



# 생물체의 Proteome 연구와 생물 정보학



이 지원  
고려대학교 화공생명공학과

## 1. Proteomics

Proteomics는 유전체(genome)에 의해 발현되는 총체적인 단백질 집합체인 단백질체(protoeme)를 대상으로 하는 연구분야로서, 다양한 환경변화에 대한 단백질체의 유동적인 변화를 정밀하게 분석하고 분석결과로부터 유용한 정보를 습득하는데 목적이 있다.

세포 내에서 정적으로 존재하는 genome에 대한 연구(genomics)는 생물체가 수행하는 기능의 이론적 예측만을 제공하는 반면, proteomics는 생물체의 주변환경에 따라, 고등 생물체의 경우에는 각 세포/조직 별로 변화하는 단백질들의 유동적(dynamic) 변화에 대한 분석이 가능하므로 생물체의 실제 기능을 이해하는데 직접적인 단서를 제공할 수 있다.

이러한 이유로 급속한 속도로 축적되고 있는 유전체 정보를 기반으로 미지의 유전자들의 기능을 신속히 좀더 정확하게 밝혀내기 위해

functional genomics와 상호 보완적인 관계의 proteomics가 부각되고 있고, 세포 내에서 일어나는 생리 현상들을 전체 단백질 단계에서 통합적으로 파악하는 수단으로 활용되고 있다.

단백질체 연구(proteomics)는 two-dimensional gel electrophoresis, mass spectrometry, yeast two-hybrid screening, computational prediction program 등의 발달을 계기로 대량 고속 검색(high-throughput screening; HTS) 기능을 확보함으로써 연구 가속화의 토대가 마련되고 있다 (1). 이러한 고속 처리 분석 시스템을 이용하여 한번에 수많은 단백질의 발현 정도 및 단백질 합성 후 변형(post-translational modification) 등을 종합적으로 자동분석해냄으로써 단백질의 구조와 기능을 이해함과 동시에 불치병의 발생 원인, 상태 및 생리 변화 등에 대한 핵심정보를 대량으로 얻을 수 있다.

일반적으로 proteomics 분야는 expression

proteomics와 cell-map proteomics로 분류될 수 있다. 주로 신약 개발에 응용되고 있는 expression proteomics는 다양한 환경에서 대상 세포나 조직의 단백질체 발현 형태나 발현 정도의 양적인 변화를 정밀 분석한다.

독성학(toxicology)에서는 disease marker를 찾아내어 특정 질병의 발병 기작과 질병의 종류에 따른 증상의 상호 연관성을 단백질의 단계에서 밝히고 있으며, 약학(pharmacology)에서는 여러 조건에서 만들어진 drug candidates의 작용이 전체 단백질체의 발현 변화와 발병기작 억제에 미치는 기능을 확인 함으로써 최적의 drug candidate를 결정하는데 이용된다.

한편, cell-map proteomics는 다양한 세포들의 각 기관에 존재하는 단백질들의 기능 및 단백질 상호간의 기능적 연관성을 파악함으로써 세포의 유전체 정보를 기반으로 세포의 종합적인 기능을 밝히는 연구분야라 할 수 있다.

이러한 proteome 연구에 의한 세포나 단백질의 주요기능에 대한 정보는 세포를 배양하여 단백질 생산하는 공정엔지니어링이나 목표지향적으로 세포의 대사기능이나 단백질의 생물학적 기능을 변화시키고자 하는 대사공학 및 분자생물공학 분야의 연구를 위해 매우 유용한 정보를 제공해 줄 수 있다.

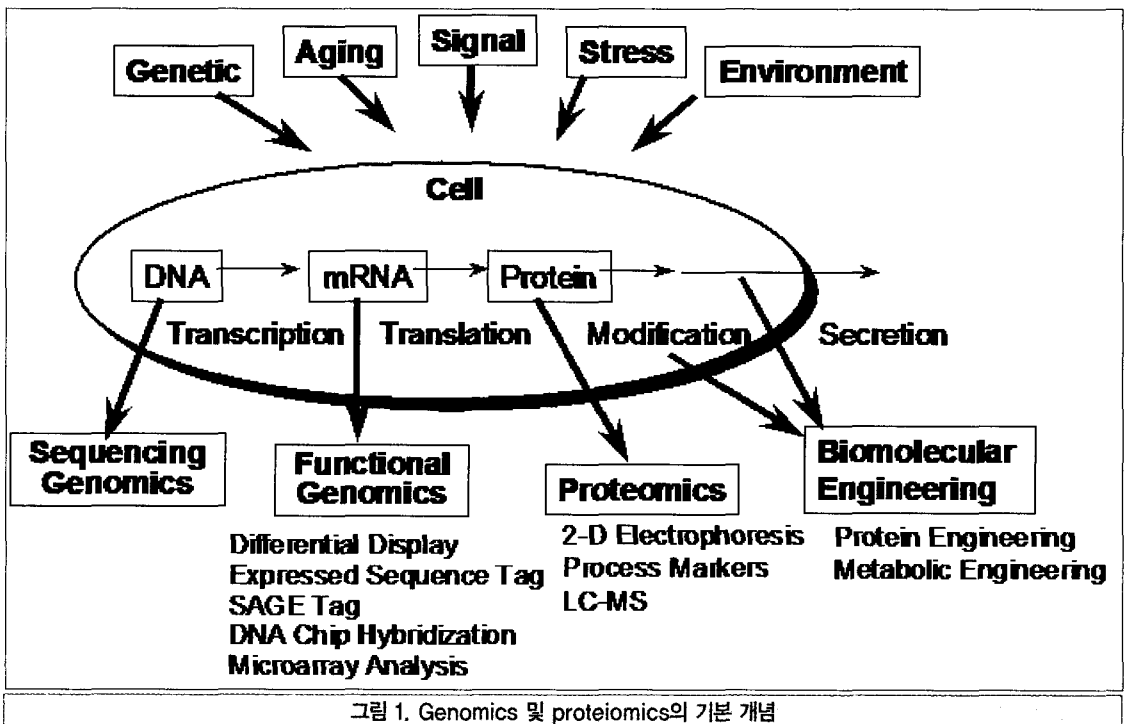


그림 1. Genomics 및 proteomics의 기본 개념



## 2. Bioinformatics: Proteomics를 위한 정보기반

90년대 후반에 들어 생물체에 대한 게놈분석이 가속화되면서 2001년 8월 현재 원핵생물(prokaryote)인 *Haemophilus influenzae*, *Mycoplasma genitalium*, *Methanococcus jannaschii*, *Escherichia coli*, *Bacillus subtilis*, *Helicobacter pylori* 과 진핵생물(eukaryote)인 *Saccharomyces cerevisiae*를 비롯하여 57종의 생물체 게놈 분석이 완료되었고 기타 253종의 생물체에 대해 분석이 진행 중이라고 한다. 2001년 2월 인간 유전자를 구성하는 30억쌍의 염기서열 중 90%의 초안이 완성되었으며, 분석 결과 인간의 염색체 위에는 약 75,000-125,000개 정도의 유전자가 존재함을 알게 되었다.

유전자 및 단백질 서열정보(sequence information) 관련 분자생물학 D/B들은 최근 급속히 성장하여 적어도 100 개 이상의 독립된 D/B가 존재하는 것으로 알려져 있다 (2).

Sequence information이 급격히 증가함에 따라 체계적인 D/B 구축, D/B의 효율적인 공공 활용을 위한 software 개발, D/B 정보를 활용한 유전자 및 단백질의 기능분석 등 bioinformatics(생물정보학)는 이미 정립된 학문분야로 그 위치를 확고히 있으며, 새로운 D/B들과 컴퓨터 기술들을 접목시킴으로써 서열이 가진 생물학 정보를 밝히는데 큰 역할을 하고 있다. 일례로서, 현재까지 효모, 초파리, 선충, 쥐 등을 모델 시스템으로 이용하여 이미 구축된 유전 정보들을 인간에게 적용하기 위한 연구가 진행 중에 있다. 각 모델

시스템에 공통으로 존재하는 특정 유전자 기능을 서로 다른 종(species) 간에 비교하는 방법은 정상과 비정상 유전자의 기능을 연구하는데 가장 좋은 수단으로 고려되고 있으며, 이미 모델 시스템의 결합 유전자를 이용하여 인간의 질병을 유발하는 몇 개의 주요 유전자들이 확인, 분리된 바 있다 (2).

또한, 이미 많은 연구가 진행되어 있는 초파리의 유전자 돌연변이 정보인 DRES(*Drosophila expressed sequence*)를 이용하여 그들과 높은 유사성을 가진 120개 인간 유전자들을 확인하고 쥐와 인간의 유전체 정보를 활용하여 각 염색체 지도상에서 각 유전자의 위치를 파악하고 유전자 발현 특성이 조사되기도 하였다 (3). 이러한 방법은 포유류에서 유전자들의 기능을 밝히는 효과적인 접근 방법이 될 것으로 평가되고 있다.

또한, amino acid sequence로부터 단백질의 3차 구조를 예측하는 functional proteomics는 생물학 정보 D/B와 컴퓨터 기술이 접목된 bioinformatics의 중요한 분야 중의 하나이다. 단백질의 3차 구조는 열역학적 안정성에 의해 결정되므로, 단백질의 3차 구조를 결정하는데 필요한 모든 정보들은 서열에 포함되어 있다고 생각할 수 있으며 in vitro 실험 조건에서 몇몇 선택적 단백질에 대해 이 법칙은 잘 적용되었다. 하지만 최근의 연구에 의해 생체 내에서의 단백질 구조 형성(in-vivo protein folding)은 훨씬 복잡하고, chaperone과 같은 다른 molecule들과 연관되어 동적인 구조를 가진 것으로 알려지고 있다. 또한 단백질 주위 환경도 단순한 열역학적 환경이 아니고, 단백질 구조 형성은 단백질을 구성하는 각 아미노산 분자들이 가지는 여러 형태의

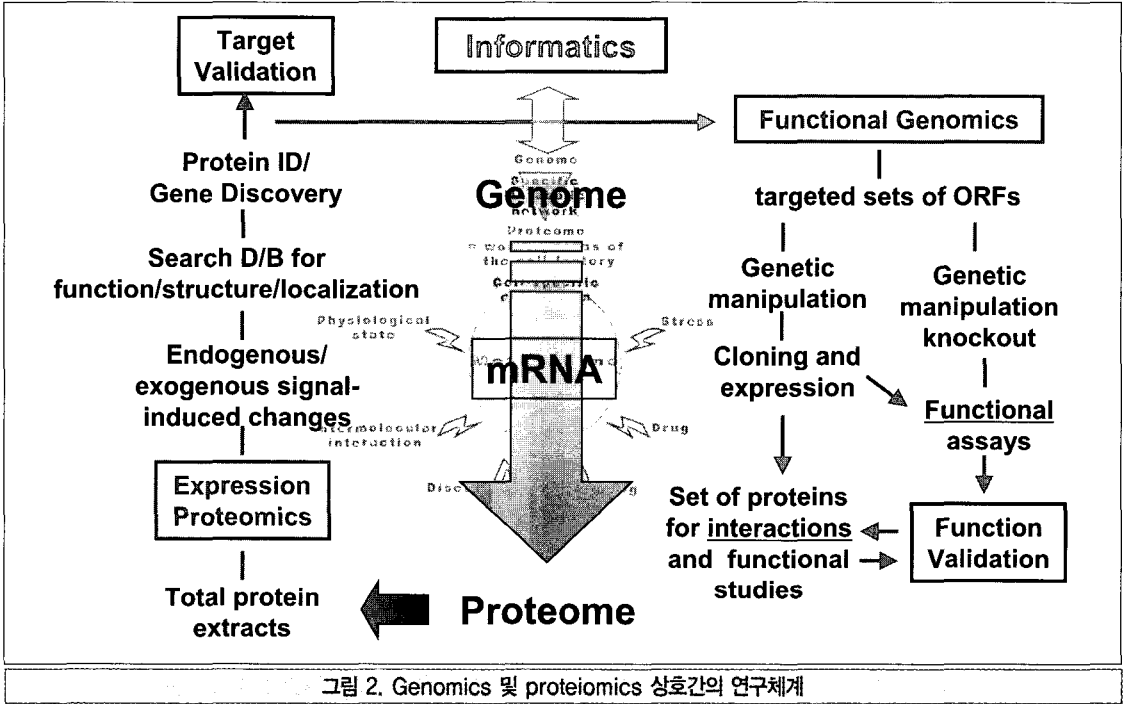


그림 2. Genomics 및 proteomics 상호간의 연구체계

interaction의 종합으로 인식되고 있다 (4,5).

따라서 간과해서는 안될 것은 실제 실험과정을 거쳐 protein-protein interaction에 관한 실험 결과가 보완되어 분자들 사이의 이러한 종합적인 interaction이 충분히 반영되지 않는 한 자연 상태에서의 protein folding을 예측하는 문제는 당분간 쉽게 해결되지 않을 것이다. 효율적인 bioinformatics tools을 이용한 체계적인 정보의 축적은 실험 결과를 바탕으로 할 때 새로운 형태의 정보를 창출해 낼 수 있으며(6), 이 같이 컴퓨터를 이용한 생물학 (silicon-based biology)과 실험 생물학 (carbon-based biology)이 효과적으로 결합할 때 생물정보학의 발전도 훨씬 가속화될 것으로 예상된다.

### 3. 최근 동향 및 향후 전망

많은 질병의 원인이 유전자와는 무관하게 단백질 folding 결함, 단백질 상호간의 interaction, malfunctioning 등의 경우이므로 genomics는 물론이고 proteomics와 bioinformatics 기술을 병용한 연구의 필요성이 부각되고 있다. 현재, 다국적 생명공학 기업들은 미정부와 계승 프로젝트 완성 경쟁을 벌이고 있는 민간기업 셀레라 제노믹스사에 거액을 투자, 지금까지 완료된 genome D/B를 사들이고 있다. 또한 세계 주요 제약회사와 화학산업계는 human genome project와는 비교도 할 수 없는 방대한 규모의 human proteome project를 준비하고 있으며 주로 분



자, 세포생물학적 방법에 의하여 질병의 원인 유전자나 기능성 유전자들을 발굴하여 진단 및 치료제를 개발하는 전략을 채택하고 있다.

한편, 유전자 발현 해석과 소프트웨어 분야는 미국 Affymetrix, Molecular Dynamics 등의 벤처기업을 주축으로 DNA chip 관련기술, 게놈 정보관리, 기능해석을 위한 소프트웨어 개발 등에 투자가 급속히 증대되고 있다. 이제 곧 맞이할 시대는 proteomics, functional genomics, bioinformatics 등의 다차원적인 접근을 통하여, 단순한 정보 축적 및 가공의 차원을 넘어서 생체 분자들의 기능에 대한 체계적이며 완벽한 분석 catalog를 제공할 수 있게 되어 단순한 sequence information으로부터 모든 유전자의 기능 및 단백질 분자구조, 반응기작 까지 예측할 수 있을 것이다.

게놈 프로젝트 연구에 있어 후발 국가인 한국 및 일본 등은 선진국으로부터 뒤쳐진 게놈 연구를 만회하기 위한 post-genome project, 즉 proteomics 기반기술 구축에 국가적인 차원의 집중 투자를 하고 있다. 미래의 신약개발의 핵심 기술은 필수 표적 단백질을 얼마나 확보하느냐가 중요한 관건이다.

다수 공개되는 유전자 서열과는 달리 신약 개발에 필요한 핵심 기능성 단백질은 미공개 내지는 특허권으로 보호될 가능성이 높은 만큼, 각국은 그 나라 국민에 특이적인 질병기작에 관련된 기능성 단백질을 밝히기 위해 독자적인 proteome 연구를 활발히 진행하고 있다. 이와 같은 흐름 속에서 생물공학 내지는 생물공정 엔지니어에게 proteomics는 매우 유용한 활용 수단이 될 수 있다.

목적산물의 생산성에 직결되는 미생물의 생리, 대사현상의 미세 변화를 종합적으로 감지하면서 좀더 완벽한 대사공학적 접근을 시도해 볼 수 있으며, in-vivo 또는 in-vitro protein-protein interaction을 정밀 분석하여 재조합 발현시스템 개발 또는 정밀 분리기술 개발에 활용할 수도 있을 것이다. 또는 미생물의 biofilm 형성 관련 기작을 규명하여 새로운 치료물질 개발이나 고생산성의 bioconsortia 시스템 개발에 기여할 수도 있는 등 proteomics를 활용한 잠재적인 발전 가능성은 무궁무진하다 할 것이다.

참고 문헌

1. Larsson M, S Graslund, L Yuan, E Brundell, M Uhlen, C Hoog and S Stahl. High-throughput protein expression of cDNA products as a tool in functional genomics. *J Biotechnol.* 80 (2000), pp. 143-57.
2. Sohaila Rastan and Lee J Beeley, Functional genomics: going forwards from the databases. *Curr Opin Genet Dev* 7 (1997) pp. 777-83.
3. Banfi S, G Borsani, E Rossi, L Bernard, A Guffanti, F Rubboli, A Marchitello, S Giglio, E Coluccia, M Zollo et al., Identification and mapping of human cDNAs homologous to Drosophila mutant genes through EST database searching [see comments] of special interest. *Nat Genet* 13 (1996), pp. 167-174.
4. Norin M and M Sundstrom. Structural proteomics: developments in structure-to-function predictions. *Trends Biotechnol.* 20 (2002) pp. 79-84.
5. Yanagida M. Functional proteomics: current achievements. *J Chromatogr B Analyt Technol Biomed Life Sci.* 771 (2002), pp. 89-106.
6. Kanehisa M. Grand challenges in bioinformatics. *Bioinformatics.* 14 (1998), pp. 309.