

Proteome 연구의 전체상

Toshiaki Isobe, Nobuhiro Takahashi

■ 머리말

게놈 정보의 흐름에 따라 진행되는 기능성 genome연구의 개요를 표 1과 그림 1에 정리하였다. 「Proteome」이란 “the entire protein complement expressed by a genome, or by a cell or tissue type” (하나의 genome 또는 특정 세포·조직·기관에서 생산되는 단백질 전체)로 정의된다. Genome은 하나의 생물에 1분자 또는 1 set밖에 존재하지 않는다. 이에 비해 proteome은 생물의 수정에서 발생, 성숙, 죽음에 이르기까지 어렵게 변동하며, 생리 상태나 주위의 미세한 환경요인에 의해서도 시간적, 공간적으로 매우 다양한 양상을 보인다. 즉 proteome은 genome의 지배 하에 있으면서 genome으로 규정되는 유전자의 수보다 훨씬 더 많다. 단적으로 말하자면 생물이 생기는 순간마다 다양한 변화를 나타내며 다른 양상으로 존재한다(그림 2). 생명활동의 어느 순간에 존재하는 단백질 set을 proteome이라고 규정하고

그 전체상을 파악하여 다음 순간에서의 전체상과 비교하여 그 변동을 해석한다. 이 경우, proteome의 전체상이란 그 속에 존재하는 모든 단백질의 종류와 존재량, 그리고 번역 후 수식을 포함하는 구조정보를 포함한다. 세포내에서 모든 단백질은 다른 단백질 혹은 생체분자와 기능적인 관련을 가지며 존재한다. 따라서 관찰된 proteome의 변동은 그 순간의 생명활동의 변화를 연출하는 기능 단백질군의 변동을 반영한 것이라고 생각할 수 있다. 이런 입장에서 proteome의 동태를 지표로 genome의 발현정보를 파악하고 생명현상을 해석하는 방법론이 proteome 연구의 기본전략이다. 생명활동의 어느 순간에 기능하는 genome 정보가 proteome이라는 형태로 구체화된다고 생각하여 그 동태를 포괄적으로 해석함으로써 생명현상을 이해하려는 시도가 기능 genome 연구 중에서 「Proteomics」라 불리는 생명과학의 새로운 영역이다(표 1).

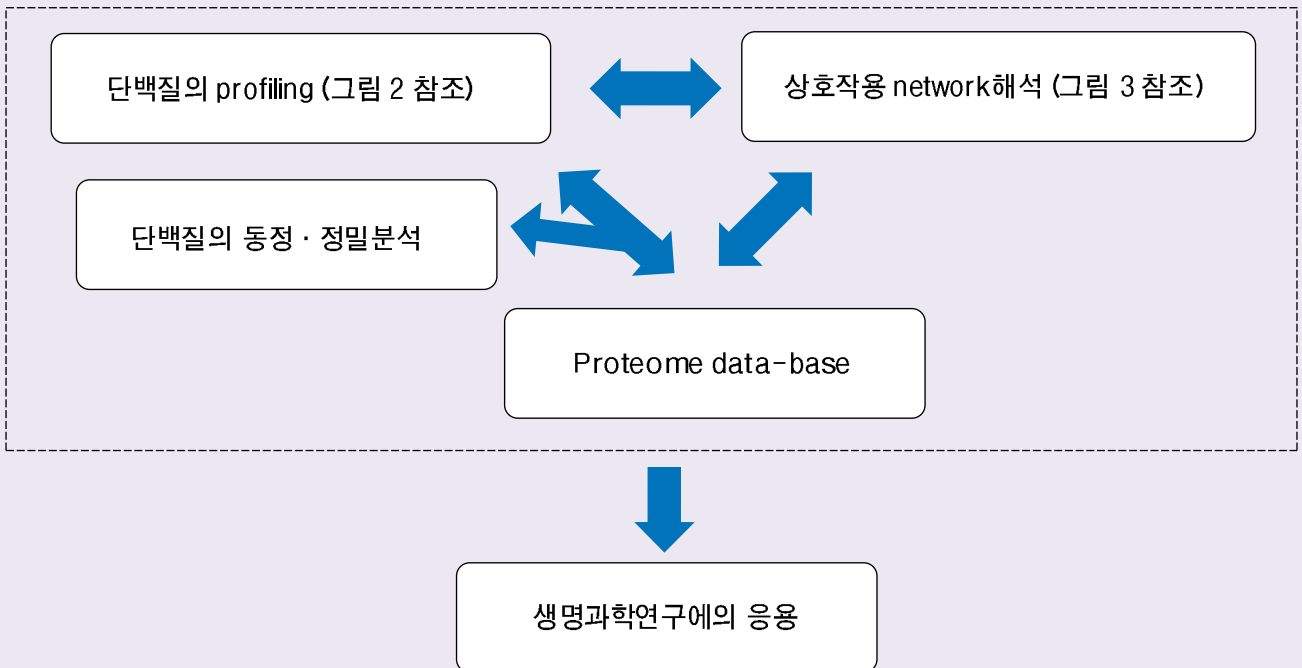


그림 1 Proteome 해석 흐름

표 1 기능 genome 해석 level

해석 level	주요 분석방법	과학영역	정보량-----증대요인
Genome DNA(유전체)	DNA 서열결정 등	Genomics	유전정보(불변)
Transcriptome mRNAs (세포·조직·기관의 mRNA 전체)	Microarray SAGE* 등	Transcriptomics	선택적 splicing 등
Proteome Proteins (세포·조직기관의 단백질 전체)	이차원전기영동- 질량분석 등 (본 protocol)	Proteomics (구조 genomics)	고차구조 형성 번역 후 수식 단백질간 상호작용 등
Metabolome Metabolites(세포·조직 ·기관의 대사중간체 전체)	NMR분석 질량분석	Metabolomics	분자간 상호작용 Energy 대사 물질대사 Homeostasis 등

* Serial Analysis of Gene Expression

■ 단백질 Profiling

세포나 조직, 기관의 proteome을 profiling하기 위해 가장 자주 사용되는 것이 이차원 전기영동법에 의한 단백질 분리와 분리된 단백질 spot의 질량분석에 의한 동정이다. 이차원 전기영동법에 의한 단백질 분리는 '70년대에 O' Farrell이 고안한 방법의 개량법이 이용된 것으로 기술적으로 혁신적인 진보가 있었다고 말할 수는 없다. 그러나 여기에서 얻은 spot 하나 하나의 동정이나 화학구조의 결정이 가능하게 된 것으로 단순한 단백질 분석법에서 의미 있는 포괄적인 image 정보로서의 가치를 지니기 시작했다. 그 배경에는 '80년대의 질량분석법의 비약적인 진보와 '90년대부터 시작된 genome 해석으로 인한 방대한 단백질 아미노산 서열 정보 data base의 정비가 있다. 질량분석 분야에서는 종래에는 불가능했던 단백질이나 펩타이드 등의 고분자물질을 기상으로 ion화하는 Electrospray Ion화법(Electrospray Ionization: ESI법)이나 Matrix 지원 Laser 이탈 Ion화법(Matrix Assisted Laser Desorption Ionization: MALDI법)이 고안됨과 함께 기상의 고분자 ion을 고정도, 고감도로 분리 검출하는 비행시간형 질량분석계(Time-of-Flight Mass Spectrometry: TOF MS)등의 고성능 질량분석계가 개발되었다. 이러한 기술혁신과 방대한 genome 정보가 종래의 이차원 전기영동법에 새로운 가치를 부여하여 현재의 proteome 연구의 맥을 이어간다고 해도 좋을 것이다.

1) 펩타이드 Massfinger print 법

이차원 전기영동과 MALDI-TOF MS를 조합한 profiling 기술에서는 목적 단백질 혼합물을 요소나 계면활성제의 존재 하에서 이차원 전기영동으로 분리하여 단백질을 염색한 후에 그 image를 scanner로 화상 처리하여 정규화한다. 이 목적에는 영동조건의 작은 변동에 따른 image의 미묘한 어긋남을 보정하고 분리한 하나하나의 spot을 정량함과 함께 다른 시료의 image를 상호 비교하기 위한 화상해석 software가 개발되고 있다. spot의 동정에는 gel에서 잘라낸 spot을 protease로 처리하여 얻은 펩타이드 조합을 생성하는 단백질을 data base에서 검색한다. 이 방법은 펩타이드 mass finger print법이라고 불리며 인터넷상에서 검색 가능한 data base(peptide mass data base)가 구미를 중심으로 정비되고 있다. 이 방법은 간편하고 효율적이며 숙련되면 1주일에 수백 spot의 동정이 가능하다. 또한 genome 해석이 완료된 생물을 대상으로 한 연구에서는 원리적으로 모든 단백질을 이 방법으로 동정할 수 있으며 genome 정보의 추적과 함께 앞으로 보다 유용한 기술로 응용될 것으로 예상된다. 현재 대부분 수작업으로 이뤄지고 있는 동정 조작의 자동화와 함께 genome 정보에서 단백질 코드영역을 정확하게 예측하여 아미노산 서열로 번역하는 software의 개량이 요구된다.

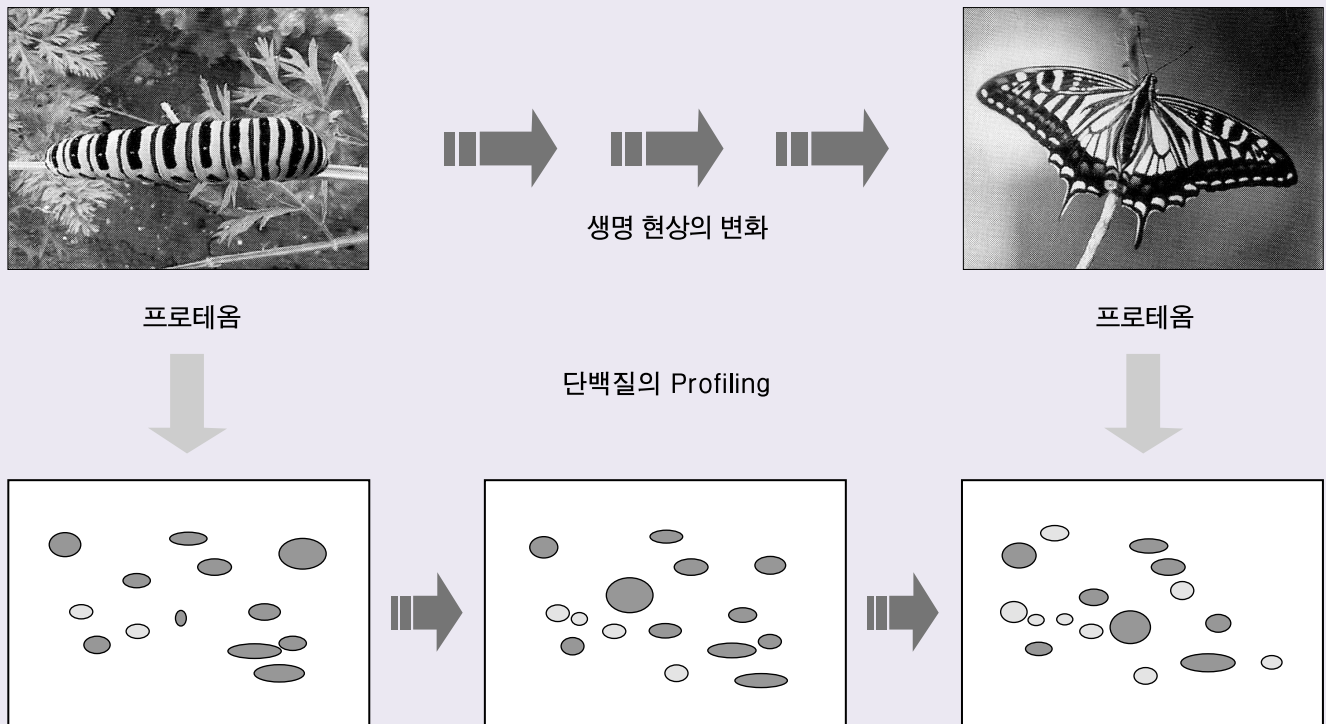


그림 2 Proteome 연구의 개념도(1)
 Proteome 연구는 생명활동에 수반되는 genome의 발현상황의 변화를 그 순간에 존재하는 단백질의 전체상(proteome)의 변화로 profiling 하는 것을 기초로 하고 있다.

2) Profiling을 위한신기술

기존의 이차원 전기영동법은 자동화가 곤란하고 번잡한 조작이 수작업으로 이루어져 재현성을 확보하기 어렵다는 문제점이 있었다. 이는 단백질의 동태 해석을 목표로 하는 proteome 연구에 있어서는 중요하다. 이 문제를 회피하기 위하여 복수의 시료혼합물을 다른 형광시약으로 표식한 후 혼합하여 한 장의 gel에 영동하여 정밀하게 비교하는 방법도 시도되고 있다. 또한 이와는 별도로 Liquid Chromatography(LC) 분리법을 이용하는 profiling 기술이 개발되고 있다. 분리양식이 다른 복수의 column을 switching valve로 연결하여 valve를 제어하고 분리를 다차원화하여 고성능화하는 것이 이차원 전기영동의 LC화원리이다. 이 시스템을 ESI interface를 장비한 질량분석계와 연결하면 on-line에서 자동해석하는 profiling 시스템 구축도 가능하다. 가용화에 계면활성제를 필요로 하는 막단백질 등의 적용에는 어려움이 예상되지만 시료 용량이 크고 분리의 재현성이나 정량성이 높은 자동화 시스템으로 앞으로 이차원 전기영동과 나란히 고성능 profiling 기술이 될 것으로 기대된다. 또한 불용성 단백질의 문제와 관련하여 세포추출액 등을 직접 protease로 처리한 펩타이드 혼합물을 LC-MS로 해석하는 것으로 추출액속의 단백질을 동정하는 것도 일부에서 시도되고 있다. 예를 들어 3,000종류의 단백질이 각각 50 종류의 펩타이드를 생성한다고 가정하면 모든 단백질의 동정에는 최대 15만개의 펩타이드 분리가 필요하다. 기존 분리과학의 개념과는 다르지만 예를 들어 동위체 label법을 병용하면 proteome의 동태해석도 가능하며, 적용범위를 막 단백질 등으로 한정한다면 흥미있는 도전이 될 것이라 생각된다. 또한 기술적으로 미흡하여 거론하지 않았지만 protein

chip 기술은 개념적으로 고려해 둘 필요가 있다. 이 기술은 개개의 단백질에 대한 항체를 DNA chip과 같이 micro chip위에 array하여 chip상에서 결합 단백질을 검출 정량하는 시도이다. 이 기술을 실현하기에는 교차반응이 없는 특이항체를 개개의 단백질에 대하여 제작하여 선정하는 무단한 노력과 계통적인 작업이 필요하며, 기능적으로 중요한 인산화 등의 번역 후 수식을 식별하기 위해서는 더욱 방대한 항체 library가 필요하다. 기반을 만들기 위한 국제적인 협력체제의 정비가 요구된다.

■ 단백질 정밀분석

Proteomics를 위한 단백질 동정법을 거론하였다. 앞으로 모든 생물이 지닌 genome의 전체염기서열이 해명되고 거기에 포함된 모든 유전자의 구조가 파악되어 정확하게 단백질의 아미노산 서열로 번역된 단계에서는 이미 「아미노산 서열 미지의 단백질」은 존재하지 않는다. 지금도 genome 해석이 완료된 생물에 대해서는 peptide mass finger print법이 그 proteome을 구성하는 단백질을 동정하는 목적에 유용하다. 그러나 human을 시작으로 mouse 등의 실험동물의 genome 정보는 불완전하며 genome 정보를 단백질 구조 정보로 변환하기 위한 지식도 충분히 축적되어 있지 않다. 따라서 실제 연구 현장에서는 여전히 단백질의 아미노산 서열 결정이 필요하다. 이렇게 얻어진 부분적인 아미노산 서열은 peptide mass finger print법과 병용하여 finger print법으로 동정의 확인과 이종 생물 간의 단백질 동정 등을 가능하게 하는 「아미노산 배열 Tag」로 이용된다. 또한 필요한 경우에는 아미노산 서열을 이용한 단백질 data base검색에 이용된다. 물론 기존 cDNA cloning을 위한

probe 설계도 가능하다. 또한 genome 과학 중에서 proteome 연구가 지니는 중요한 역할로서 세포내에 존재하는 단백질의 실상 즉, 번역 후에 받는 인산화나 당쇄의 부가, protease에 의한 processing 등의 수식의 실패를 알 수 있다. 이런 수식은 생합성 후의 단백질이 세포내에서의 위치나 기능발현에 있어 매우 중요하며 기능 genome 연구 중에서도 proteome 연구에서만 얻을 수 있는 정보이다.

■ 단백질의 linkage 분석

기능 genome 연구에 있어 proteome 해석의 또 하나의 중요한 측면은 proteome 해석에 의해 얻어진 단백질 변동을 어떻게 하여 생명현상과 관련 짓느냐 하는 점이다. 단백질 기능발현은 모든 단백질이 다른 단백질이나 생체분자와 상호작용하는 것에 의해 이루어진다. 이 같은 이유로 proteome 해석에 있어서 단백질 상호작용 network를 밝히는 중요성이 있으며 genome 기능해석에 있어 proteome 연구가 유망하다고 생각된다. 저자는 생물의 일생이 한편의 이야기라고 한다면 genome(유전자)은 모든 단어(단백질)를 망라한 사전, proteome은 어떤 이야기의 문장의 한 구절이라고 표현하고 있다¹⁾. 이 비유에 따르면 현시점에서의 proteome 해석은 한 구절의 문장을 해독하기 위한 초기 단계이며, 하나의 proteome에는 어떤 단어가 사용되어지고 있는가를 망라적으로 동정하는 작업이 단백질 profiling이라고 생각할 수 있다. 다음 단계는 단문의 이해작업이지만 그 전에 단어를 조합하여 문장으로

만드는 규칙 즉, 문법을 알 필요가 있다. 단문은 proteome에 포함되는 단백질의 상호연결고리이며 그 가운데 일정한 문법이 포함되어 있다. 단백질 상호작용 network의 해석법은 proteome연구의 중핵을 이루는 가장 중요한 기술 중 하나이다. 즉 세포와 조직에서 관찰된 proteome은 단백질 상호작용의 해석에 의하여 상호의 관련(linkage)이 이해되어 기능 단백질 set로서 파악할 수 있다(그림 3). 현재 단백질간의 상호작용 해석을 목적으로 효모균을 이용한 two-hybrid analysis가 계통적 또한 대규모로 이루어지고 있다²⁾. 또한 이러한 단백질 간 상호 작용 network는 생체분자 전체를 포괄하는 더욱 고차원의 「세포정보 network」가운데에서 복잡한 생명현상을 규정하여 세포나 조직간의 조화를 만들어내는 것으로 생명활동을 전체로 조절하고 있다고 추정되지만 그 해석은 현시점에서는 Bioinformatics에 의한 구조 genome학의 분야에서의 접근에 의해서만 시도되고 있다³⁾.

■ Proteome data base

위와같은 연구에서 얻을 수 있는 proteome에 관한 정보는 data base로서 인터넷상에 공개되어 열람할 수 있게 되었다. 현재의 data base는 이차원 전기영동을 기반으로 다양한 미생물과 동식물의 세포, 조직, 기관마다 작성되어, genome과 유전자, 단백질의 data base와 link되어 생명과학정보의 방대한 network를 형성해 나가고 있다. 이러한 정보의 공유와 통합은 genome 과학을 중심으로 하는 생명과학연구의 발전에 대단히 중요하다.

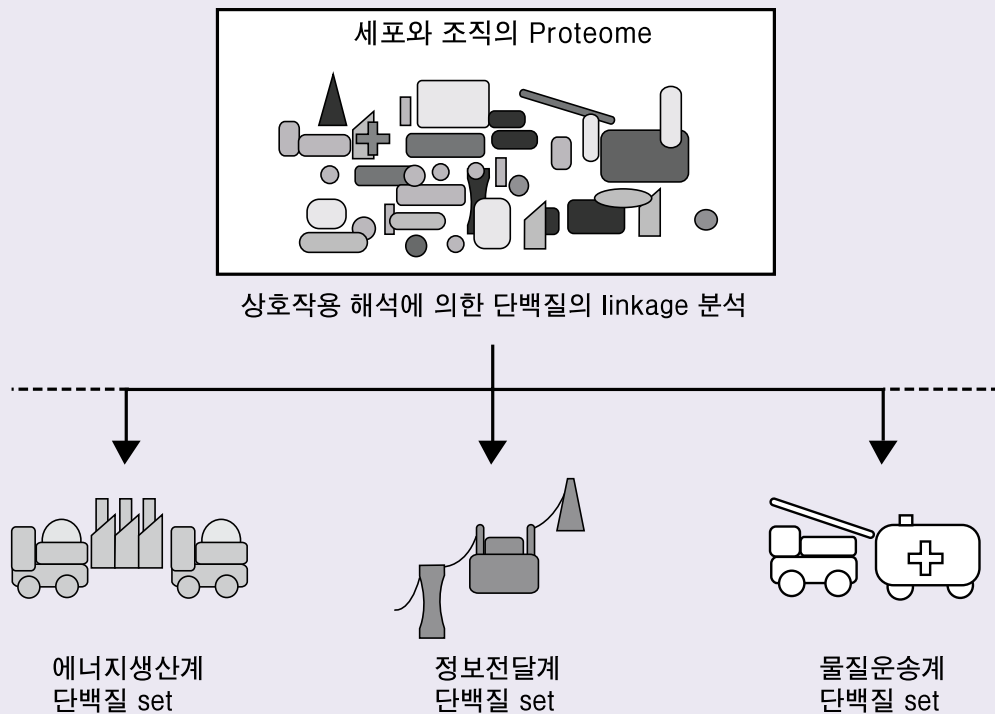


그림 3 Proteome 연구의 개념도(2)
Proteome 연구는 어떤 순간의 proteome에 포함된 단백질 상호 연결(linkage)을 밝히는 것으로 생명현상을 포괄적으로 이해하는 것을 목표로 하고 있다.

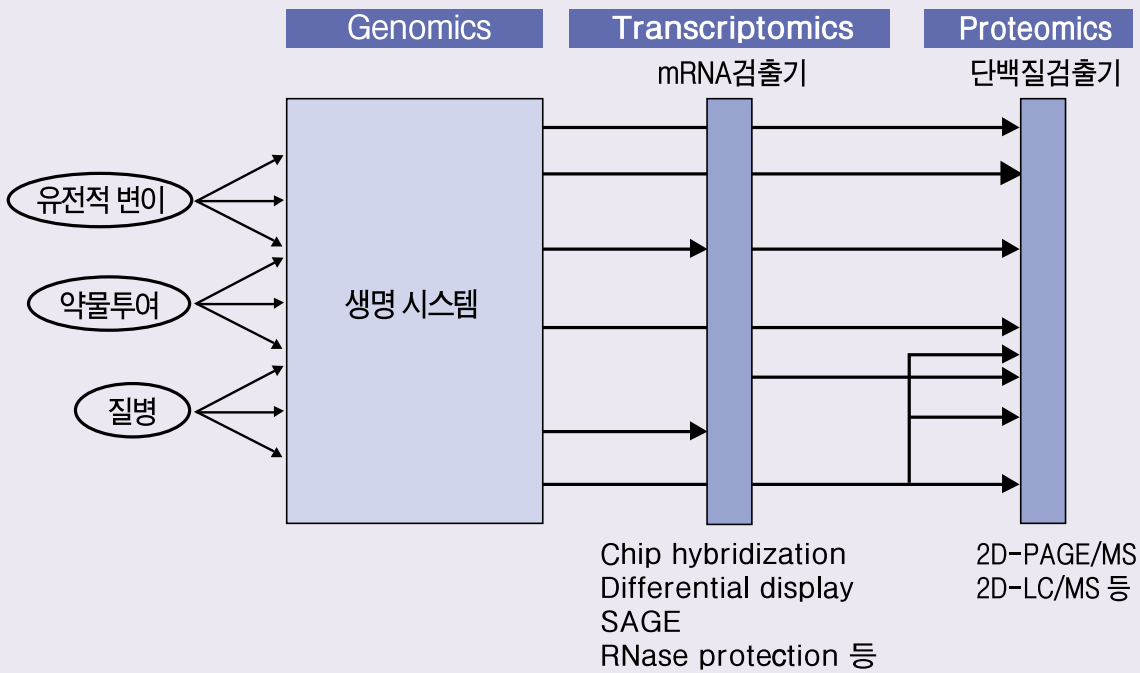


그림 4 생명 시스템의 검출기로서의 기능 genome 해석
 기능 genome 해석 방법은 외부의 스트레스에 의한 생체의 출력 신호(genome 정보의 변화)를 수신하는 검출기로 이용된다. Proteomics 방법은 단백질의 번역 후 수식의 변화 등도 검출할 수 있는 가장 정보량이 많은 검출기이다. (Anderson, N.L. : *Electrophoresis*, 19: 1853-1861, 1998을 일부변경)

■ Proteome 연구와 Life Science

Proteome 연구의 개념과 방법론은 생명과학연구에 새로운 시점을 가져오는 것이다. 이런 관점에서 genome 염기서열해석이 완료된 고초균, 란조, 선충과 이와 유사한 해석이 급속도로 진행되고 있는 식물(벼, Arabidopsis)의 proteome 해석의 실례를 소개하고, 또한 같은 목적에서 앞으로 proteomics가 새로운 발전을 가져올 것으로 기대되는 생명과학의 기초와 의학·약학 영역으로의 응용에 대하여 일본의 선구적인 연구를 소개할 예정이다. 한 예로, 구미 기업을 포함한 연구기관에서는 proteomics의 방법이 대장암 등의 종양의 진행이나 약물투여의 영향을 monitor하기 위한 임상진단기술로서도 응용되고 있다. 이 경우 proteome은 가장 많은 정보를 갖는 생명현상의 예민한 검출기라고 생각할 수 있다(그림 4). proteome 연구가 지니는 새로운 가능성에 대해서 생각해 주길 바란다.

마지막으로 proteome 연구를 시작으로 하는 functional genome 연구를 21세기 중심적인 생명과학으로 발전시키기 위해서는 genome에서 metabolome에 이르기까지 광대한 실험과학의 연구성과는 생명의 이야기를 이해하기 위한 정보를 이끌어내기 위한 컴퓨터 과학 즉 bioinformatics와의 융합에 의해 집약될 수 밖에 없다(표 1). 생물을 시스템으로서 이해하기 위해서는 방대한 정보의 통합이 필수적이라는 것은 명백하며 생명과학이 정보

과학으로 변모하고 있는 것은 실험과학을 주체로 하는 우리들도 받아들이지 않으면 안 되는 상황이 되고 있다. Bioinformatics에 의한 생물 시스템 model의 이론적 구축과 실험과학에 의한 검증, model의 수정과 실험적 검증의 순환작업은 앞으로 생명과학의 방향이며 proteomics의 입장에서도 이 분야의 육성을 향한 발빠른 대응이 요구된다.

【참고문헌】

- 1) 安樂泰宏 : *Bioscience와 Industry*, **55** : 837-838, 1997
- 2) Oliver, S. : *Nature*, **403** : 601-603, 2000
- 3) Sali, A. : *Nature*, **402** : 23-26, 1999

다카라 “유전자 해석센터”

- 유전자 염기서열 해석 서비스
- Library 제작 서비스
- Rapid cDNA Screening 서비스
- Genome Project
- 유전자 해석 연구 수탁 서비스

