

를 통해서 얻어낼 수 있는 긍정적인 효과이다. 단백질의 특징에 따라 선별법은 다양해야 하기 때문에 자신의 목적 단백질에 적합한 선별 기술을 개발하여야 한다. 이러한 분자 진화 기술은 Post

genome 시대에 보다 많은 효소들의 발견과 더불어, 유용한 단백질을 생산하는 기본 기술로 받아들여질 것이다.

프로테오믹스 연구의 최신동향과 활용

주 현 · 서진호*

(주)엠에이치투바이오테크놀로지, *서울대 농생명공학부 및 협동과정 생물화학공학 전공

hyun@cheme.caltech.edu, jhseo94@snu.ac.kr

프로테오믹스(Proteomics)란 유전자에 암호화되어 발현되는 단백질 전체를 세포기능별로 확인하고자 하는 학문이다. 프로테오믹스(Proteome)이란 단백질(Protein)과 전체(-ome)의 합성어로서 하나의 생물 종에서 만들어지는 모든 단백질에 대한 분석을 의미한다. 이는 게놈(Genome)이 전체 유전자 집합을 의미하는 것과 동일하게 해석된다. 그러나 유전체와는 달리 단백질은 시간과 생체 환경에 따라 변화한다. 즉 단백질의 발현양상은 질병특이적 혹은 조직특이적인 차이를 반영하기에 생체활성을 제어하는 직접적인 원인체가 되는 것이다. 프로테오믹스 연구는 단순한 유전정보만을 지닌 정적인 유전자 수준을 넘어 매우ダイナミック하게 변화하는 직접적인 생명제어의 도구인 단백질을 세포전체 수준에서 분석하고자 하는 학문 분야이다.

프로테오믹스는 포스트게놈시대의 기반기술로서 의료기술의 진보, 농업 및 바이오, 에너지 분야에 혁신적인 활용이 예상된다. 프로테오믹스 연구는 마크 윌킨스에 의하여 처음으로 제안된 후 최근 10년 기간 동안 눈부신 발전을 이루어왔다. 그러나 실제로 단백질 단위를 분석하는 2차원 전기영동 장치는 30여년 전에 도입되었다. 시간적 공

백에도 불구하고 최근 프로테오믹스가 학문적 혹은 주요 바이오산업의 연구기법으로 각광을 받게 된 커다란 원인은 단백질의 직접적인 서열분석과 동정을 신속하게 수행할 수 있는 질량분석 기법의 도입에 기인된다. 다음에 보여주는 [그림 1]은 세포 내 발현 단백질을 하나의 2차원 전기영동 겔에 등전위점(pI, isoelectric point)과 분자량에 따

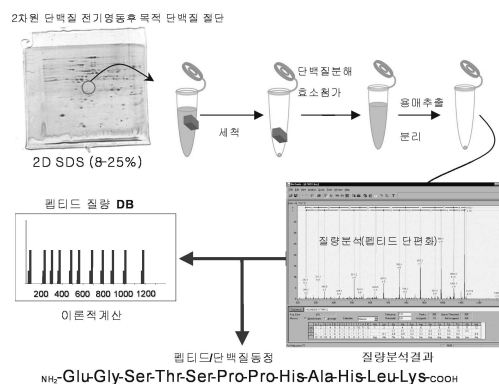


그림 1. 프로테오믹스 기법을 이용한 미지의 단백질 동정과정. 2차원 전기영동을 통하여 단백질을 그 고유 특성인 pI 값과 분자량 별로 분리 후, 세척과 작은 펩티드로의 절단과정을 거쳐 질량분석기에서 구성 아미노산 서열정보를 획득한다. 에드만 분해법과 비교하여 약 1,000배 이상의 감도와 100배 이상의 서열 결정 속도를 보여준다.

라 분리 후 대상 단백질을 질량분석 기법을 이용하여 분석하는 과정이다. 이러한 일련의 과정을 “작은 의미의 프로테오믹스”라 한다. 단백질 자체를 동정하는 이 기법의 출현으로 인하여 생체 내 원하는 단백질을 기존의 유전자 서열분석법과 필적할 만한 속도로 그 아미노산 서열 획득이 가능하다.

생체가 지닌 정보 해석은 두가지 경로를 통하여 가능하다. 하나는 직접적인 유전정보를 암호화하는 유전자의 염기서열을 측정하는 것이다. 이 경우 염기하나에 약 1달러 정도의 비용을 투입하면 최대 600~1,000bp의 DNA정보를 불과 서너 시간만에 확인할 수 있다. 이와 같은 기술적 혁신으로 인하여 인류는 현재 무려 30여 종 이상의 생물체에 대한 유전체 서열 정보를 획득하고 있다. 그러나 단백질의 경우 서열결정에 많은 시간이 걸리며 요구되는 작업 또한 노동 집약적이었다. 단백질이란 20개의 아미노산이 결합되어 만들어진 생체 고분자 사슬이다. 4개의 유전자 코드를 가지는 DNA와 비교하여 20개의 다양한 아미노산으로 구성된 단백질은 유전자보다 적은 길이를 가지는 사슬 안에 보다 함축적이고 다양한 정보를 기록하고 있다. 단백질이 생체내에서 기능을 발휘할 수 있는 이유도 바로 이 다양성에서 비롯된다.

그렇다면 생체에서 동 시간대에 존재하는 단백질의 수는 얼마나 될까? 일반적으로 인체에서는 약 34,000개의 단백질이 발현되고 있다고 알려져 있으며, 선택적 스플라이싱(alternative splicing)과 번역 후 변형 과정 등을 거쳐 구조적으로 다양화된 형태는 무려 50만 종 이상이 존재한다고 한다. 이들 복잡한 형태의 단백질들을 연구하기 위해서는 실험실 수준에서의 확인작업뿐만 아니라 최근 각광받기 시작한 단백질 정보학의 도움을 얻는 방법이 필수적이다.

21세기 들어 완성된 인간게놈연구를 시발점으로 바이오 관련 연구가 급증하고 있는 상황이다.

이는 실로 어마어마한 학문적 산업적 가치를 내재하고 있다. 그러나 전통적인 실험적 접근법과는 달리 생체 안에서 시시각각으로 변화하는 이들 단백질들의 거동을 빠른 시간안에 분석하고 주요 타겟물질을 발굴하기 위해서는 프로테오믹스와 연관된 단백질 정보학의 도입이 필수적이다. 지금까지는 대부분 핵분열 시뮬레이션과 기류의 흐름 등을 연구하는 유체역학 등의 빠른 연산에 이용되던 슈퍼컴퓨터들이 프로테오믹스 산업과 연관된 바이오 연구에 널리 이용되기 시작하였다. 작년 한해 미국내 바이오인포매틱스 한 분야에서만 150억\$의 매출이 이루어진 사실만 보아도 결코 무시할 수 없는 산업분야의 하나로 자리매김하고 있다. 매출신장률 또한 30% 이상을 상회한다.

프로테오믹스 연구에서 바이오인포매틱스와의 융합이 왜 중요한지에 대하여 살펴볼 필요가 있다. 단백질들이 생체내에서 작용하는 모든 기작을 파악하고 제어타겟과 연결된 생체실험을 수행하기 위해서는 천문학적인 연구비가 소요된다. 대다수 제약사들의 고민은 매년 투입되는 신약 R&D 비용이 날로 심각하게 증가되고 있다는 사실이다. 이를 해결하기 위한 방안중의 하나로 인체 혹은 생물체 안에서 작동하는 단백질들의 움직임을 거시적이고 동시에 미시적인 수준에서 빠른 시간 안에 예측할 수 있는 시스템 수준의 연구가 프로테오믹스 연구에 있어 앞으로 이루어져야 할 부분이다. 이는 곧 생체네트워크를 구성하기 위한 가장 기본적인 작업이 된다. 미국 IBM사가 생명공학 부문에 막대한 자금을 투입하며 진행중인 “블루진 프로젝트”는 향후 1~2년안에 초당 1,000조회의 부동소수점 연산이 가능한 슈퍼컴을 개발하는 것이다. 블루진에는 중앙처리장치인 CPU가 100만 개 정도 사용되지만 크기는 냉장고 2대만하다. 담당자인 코백의 말을 빌리자면, 앞으로 이들 슈퍼컴퓨터를 가장 원하는 주 구매자는 물리학자도 아

닌 생명공학 연구자들이라 한다. 이러한 연구분위
기 조성은 바로 10년도 안되는 역사를 가지는 프
로테오믹스의 등장으로 인한 것으로 보아도 과언
이 아니다.

왜 단백질 연구는 과거의 생명공학 연구가 진행
하여오던 기술 패러다임을 벗어나 이토록 방대한
양의 정보처리가 요구되는 거대학문분야로 발전
될 수 밖에 없었는가? 유전자는 그 길이에 무관하
게 A(아데닌), T(티민), G(구아닌), C(시토신)
라는 4가지의 유전적 코드를 이용한다. 반면 단백
질의 경우 20가지 아미노산으로 구성된 고분자이
다. 일부 RNA촉매 기능을 제외하고는 N개의 서
열을 지니는 유전자가 지닌 이들 4개의 물질은 4^N
개의 구조를 지닐 수 밖에 없어 생체 내에서 비트
에 해당되는 정보저장 역할을 제외하고는 그 이용
이 다분히 제한적일 수 밖에 없다. 반면 단백질의
경우 20^N 에 해당하는 기하학적 도형체를 만들 수
있다. 예로, 암 세포 전이에 필수적인 외부세포막
(ECM)을 녹이는 효소 MMP를 생각해 보자. 이
단백질은 세포와 세포 사이의 장벽물질인 콜라겐
을 선택적으로 녹일 수 있는 기능을 수행한다. 콜
라겐을 구성하는 특이 펩티드 서열만을 공격하는
것이다. 인간이 합성한 촉매 중에서 다른 단백질
들의 펩티드 결합 서열을 파괴하지 않고 콜라겐만
선택적으로 녹일 수 있는 촉매는 존재하지 않는다.
이는 100억분의 1의 확률 중에 단 하나만을 선택
적으로 골라낼 수 있는 기하학적인 다양성이 확보
되어 있기 때문이다. 인기있는 Lotto가 이와 같은
확률을 지니고 있다면 범 국민적인 열풍은 벌써
사라졌을 것이다. 실제 생체를 구성하는 단위체와,
기능을 유지하는 부분을 차지하는 대다수는 단백
질이다. 이는 앞서 살펴본 바와 같이 그 분자구조
적 형상의 다양성에 기인된다 할 수 있다. 즉, 서
로간의 간섭과 혼돈을 최소화하는 방향으로 생물
체가 진화된 것이다.

프로테오믹스 연구의 주요 영역으로는 다음을
들 수 있다.

새로운 프로테옴 분석도구의 개발

프로테옴 분석의 가장 큰 장애물은 단백질 검출
의 민감도이다. 물론 현재까지 개발된 도구들이
상당부분 세포내 전체 반응 및 질병과의 연결고리
를 어느정도 확인 가능하게 하지만, 낮은 용해도
를 지니는 막단백질이나 저 발현율을 지니는 단백
질은 아직도 그 검출에 많은 어려움을 수반된다.
이를 극복하기 위한 기술들로는 예를 들면 LC-
QIT-TOF 질량분석기법의 도입이나 대기압조건
에서 작동이 가능한 APMALDI- α TOFMS를
들 수 있다. 이들 분석도구의 도입은 현미경 수준

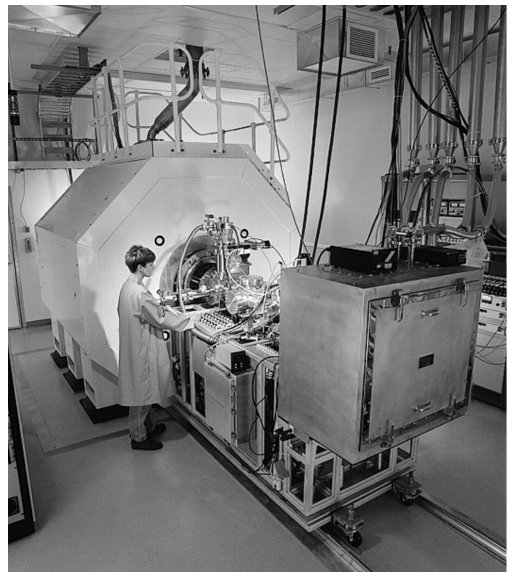


그림 2 115 테슬라급의 푸리에 변환 이온사이클로트론 공명 질량분석기(FTICR). 미국 태평양 북서 국립연구
소소가 소유한 고감도 질량분석기로 프로테오믹스와
생화학연구에 이용되고 있다. 자료사진은 Pacific
Northwest National Laborator의 허락하에 수록되었
다. 자속 밀도 1 테슬라(T)는 1쿨롱(C)의 전기량을 갖
는 전하가 1m/s의 속도로 수직으로 작용할 때, 1 뉴
턴(N)의 힘이 미치는 세기를 의미한다.

의 분해능을 벗어나 실제적인 생체 내부 네트워크를 볼 수 있게 한다. [그림 2]는 11.5 테스라의 장을 지나는 이온 싸이클로트론 형태의 질량분석기로 소수점 이하 6자리까지 단백질 질량분석이 가능하다.

대량의 프로테옴 분석을 위한 스토리지 기법과 유전체 단위로의 확장검색이 가능케 하는 고성능 온-스톱 단백질 발굴시스템의 개발

의료분야로의 확대를 예를 들면 임상단계에서 신속한 분석과 진단이 가능케 하는 하이브리드 타입의 large-scale 프로테오믹스 기법의 개발과 시스템 수준의 진단이 가능한 세포 및 기능모사 시뮬레이터의 개발을 들 수 있다. 여기서 하이브리드 타입이란, 정성적인 프로테옴 프로파일링 분석이 주된 분석기법에 정량적 변화의 의미를 부여함으로써 세포분자 수준에서 정확한 변화추적이 가능케 하는 점이다. 기존의 프로테오믹스란 세포내 거동 분자의 유, 무 혹은 절대치 이상의 발현을 보이는 단백질들만이 추적이 가능하였다. 이는 원인체 발굴을 위한 분자디스커버리 영역에서 많은 장점을 제공하여 왔던 점이 사실이나, 당뇨 등의 대사성 질환과 같이 미묘한 생체 내 변화가 수반되는 경우에는 그 해석에 많은 어려움을 지녀왔다. 이에 대한 기술적 대안으로는 단백질 칩 등의 개발이 제안된다. 항체나 특이 서열 인식분자의 도입을 통하여 빠른 시간내에 주어진 배열속에 존재하는 단백질 분자들을 분석이 가능케 하고 이에 대한 정량적 기능정보 등을 획득이 가능케 하는 기술분야이다.

세포분자 디스커버리를 위한 프로테오믹스

앞서 간단히 언급된 바와 같이 프로테오믹스 연구가 지닌 가장 강력한 장점중의 하나는 생물학적 기능에 따른 발현 단백질군의 프로파일링을 거시

적으로 분석할 수 있다는 점이다. 이는 현재 다른 방법들을 통하여 분석이 어려웠던 세포 전체의 거동을 실시간으로 사진기에 담아두는 기술이다. 즉, 필요한 데이터를 시간대 별로 'freeze' 시킨 후, 필요에 따라 그 사건 파일을 다시 열어볼 수 있는 아주 매력적인 기술이다. 대다수의 생명공학 회사들이 신약 탐색 위주의 성장 전략을 추구하고 있는 현 실정을 감안할 때 신약 발굴을 위한 본 기술의 활용도는 매우 높은 편이다. 생명공학 개발의 최전선을 달리는 미국 캘리포니아주의 바이오텍 회사들의 예를 살펴보기로 하자. 이들 회사에서 가장 많이 연구되는 분자디스커버리 분야를 살펴보면, 우선 항암 치료제 발굴을 들 수 있다. 2004년 미국 NCBI에 등록된 단백질 정보 중 암과 관련이 있는 단백질 수는 무려 15만종 이상이 보고되어 있다. 물론 이중에는 동소체(isoform)와 다른 동물 종의 단백질도 포함하고 있지만, 이 단백질 모두가 궁극적으로 연구대상이 될 수 밖에 없다. 그 누가 이들 단백질 중 어느 하나도 암 발생과 무관하다고 단언할 수 있는 연구자는 한 명도 없다. 두 번째로는, 감염성, 바이러스성 질환을 들 수 있으며, 세 번째로는 중추신경계(CNS)와 연관된 뇌, 신경질환에 대한 연구이다. 네 번째로 사망률 1, 2위를 달리고 있는 심혈관질환치료제 연구를 꼽을 수 있다.

세포분자 디스커버리를 위한 프로테오믹스는 실로 방대한 양의 데이터 처리가 요구되는 영역이다. 이 분야의 연구는 우리가 잘 알고 있는 5대-OMICS를 모두 연결하는 고리를 요구하기 때문이다. 현재까지 구축된 게노믹스 영역의 데이터와 생리체 정보, 그리고 당 구조 및 특성 정보, 화학 구조 및 물성 정보등과도 연결되어야만 진정한 의미의 분자 디스커버리가 가능하기 때문이다. 즉, 새로운 타겟 발굴이 이루어지려면 프로테오믹스를 포함한 이 5가지 영역별 정보가 유기적으로 통

합되고 가공되어야 하기 때문이다. 5가지 정보란 게놈정보, 단백질정보, 전사체정보, 생리체정보, 대사체정보를 의미한다. 이들 영역별 방대한 정보를 가공하고 처리하는 분야가 바로 생물정보학 영역에 속한다. 특별히 본 고에서는 단백질 정보학이라 부르지만 SNP, 유전자서열로부터 구조적 모티프, 기하학적인 단백질의 3차구조 좌표정보 획득, 2차원 전기영동상에서 보여지는 pI와 분자량 정보, 단백질 서열확인용을 위한 펩티드 이온 단편화 정보 등이 모두 해당된다.

미니 단백질, 펩티돔(Peptidome)이란?

기능을 부여하는 단백질 구조에서 가장 핵심이 되는 소단위 아미노산 서열부분만을 이용하고자 하는 것이다. 대표적인 예가 항체의 인식부위인 에피토프(항원인식기) 서열이다. 이 항원인식기는 작게는 15개에서 25정도의 아미노산 서열을 이루는 펩티드 단편이다. 거추장스러운 몸통 부분을 제거하고도 이들 펩티드들은 여전히 높은 선택도를 유지하면서 원하는 항원분자에 결합이 가능하다. 실제 생체 내 활동을 위해서는 전체적인 단백질 구조가 목적하는 기능 및 촉매작용에 요구되기도 하지만, 일부 분자체들은 거추장스러운 몸통 부분들을 제거하고도 세포내 기능을 충분히 발휘하는 경우가 있다. 대표적인 예들들면, 프로오피오멜라노코르틴은 뇌하수체 전엽과 중간엽에서 조직 특이적으로 변형이 이루어져 중간엽에서는 α -멜라닌 세포자극 호르몬이 합성되고 뇌하수체 전엽에서는 부신피질 호르몬과 β -엔도르핀이 만들어진다. 이들은 모두 작은 펩티드성 호르몬이다.

펩티돔은 주로 바이오마커 개발, 치매, 항암제 등의 발굴 및 치료, 심혈관 질환과 연관된 뇌 및 신경조절 물질 발굴 등에 널리 이용되고 있다. 펩티드들은 생체내에서 단백질의 합성 기작으로 발생되는 경우가 드물다. 이는 거의 화학적인 합성

경로를 지니고 있어 기존의 화학-단백질간의 연결고리를 제공하기도 한다. 실제로 생체 안에는 이러한 작은 단위체의 펩티드절편들이 중요기능을 수행하는 많이 예가 존재한다.

안정성 검토 및 약물효과 검증을 위한 도구

FDA는 PAT(Process Analytical Technology)라는 부분을 통하여 약물의 연구와 제조사로 하여금 임상학적인 결과에 다변량 해석기법을 도입하도록 하고 있다. 이를 통한 임상학적 결과들은 거대 스토리지에 저장되고, 피드백되어 신속하게 연구 및 제조 공정에 직접적인 수정이 가해질 수 있는 근간을 제공하고 있다. 이러한 관점에서 프로테오믹스는 그동안 거대제약사들에 의하여 베일에 가려져 있던 약물독성, 효과 등에 대한 객관적인 정보를 소비자에게 제공하는데 커다란 기여를 할 것으로 예상된다. 한 예로, 지금까지는 독자적이고 개별적인 데이터 제공이 없이는 알 수 없었던 증상들을 2차원 전기영동을 통한 간단한 실험 기법만 가지고도 세포조직 혹은 생체 내에서 이들 약물이 부여하는 독성 및 효과들을 한눈에 확인할 수 있게 하여준다. 앞으로는 과학적인 임상자료 제공 및 객관화에 더 큰 기여를 할 것이다.

프로테옴 정보학

앞서 간단히 언급된 바와 같이 IBM이나 컴팩사들이 바이오연구에 이용하기 위하여 슈퍼컴퓨터 개발에 수 조원의 연구비를 투입하고 있는 이유는 명백히 단 한가지이다. 단백질이 세포 안에서 작용하기 직전, 혹은 작용 후 변화를 실시간으로 분석하기 위함이다. 보다 현실적인 예측이 주목적이 된다. 수천개 혹은 수만개의 매개변수를 포함한 미분 방정식의 계산은 말할 것도 없이, 이를 위해서는 3차 구조적 Protein Dock 혹은 단백질-단백질, 그리고 단백질-리간드 간의 상호결합

력 분석 등이 병행되어 진행되어야 한다. 그러나 사설 연구기관이나 소단위 연구실에서 이를 위하여 매년 새로운 슈퍼컴퓨터로 교체한다는 것도 다소 무리다. 가장 중요한 점은 어느 일정규모 이상의 시스템이 구축된다면 연구대상에 맞는 유기적인 운영체제와 스마트한 분석법이 실제로 더 요구된다. 최근 바이오-그리드(Bio-Grid) 기법을 이용하는 슈퍼컴의 도입은 실로 많은 기술적 혁신을 가져오게 되었다. 빈번한 변종 출현으로 고민중인 에이즈 치료제 개발에 있어 슈퍼컴의 도입은 필수적이다. 연산속도 증진은 활성분자 디스커버리에 크게 두 가지 의미를 부여한다. 첫째는 에이즈 치료제와 같이 다수의 변이체가 생성되는 경우 이들 모든 변이체의 특정 부위와 결합이 가능한 화학분자들을 일일이 매칭시킬 때 소요되는 시간적 의미와, 또 다른 하나는 모든 변종에 공통적으로 존재하는 골격을 탐색 후 *in silico* 상에서 가상적으로 재현하는 작업이다. 즉, 보이지 않는 에이즈 바이러스의 기원을 찾아보는 일이다. 실제로 SGI사의 그리드-컴퓨팅을 이용하여 에이즈의 기원이 1930년대에 어느 한 HIV “Eve”로부터 출발하였음을 알 수 있었다. 이 서열은 변이체 모두에 작용되는 약물 탐색점의 출발점이 될 것이다. 이처럼 방대한 데이터를 고속으로 처리할 수 있는 정보처리 기술과 융합된 미래의 프로테오믹스는 현재 전세계에서 6초당 1명꼴로 발생하는 에이즈 환자를 공포로부터 구할 것으로 기대된다. 현재까지 3차 구조 좌표가 확보된 단백질은 그 수가 수천 종 밖에 되지 않는다. 전세계 60여개의 X-선 결정연구실들로부터 지난 30여 년간 노력 끝에 얻어진 결과이다. 인간유전자 지도가 완성되고 ORF검사를 통하여 얻어진 전체단백질 수는 무려 3만에서 5만 종에 상당한다. 물론 단백질의 번역 후 변형과정까지를 포함한다면 그 수는 수십만에서 수백만 종에 이르게 된다. 이들 모두를 실험을 통하여 그 기

능과 구조를 파악하기란 실로 많은 시간과 인력이 소요된다. 최근 일본의 RIKEN은 국가단위의 거대 프로젝트의 일환으로 Spring-8을 구축하여 결정화가 빠르게 얻어질 수 있는 단백질들은 X-ray 혹은 방사광 가속기를 이용하여 그 구조를 파악하고, 막 단백질이나 독성이 강하여 발현자체가 어려운 단백질들은 무세포 단백질 발현 시스템을 통하여 발현물을 얻어 바로 핵자기 공명 분석을 시도하는 연구를 작년부터 가동하고 있다. 지난 1년간 이를 통하여 얻어진 단백질 3차 구조의 수는 무려 250종 이상이며, 향후 3차구조 해석에 혁신을 가져올 것으로 생각된다.

미생물 프로테오믹스

많은 생명공학 관련 기업과 학계는 미생물학 분야에의 프로테오믹스 이용이 의학이나 질병치료용 타겟 발굴용도보다 상대적으로 그 활용도가 떨어진다고 생각한다. 그럼에도 불구하고 현재 많은 바이오 연구자들이 프로테오믹스에 의존하는 빈도가 높아지고 있는 실정이다. 많은 화학자 혹은 생물공학자들은 지금까지는 단순히 질량 측정만이 그 주요 목적으로 알고 있었던 질량분석 기법이 최근 그 이온화 방법에 따라서 전혀 다른 분석도구로 사용될 수 있다는 것을 잘 알고 있다. 100년이 넘는 전통을 가지는 발효공학 분야와 바이오 촉매전환기술 부분에서 신규 효소자원의 확보는 기업의 사활을 좌우한다. 가장 성공적인 탐색의 예를 살펴보면 (주)엠에이치투 바이오케미컬(한국)을 들 수 있다. 이 회사에서 생산하는 아미노산 유도체의 경우 생체 내 요소(urea) 회로에서 합성되는 산물이다. 일본 아지노모도의 생산량이 130g/L인데 반하여 이 회사는 고속 대용량 생물 정보 클러스터 시스템을 이용하여 미생물 중에서 가장 활성이 뛰어난 18개의 관련 유전자를 확보하였고 이들 후보 미생물로부터 유전자발현 및 활성

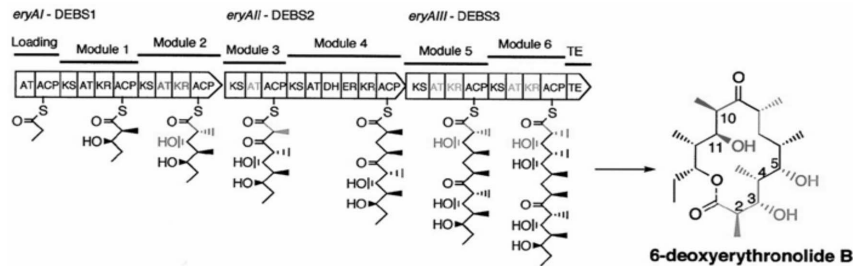


그림 3 전형적인 폴리케타이드 합성의 예. 하나의 세포 안에 코딩된 개별적인 모듈 유전자는 일반적으로 하나의 화학반응을 이끄는 촉매를 발현시키는 정보를 지니고 있다. 6-deoxyerythronolide B 화합물을 합성하기 위하여 8개의 모듈러가 이용된다. 모듈 유전자의 삽입과 삭제 혹은 변형을 통하여 상이한 구조의 화합물들을 다양하게 합성이 가능하다.

실험을 통하여 200g/L의 시트룰린을 3시간이 내에 100% 전환하는데 성공하였다. 공정상의 개선이 필요 없이 무려 70% 가량의 수율증대를 꾀한 것이다. 미생물에서 생산되는 주요 약물들은 그 수를 헤아릴 수가 없다. 예로 녹내장 치료제로 판매되는 CAI약물(제네카)은 *Neurospora crassa* 곰팡이 균주에서, 협심증 치료제인 딜티아젬(Diltiazem, 일본 다나베)은 *Candida cyclindracea*로부터, 파킨슨병 치료제인 L-DOPA(일본 아지노모토)는 *Erwinia*로부터 생산된다. 특히 최근 각광 받고 있는 폴리케타이드(polyketide) 계열의 화합물들은 대표적인 화학유전체 연구분야에 속한다. 이는 원하는 화합물의 조합을 바로 균주 내에 존재하는 바이오 촉매인 효소들의 조합에 의하여 제각기 다른 구조로 만들 수 있기 때문이다. 기존의 화학자가 각 과정별로 반응을 수행하며, 원하는 구조를 합성하던 작업이 생체 내에서 필요한 유전자들의 삽입만으로 얻어질 수 있기 때문이다. 궁극적으로 하나의 세포가 수십에서 수백만 종의 각기 다른 공정 유닛을 지니는 화학콤비나트로 활용될 수 있다. 단위생산량을 증대시키기 위해서는 단지 기하급수적으로 증가하는 세포 수를 늘리는 것만으로도 스케일-업이 가능하다.

대사공학연구를 위한 프로테오믹스

대사공학적 연구에의 활용에 프로테오믹스가 매우 중요한 도구가 될 수 있다. 대사공학자들은 항상 많은 딜레마를 안고 살아왔다. 이유는 많은 생물체들은 항상 스스로를 일정하게 유지시키려는 항상성(homeostasis)을 가지고 있기 때문이다. 일례로 우선 실험용 쥐에게 설탕을 5스푼을 먹여 보자. 순간적인 당 농도의 변화에도 불구하고 쥐가 이로 인한 쇼크를 받지 않는다는 것이다. 즉, 특별히 별다른 소화산물을 내어놓지 않는다는 것이다. 우리 몸에 전기망처럼 얹혀있는 단위 반응 유닛들이 서로간에 피드백을 받으면서 일정한 수준을 유지하려 하기 때문이다. 우리가 잘 알고 있는 “멘델의 유전법칙”을 살펴보기로 하자. 우선 노란완두콩과 녹색 완두콩의 유전비가 우성:열성의 경우 3:1 인 것은 잘 알고 있다. 그렇다면 쭈글쭈글한 완두콩과 탕탕하고 먹음직스럽게 보이는 완두콩의 차이가 왜 발생할까? 대사공학적 해석이 가능하면 직접적으로 농생명과학 분야에 혁신을 가져올 수 있을 것으로 많은 연구자들이 예상하였다. 이에 대한 해석을 가져다 준 것이 바로 프로테오믹스 기술이다. 쭈글쭈글한 완두콩과 탕탕한 완두콩의 유전형 비 즉, RR/rr의 차이를 가져오는 두 대상체간의 전체 단백질 프로파일링 데이터가 확보

되었을 때 대사공학자들은 놀라움을 금치 않았다. 그 이유는 대사공학적 차이를 가져오는 인자는 미처 발견하지 못하였던 특정 key 유전자가 아니라, 발현 단백질 대부분이 모두 3배 이상의 발현율 차이를 보임에 따라 표현형(phenotype)의 비가 달라진 것으로 확인되었기 때문이다. 인류는 멘델의 발견 이후 많은 시간이 흘렀음에도 불구하고 결국은 프로테오믹스의 도입을 통하여 이제서야 “완두콩의 미학”에 대한 궁금증을 알 수 있었다.

우리는 항상 새로운 진보를 원한다. 앞서 기술된 프로테오믹스 연구의 현황 및 미래는 항상 끊임없이 도전하는 연구자들의 손에 달려있다. 이를

위해 고효율 단백질 탐색 및 해석기법을 독자적으로 구축하고, 대량 정보처리에 기반한 단백질 기능해석이 가능케 하는 단백질 정보학과 프로테오믹스의 융합적인 발전을 꾀할 수 있다면, 향후 현재의 기술역량을 선진국 수준으로 끌어올릴 수 있을 것이다. 단백질 기능해석 및 새로운 단백질의 발견은 신약의 개발, 고부가가치 생물소재의 생산으로 연결되어 생물산업의 발전에 크게 기여할 것이다. 특히 화학공학적인 기법의 활용으로 생물산업에서의 단백질체학의 응용은 더욱 가속화될 것으로 기대된다.