

Proteomics의 최근 연구 기술동향

양혜정, 권대영
식품기능연구본부

최근 들어 게놈기능연구는 주요국가의 새로운 국가적 연구표적으로 지정되면서 유전자기반 생물산업의 핵심으로 부각되고 있다. 이러한 발전은 인간(2001년 2월 인간 게놈의 초안 발표)을 비롯한 생물체의 게놈구조가 규명되어 이 유전자구조 정보를 web상에서 쉽게 알아낼 수 있는데서 비롯된다. 포스트게놈시대의 게놈기능연구를 총괄적으로 '기능유전체학 (Functional Genomics)'이라고 하며 여기에는 핵산(DNA나 RNA)을 표적으로 게놈기능을 연구하는 genomics(유전체학, RNA발현을 대상으로 하는 transcriptomics(전사체학) 포함), 총체적인 단백체를 대상으로 유전자기능을 연구하는 proteomics(단백질체학) 및 대사물질을 대상으로 하는 metabolomics(대사체학), 이들 분야를 공통적으로 지원하는 bioinformatics(생물정보학)로 구분된다. 본 고에서는 프로테오믹스분야를 중심으로 소개하고자 한다.

1. 기능유전체학 (Functional Genomics)

먼저 그림 1에 기능유전체학의 카테고리를 간

단하게 요약하였다. 그림 왼쪽 부분은 유전자에서 전사체를 거쳐 단백질에 이르는 유전자 발현 과정을 나타내고, 삼각형은 대규모 실험법에 해당하는 영역을 표시한다. 유전체학(genomics)은 모델생물의 염색체 서열을 결정하고 개체간의 서열차이를 분석한다. 다음 단계인 전사체학(transcriptomics)은 전사체(mRNA)의 양을 비교하기 위해 microarray를 이용한다. 그 다음 단계에서 다루는 단백질은 아미노산의 서열과 변형, 3차원 구조, 농도, 다른 단백질과의 관계, 외부 환경에 따라 다르게 기능하기 때문에 단백질체학(proteomics)은 단백질의 발현 수준 외에도 구조와 변형, 위치, 상호작용을 연구한다. 소분자의 양을 결정하는 효소 기능을 조절하거나 효소로 작용하는 많은 단백질을 다루는 대규모 실험법은 공정과 대사물 흐름 분석에 있어 가장 앞서 있으며 종종 대사체학(metabolomics)이라 불린다. 이러한 정의가 절대적인 것은 아니며 명명법과 설명에 대한 일반적 합의가 없고 실험 방법도 종종 중복된다.

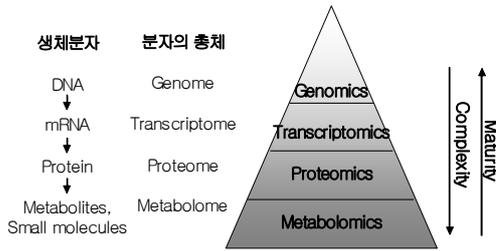


그림 1. 기능유전체학의 분류

2. 프로테오믹스(Proteome)과 프로테오믹스(Proteomics)

프로테오믹스(Proteome)는 어원적으로 단백질체 (Protein-body)라고 할 수 있으나, 실제로는 게놈(genome)의 상대어로서 <PROTEin expressed by a genOME>의 합성어로 인식되고 있는 용어로, 프로테오믹스는 프로테오믹스(proteome)의 어미에 ~학, ~론을 의미하는 접미사 -ics가 붙어 프로테오믹스를 연구하는 방법과 기술을 포괄적으로 의미하는 말로, 굳이 번역한다면 '단백질체분석학' 이라고 해야 할 것이다. 즉, 단백질의 성질을 발현, post-translational modification(해독후 변형), 다른 단백질과의 결합에 초점을 두어 연구함으로써 세포내 변형과정과 네트워크 형성을 질병의 진행과정과 연계시켜 총괄적으로 이해할 수 있는 연구분야를 뜻한다. 한 개의 단백질을 순서적으로 분석하던 통상의 단백질 생화학과는 연구대상의 볼륨, 속도, 분리수단의 자동화 및 게놈/단백체 DB정보와의 연동성 측면에서 매우 다른 특성을 지닌다. 프로테오믹스는 세포내 전체단백질을 연구하는 대형 스케일의 다단계 고속분석기술이므로 연구대상도 단백질의 발현(expression), 기능(function), 구조(structure) 및 생합성후 구조변형, 관련 단백질간의 결합성(protein-protein interaction)에 초점을 두고 질병의 진행과정과 연계시켜 총괄적으로 분석하는 기술이므로 유전체학보다 복잡하며 데이터양도 매우 크다.

프로테오믹스는 크게 세 부분으로 나눌 수 있다(그림 2). 첫째가 구조적 프로테오믹스(functional proteomics)로써 단백질의 고차구조에 관계하는 접합(folding)과 안정성(stability) 또한 이들 사이의 상관관계(interaction)를 밝히는 분야이다. 관련 기술로는 X-ray crystallography 및 NMR 등이 있다. 둘째는 발현 프로테오믹스(expression proteomics)로 발현양상, 신규 단백질의 규명, 단백질 생성과 소멸의 역학관계이며 관련기술로는 이차원 gel-mass spectrometry 등이 있다. 셋째는 기능적 프로테오믹스(functional proteomics)로 단백질구조의 이차적 변화, 활성화를 위한 단백질의 생분해 및 단백질과 그 외의 다른 인자(protein, ligands, DNA)와의 상관관계를 밝히는 분야이다. 관련기술로는 yeast two hybrid system, SPR(surface plasmon resonance), microarray protein chip 등이 있다.

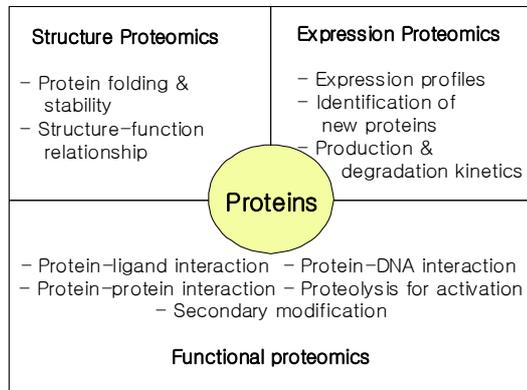


그림 2. 프로테오믹스의 분야

프로테오믹스와 지노믹스는 공히 유전자의 기능을 밝힌다는 점에서 목표는 같으나 방법이 매우 다르다. 기술적으로는 실험의 출발점도 전자가 단백질 차원에서, 후자는 유전자차원에서 각각 유전자의 기능을 연구하는 기술이므로 연구형태가 매우 다르다. 그러나 이 둘은 실제 유전자구조를 이용한다는 점에서 한 뿌리이며 상호 보완적인 연

구분야이다. 가령 지노믹스가 세포의 핵산 (DNA, RNA, 또는 이들의 유사체)을 대상으로 각 유전자 발현의 정도차이를 분석하고, 유사유전자들의 수집 및 분류, 특정 생리조건에 반응하는 유전자들의 대량분석(DNA칩 및 올리고 칩), 특정 유전자의 발굴(mining)을 수행하는 연구분야로서, 이미 알고 있는 유전자구조를 근간으로 연구하는 기술인데 비해, 프로테오믹스는 일단 유전자의 산물인 단백질을 대상으로 조직이나 세포, 체액과 같은 개체의 실시간 세포 생리현장에서 얻어진 시료를 대상으로, 이들을 대량으로 분석하고, 상호기능관계의 지도를 작성하며, 구조분석을 통하여 궁극적으로 특정 단백질과 이를 만드는 유전자의 기능을 동시에 밝혀 내는 기술이다. 실제로, 특정 유전자의 발현과 단백질의 발현차이에는 밝혀진 유전자를 기준으로 30% 이상이 상관성이 거의 없는 것으로 알려져, 그만큼 지노믹스의 수단으로 단백질과 유전자의 발현(mRNA) 상관성을 분석하고 예측하기는 불가능하다. 이 두분야에 속한 공통분야가 있는데, 구조유전체학 (sturctural genomics) 또는 구조프로테오믹스(structural proteomics)라고도 불리는 단백질 구조분석 연구이다. 이 분야는 프로테오믹스에 의해 밝혀진 특정 단백질이나 이미 알려진 게놈 DB상의 특정유전자 정보를 바탕으로 단백질의 1 차 구조를 예측하고, 구조와 기능간의 관계를 유추하여 궁극적으로는 생리적 활성상태인 가상적인 3-D 구조를 추측해 낸다. 이러한 결과를 사용하여 drug ligand나 receptor의 공간적 관계를 추정할 수 있으며, 이를 drug design에 응용할 수 있다.

아무튼 한 유전자가 최종적으로 세포 내에서 어떠한 기능을 하는가는 얼마나 정교하고, 적절하게 단백질 합성 후 변형되는가에 달려 있어서, 프로테오믹스를 이용해 최종적으로 완벽한 모양과 기능이 갖추어진 단백질을 분석하지 않고는 그 유전

자의 세포내 기능을 알 방법이 없기에 프로테오믹스의 독특한 가치가 여기에 있는 것이다.

3. Proteomics의 기술

Proteomics는 ①단백질발현, 목록 및 양 (expression, catalog, quantity), ②단백질구조 (structure), ③단백질의 해독후 변형 (post-translational modification), ④단백질의 세포내위치 (subcellular localization), ⑤단백질-단백질의 상호작용 (protein-protein interaction) 등 단백질을 포괄적으로 분석하는 기술을 포함한다.

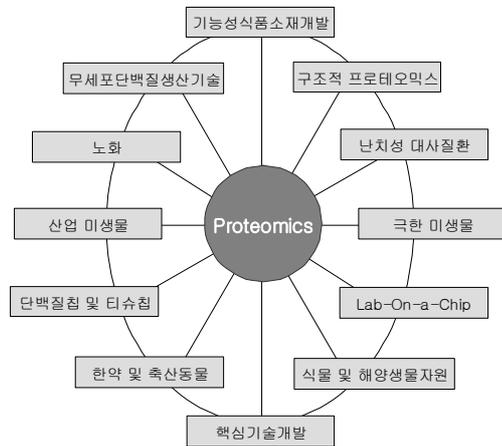


그림 3. 프로테오믹스의 응용분야 예시

3.1. 단백질 프로파일링

각 세포마다 전체 단백질의 발현 패턴은 고유의 특성을 갖고 있다. 한 종류의 세포 내에서 성장과정 중 일어나는 환경적 변화와 질병이 생겼을 때, 단백질이 이러한 변화를 어떻게 적응하는지를 이해한다면 분자수준에서 생물학적 과정을 설명하는데 중요한 단서를 제공할 것이다. 최근에는 세포 내 다수 단백질을 확인하고 조건변화에 따른 단백질 발현의 차이를 전체적으로 감지하기 위한 방법들이 개발, 사용되고 있다.

3.1.1. 2차원 전기영동 (2DE)

2차원 전기영동 방법은 여러 조건에서 프로테오ムの 발현수준을 탐지하기 위해 오래 전부터 사용되어 왔다. 단백질을 먼저 isoelectric focusing으로 일차적으로 분리한 후 직각 방향으로 SDS-PAGE를 걸어 단백질들의 분자량에 따라 분리한다. 이 방법은 수천개의 단백질을 하나의 젤에서 분리할 수 있지만, 다루기 힘들고 dynamic range가 좁으며 발현량이 많은 수용성 단백질에만 적용가능하다는 문제가 있다. 또한 분리된 단백질들의 위치만 갖고는 단백질 확인이 불가능하다. 2DE의 기술적 한계성(표 1)은 현대 proteomics에서 요구되어지는 대량 프로테오మ్을 빠른 시간내에 효율적으로 분석할 수 있는 high-throughput 실험방법이 아니라는 것을 명확하게 나타내고 있다. 그리고 2DE를 이용한 프로테오మ్의 분석에서 가장 어려운

점은 2DE 이미지의 분석과정을 들 수 있다. 일반적인 2DE 이미지 분석 프로그램의 경우 spot detection, pairing 등의 매우 중요한 이미지 분석 과정에 분석자의 주관적인 판단이 필수적으로 필요하기 때문에 이미지 분석에 많은 시간이 소요되며, 분석된 결과의 신뢰도가 떨어지는 결과를 낳는다. 그리고 2DE 분석과정 중의 오류는 데이터 베이스 형태로 모아서 분석할 경우 적절한 필터링 과정이 없으면 오류가 중첩되어 정상적인 분석이 매우 어렵다.

3.1.2. 질량분석기 (MS)

MS와 연결된 단백질 분리방법들이 개발되어 단백질 확인과 복합체 분석이 널리 사용되고 있다. 특히 단백질을 파괴하지 않고 이온화시키는 Matrix-Assisted Laser Desorption Ionization(MALDI)

<표 1> 2DE의 문제점

2DE의 문제점	내용
지질막 단백질분리	소수성 지질막 단백질의 분리가 용해성 문제로 잘 안됨. 용해용 detergent 카테일이 제공되는데 이들을 이용해도 분리가 어려움.
희소성 단백질 분리	Gel에 load 할 수 있는 샘플양의 한계로 희소성 단백질은 대부분 다른 단백질과 함께 이동하여 염색되는 게 보통이므로, 이들만을 별도로 분리하기 위해서는 pre-fractionation이 필요함.
호알카리성이나 호산성 단백질 분리	현재 비교적 광역대의 알카리성 영역대의 IPG strip(pH 6-12)이 나와 있긴 하지만 buffer 용액 조건 등의 문제로 아직도 개선이 필요함.
분자량이 큰 단백질의 분리	최근에 GradPore™라는 장치로 사전에 분자량별로 분획하는 기술이 도입이 되었으나, 단백질의 용해성과 맞물려 있어서 이 역시 최적조건의 개발이 필요함.
낮은 dynamic range	106이 넘는 프로테오మ్의 dynamic range를 수용하지 못함.
낮은 재현성	실험별로 동일한 샘플을 분석하고자 할때, buffer, IPG strip의 batch, gel size 등에 따라 spot의 분리양상이나 상대적인 위치가 실험 batch 별로 재현되는 것이 매우 떨어짐.
자동화, high throughput의 장애	일부 회사(예, Tecan, OGS, Large Scale Biology 등)에서 2DE 전과정을 자동화할 수 있는 장비를 판매하고 있으나, 도입단가가 매우 높다는 문제가 있음.

방법과 Electrospray Ionization(ESI) 방법이 Time-Of-Flight(TOF) MS와 함께 사용됨으로써 펩타이드와 단백질 같은 생체고분자 분석이 가능해졌다. 2차원젤이나 다른 방법으로 정제된 단백질을 트립신과 같은 아미노산 서열 특이적인 프로티아제로 자른 후 나온 펩타이드를 MS로 분석한다.

MALDI를 사용할 경우, 시료를 matrix와 섞어 말린 후 레이저를 쏘아주면 matrix가 흡수하여 기체화되면서 시료를 기체로 만들며 전하를 띠도록 한다. 강력한 전기장에 의해 전하를 띤 시료는 mass-to-charge(m/z) ratio에 따라 다른 속도로 MS로 주입된다. MALDI-TOF의 장점은 쉽게 단백질이나 펩타이드를 확인할 수 있고 중간정도의 시료 처리 능력을 갖고 있다는 것이다. 이 방법은 한 단백질을 형성하는 여러 펩타이드들의 질량 정보를 정확하게 주기 때문에 유전체에서 만들어지는 모든 단백질에 대한 펩타이드 질량 정보를 자동 검색하여 대상 단백질을 확인할 수 있다.

ESI의 장점은 시료를 액체에서 기체상으로 바꾼다는 것이다. 즉, 분무장치에 의해 만들어진 시

료의 수용액 방울이 MS내에서 증발되어 시료가 전하를 띤 기체상태로 되게 한다는 것이다. 일반적으로 ESI가 triple quadruple, ion trap 또는 hybrid TOF 중 하나와 연결되어 사용된다. MALDI와 비교하여 ESI는 LC와 HPLC와 쉽게 연결되어 펩타이드나 단백질 혼합물을 고효율 분석과 on-line 분석이 가능하다는 장점이 있다. 단백질 혼합물은 LC에 의해 분리된 후 펩타이드로 만들어 탠덤 MS(MS/MS)에서 분석된다. 첫 번째 MS에서 m/z에 따라 분리하여 m/z 리스트를 만들고 두 번째 MS에서는 특정 m/z를 갖는 펩타이드를 선택하여 collision cell에서 아르곤 기체와 충돌시켜 아미노산 한개의 질량 차이가 나는 ladder 형태의 조각으로 만든 후 펩타이드의 아미노산 서열을 결정하게 해준다. 이 서열을 데이터베이스에서 비교하여 단백질을 찾을 수 있다.

높은 질량 민감도를 갖는 MS로 Fourier transform ion cyclotron resonance MS (FTICR-MS)가 있는데 혼합물에 섞여 있는 낮은 농도의 화합물이나 단백질을 확인하고 비슷한 m/z ratio를 갖는 물질들을 분석하는데 사용된다. 500Da 미만의 단일화

<표 2> 질량분석 방법의 특성비교

	MALDI-TOF MS	NanoES-MS/MS	NanoLC-MS/MS	LC/LC-MS/MS
No use of gels	Very poor	poor	Good	Very good
Use with 1D gels	Good	Good	Very good	Fair
Use with 2D gels	Very good	Very good	Good	Poor
Throughput	Very high	Poor	High	High
Sensitivity	Very high	Very poor	High	Fair
Dynamic range	Poor	Fair	Good	Very good
Mixture analysis	Fair	Good	Very good	Very good
Robustness	Very high	Fair	High	Fair
Special expertise	Low	Fair	High	High

(Eur. J. Biochem 2003)

합물을 분석할 수 있고 단백질 혼합물에서 펩타이드의 질량을 정확히 측정할 수 있다. 표 2에 각 질량분석 방법의 특성을 비교하여 나타내었다.

3.1.3. Protein microarray

단백질 microarray는 단백질이나 펩타이드, 항체, 항원, 탄수화물, 소분자 등의 리간드를 표면에 적어서 사용하며 단백질의 발현 양상 분석이나 임상적 진단에 쓰이고 있다. 인간의 수많은 단백질에 대한 항체를 생산하기가 어렵기 때문에 대부분 사이토키인과 재조합 단백질에 이용이 제한되어 왔다. 이 외에 단백질이 상호작용 능력을 유지해야 하고 생체 내 실험이 뒤따라야 한다는 단점이 있지만, 최근에는 약물-결합능력 분석을 통한 신약개발 및 검증, 재조합 단백질과의 반응을 통한 항체의 특이성 조사에도 쓰이고 있다.

항체 microarray는 적은 수의 단백질을 고효율 방식으로 프로파일링하는데 적합하다. 유리 슬라이드 표면에 항체를 붙인 후 세포추출물이나 혈액 같은 혼합물들을 통과시켜 항원이 항체에 붙도록 한다. 붙은 항원에 대한 탐지는 형광물질이나 방사성동위원소를 이용한 세포추출물을 이용하거나 각 항원에 대한 2차 항체를 이용하여 수행한다. 이 방법의 최대 걸림돌은 항체의 특이성이다. 항체와 항원간의 특이성에 대한 조사 결과, 30% 정도만이 특이성을 띠는 것으로 나타나 대부분의 항체는 정량분석에 적합하지 않다. 그러나 특이성을 띠는 나머지 항체를 이용해 단백질 혼합물에서 항원의 정량적 분석은 가능하다. 또한 array 표면에 정착되는 물질은 항체에 국한되지 않고 펩타이드, aptamer, allergens, 탄수화물, 저분자 합성물질 등을 정착하여 세포나 조직에 생물학적 활성을 프로파일링 할 수 있다.

3.1.4. 구조와 이미지

한 단백질이나 도메인의 포괄적인 구조를 분석

하는 구조 단백질체학에서는 주로 X-ray 결정법이 쓰이고 있으며, 거의 모든 과정이 자동화되어 있다. 그러나 여전히 단백질 발현이나 결정화 과정이 어렵고 다른 분야처럼 대규모로 이루어질 수 있을지가 불분명하다. 핵자기공명법 (NMR)과 이차원 전자현미경 (2D EM)이 쓰이고 있지만 많은 시간과 노력을 요하거나 해상도가 낮다는 단점이 있다. 따라서 대규모로 고해상도의 결과를 얻을 수 있는 방법을 개발하기 위해, 급속 냉동된 한개의 비염색 세포에서 단백질의 3차원 구조를 볼 수 있는 전자 단층 촬영법이 주목받고 있다. 형광 단백질을 이용한 광학 이미징법은 살아 있는 세포에서 단백질을 분석할 수 있는 방법으로 FRET(형광 공명 에너지 전달)이 대표적이다. 이 방법은 단거리 내의 형광물질 사이에 일어나는 에너지 전달을 이용하는데 생리활성 상태의 세포에서, 그리고 단백질의 정확한 세포 내의 위치에서 정보를 얻을 수 있다는 장점이 있다.

3.2. 유전자와 단백질 발현

특히 질병 연구에 중요한 정보를 제공하는 모든 유전자의 절대적, 상대적 발현 수준을 분석하기 위해 microarray가 쓰이고 있지만 기술적 문제와 발현 조절 과정으로 인해 전사체와 단백질의 발현 수준이 상관성을 갖는지에 대해서는 논란이 많다. 현재 각광받고 있는 분야 중 하나인 혈청 단백질체는 훌륭한 마커지만 대량으로 존재하는 알부민과 면역글로불린으로 인해 아직 초기 단계에 머물고 있다. MALDI를 이용하는 또 다른 기술은 간편하고 임상에 적용할 수 있지만 역시 대량 단백질의 문제를 해결해야 한다. 대규모의 항체 생산 프로젝트를 통해 조직 arrays를 이용하고 있으며 암세포주의 단백질 발현 수준을 연구하기 위해 세포 용해물 array도 사용되고 있다.

3.3. 단백질 해독 후 변형(Post-translational modification: PTMs)

단백질 해독 과정이나 해독 후 일어나는 단백질 구조에서의 공유결합에 의한 변형은 단백질활성을 조절하는데 중요한 역할을 한다. 변형의 형태와 위치를 확인하면 생물학적 전달체계에서 특정 단백질의 기능과 조절에 대한 중요한 정보를 얻을 수 있다. 많은 단백질의 활성을 조절하는 해독 후 변형에는 주로 인산화, 아실화, 메틸화, 당질화, 편재화 및 지방 변형 등 신호전달과 세포조절에 관여하는 것으로 알려진 것만 200가지가 넘는다. 이러한 변형은 해당 아미노산의 분자량을 바꾸기 때문에 PTMs 분석에도 MS가 널리 쓰이고 있다. 한 번의 실험으로 여러 형태의 단백질 변형을 확인하기 위해 최근 shotgun MS 방식이 사용되는데, 이 방식은 다차원 LC/LC, 탠덤 MS/MS, 데이터베이스 탐색 프로그램을 연결하고 사용된다. MS는 해독 후 단백질 변형을 확인하는데 중요한 역할을 수행하며 단백질 마이크로어레이는 단백질 변형을 일으키는 효소를 동정하고 그런 효소들의 기질 특이성을 연구하는데 중요한 수단이 된다. 그러나 특정 변형을 포함하고 있는 펩타이드의 양이 매우 적을 수 있기 때문에 여러 가지 단백질 가수분해효소와 반복 질량 스펙트럼을 이용함에도 불구하고 PTMs 동정은 기술적으로 여전히 어렵다.

3.4. 단백질의 세포내 위치 확인 (Subcellular localization)

단백질의 세포 내 위치 데이터는 진핵생물 단백질 기능을 밝히는데 중요한 정보를 제공한다. 트랜스포존을 이용하는 Cre-lox 기술을 사용하여 epitope tag된 유전자를 염색체에 넣어 단백질 발현을 한 후 그 세포 내 위치별 발현 양상을 고효율적 immunostaining으로 확인 가능하다. 효모 유

전자들중 2000여개의 유전자에서 만들어지는 단백질들의 세포 내 위치가 이 방법을 사용하여 확인되었다. 이 방법의 단점은 아직 모든 유전자들이 이 시스템으로 전환되어 있지 않다는 것이며 효모의 경우 60%만 전환되어 있다. 또한 tag된 단백질의 발현은 유전자 자신의 프로모터에 의해 유도되기 때문에 위치정보는 발현이 많이 되는 단백질에 집중되어 있다.

3.5. 단백질 상호작용과 복합체 (Protein-protein interaction)

단백질들은 생체 내에서 일반적으로 다른 단백질들과 상호작용을 통해 기능을 한다. 같은 세포 기능을 담당하는 단백질들은 서로 상호작용한다는 것이 잘 알려져 있다. 이러한 결과를 근거로 새로운 단백질의 기능을 연구하는 한 방법은 이 단백질과 상호작용을 하는 단백질을 찾아내는 것이다. 대규모 연구를 통해 단백질간 상호작용을 지도화하면 단백질 기능에 대한 예측과 세포기능의 분자 메커니즘을 설명할 수 있는 기능전달체계를 모델링하는데 용이하다.

3.5.1. Two-hybrid 연구

효모 two-hybrid 방법은 단백질간의 상호작용을 지도화하는데 가장 체계화된 생체 내 연구 방법들 중 하나이다. 연구대상 단백질(bait)은 DNA 결합 도메인에 융합되어 있다. 전사활성 도메인에 융합되어 있는 다른 단백질(preys)들을 스크리닝하여 bait와 물리적 상호작용 단백질을 전사활성을 이용한 탐지분석법을 이용해 발굴한다. 이 방법은 대규모로 자동화가 가능하다. 비록 two-hybrid 방법이 몇 가지 단점이 있지만 다른 방법을 통해 얻은 단백질 상호작용 데이터와 통합하여 사용하면 매우 유용한 연구접근 방법이 된다.

3.5.2. Affinity tagging 및 질량분석기

Affinity를 이용한 정제 방법은 세포 추출액에서 단백질 상호작용을 밝히고 검증하는 소규모 단백질 연구에 매우 유용하게 사용되어 왔다. Tandem affinity purification(TAP) tag를 이용하거나 FLAG epitope tag를 이용하여 MS로 분석하는 방법이 발표되었다. 두 방법은 tag된 단백질의 발현 방법의 차이로 결과에 차이가 있지만 이 두가지 정보를 통합하여 사용함으로써 상호작용 정보의 정확도가 크게 증가 되었다.

3.5.3. Network 만들기

대규모의 상호작용 결과는 단백질 네트워크를 연구하는 생물정보학자들에게 매우 유용한 자료이다. 이 네트워크에서 단백질의 위치는 그 중요도와 상관성을 나타낸다. 단백질의 상호작용은 너무나 방대해서 두 가지 효모 질량 스펙트럼 결과가 겨우 40%만 중복될 정도이며 이를 극복하기 위해 MS 민감도와 서열결정 속도를 향상시키고 있다. Bayesian 네트워크와 같은 좀더 정교한 수학적인 방법들은 네 개의 효모 상호작용 결과, 유전체 문맥법, 공동발현 결과에 적용한 실험은 이 분석법의 가능성과 한계점을 동시에 보여준다.

3.6. 단백질의 생화학적 활성분석

단백질의 기능과 조절을 밝히는 가장 직접적인 방법 중 하나가 그 단백질의 생화학적 활성을 결정하는 것이다. 지난 몇 년간 효모 two-hybrid, affinity tagging, chip을 이용한 어레이 방식 등이 대규모의 생화학적 활성 분석에 사용되었다. 이 방법들은 특정 생물체에서 유전자를 확보하여 과발현된 단백질들의 생화학적 활성 스크리닝에 사용되는데, 유전자들을 효율적으로 클로닝하여 적당한 숙주에 넣어 과발현 시키는 방법의 발달로 대규모 생화학적 분석 연구가 시작되었다.

3.6.1. Pooling 전략을 통한 생화학적 활성 분석

이 방법은 단백질발현 유전자를 제조함으로써 클로닝한 후 각 유전자 발현 숙주를 pooling하여 단백질 정제를 하고 생화학적 활성을 스크리닝함으로써 신속하고 정확한 결과를 얻을 수 있다. 단점은 여러 단계가 필요하고 인산화효소와 탈인산화효소들 같이 흔히 나타나는 효소나 활성을 스크리닝할 때는 작은 pool을 만들어 사용해야 한다는 것이다.

3.6.2. 기능적 단백질 microarray

특정 생화학적 활성도에 대한 대규모 발굴에 더 직접적으로 사용되는 방법은 기능적 단백질 마이크로어레이 방식이다. 이 방법에서는 단백질 세트나 전체 프로테오미를 과발현시켜 정제하고 어레이 형태로 분포시켜 분석을 수행한다. 특정 활성도 스크리닝 뿐만 아니라 해독 후 변형 검사 그리고 저분자, 단백질, 항체, 약물과의 결합을 감지하는데 사용될 수 있다. 이 방식의 한계는 분석할 단백질을 준비하는 것이다. 기능적으로 활성을 가진 많은 수의 단백질을 확보하려면 고효율적이고 포괄적인 발현 라이브러리와 방법을 필요로 한다. 하지만 이러한 연구가 *in vitro* 분석이긴 하지만 실험 조건을 조절하여 연구를 수행할 수 있으며, 실험에 정제된 단백질을 사용하므로 단백질 양을 비슷하도록 조절할 수 있다는 장점이 있다. 더욱이 감지 방법을 보유하고 있다면 신호전달체계에 서 하부 타겟 단백질도 발굴할 수 있다. 단백질 대신 펩타이드가 사용되는 펩타이드어레이 방식은 효소활성을 위한 기질로 사용되거나 단백질이나 다른 분자로 탐색할 때 리간드로 사용될 수 있다.

3.7. Proteome informatics

Proteomics의 기술적 발전과 이의 적용 확대에

의해, proteomic 분석 결과의 양이 대폭 늘어나게 되었고, 또한 이렇게 늘어난 정보를 어떻게 처리하여 필요한 정보만을 추출해 내는 proteome informatics가 매우 큰 이슈로 등장하게 되었다. 또한, 다양한 실험적 접근 방법 간의 데이터 비교, 링크 및 샘플과의 연관성 파악 등을 수행하기 위하여 sample documentation, XML 정의, controlled vocabulary가 정의가 필요하게 되었다.

프로테오믹 데이터들은 정적인 genome 결과에 비해서 매우 dynamic하며, 샘플의 조건에 따라 변화의 양상이 크기 때문에 샘플의 정확한 documentation이 매우 중요하며 이의 database가 필요하다. 또한 XML (eXtensive Markup Language)로 구성되어, database간의 데이터 호환성을 크게 높일 필요가 있으며, 다양한 필드에 대한 정의를 반영한 새로운 XML format이 필요하다. 그리고, 이러한 모든 정보들은 모두 사전에 약정되어진 controlled vocabulary를 사용하여, 서로 다른 vocabulary를 사용할 때 나타날 수 있는 중복성 검토작업을 회피할 수 있어야 한다.

4. 기능단백질체학을 이용한 신약 표적 단백질

4.1. 단백질 상호작용 연구의 중요성

미국의 경우 전문기술기반의 바이오테크 기업을 중심으로 신약표적 단백질 탐색이 활발히 진행되어 방대한 정보축적과 요소기술이 발전하고 있으며, 또한 바이오테크 기업의 가치 증가에 따라 대형제약사와의 제휴도 급증하고 있는 추세이다. 단백질 상호작용 분석은 표적단백질 발굴을 위한 유전자 기능분석 방법 중 가장 효용성이 높은 방법이며, 그 중요성이 강조되는 이유는 생체내에서 ①단백질은 다른 단백질과의 결합을 통해 기능을 나타내고, ②유전자 및 단백질 발현의 변화, 세포

내 위치의 변화, 해독후 변형을 통한 구조의 변화는 단백질 결합의 변화를 야기시키고 ③이는 궁극적으로 세포내에서 일어나는 대사 및 신호전달과정의 활성화, 제어의 변화를 초래하고 ④또한 유전자 변이(mutation)에 의한 단백질 상호작용의 이상은 질병과 직결된다는 점이다. 그러므로 생물정보학(bioinformatics)적 기술을 이용하여 단백질들 간의 상호작용 데이터베이스 및 글로벌 상호작용 네트워크를 구축하면 단백질의 기능을 네트워크상의 위치파악을 통해 알아낼 수 있는 토대가 될 수 있다. 그리고 이를 기반으로 대사 및 신호전달 경로에 대한 총괄적 지도는 질병과 관련된 표적단백질 발굴 및 신약개발을 가속화시킬 것이다.

4.2. 단백질 상호작용 연구의 대상 및 접근 방법

총괄적 개념으로 생체내 모든 생체분자의 상호작용 총합을 interactome이라 칭하고 이를 연구하는 방법과 기술을 포괄적으로 interactomics라 한다. 현재 상호작용 연구가 단백질에 편향되어 있기 때문에 일반적으로 interaction은 단백질-단백질 상호작용의 총합으로 인식되고 있다. 단백질 상호작용 연구는 상호작용관계에 있는 단백질을 확인하고, 단백질 복합체에서 각 단백질의 화학양론적 특징을 분석하고, 구조 분석방법 등을 통해 접촉면의 특성을 파악하며, 또한 열역학적 속도론적 매개변수를 분석하며, 이를 바탕으로 포괄적 상호작용 네트워크를 구축하는 연구로 나눌 수 있다.

4.3. 단백질 상호작용 검색을 위한 주요 연구기술

단백질 상호작용의 확인은 X-ray crystallography 등을 이용하여 원자수준에서 상호작용에 관련된

아미노산 잔기 또는 원자를 규명할 수 있고, yeast two-hybrid 시스템을 이용하여 두 단백질간의 이성분 상호작용을 확인할 수 있고, 면역침전 및 질량분석을 통한 복합체 상호작용을 파악할 수 있으며, 이를 기반으로 활성측정방법을 통해 세포수준에서 상호작용 효과를 이해할 수 있다. 현재 단백질 상호작용 네트워크 작성을 위해 고속다중검색을 통한 대규모 단백질 상호작용 분석 등 다양한 방법들이 시도되고 있다(표 3).

<표 3> 단백질 상호작용 규명에 사용되는 연구기술

연구기술	A	B	C	D
X-ray crystallography				
NMR				
Yeast two-hybrid system				
Phage display				
Ribosomal display				
Protein chip				
Microfluidics				
BIAcore				
Immunoprecipitation				
GST-pulldown				
Mass spectrometry				
FRET analysis				
Chemical cross-linking				
GFP proximity imaging				
Co-sedimentation				
Native gel electrophoresis				
Far western(ligand overlay)				
Immunofluorescence				
SELDI-TOF				

(A: atomic interaction, B: binary interaction, C: complex interaction, D: cellular interaction)

4.3.1. Co-purification

대부분의 단백질은 생체내에서 다른 단백질과 결합을 통해 기능을 수행하기 때문에 특정 단백질을 정제하면 결합하는 다른 단백질도 같이 정제가 된다. 그러므로 단백질 복합체 분석은 단백질 정제와 확인 과정으로 나누어진다. 우선 단백질 복합체를 적절한 방법으로 정제한 후 복합체 구성단백질을 1D나 2D, LC를 이용해 분리한다. 그 이후 matrix-assisted laser desorption/ionization time-of-flight mass spectrometry(MALDI-TOF-MS) 등을 이용한 펩타이드 질량분석, nanoelectrospray ionization(nanoES)-MS/MS, nanoLC-MS/MS 및 LC/LC-MS/MS 등의 tandem MS(MS/MS) 방법을 이용해 단백질을 확인할 수 있다(표 2참조).

최근 질량분석 기술의 급진적으로 극소량(100fmol 이하수준)의 단백질만으로도 단백질 확인이 되므로 단백질 확인보다는 단백질 정제과정이 현실적인 문제로 대두된다. 단백질 정제에 있어서는 간편하고 정확하게 단백질을 정제할 수 있는 친화 크로마토그래피 방법이 주로 사용된다. 이 방법은 두 단백질 사이에 다른 단백질을 매개로 하여 결합을 하더라도 같이 정제되므로 복합체 상호작용을 이해하기에 적합한 방법이다. 크게 내재 단백질 복합체(endogenous protein complex)를 직접 항체로 정제하거나 항원결정기 표식(epitope tagging)을 한 후 항체로 정제하는 면역침전법(immunoprecipitation), GST-침전법(GST-pulldown)등의 비면역적 친화 정제법(affinity purification), 비특이적 반응 억제와 회수를 및 재현성을 높이기 위해 2개 이상의 표식을 이용한 Tandem Affinity Purification(TAP) 시스템 등의 다표식 친화정제법(multi-tag affinity purification) 등으로 나눌 수 있다.

4.3.2. Yeast two-hybrid system

DNA 결합도메인(DBD; DNA Binding Domain)

과 전사활성도메인(TAD; Transcription Activation Domain)을 유전자 재조합 방법으로 분리하고 상호작용하는 두 단백질의 유전자를 각각 DBD와 TAD에 클로닝한 후, 이들을 효모내에서 발현되도록 유도하면 DBD와 TAD가 물리적으로 분리된 단백질이 발현되지만 단백질 상호작용에 의해 전사인자로 재구성되어 유전자 발현이 가능해진다. 유전자조작기술을 기반기술로 하여 단백질의 특성에 관계없이 동일한 실험방법으로 상호작용을 검색할 수 있기 때문에 대규모 상호작용 분석에 용이하다. 그러나 지시유전자의 전사활성 측정은 단백질 상호작용의 간접적 확인이므로 위양성율이 높은 단점이 있다.

4.3.3. Protein chip

생체분자(biomolecule)간의 상호작용을 측정할 수 있는 Biacore의 센서칩이 상용화된 후 DNA칩의 등장과 함께 단백질칩개발이 가속화되었다. 단백질칩 기술은 질병진단, 단백질활성연구, 신약후보물질 검색, 생체분자 상호작용 검색 등의 분야에 응용될 수 있기 때문에 단백질체학의 핵심기술로 부각되고 있다. 슬라이드글라스에 단백질을 고정화시킨 후 광학현상(Surface Plasmon Resonance; SPR)분석, 질량분석, 형광분석, 전기화학적 분석 등을 통해 단백질 상호작용을 분석할 수 있다. 특히, 이온강도, pH 등의 다양한 조건에서 실시간으로 단백질간의 결합친화력 측정이 가능하고, 결합에 관여하는 단백질상의 위치 등 특이적 결합 여부를 확인할 수 있으므로 여러 매개변수 변화에 따른 상호작용의 변화 등 상호작용 기전을 이해하는데 필수적 기술로 인식된다.

4.3.4. Phage-display library

재조합항체를 이용한 면역치료제 개발의 핵심기술로 항체라이브러리를 이용하여 항원-특이적 항체를 선별하는 방법이 가장 많이 이용되고 있다.

최근에 펩타이드 라이브러리를 이용하여 특정 단백질에 결합하는 공통서열(consensus sequence)을 찾아낸 후 생물정보학기술로 후보단백질을 선별하여 상호작용 지도를 작성하고 이를 yeast two-hybrid 시스템으로 검증하는 방법이 사용되고 있다.

4.3.5. 생물정보학적 방법을 이용한 단백질 상호작용의 예측

게놈 서열과 단백질 구조 정보의 축적은 단백질 상호작용을 예측할 수 있는 생물정보학적 접근방법의 개발을 촉진하였다. 현재 게놈 서열 정보를 기반으로 계통분류학적 접근이나 다중서열정렬을 통해 접근하는 방법 등이 이용되고 있다. 또한 단백질 구조정보를 기반으로 상호작용을 추론하거나 단백질 상호작용 기전을 분자수준에서 예측하려는 연구도 진행되고 있다.

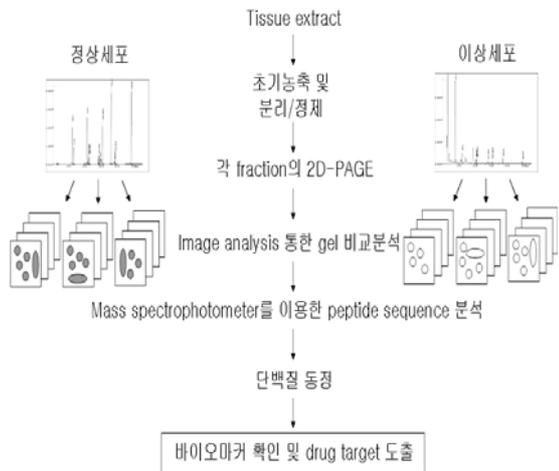


그림 4. 질병특이성 바이오마커 개발을 위한 proteomics 응용에

4.4. 시장현황

Proteomics를 기반으로 한 시장은 2001년 이래 연간 성장률 30%를 기록하면서 그 시장의 크기

또한 7.7배나 커지고 있다. 세계시장 규모는 2001년 565 million 달러수준이었던 것이 연간 40%의 성장을 기록하며 2006년에는 3.3 billion 달러에 이를 것으로 보여진다. 시장의 구성은 기반기술이 절반을 차지하고, 나머지 protein chip, 신약개발, 정보기술, 용제 등이 각각 별 차이없이 시장을 점유하고 있던 초기시장 형태와는 달리 해를 거듭하면서 신약개발 및 protein chip에 대한 관심과 투자가 증대되어 2006년에는 protein chip의 경우 시장점유율이 20%를 웃돌 것으로 예상되어 진다(표 4).

<표 4> Proteomics 시장의 년도별 분야별 점유율

	% In 2002	% In 2004	% In 2006
Drug discovery	10.9	10.6	12.1
Informatics tools	10.5	8.1	7.5
Platforms	55.2	54.4	50.7
Protein chips	14.3	18.1	21.6
Reagents	4.0	3.1	2.7
Services	5.1	5.7	5.3

4.5. 시사점

신규 유전자의 기능 분석은 신약표적단백질 발굴을 위해 필수적이며 단백질 상호작용 분석은 유전자 기능분석을 위한 유용한 방법이다. 뿐만 아니라 단백질-단백질 상호작용 자체가 신약의 표적으로 엄청난 잠재력을 가지고 있다. 단백질 상호작용 연구는 단백질 기능분석을 통한 신약표적단백질을 발굴할 수 있고, 단백질 결합 및 상호작용 그 자체가 바로 신약의 표적이 될 수 있으며, 다양한 바이러스, 박테리아, 곰팡이 및 기타 감염세균에 의해 발생하는 질환의 치료제 개발에 유용한 정보를 제공할 수 있다. 그러므로 새로운 대규모

단백질 상호작용 분석 및 검증 기술 개발, 복잡한 데이터베이스 구축 및 생물정보학적 기술혁신 등에 대한 연구가 지속적으로 필요하다.

5. 결론

지난 수년간, proteomics의 기술 발전의 양상은 단위 시간당 더 많은 proteom 결과를 얻어 낼 수 있는 high-throughput system으로 요약되어 질 수 있다. 또한, 단순한 단백질의 cataloguing 작업이 아닌, protein-protein interaction, post translational modification, 특정 조건에서의 proteom의 변화양상 등 genome level에서 관찰하지 못하는 proteom의 특징적인 부분을 확인하는 방향으로 모든 기술적 지향성의 초점이 맞추어져 진행되고 있다.

단일 proteomics 수행방법의 기술적 응용도를 높여가고, 또한 개별적인 proteomics 수행방법들간의 결합 및 단백질에 대한 분리 차원 (dimension)을 획기적으로 늘리는 방법 등을 통하여 전체적인 proteom에 대한 profile을 분석하고 있다. 예를 들어, 다양한 Mass spectrometer의 결합 (tandem mass, Quadrupole time-of-flight, Fourier transform ion cyclotron resonance mass spectrometer (FT-MS)) 및 이들의 응용은 보다 다양한 proteom의 초고속 분석에 널리 사용되어질 예정이다.

다만 이러한 새로운 기술이 적용된 기기의 단가가 매우 높으며, 이의 운영을 위해서는 잘 교육되어진 분석전문가가 필요하다. 일반적으로 proteomics에서 중요한 연구결과를 내고자 하는 대학 실험실 혹은 벤처의 경우에는 이러한 고가의 장비와 전문인력을 지원할 수 있는 연구비의 규모가 될 수 없다. 대신, 이러한 일반 대학 실험실 및 벤처, 더 나아가 proteomics를 이용하고자 하는 중견산업계(제약회사, 화장품회사, 식품회사 등)의 고가 첨단 proteomics 기기의 이용 및 분석을 도와줄 수 있

는 국가적 차원의 'proteomics 연구센터'의 구축 및 운영이 필요하다.

프로테오믹스는 향후 기능유전체학의 주요한 연구부분으로 계속 성장할 것이며, 이의 직접적 혹은 간접적 영향을 통하여 우리나라 전체 BT 연구부분의 성장동력으로 충분한 역할을 할 것이다.

6. 참고문헌

1. Proteomics, Annu. Rev. Genomics Hum. Genet. 2004, 5: 267-293
2. Proteomics, Annu. Rev. Biochem. 2003, 72: 783-812
3. 세포생물학과 Proteomics의 응용, Kor. J. Microbiol, 2001, 37(2): 109-113
4. Peptidomics, J. Chromatography B., 2004, 803: 3-16
5. 단백질체학을 이용한 신약표적단백질 발굴동향, 보건산업기술동향, 2003
6. Genomics, proteomics and nutrition, Nutri. Sci. 2002, 5(3): 129-133
7. 프로테오믹스의 연구방향과 국내프로테오믹스 연구의 활성화, 한국분자생물학회뉴스지, 2000, 12: 20-29
8. Application of plasma-polymerized films for isoelectric focusing of proteins in a capillary electrophoresis chip, Analysis, 2003, 123: 237-244

