

## Cancer and Proteomics-What Can You Do with Mass Spectrometry?

경상대학교 생명과학부

임 동 빈

Hillenkemp 그룹에서 처음으로 MALDI (Matrix-assisted Laser Desorption Ionization)를 이용한 단백질의 이온화에 성공한 후 Mass Spectrometry의 생물학적 응용 기술이 속속 개발되어 이제 MALDI는 생물학 연구의 필수적인 기기가 되었다. 이 글은 질량분석기에 익숙하지 않은 생명 과학자들에게 Mass Spectrometry가 단백질 분석에 어떻게 이용되고 있는가를 간단히 설명함으로써 그들이 Mass를 이용한 실험을 디자인하는데 도움이 되고자 하였다.

### 단백질의 동정

프로테오믹스의 대중화에 힘입어 가장 널리 알려진 질량분석기의 대표적인 응용분야이다. 트립신과 같이 특정 아미노산을 인지하는 단백질분해효소를 사용하여 단백질을 절단한 후 그 생성물인 펩타이드 조각들의 분자량을 측정한다. 이들의 질량은 그 펩타이드를 구성하는 아미노산에 따라 결정되므로 어떤 단백질에서 유래한 펩타이드 조각들의 질량 값은 아미노산 조성이나 아미노산 서열처럼 그 단백질의 고유 속성이다(따라서 이를 단백질의 펩타이드 질량 지문 -Peptide mass fingerprint- 이라 부른다.). 단백질 동정은 이 고유속성을 데이터베이스 내의 모든 단백질에 대하여 트립신 처리할 경우 나타날 수 있는 펩타이드 질량지문과 비교하여 서로 일치하는 단백질을 찾아 이루어진다. 펩타이드 질량지문법의 가장 큰 장점은 실버 염색으로 검출되는 양의 시료로부터 빠르고 쉽게 단백질을 동정할 수 있다는 점이다. 그러나 이 방법은 시료 단백질의 아미노산 서열이 이미 알려져 있는 경우만 적용할 수 있다. 필자는 T.reesei라는 곰팡이로부터 유래한 100여개의 단백질을 펩타이드 질량 지문법을 사용하여 동정을 시도한 바 있으나 단지 하나의 단백질만 동정에 성공한 경험을 가지고 있다. 이는 이 곰팡이의 게놈 데이터가 워낙 불완전하기 때문일 뿐 만 아니라, T. reesei가 효모와 같이 게놈 구조가 알려진 곰팡이와의 상동성이 매우 낮기 때문이다. 포유류인 캥거루의 경우 게놈데이터가 취약하긴 마찬가지이나, 인간이나 생쥐 등의 유전자와의 높은 상동성을 보이기 때문에 거의 80% 정도의 단백질이 펩타이드 지문법으로 동정되는 것을 확인한 바 있다. 또 질량이 일만 달톤 이하의 경우 프로테아제 처리로 생성되는 펩타이드의 개수가 적기 때문에 정확한 동정이 쉽지 않다. 이런 경우 트립신 산물 뿐만 아니라 V8 프로테아제 등 다른 효소 산물도 분석하여 동정의 신뢰도를 높일 수 있다.

## 단백질의 변이/수식에 대한 정보

세포내 실제로 존재하는 단백질은 다양한 번역 후 수식을 받는데 질량분석기를 이용하면 이들에 대한 정보를 얻을 수 있다. 예를 들어 단백질 내의 cysteine은 -SH, disulfide bond 등 다양한 형태로 존재하며, methionine은 methionine의 산화물인 methionine sulfoxide와 가역적으로 전환되면서 종종 단백질의 기능을 조절한다고 알려져 있다. Peptide mass fingerprinting를 수행하면 이들의 산화-환원 상태를 간단히 알 수 있다. 예를 들어 methionine을 함유한 펩타이드 조각의 측정 질량 값이 이론상 계산치 보다 18단위 높으면 이 methionine은 산화된 상태인 methionine sulfoxide로 존재함을 보여준다. 이와 마찬가지로 단백질에 phosphorylation이 일어나면 이론상 질량 값보다 80 증가하므로 특정 펩타이드의 phosphorylation 여부를 알 수 있겠으나, 인산화 여부를 아는 그리 간단치 않다. 우선 (1) 전체 단백질의 일부분만 인산화되므로 인산화 되지 않은 조각으로부터 인산화된 조각을 분리해야 하고, (2) 인산화는 펩타이드에 음이온을 첨가하는데, 펩타이드의 MALDI에서는 음이온 비효율적이기 때문에 다량의 시료가 필요하다는 점 때문이다. 따라서 인산화된 펩타이드를 metal chelate column 등을 이용하여 분리 농축한 후 아미노산 서열을 분석하여 인산화 자리를 알아내는 방법이 사용돼 왔으나, 최근에 인산화 분석을 위한 획기적인 방법이 개발되어 앞으로의 응용이 기대된다(아래 참조).

## 단백질의 아미노산 서열분석

Mass를 이용한 아미노산 서열 분석은 주로 MALDI-TOF보다는 Q-tof 또는 Qstar라는 hybrid type의 Mass spectrometer를 사용한다. 시퀀싱하고자 하는 펩타이드 조각을 질량차이를 이용하여 다른 펩타이드들로부터 선별한 후 불활성 기체와 충돌시키면 그 펩타이드는 여러 조각으로 깨지게 된다. 이때 생성되는 분해조각들의 질량을 측정하여 원래 펩타이드의 아미노산 순서를 추론해 내는 방법이다. 이 방법은 실버 염색으로 검출되는 단백질의 아미노산 순서도 알아낼 수 있는 에민도를 가지고 있고, Mass spectrometer 자체가 목표 펩타이드 조각을 분리하여 시퀀싱하므로 Edman sequencing에서 요구하는 펩타이드 조각의 분리 등이 필요 없다는 장점이 있다. 이 기술의 가장 큰 단점은 상당량의 펩타이드가 너무 복잡한 스펙트럼을 보여줘 현재의 기술로는 시퀀스를 읽기가 힘든 경우가 꽤 많다는 점이다. 이 문제는 조만간 spectrum을 읽어내는 좀더 나은 프로그램이 개발되면 극복될 것으로 여겨진다. 그러나 질량이 동일한 Leu과 Ileu은 원칙적으로 구별할 수 없고, 질량이 비슷한 Gln과 Lys도 현재 기기의 해상도로는 구별하기 힘든 단점도 있다.

## Quantitative Proteomics

Mass Spectrometer는 mass analyzer내로 들어온 이온의 질량을(실제로는 질량대 전하의 비) 측정한다. 따라서 mass spectrum 내의 peak intensity는 그 물질의 절대량과도 관계있지만 사용한 실험조건에서 그 물질이 얼마나 잘 이온화되느냐와 더 큰 상관성이 있다. 이온화의 정도는 물질의 화학적 성질에 따라 다르고, 또 실험과 실험사이의 정량적인 재현성이 매우 낮기 때문에, MLADI로는 무슨 물질이 있는가는 알 수 있어도 이 물질이 얼마나 있는가는 전혀 알 수 없다. 양적 비교를 위해

서는 화학적 성질은 동일하나 질량분석기가 분별할 수 있는 무겁고 가벼운 동위원소로(수소와 중수소, 또는 무겁고(N15) 가벼운(N14) 질소 등으로) 표지(labeling)하여야 한다. 질량차이에 대한 표지는 전통적으로 배지 내에 동위원소를 넣어 줌으로서 생체내(in vivo)에서 이루어져 왔으나, 1999년 in vitro에서 화학 반응을 통한 표지방법이 개발되어 Mass의 새로운 지평을 열었다. 두 가지 서로 다른 시료로부터 단백질을 추출한 후, 각각 중수소 또는 수소로 표지된 biotin과 반응시키면 단백질은 cysteine 잔기에 가볍거나 무거운 biotin으로 tagging 된다. 두 시료를 섞어 트립신 처리한 후 Avidin-affinity column을 통과시키면 biotin이 달린 펩타이드만 골라낼 수 있다. 이들의 Mass를 찍으면 동일한 단백질은 항상 한 쌍의 펩타이드 피크(가벼운 biotin이 붙은 것과 무거운 biotin이 붙은 것)를 보이게 된다. 이 쌍으로 나타나는 두 펩타이드의 화학적 성질은 동일하기 때문에 이온화 정도의 차이는 없을 것이고 따라서 두 펩타이드의 peak intensity 차이는 시료내 단백질의 양적 차이를 나타낸다. 이렇게 하면 두 시료간에 차이를 보이는 단백질을 쉽게 찾아 동정할 수 있을 뿐 만 아니라 서로 다른 조건에서 특정 단백질의 발현이 몇 배나 차이 나는가 하는 정량적 비교도 가능하다.

### Phosphoproteomics

만일 biotin을 단백질의 phospho amino acid에 특이하게 tagging시키고 트립신 처리 후 Avidin column을 통과시키면 인산화된 펩타이드 조각만 선별해 낼 수 있을 것이며, 이들의 아미노산 순서를 알아내면 어떤 단백질의 어디에 인산화가 됐는지를 알 수 있을 것이다. 전술한 Quantitative proteomics에서 biotin tagging 기술이 발표된 후 여러 그룹이 phospho aminoacid-specific tagging 방법의 개발에 열중하였던 바, 지난 4월 phosphoproteomics에 관한 세계의 논문이 일제히 보고되었다. 세 논문의 원리는 그림 1에 보여준 바와 같으며, phosphoserine/phosphothreonine/phosphotyrosine에 특이하게 biotin을 붙일 때 사용한 화학반응에서 서로간에 차이를 보인다. Phosphoaminoacid에 biotin을 tagging할 때 사용되는 현재의 화학반응이 복잡하기 때문에 최종 수율이 떨어지고, 따라서 아직은 sensitivity 등에서 극복해야 할 문제점이 있으나 앞으로의 개선여부에 따라 이 방법의 생물공학에 대한 기여는 막강해지리라 예상된다.

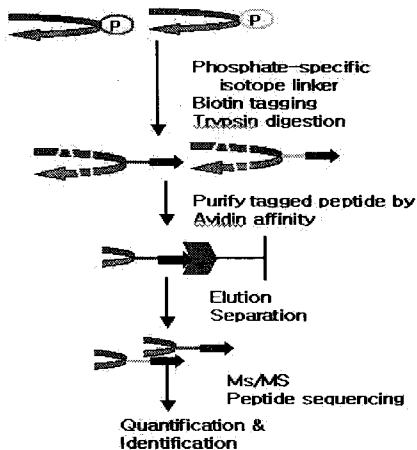


그림 1. Quantitative phosphoproteomics. 두 시료로부터 총 단백질을 추출한 후 한 시료는 무거운 링커(빨간색 막대기)로, 다른 시료는 가벼운 링커(노란색 막대기)를 사용하여 인산화된 아미노산에 특이하게 결합시킨다. 두 시료를 섞은 후 링커와 Biotin (파란색 화살표)을 결합시키고 트립신을 처리한다. Avidin column (빨간색)으로 biotin이 tagging된 펩타이드만 분리한 후 Mass를 찍어 쌍으로 나타나는 펩타이드의 양적 비교를 수행하여 두 시료 내에 있던 인산화 단백질의 상대적 양을 알아낸다. 이 펩타이드의 아미노산 순서를 결정하여 어느 단백질인지 알아낸다. 이렇게 하면 주어진 조건에 어떤 단백질이 얼마나 인산화 됐는지 알아낼 수 있다.

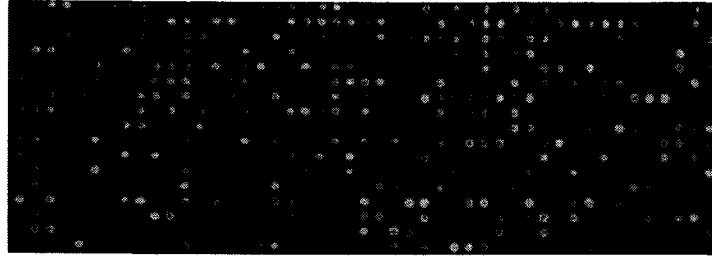


그림 2. 미래의 Quantitative phosphoproteomics 데이터. 각 스팟은 각각 서로 다른 인산화된 단백질을 나타내며 두 생리 조건에서 인산화 정도의 차이는 색의 진하기로 나타내었다. 노란색은 성장인자를 처리하기 전, 붉은 색은 성장인자를 처리한 후의 인산화 정도를 나타낸다.

## 결 론

현재의 Mass 분석은 80년대 초의 DNA sequencing 기술과 여러 면에서 유사점을 보인다. DNA Sequencing이 파플러해지기 전까지는 유전자의 클로닝 및 약간의 제한효소데이터만으로도 좋은 논문을 쓸 수 있었으나, 어느날 부터인가 유전자의 염기서열 데이터도 요구해와 논문 발표에 고생을 겪은 사람이 많을 것이다. 이와 마찬가지로 현상이 앞으로의 단백질 연구에도 나타날 것이다. 즉 단백질 관련 논문의 경우 활성을 보이는 단백질과 게놈데이터를 연결하는 peptide mass fingerprinting 결과를 기본적으로 요구하는 날이 향후 일이년 내에 반드시 오리라 믿는다. 또한 quantitative proteomics와 phosphoproteomics에서 사용하는 방법은 근본적으로 자동화가 가능한 기술이기 때문에 대규모적 분석에 대한 기술 개발이 이루어질 것이고, 따라서 현재 microarray 데이터를 보여주기 위하여 사용하는 그림 2와 같은 데이터 프레젠테이션 방법이 프로테오믹스 논문에서 심심치 않게 등장하는 날이 앞으로 4~5년 내에 오리라 확신한다.

## REFERENCES

1. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* 1999 Oct;17(10):994-999.
2. Oda Y, Nagasu T, Chait BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol.* 2001 Apr;19(4):379-382.
3. Ahn NG, Resing KA. Toward the phosphoproteome. *Nat Biotechnol.* 2001 Apr;19(4):317-8. No abstract available
4. Zhou H, Watts JD, Aebersold R. A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol.* 2001 Apr;19(4):375-378.
5. Goshe MB, Conrads TP, Panisko EA, Angell NH, Veenstra TD, Smith RD. Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. *Anal Chem.* 2001 Jun 1;73(11): 2578-2586