

## 프로테오믹스 기술의 현황 : 정량 비교 프로테옴 분석들

● 최 중 순(한국기초과학지원연구원 프로테오믹스팀 팀장)

### 서론

유전체 기술이 도입되면서 유전학 정보에 대한 우리의 지식이 확장되면서 유전학적 방법에 의한 맞춤형 생물체 제작이 가능해 졌다. 유전체 자체는 우리에게 일종의 생명현상에 대한네비게이션을 제공함으로써 우리가 어디에서 출발하여 어디로 가야하는지를 알려준다. 실제 생명 현상은 세포내 단백질들 간의 조화와 상호작용에 의해서 일어난다. 그러나 유전체 서열과 세포내 행동에 대한 교량을 연결할 기술적 격차는 여전히 존재한다. 그러므로 어떤 단백질들이 어떤 시간, 어떤 공간상에서 발현되는지를 결정하는 ‘단백질 스냅 사진’으로부터 세포들이 내부에서 실제 일어나는 ‘단백질의 일생’을 보여주는 영화에까지 확장될 수 있다. “프로테옴” 그리고 “프로테오믹스”라는 용어는 초기에 Siena 이차전기영동학회에서 Mark Wilkins에 의해서 각각 “유전자”와 “유전체”에 상응하는 용어로 사용되었다. 프로테오믹스(단백질체학)의 출현에 대한 기술적 혁명은 (i) 유전자와 전체 단백질 서열 정보의 풀 세트, (ii) 사용자가 광범위한 생물 데이터를 처리할 수 있는 사용하기 편리한 생물정보 도구의 발전, (iii) 유전체 전체의 발현에 대한 해석을 위한 마이크로칩 어레이 기반 기술의 발전 등을 들 수 있다. 현재, 다수의 유용한 도구들은 우리 손에 놓여 있고 전체 생명계를 들여다 볼 수 있어서 생물학자들이 보다 광범위한 방법으로 생각하기 시작하게 되었다(그림 1). 현 기고문에서는 이러한 독특한 프로테옴 기술들에 대해서 알아보고 정의하고자하였고 언젠가 산업적으로 적용할 수 있는 통합학문으로서의 가능성에 대해 알아보하고자 한다.

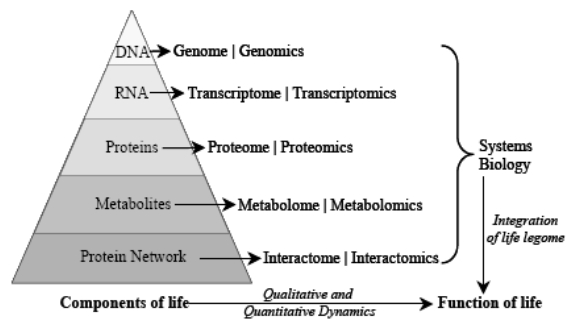
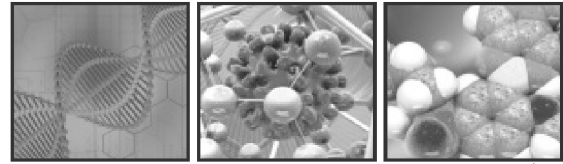


그림 1 생물 구성성분의 계층적 구조

### 생물계 내부에 대한 고찰

프로테오믹스는 종종 과학이라기 보다는 기술에 더욱 가깝다고 한다. 프로테오믹스 기술들은 강력한 프리즘 역할을 함으로서 세포내 다중 단백질들을 스크린하고 단일 단백질 분자의 특성을 규정하는데 그치지 않고 전체 생물계의 거동을 살펴보는데 있다. 이러한 관점에서 프로테오믹스를 연구하는 사람들의 궁극적인 목적은 단백질 화학을 연구하는 방법을 적용하는 것과는 전적으로 다르다. 단백질 화학은 세포내 단백질의 기능을 결정하기 위하여 표적 단백질들의 완전한 서열에 관심이 있으며 입체구조 결정을 통해서 추론할 수 있다. 전통적인 단백질 화학과는 다르게 프로테오믹스는 부분적인 서열 분석을 통하여 고속 발현 단백질들을 동정하고 궁극적으로 세포 기능의 체계적인 해석을 위한 단백질의 정량에 관심을 두고 있다. 그러므로 프로테오믹스는 생물 환경에서 프로테옴의 세포내 동력학적 거동에 대하여 조사하는데 초점을 맞추고 있으며 그러한 특징들은 유전체학적인 기술에 의해 정적이고 고정된 것과는 전적으로 다



른 것이다. 세포 유형에 관계없이, 관찰된 단백질들은 DNA 복제, 단백질 합성, 그리고 에너지 대사와 같은 생물학적으로 필수적인 기능을 갖는 항상 공통적으로 발현되는 것들이다. 레티날 상피세포에서 로돕신 같은 세포 특이적인 기능을 가지는 단백질들은 특화된 바이오 리소스에서 특이적인 기능을 가지는 표적 단백질 바이오마커가 될 수 있다. 특히, 신호전달, 전사인자, 그리고 세포주기 조절과 관련된 단백질들은 빠르게 발현이 변하며 특화된 세포에서 특이 기능에 대한 기능을 수행한다. 올리고유전자 칩 분석 툴의 개발은 세포내에서 유전체 수준에서의 발현 해석을 가능하게는 하지만 해당 단백질에 대한 발현 및 기능에 대해서 직접적으로 설명할 수는 없다. 전사체와 단백질체 간의 차이가 나는 세 가지 가능한 이유는 (i) mRNA의 안정성과 (iii) 전사 후 번역의 효율 등이 단백질 발현에 정성적, 정량적으로 영향을 주기 때문이다. 게다가 (iii) mRNA 수준 자체만으로는 주어진 세포가 조절되는 방식으로 자세하게 결정하기에는 충분하지가 않다. 이전 문헌에서 mRNA copy 수와 coding 단백질의 양적 상관성이 떨어진다는 상당한 수의 증거들이 축적되어 보고되었다. 유전체학과 전사체학은 우리들에게 각각 무언인가가 일어날지 모른다는 것과 무엇인가가 일어날 수 있다는 정보를 제공해주는데 반하여, 프로테오믹스는 세포내에서 지금 무엇인가가 일어나고 있는 것을 알려준다. 예를 들어, 특이하게도 인간의  $\alpha$ -antitrypsin은 인간 혈장내에서 22가지의 서로 다른 유형의 isoform들을 보이고 있다. 이러한 복잡한 번역후 수식화 현상은 유전체, 전사체 연구에서는 볼 수 없으며 단백질의 진정한 복잡성은 프로테오믹스 연구를 통해서만이 결정할 수 있다. 전체 인간 유전체 염기서열 결정이 완료되고 나서 많은 환원주의자들은 1013개의 세포를 가지고 있는 인간의 유전자 수가 ~26,000여개에 불과하며 959 세포로 이루어져 있는 예쁜 꼬마선충 (*Caenorhabditis elegans*)이 18,000여개의 유전자를 갖고 있다는 사실 (OnLine Database at <http://wit.integratedgenomics.com/GOLD>)에 매우 당황하였다. 그러므로 과학자들은 그

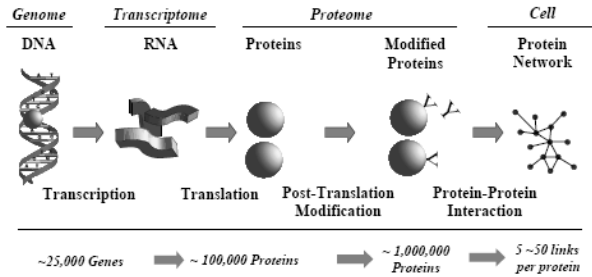


그림 2 세포내 복잡성의 기원

림 2에서 보는 것처럼 복잡성의 높은 차원은 유전자의 수만이 아니라 다양한 번역후 단백질의 수식화 과정이나 스케일 없는 단백질체 society의 네트워크에 기인한다고 생각한다. 따라서 프로테오믹스는 단백질을 동정하고 정량하며 궁극적으로 단백질의 위치와 수식화, 상호작용, 활성 그리고 세포내 기능을 연구하는데 그 목적이 있다.

### 이차원 젤에 기초한 단백질의 분리

분석 프로테오믹스는 대개 단백질 또는 펩티드의 분리로부터 시작된다. 그림 3에서처럼 질량분석기에 의한 프로테오믹스 분석의 단순한 전략은 단백질 혼합체를 분리하여 trypsin 분해과정과 단백질 혼합체의 전체 효소 분해 후, 온라인 MS에 의한 단백질들의 동정을 위한 분리과정으로 구분된다. 전자의 경우는 이차원 전기영동에 의해 단순히 수행되며 후자는 shot-gun proteomics 방법으로서 이차원 액체크로마토그래피가 연결된 온라인 질량분석기 (Liquid Chromatography-Mass Spectrometer)에 의해서 이루어진다. 그림 4에서 보는 것처럼 2-DE는 잘 알려진 단백질 분리 방법으로서 단백질들의 등전점 (isoelectric point, pI)과 분자량 (molecular weight, Mr)의 성질에 따라 분리한다. 상업적으로 pH 고정 구배 젤 (IPG strip gel)이 존재하여 이차원 젤 실험을 위한 안정된 pH 구배를 제공한다. 2DE의 장점은 이차원 젤을 이용하여 단백질 프로파일을 시각적으로 단순하게 전개하여 보여주며 번역 후 수식화와 같은 단백질 이성체들을 측정할 수 있게 해 준다. 대개, 인산화와 비인산화 단백질들은 2D-젤

# 기획특집

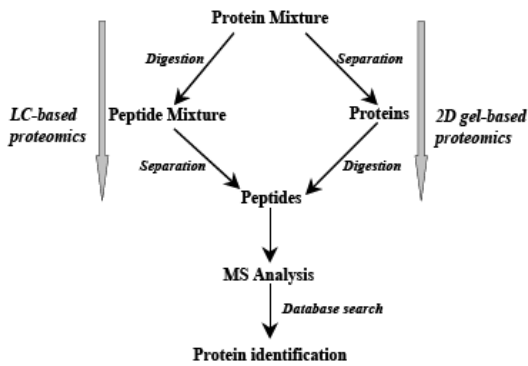


그림 3 두 가지 프로테옴 분석 워크플로우

상에서 수평으로 전개되는데 이것은 인산화 과정에 의한 pI 값들의 변화 때문에 일어난다. 최근에 향상된 2DE 방법으로서 차별화 젤 전기영동 (differential gel electrophoresis, DIGE) 기술이 소개되었다. DIGE는 두 가지 서로 다른 형광염료 (Cy5와 Cy3)를 이용하여 염색한 후, 단백질 시료들을 동일한 하나의 gel에 올려 순차적으로 분석하는 기술이다. 형광 염료는 N-말단 단백질의 amine 특이적 또는 lysine 기의 ε-amine side chain에 부착되어 동일한 단백질에 부착된 Cy5와 Cy3

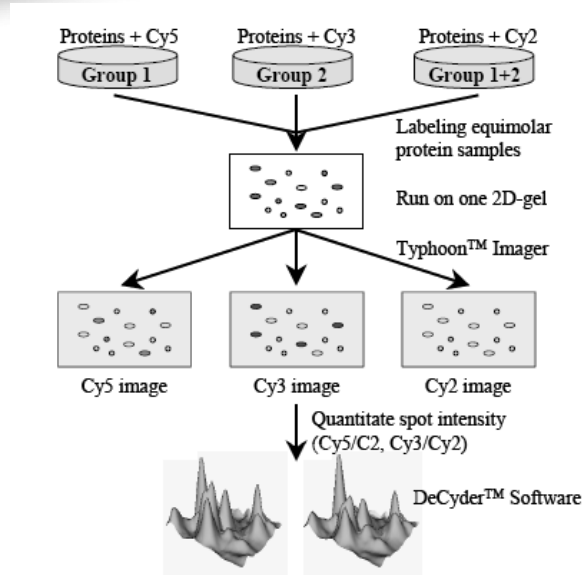


그림 5 DIGE 분석 과정의 모식도

의 형광 정도를 측정, 비교하는 것이다 (그림 5).

비록 2-DE에 기초한 프로테오믹스 방법은 단순하고 저비용의 방법이지만 노동 집약적이고 시간이 많이 소비되며 특히, 자동화가 취약한 단점이 있다. 또한, 강산, 강염기 pI 단백질들과 소수성 단백질들은

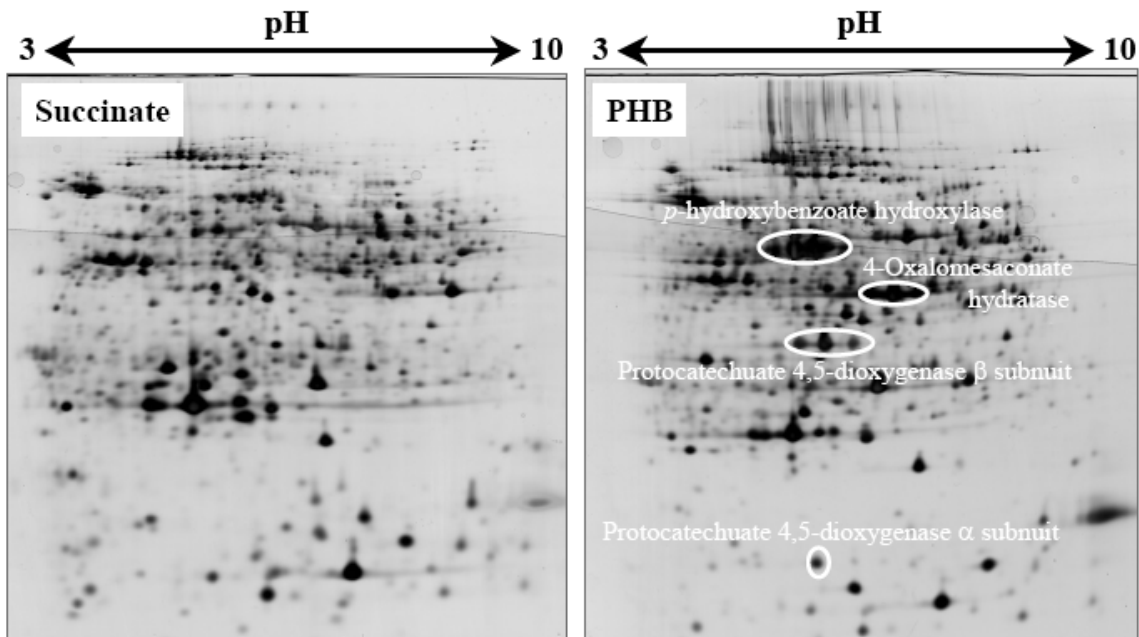


그림 4 토양 미생물 Pseudomonas putida KT2440으로부터 수용성 단백질의 발현 차이. (A) succinate, (B) benzoate로 배양한 세포의 이차원 젤 사진



# 기획특집

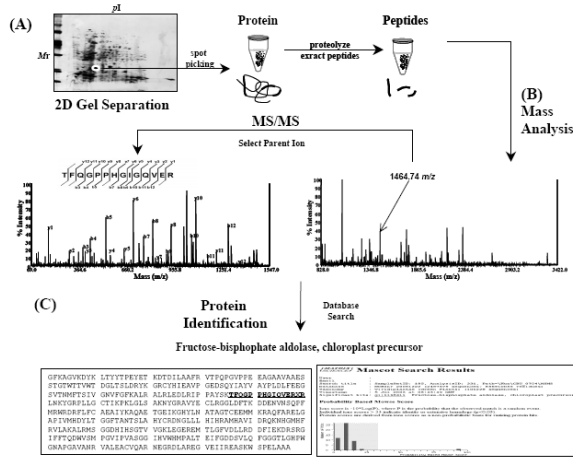


그림 8 MS/MS 단편화 과정에 의한 단백질 동정과정의 모식도

스 용액에 적용할 수 있다. 혼합액들은 MALDI plate에 적용하여 형성된 결정들은 nanosecond laser pulse로 이온화한 다음, 이온화된 펩티드들은 진공관을 통해 이동하여 detector에 도달하게 한다. 분자량은 질량 대 전하비 (mass-to-charge,  $m/z$ ) 형태를 주어진 역학에너지  $E=mc^2$ 로부터 시간비행 (time-of-flight) 분석을 통하여 계산할 수 있다. 축적된  $m/z$  데이터는 다시 Mascot (<http://www.matrixscience.com>) 검색엔진을 통하여 최대 스코어 순으로 PMF의 통계학적인 요약 형태로 정리하게 된다. Edman 서열분석법이나 PMF와 같은 동정 방법은 주어진 생리학적 조건하에서 상보적인 단백질 정보를 제공할 수 있다.

위에서 언급한 단백질 동정방법으로 해결되지 않는 경우에는 단백질 정보의 더 고차원적인 방법이 사용되는데 가장 많은 parent 이온물질의 추가적인 단편화 과정을 통해 결과적으로 daughter ion에 대한 정보를 얻을 수 있다. 이러한 MS/MS 분석을 통하여 한 단일 펩티드의 양방향 펩티드 서열은 b- 또는 y-이온 시리즈 분석을 통해서 얻을 수 있다 (그림 9). MS/MS 분석을 통해 얻는 de novo 펩티드 서열분석 데이터는 핵산과 단백질 데이터베이스를 통한 펩티드 서열분석 검색을 통하여 고신뢰 단백질 동정에 이용될 수 있다. Edman 서열분석처럼 MS/MS 분석을 통한 단백질 동정은 불완전한 게놈 서열을 가지고 있는 데이터

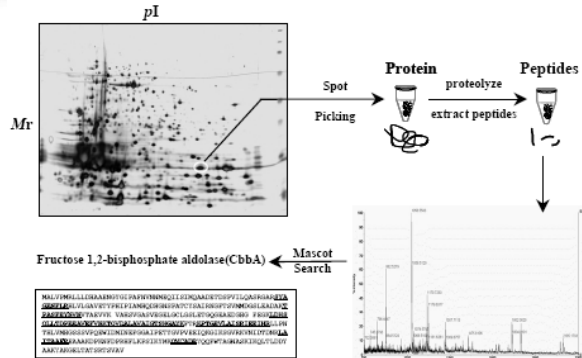


그림 9 펩티드분해과정 모식도

베이스용 생물이료 분석에 적극 활용할 수가 있다. 대량의 MS/MS를 통한 자동화 단백질 동정법은 온라인 나노스케일의 전자분무이온화 사중극자 TOF MS 나 오프라인 LC에 연결된 MALDI-TOF2 MS 시스템에 연결하여 사용된다. 이러한 연속 MS/MS 분석은 MudPIT을 포함하여 shot-gun 프로테오믹스 분석기술에도 필요한 요소들이다.

단백질 인산화와 같은 프로테오믹스 분석은 지금까지 개발된 기술 중에서 가장 도전받는 새로운 기술 중의 하나이다. 단백질 인산화는 가장 광범위하게 사용되고 가장 중요한 번역후 수식화 과정중의 하나로서 인산화는 생물학적 기능을 조절하는 것으로 잘 알려져 있다. 인산화는 세포분열, 대사조절, 그리고 신호전달과 같은 세포내 역동적인 다양한 과정의 하나이다. 신호대사과정에 대하여 인산화된 단백질들은 전체 프로테오믹스의 1~2%를 차지하며 kinase와 phosphatase 작용에 의한 역동적인 특징을 잘 보여준다. 전통적인 인산화 단백질을 측정하는 방법으로는 방사선동위원소 무기 인으로 인산기를 표지하여 자가방사선량을 측정한다. 다른 인산화단백질 동정 방법으로는 anti-phosphoserine, -phosphothreonine, -phosphotyrosine과 같은 인산화 특이 항체를 이용한 Western blot 방법이 있다. 그러나, 질량분석기를 통한 직접적인 인산화 부위의 결정은 인산화 단백질의 농축과 친화 크로마토그래피를 통한 인산화 펩티드 분획과정이 필요하다. 이러한 인산화단백질 또는 인산화펩티드의 가장 일반

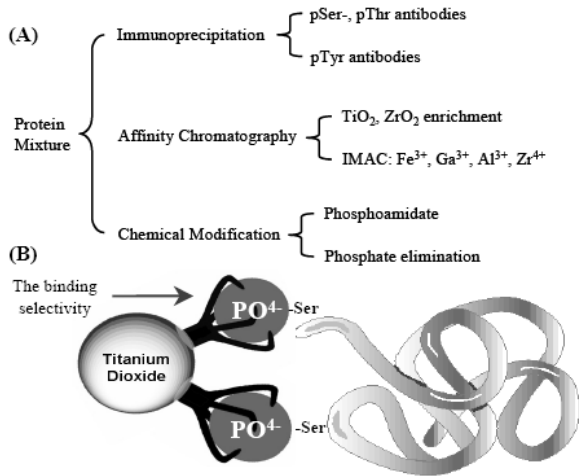
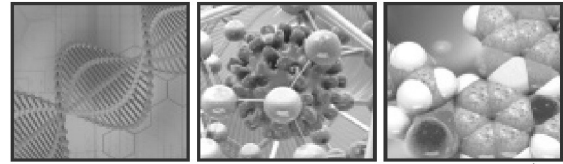


그림 10 인산화 단백질/펩티드의 분석방법

적인 농축방법은 고정화 금속이온 친화 크로마토그래피 (immobilized metal-ion affinity chromatography, IMAC) 방법으로서 histidine 표지 단백질을 선택적으로 농축하는데 사용된다. 금속이온으로는 Fe<sup>3+</sup>, Ga<sup>3+</sup>, Al<sup>3+</sup>, Zr<sup>3+</sup> 등이 인산화펩티드를 고순도로 정제하거나 농축하는데 사용된다, 다른 가능한 농축기술로는 titanium oxide 나노입자를 이용하여 산성 펩티드의 불순물을 제거하여 효율적인 인산화펩티드를 동정하는데 사용한다. 그림 10에서 보는바와 같이 인산화 단백질과 인산화 펩티드에 대한 농축 전략을 모식화하였다. 인산화 단백질 분석에 대한 방법들 중에서 인산화

펩티드의 선택적 정제 및 동정을 표준 인산화 단백질을 이용하여 검증하였다 (그림 11).

### 체계적인 정량 프로테오믹스 분석기술

프로테오믹스의 체계적인 정량분석법은 그림 3에서처럼 gel-기초한 분석방법과 LC-기초한 방법으로 구분된다. DIGE를 포함한 이차원 젤에 기초한 비교 프로테오믹스 분석법은 몇 가지 제한을 받는데 완전 자동화가 어렵고 미량 단백질들은 저평가되며 극산성, 극염기성 단백질들과 소수성 단백질들은 분리하기가 어렵다. 젤에 기초한 프로테오믹스 기술과는 다르게 LC에 기

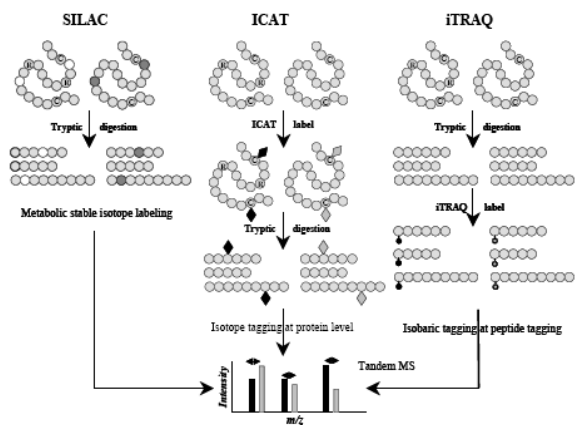


그림 12 정량분석기에 의한 정량분석 방법의 모식도

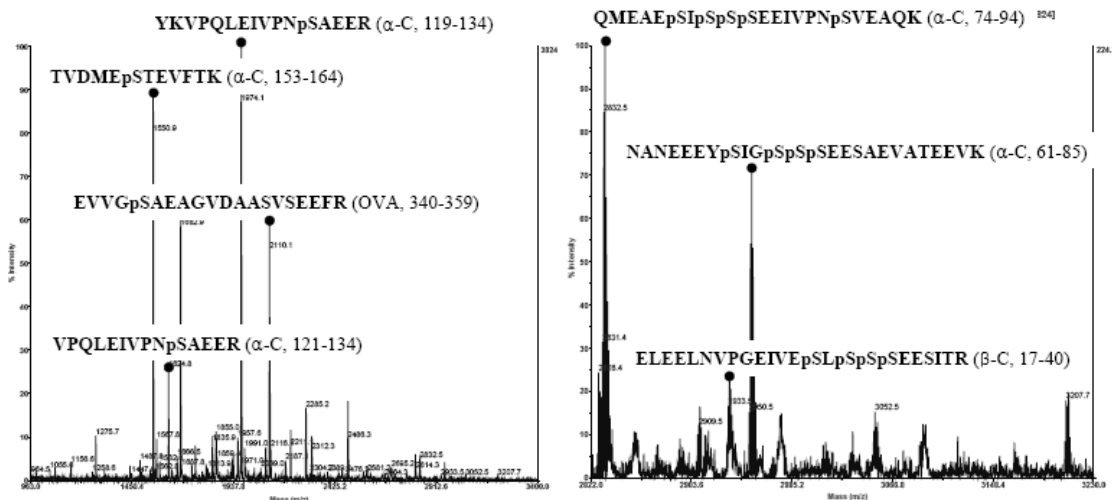


그림 11 인산화 펩티드 (표준 인산단백질, α-casein, β-casein, ovalbumin)의 농축과 결정

# 기획특집

초한 프로테옴 정량 비교 분석법은 다양한 접근이 가능하며 LC 분리를 통한 적절한 단백질 또는 펩티드의 크기로 downsizing하여 고속 자동화프로테옴 분석이 가능하다. 다중 LC 분석법은 MDLC (Multi-Dimensional Liquid Chromatography)로 약칭해서 부르는데 기본적인 분석 틀은 두 세포상태의 프로테옴의 체계적인 분석 방법으로 적용할 수가 있다. 두 가지 프로테옴을 단백질과 또는 펩티드 레벨에서 화학적으로 표지하여 LC-MS 분석을 통한 직접적인 상대 정량 분석이 가능하다 (그림 12).

화학적 표지의 prototype은 Aebersold 등에 의해서 개발된 isotope-coded affinity tag (ICAT)이 있다. ICAT 시약은 avidin column에 결합하는 biotin group과 8개의 수소 ( $D_0$ )나 중수소 ( $D_8$ )로 이루어진 polyether linker 그리고 cysteine 잔기에 특이적으로 결합하는 활성반응 group으로 구분 된다 (그림 13). 각각 시료조건에서 light ICAT과 heavy ICAT으로 표지된 혼합 프로테옴들은 트립신 분해가 가능하고 biotin-포함 펩티드를

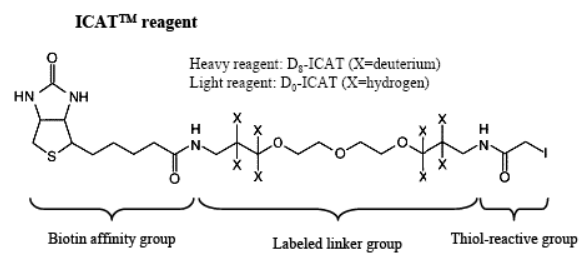


그림 13 ICAT 시약의 화학적 구조

avidin column에서 포획이 가능하다. 비결합 펩티드들은 완충용액으로 세척한 후, 표지된 cysteine-포함 펩티드들은 질량분석기를 통해 용출하여 정량화 할 수가 있다. 8 Da의 이중 피크는 두 가지 상태의 세포군에서 상대적 양을 나타내는 피크 강도의 정량을 위한 표적이 된다. 표적 피크의 단백질 동정은 양방향의 MS/MS 서열분석을 통하여 읽음으로서 가능해진다. 최근에는 새로운 절단 가능한 ICAT 시약이 개발되었는데  $^{13}C$  동위원소 linker 분자를 이용하여 산 분해가 가능한 biotin 그룹이 개발되어  $D_0$ 와  $D_8$ 이 같이 용출되던 문제

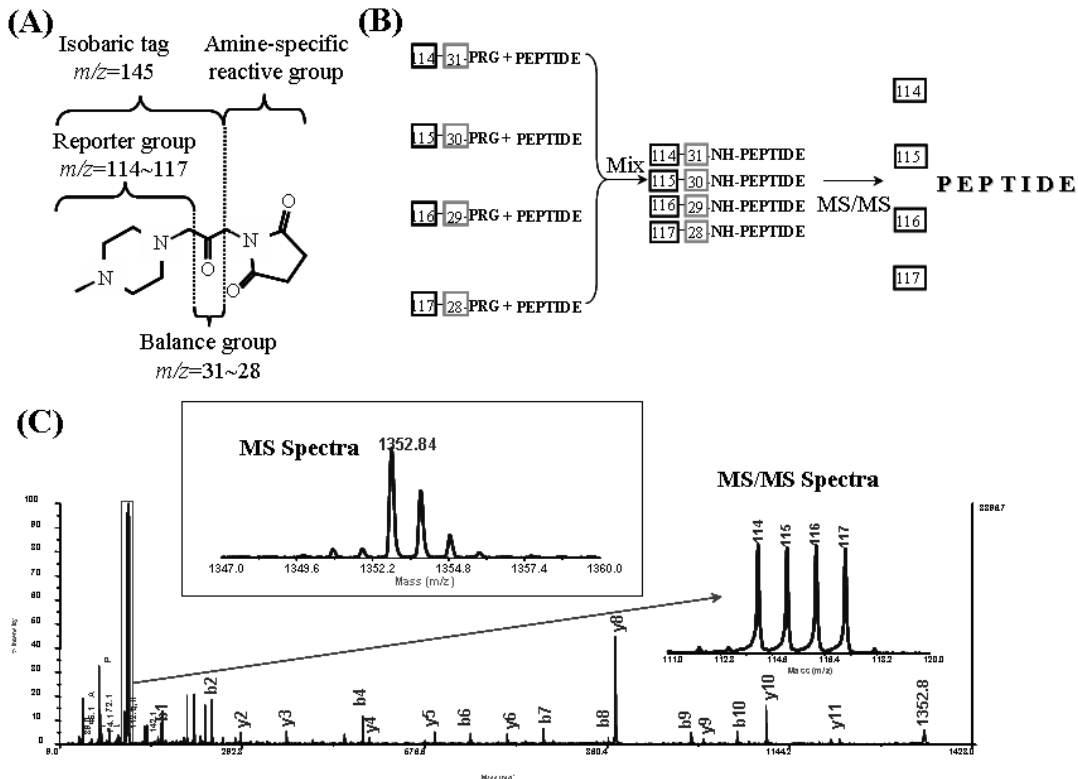


그림 14 4중-염색 iTRAQ™ 시약의 화학적 구조와 원리

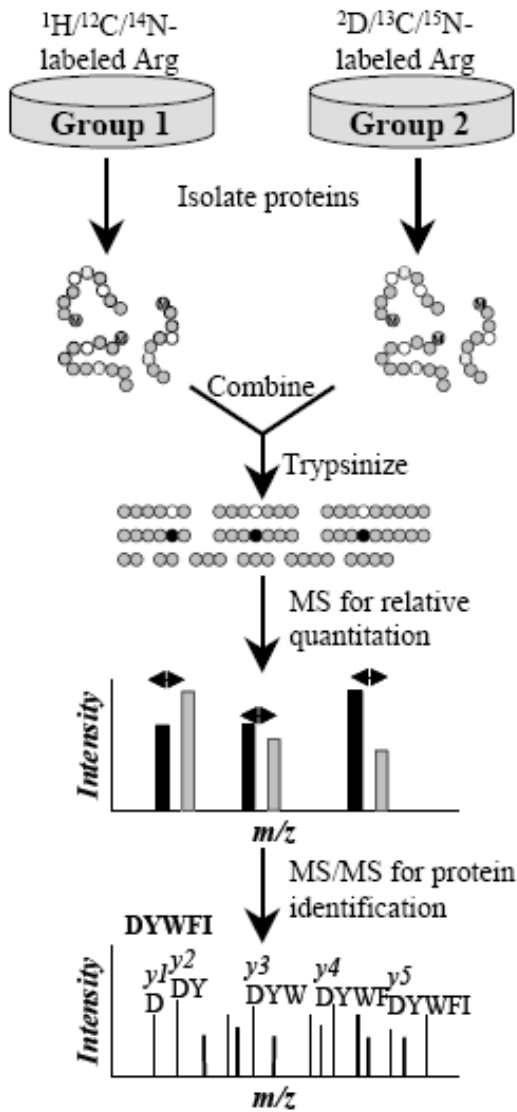
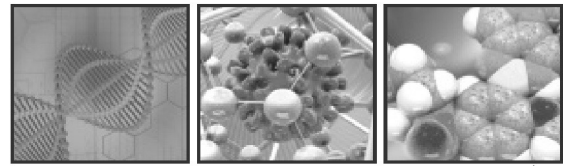


그림 15 SILAC *in vivo* 표지 방법 모식도

와 biotin group의 masking 효과를 없앨 수 있는 획기적인 방법으로 평가받고 있다.

또 다른 획기적인 정량 비교분석법으로서 iTRAQ (isobaric tags for relative and absolute quantitation) 방법이 개발되어 4가지 서로 다른 조건에서의 단백질 비교분석이 amine-특이적인 isobaric tag ( $m/z$  145) 이 가능해서 이론적으로 iTRAQ 시약이 전체 단백질을 표지할 수가 있다 (그림 14). iTRAQ 표지는 N-말단이나 lysine의 측쇄 amine group에 펩티드 표지가 가능하며 트립신 분해 후, 모든 단백질 표지가 가능

해진다. 4가지 isobaric tag으로 표지된 펩티드들은 직접적으로 질량분석기로 단백질을 동정하는 것이 아니라 4가지 조건의 단백질을 혼합하여 MS/MS 분석을 통한 signature ion peak들을 분자량 114~117까지의 이온들을 비교하여 정량분석 데이터를 획득하게 된다. MS/MS 스펙트럼들은 동시에 표적 단백질들을 동정하기 위하여 제공된다. 화학적 표지방법은 ICAT, cICAT 그리고 iTRAQ 등은 *in vitro* 상태에서 단백질 또는 펩티드를 표지하는데 사용되며 표지, 농축, 정제 과정을 거쳐 MS 또는 MS/MS 분석을 거치게 된다. 연속 화학적 반응 과정은 재현성이 떨어지고 낮은 신뢰도를 줄일 수 있다. 따라서 이러한 *in vitro* 표지법의 한계를 극복하여 개발된 방법이 필수 아미노산에 의한 동위원소 표지법 (stable isotope labeling by essential amino acid culture, SILAC)이 새로운 방법으로 개발되어 *in vivo*에서 자연표지에 의한 *in vitro*의 추출과정을 최소화하였다. 게다가 SILAC은 대사적 침투 방법으로서 살아있는 세포에 직접 표지되어 두 상태의 단백질의 변화를 배양세포로부터 단백질들을 직접 정량분석할 수가 있다 (그림 15).

통합 생물학을 향하여

빠른 단백질 분석기술 덕분에 특히, 질량분석기의 눈부신 발전으로 생물학자들은 이제 단백질 레고 조각으로부터 내부 세포기능의 동력학적 변화 상태를 추적할 수가 있게 되었다. 이것은 과거의 기술과 비교하여 중요한 진일보로서 세포 자체가 정적인 상태가 아닌 동적인 대상으로 입체적으로 이해하기 시작하였다. 현 기술들은 정량적인 정보 (단백질 동력학), 단백질 위치 (단백질 하부 아틀라스), 단백질-단백질 상호작용 (단백질 사회구성)의 다양한 데이터들을 수집하여 세포내에서의 역동적인 변화를 분자 네트워크, 수학적, 생물정보학적 기술과 접목되어 체계적인 해석이 가능해졌다. 과거 기술의 한계를 뛰어넘어 다양한 기술적 breakthrough들은 현 단백질 기술을 활용하여 생물학적 기능을 수치적



# 기획특집

인 데이터로 표현 가능한 정량 기술적 플랫폼을 제공하기 시작하였다. 그러므로 우리는 과거 다루기 힘들었던 세포내 수수께끼들을 수학적 틀로 적용하여 첨단 정량 프로테오믹스 분석 방법으로 풀기 시작했으며 이것을 우리는 통합 생물학 또는 시스템생물학이라는 신조어를 만들게 된 계기가 되고있다. 미래에는 생명현상의 단백질들의 동력학적 변화를 모니터링하면서 맞춤형 단백질 기반 의약품 제작이 가능해질 것이다.

## 참고문헌

1. Choi, J.S., Kim, D.S., Lee, J., Kim, S.J., Kim, S.I., Kim, Y.H., Jong, J., Yoo, J.S., Suh, K.H., and Park, Y.M., (2000) Proteome analysis of light-induced proteins in *Synechocystis* sp. PCC 6803: Identification of proteins separated by 2D-PAGE using N-terminal sequencing and MALDI-TOF MS. *Mols. Cells* 10, 705-711.
2. Choi, JS, Hung KY, Woo, SH. (2006) Quantitative proteomics towards understanding life and environment. *Kor J Environ Agr* 25, 371-381.
3. Kim, SI, Choi, JS, Khang HY. (2007) Proteomic strategy for the analysis of bacterial biodegradation pathways. *OMICS A J Int Biol* in press.

## 안내사항

정량 비교프로테오믹스에 대한 분석이 필요하시면

저희 기초(연) 프로테오믹스팀에서 연구지원을 해 드립니다. 단순지원에 대해서는 <http://www.kbsi.re.kr/dept/pat/>로 접속해서 안내를 받으시기 바랍니다. 공동연구에 대해서는 아래와 같은 분석이 가능하며 연구기간 및 비용 및 논문작성방법에 대해서 연구책임자간의 협의에 의해서 가능합니다. (☎ 042-865-3428, 019-207-2713, jschoi@kbsi.re.kr)

## \* 정량 프로테오믹스 안내

- 2-Dimensional Gel Electrophoresis/Differential Gel Electrophoresis/MALDI-TOF2 or ESI-QTOF MS: 이차원전기영동 (형광염료 표지/통계비교분석): 그룹당 1 mg 정도 단백질 필요
- 1D-SDS-PAGE/nESI-FT MS : 1D-SDS PAGE gel을 이용한 단백질을 전기영동한 후, in-gel digestion한 후, 고속 단백질 분석. 시료당 250~500 ug 소요
- iTRAQ/MS : 최대 4개 group 단백질 시료를 표지하여 정량 비교분석. 시료당 250 ug 내외 소요
- Phosphoproteome 분석 : 1D-gel을 이용한 단백질을 형광 인산화 단백질 염색 후, Ser, Thr, Tyr 잔기의 인산화 부위를 nESI-FT-MS 결정. 단백질량 최소 1 mg 소요
- 막단백질 순수 분리 및 막단백질 동정 : Aqueous polymer Two-phase를 이용한 막단백질 (plasma membrane, outer membrane 등) 순수분리 후, gel-base 또는 solution-base 막단백질 고속 분석. 시료량 1 mg이상 필요