 직접 동정할 수도 있다. 이런 protein complex의 어떤 알려진 subunit에 myc (19,20), protein A (21) 또는 TAP (22)과 같은 affinity tag를 붙이면 이런 complex를 쉽게 정제할 수 있다. 그리고 세포내 적은 양으로 존재하는 단백질의 농축도 가능하다. 이런 protein complex는 protease로 분해되고 peptide들의 혼합물은 tandem MS (LC/MS/MS)에 의해 분석이 가능하다. 얻어진 tandem mass spectrum이 computer algorithm (예: SEQUEST)를 사용하면서 amino acid sequence database를 검색하는데 사용된다. 이러한 일종의 단백질 shotgun identification은 protein-protein interaction의 동정에도 이용될 수 있다(6). 이러한 shotgun identification 방법은 genome sequence가 밝혀진 세포에서 응용력이 크다.

7) Protein chips

최근에 protein microarray 기술을 이용한 단백질들의 large-scale 분석이 보고되었다(23-25). Protein microarray란 정제된 적은 양의 단백질을 함유하는 높은 밀도의 grid들인데, 이를 통하여 단백질들의 생화학적 활성, protein-protein 및 protein-DNA(RNA) interaction 등이 검색될 수 있다. 예를 들어 Zhu 등 (25)은 protein microarray를 통하여 효모 유래 protein kinase의 분석을 시도하였다. 효모 protein kinase 유전자들을 cloning한 후, 17 개의 다른 기질을 사용하면서 119개의 protein kinase가 분석되어졌다. 먼저 기질이 chip 위에 결합되고 그 후 효모 kinase들과 radiolabeling된 ^{33}P - γ -ATP가 각각의 microwell에 첨가되었다. Chip의 washing 후 phosphorimager를 사용해 phosphorylation된 기질이 검색되어졌다. 이런 분석을 통해 그들은 효모에 약 27개의 potential tyrosine kinase가 있음을 밝혔다.

이제까지 몇가지 대표적인 proteomics의 응용 분야에 대하여 살펴보았다. 이외에도 nervous system의 연구라든가(26) metabolic network의 연구(27), cardiovascular study (28), GTPase signaling network의 분석(29) 등 많은 분야에서 proteomic analysis가 이루어지고 있다.

맺는 말

Proteomics는 genomics와 더불어 세포내 기능을 총체적으로 이해하려는 종합적 분석 방법이다. 최근 몇 년 사이에 post-genome 시대의 기술로서 proteomics의 연구가 급진적으로 이루어져 왔다. 그러나 proteomic analysis를 효과적으로 수행하기 위해서 sample의 분리, 2D-GE, MS 등에서 아직도 많은 기술의 발전이 요구된다. 특히 발현량이 적은 단백질, 소수성 단백질, 염기성 단백질 등을 분리하기 위한 분리과학의 연구가 뒷받침되어야 한다. 그리고 genomics와 bioinformatics 연구에서의 진전 또한 상호 보완적인 측면에서 중요한 요소들이다. 이러한 제한적 요소가 있음에도 불구하고 지난 몇 년간 이루어진 proteomics에서의 혁신적 연구 성과는 genomics에서의 업적과 더불어 생물학의 기초연구에는 물론이고 응용연구, 예를 들어 질병 치료를 위한 신약의 개발 등에도 큰 기대감을 주고 있다. 21세기는 세포라는 소우주의 시간적, 공간적 기능 연구를 위해 proteomics의 응용이 더욱 활기를 띠 것이다.

참고문헌

1. Regula, J.T., B. Ueberle, G. Boguth, A. Gorg, M. Schnolzer, R. Herrmann, and R.R. Frank. 2000. Towards a two-dimensional proteome map of *Mycoplasma pneumoniae*. *Electrophoresis* 21, 3765-3780.
2. Santoni, V., M. Molloy, and T. Rabilloud. 2000. Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* 21, 1054-1070.
3. Wu, C.C., J.R.III Yates, M.C. Neville, and K.E. Howell. 2000. Proteomic Analysis of Two Functional states of the Golgi Complex in mammary epithelial cells. *Traffic* 1, 769-782.
4. Taylor, R.S., C.C. Wu, L.G. Hays, J.K. Eng, J.R.III Yates, and K.E.

- Howell. 2000. Proteomics of rat liver Golgi complex: Minor proteins are identified though sequential fractionation. *Electrophoresis* 21, 3441-3459.
5. Rabilloud, T., S. Kieffer, V. Procaccio, M. Louwagie, P.L. Courchesne, S.D. Patterson, P. Martinez, J. Garin, and J. Lunardi. 1998. Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: toward a human mitochondrial proteome. *Electrophoresis* 19, 1006-1014.
6. Yates, J.R.III. 1998. Mass spectrometry and the age of the proteome. *J. of Mass Spectrometry* 33, 1-19.
7. McDonald, W.H. and J.R.III Yates. 2000. Proteomic tools for cell biology. *Traffic* 1, 747-754.
8. Oda, Y., K. Huang, F.R. Cross, D. Cowburnand, and B.T. Chait. 1999. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* 96, 6591-6596.
9. Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994-999.
10. Krishna, R.G. and F. Wold. 1993. Post-tranlational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* 67, 265-298.
11. Wilkins, M.R., E. Gasteiger, A.A. Gooley, B.R. Herbert, M.P. Molloy, P.-A. Binz, K. Ou, J.-C. Sanchez, A. Bairoch, K.L. Williams, and D.F. Hochstrasser. 1999. High-throughput mass spectrometric discovery of protein post-translational modifications. *J. Mol. Biol.* 289, 645-657.
12. Wu, C.C., K.E. Howell, M.C. Neville, J.R.III Yates, and J.L. McManaman. 2000. Proteomics reveal a link between the endoplasmic reticulum and lipid secretory mechanism in mammary epithelial cells. *Electrophoresis* 21, 3470-3482.
13. Wu, C.C., R.S. Taylor, D.R. Lane, M.S. Ladinsky, J.A. Weisz, and K.E. Howell. 2000. GMx33: A novel family of trans-Golgi proteins identified by proteomics. *Traffic* 1, 963-975.
14. Nouwens A.S., S.J. Cordwell, M.R. Larsen, M.P. Molloy, M. Gillings, M.D.P. Willcox, and B.J. Walsh. 2000. Complementing genomics with preteomics: The membrane subproteome of *Pseudomonas aeruginosa* PAO1. *Electrophoresis* 21, 3797-3809.
15. Langen, H., B. Takacs, S. Evers, P. Berndt, H.-W. Lahm, B.D. Wipf, C. Gray, and M. Feountoulakis. 2000. Two-dimensional map of the proteome of *Haemophilus influenzae*. *Electrophoresis* 21, 411-429.
16. Wasinger, V.C., J.D. Pollack, and L. Humphery-Smith. 2000. The proteome of *Mycoplasma genitalium*. *Eur. J. Biochem.* 267, 1571-1582.
17. Wigge, P.A., O.N. Jensen, S. Homles, S. Soues, M. Mann, and J.V. Kilmartin. 1998. Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass apectrometry. *J. Cell Biol.* 141, 967-977.
18. Rout, M.P., J.D. Aitchison, A. Suprpto, K. Hjertaas, Y. Zhao, and B.T. Chait. 2000. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* 148, 635-651.
19. Sacher, M., Y. Jiang, J. Barrowman, A. Scarpa, J. Burston, L. Chang, D. Schieltz, J.R.III Yates, H. Abeliovich, and S. Ferro-Novick. 1998. TRAPP, a highly conserved novel complex on the cis-Golgi that mediates vesicle docking and fusion. *EMBO J.* 17, 2494-2503.
20. Kim, D.W., M. Sacher, A. Scarpa, A.M. Quinn, and S. Ferro-Novick. 1999. High-copy suppressor analysis reveals a physical interaction between Sec34p and Sec35p, a protein implicated in vesicle docking. *Mol. Biol. Cell* 10, 3317-3329.
21. Sacher, M., J. Barrowman, D. Schieltz, J.R.III Yates, and S. Ferro-Novick. 2000. Identification and characterization of five new sub-units of TRAPP. *Eur. J. Cell Biol.* 79, 71-80.
22. Rigaut, G, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030-1032.
23. Williams, D.M. and P.A. Cole. 2001. Kinase chips hit the proteomics era. *Trends in Biochem. Sci.* 26, 271-273.
24. Zhu, H. and M. Snyder. 2001. Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* 5, 40-45.
25. Zhu, H., J.F. Klemic, S. Chang, P. Bertone, A. Casamayor, K.G. Klemic, D. Smith, M Gerstein, M.A. Reed, and M. Snyder. 2000. Analysis of yeast protein kinase using protein chips. *Nat. Genet.* 26, 283-289.
26. Husi, H. and S.G.N. Grant. 2001. Proteomics of the nervous system. *Trends in Neurosci.* 24, 259-266.
27. Ideker, T., V. Thorsson, J.A. Ranish, R. Christmas, J. Buhler, J.K. Eng, R. Bumgarner, D.R. Goodlett, R. Aebersold, and L. Hood. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929-934.
28. Arrell, D.K., I. Neverova, and J.E. Van Eyk. 2001. Cardiovascular proteomics: evolution and potential. *Circ. Res.* 88, 763-773.
29. Alton, G., A.D. Cox, L.G.3rd Toussaint, and J.K. Westwick. 2001. Functional proteomics analysis of GTPase signaling networks. *Methods Enzymol.* 332, 300-316.

(Received May 25, 2001/Accepted May 31, 2001)

ABSTRACT : Proteomic Application in Cell Biology

Dong-Wook Kim (Boyer Center for Molecular Medicine and the Department of Cell Biology,
Yale University School of Medicine, New Haven, CT 06510, USA)

As the complete genomic sequences accumulate, the use of global techniques became possible. DNA microarray is a powerful technology for measuring global mRNA levels. This method, however, does not provide information on post-translational modifications of proteins. In addition, mRNA levels do not strictly correlate with protein concentrations, especially for lower-abundance proteins. Therefore, studies at the level of transcription are not sufficient to understand cellular activity. Proteomic techniques to analyze protein expression and function at the large-scale have been developed and used. This review introduces a simple explanation for proteomic analysis and examples of how proteomics is applied in cell biology.