

생물시스템의 네트워크 모델링과 시스템즈 생물학

이민영, 유창규, 이인범
포항공과대학교 화학공학과

1. 서론

생물공학분야에서 전통적인 접근방식은 분자수준의 생체물질(단백질, 핵산, 대사물질)이나 특정 유전자의 효율적인 기능 연구에 중점을 두었다. 지난 수십 년간 재조합 DNA 기술 등으로 유용한 물질을 성공적으로 생산해냈지만 개별 생체물질이나 유전자의 조작만으로는 생산성과 이용범위에 한계가 있다. 그러나 다양한 생물의 유전자 정보가 공개된 수 생물의 정성적인 특성을 이해하는 것을 넘어서 시스템적으로 분석하고 정량적인 모델을 구축하는데 필요한 다량의 생물정보가 축적되었다. 본 기술소개에서는 최근 생물을 시스템적으로 분석하는 방법으로 각광받고 있는 생물 네트워크 컴퓨터 모델 구축 방법 및 구성된 네트워크가 사용되는 분야에 대하여 소개하고자 한다. 네트워크 모델은 네트워크를 노드와 링크로 표시한다. 예를 들어, 대사 네트워크에서는 노드가 대사물질이고 링크는 대사반응을 나타낸다. 단백질-단백질 네트워크에서는 각 단백질이 노드이고 링크는 단백질 간의 물리적인 접촉을 나타낸다. 특별히 다른 물리/화학적 네트워크 모델과 달리 생물 네트워크 모델은 상호작용을 확인 가능케 하는 실험방법론 개발과 잡음 제거 모델의 개발이 매우 중요하다. 또한 본 기술소개에서는 생물 네트워크에 관한 활발한 연구가 진행 중인 시스템즈 생물학을 개략적으로 소개한다.

1.1. 대사 네트워크 (metabolic network)

지난 수십 년간 대사공학 (metabolic engineering)은 미생물 및 발효공정등에서 산업에 응용되어왔다. 대부분의 경우 특정 유전자를 변형하는 방법을 채용했지만 이러한 방법으로는 높은 수율과 높은 전환율을 동시에 달성하지는 못했다. 이는 생물의 대사에는 다양한 효소가 참여하고 다른 세포내 네트워크와 연결되어 있기 때문에 개별적인 유전자나 효소의 조절로는 대사를 효과적으로 다룰 수 없기 때문이다. 따라서 이러한 다양한 수준에서의 상호작용을 수학적 모델링으로 정량화하는 작업이 필요하다. 현재까지 대사 네트워크 데이터는 비교적 세부적인 부분까지 KEGG등의 데이터베이스에 축적되어 있어 이를 이용한 대사 네트워크 예측 모델링이 연구되고 있다. 축적된 데이터가 많은 종의 기존의 데이터베이스 정보를 기반으로 다른 생물종의 네트워크를 예측하는 방법, 유전자 시퀀스 등을 이용한 모델링 방법 등에 관한 연구가 효모 (*Saccharomyces cerevisiae*) 나

인간의 암을 대상으로 활발히 진행되고 있다.

1.2. 신호전달 네트워크 (signal transduction network)

최근 유전자관련 기술의 발달로 신호전달 네트워크를 좀 더 세부적인 부분까지 재구성할 수 있게 되었다. 신호전달 네트워크는 친식, 암, 당뇨 등 다양한 질병과 연관되어 있기 때문에 이러한 기술은 네트워크의 기능을 체계적으로 이해하는데 매우 중요한 역할을 하고 있다. AfCS, CSNDB, TRANSPATH 같은 신호전달 관련 데이터베이스가 만들어져 다양한 정보를 제공하고 있고 신호전달 네트워크를 예측하는 컴퓨터 모델링 기법 또한 활발히 연구되고 있다. 현재 연구 동향은 박테리아의 chemotaxies, 인간의 암등을 대상으로 네트워크 구조 뿐 만 아니라 각 네트워크의 세포내 기능이 네트워크의 구조와 어떻게 연관되어 있는지 연구되고 있다.

1.3. 단백질-단백질간의 네트워크 (protein-protein interaction network)

많은 단백질은 신호전달 네트워크에서 중요한 역할을 담당한다. 신호전달 네트워크는 단백질간의 상호작용에 의한 생화학 반응을 수반하기 때문에 단백질-단백질 상호작용에 관한 연구가 활발히 진행되었다. 단백질간의 상호작용은 주로 yeast two-hybrid screen, in vivo pull down 등의 방법으로 전체 게놈을 대상으로, 또는 생물시스템의 특정 프로세스를 대상으로 조사되었다. 다양한 종에 관한 데이터가 IntAct, MINT, The GRID 등의 데이터베이스에 축적되어 있다. 또한 마이크로레이 기술이 프로테오믹스 연구로 확장되어 최근 단백질의 기능적 연구를 위한 단백질 칩 (protein chip)이 만들어지고 있다. 단백질 칩은 한 번의 실험으로 대량의 데이터를 얻을 수 있어서 단백질-단백질 네트워크 모델링에 직접적으로 이용될 수 있다. 또 시퀀스나 구조분석, 비교 유전체학에 근거하여 단백질-단백질 상호작용을 수학적으로 예측하는 연구도 진행되고 있다.

1.4. 조절 네트워크 (gene regulation network)

다양한 세포는 수천 개의 유전자가 특이적으로 조절되어 만들어진다. 또한 세포 주기 (cell cycle), 다양한 환경 자극, 성장 등의 과정에서 유전자 조절이 변화하기 때문에 조절 네트워크는 시스템으로서의 생명체를 이해하는데 중요한 역할을 한다. 지난 수십

년간 개별 유전자의 기능을 한 번에 하나씩 결정하는 것이 주된 연구였지만 이러한 방법으로는 다양한 상황에서의 유전자 조절 현상을 완전히 해석할 수 없었다. 최근 네트워크 모델링 기법을 이용하여 yeast 등을 대상으로 transcription factor의 location 데이터와 유전자 발현 마이크로어레이 데이터를 함께 사용하여 많은 연구가 진행되고 있다. Transcription factor binding location 정보는 Gel shift, DNA footprinting, chromatin immunoprecipitation (ChIP) 등의 방법으로 TRANSFAC, RegulonDB 등의 데이터베이스에 축적되어 있다.

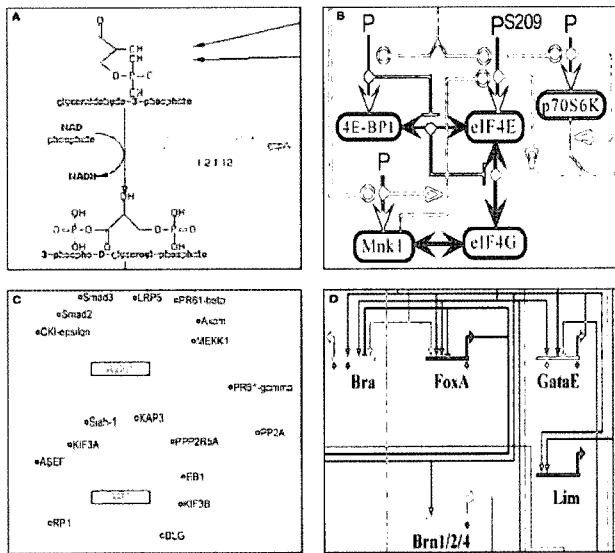


그림 1. (A) 대사 네트워크: EcoVv의 glycolysis 1 네트워크의 부분. (B) 신호전달 네트워크: eMIM의 hypoxia-responsive 유전자의 조절 네트워크. (C) 단백질-단백질 네트워크: HPRD의 WNT 네트워크의 부분. (D) 조절 네트워크: BioTapestry viewer로 나타낸 endomesoderm gene네트워크의 부분. (FEBS letter, 2005).

위와 같이 다양한 네트워크를 모델링하는 방법으로 correlation metric construction, Boolean networks, Bayesian networks 등이 사용되고 있다. Boolean network은 노드를 이진상태 (binary state)로 나타내어 on/off, active/inactive 등을 0 과 1로 표시한다. 다양한 이진 변수를 입력(input)으로 받아서 하나의 이진변수를 출력(output)으로 만들기 때문에 시스템의 다음 상태(state)를 예측할 수 있다. 하지만 이는 시스템의 불확실성을 반영하지 못하기 때문에 확률 불연산 네트워크 (probabilistic Boolean network)가 적용되기도 하지만 계산이 어렵고 여전히 생물시스템을 이진 변수로 가정하기 때문에 근본적인 한계가 있다. 최근의 연구들은 포괄적인 모델링으로 불연산 네트워크를 사용하는데 주요점을 두고 있다. 최근 특히 주목받는 모델링 방법은 베이시안 네트워크 (Bayesian network)

이다. 베이시안 네트워크는 변수들 간의 조건부 독립(conditional independence)을 반영하는 다변량 확률분포 (joint multivariate probability distribution)를 그래프로 나타낸다. 이 방법은 복잡한 확률 프로세스를 나타낼 수 있고 가능한 모든 네트워크 모양을 점수로 매겨 가장 점수가 높은 구조를 시스템의 네트워크로 결정하는 방법이 널리 쓰인다. 또한 동적 베이시안 네트워크를 이용하면 시간에 따른 확률 변수의 변화를 모델링 할 수 있다. 하지만 베이시안 네트워크는 모든 샘플이 같은 분포를 갖는다고 가정하기 때문에 다양한 환경자극에 대한 반응을 모델링하기에는 적합하지 않는 단점이 있다. 한편 생물시스템에 체계적으로 다양한 자극을 주고 이에 대한 데이터를 마이크로어레이 등으로 얻어서 네트워크 구조를 결정하는 방법도 제안되었다. 이와 같은 방법은 drug target discovery등에 직접 응용되고 있다.

1.5. Network for drug target discovery

신약개발 과정의 주요한 과제 중 하나는 후보약의 직접적인 target을 찾는 일이다. 생물시스템의 다양한 분자들은 네트워크로 상호 연결되어 있기 때문에 후보약의 직접적인 target이 간접적인 영향에 의해 변화되는 다른 분자들의 변화에 가려 찾아내기 힘든 경우가 많다. 최근 promoter insertion, gene deletion, drug treatment에 의한 유전자 발현을 측정된 마이크로어레이 데이터를 역공학 (reverse-engineering) 기법으로 네트워크를 구성하여 후보약의 target을 찾는 연구가 있다 (그림2). 하지만 대부분의 약이 단백질에 직접 작용하고 유전자 발현에는 간접적인 영향을 주기 때문에 유전자발현 데이터만을 이용하면 예측성능에 한계가 있다.

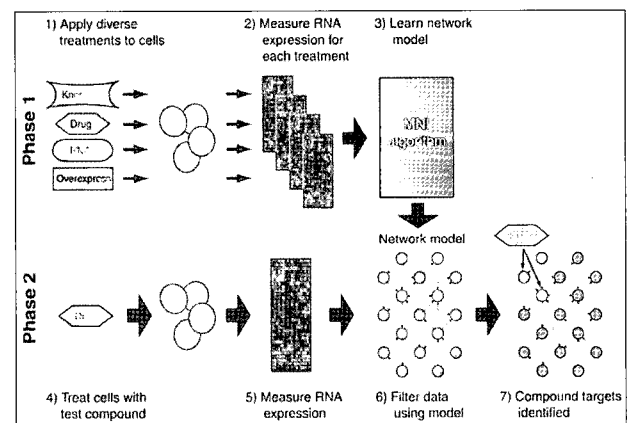


그림 2. Reverse-engineering network modeling 방법 (Nat. Biotechnol., 2003).

한편 지금까지 알려진 유전자 수준의 데이터는 다양한 연구에서 많은 false positive를 나타낸다고 알려져 왔고 모든 데이터들의 통합만이 가장 정확한 생물시스템 모델링 및 네트워

크 모델링이 가능하다고 알려져 왔다. 따라서 보다 정확한 생물 시스템 모델링을 위하여 이용가능한 모든 데이터(pathway & process data)와 정보시스템(information systems)을 통합하여 네트워크를 전산 모델링(computational model)하고 이를 바탕으로 실험과 모델링을 반복하는 반복적인 생물시스템 모델링 방법이 제시되었다(그림 3). 여기서 모델기반 실험 계획에 기반을 둔 실험결과는 반드시 모델 결과와 비교되어야 하고 그 후 모델 보정에 사용되어야 한다. 이러한 생물 시스템 모델링을 통해 우리는 생물시스템(조절, 대사, 단백질, 신호 전달 네트워크)내에서 주요 영향 인자가 무엇인지? 시스템 구성 성분들 간의 가장 중요한 관계가 무엇인지? 환경으로부터의 변화에 따른 시스템 생리학적 효과가 무엇인지?와 같은 문제에 대한 답을 얻을 수 있다.

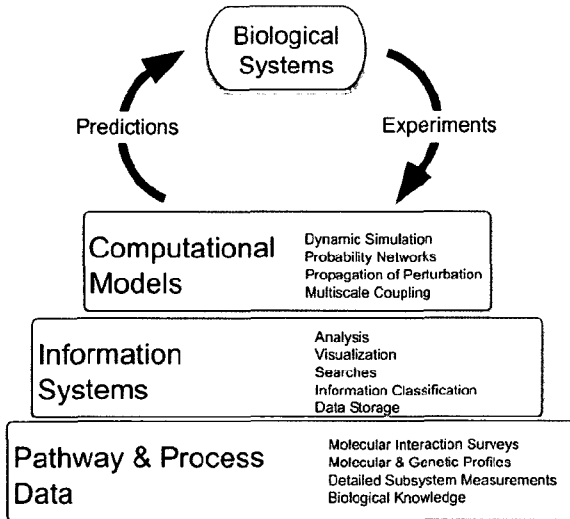


그림 3. 반복적인 생물시스템 모델링 (FEBS letter, 2005).

1.6. 시스템즈 생물학 (systems biology)

기존의 많은 생물시스템 모델링의 경우 세포내 반응이나 일부 상호작용을 이해하기 위한 실험에 많은 관심을 가져왔다. 최종적으로는 생물시스템의 환경변화 예측, 약의 예측 등은 다양한 차원에서 실험 및 결과들은 전부 조합해야만 설명이 가능하다고 알려져 왔다. 시스템에 대한 이러한 상황에서 생물시스템 내에서 상호 작용하는 성분들의 동력학적 특성을 거시적인 시스템적인 방법으로 모델링하고 분석하는 시스템즈 생물학(systems biology)이 최근 많은 관심을 받고 있다. 이러한 시스템즈 생물학에서 전체 생물시스템의 상호작용들의 조합으로 재구성하고 해석하는 동력학적 네트워크 모델은 시스템즈 생물학에서 가장 중요한 하다고 알려져 왔다. 여기서 생물시스템의 주요 특성인 다양한 환경변화에 강건한 모델(robust network model)과 많은 작은 네트워크들의 통합하여 핵심모델

(core network model) 구축이 중요하다고 할 수 있다. 대사 네트워크, 신호전달 네트워크, 조절 네트워크 등 다양한 생화학적 반응 네트워크들은 시스템즈 생물학의 토대가 되고 다양한 네트워크 관련 데이터베이스의 통합을 위해 2000년에 SBML(Systems Biology Markup Language)과 CellML이 제안되었다. 이는 다양한 반응 네트워크에 사용되는 모델과 데이터베이스의 형식과 언어의 통합을 위해 현재까지 계속 발전되고 있다.

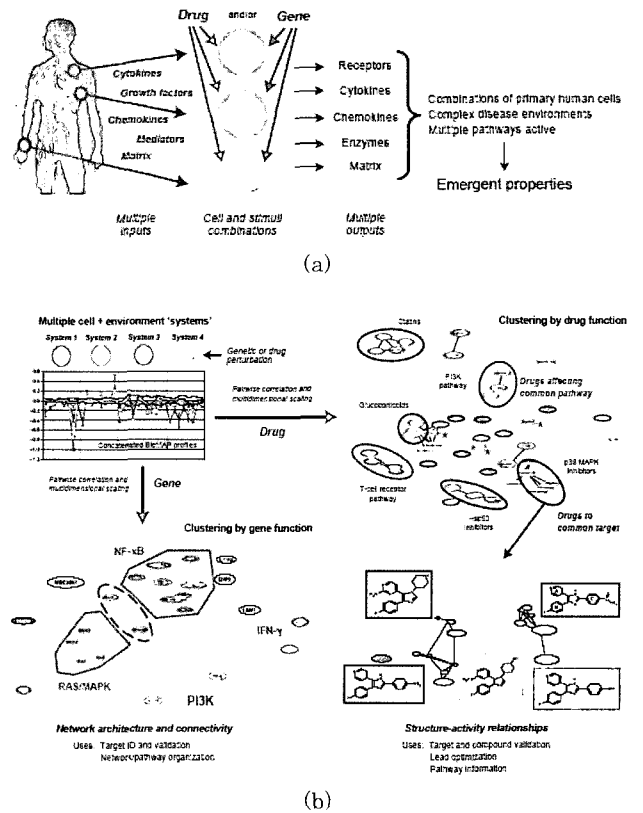


그림 4. 신약개발에서 세포의 시스템즈 생물학(Nat. Biotechnol. 2004).

특히 신약개발에서 시스템즈 생물학의 영향이 두드러지고 있는데 시스템즈 생물학은 복잡한 신호전달의 여러 단계 중 신약개발에 가장 효과적인 부분을 선택하여 신약후보물질의 치료 효과뿐만 아니라 부작용과 관련된 부분까지도 예측할 수 있다고 보고되어 있다. 최근 보고서에 따르면 1998년부터 미국을 비롯한 여러 나라의 학계와 회사에서는 시스템즈 생물학을 이용한 연구가 진행 중이며 Beyond Genomics와 같은 시스템즈 생물학 전문 벤처기업과 시스템즈 생물학 연구소(미국)가 설립되어 면역체계의 시스템 생물학적 접근 방법을 도입하였다. 미국, 일본 및 다국적 기업에서는 시스템즈 생물학을 이용하여 바이오 신약개발, 단백질 3차 구조 규명(Protein3000),

대장균 내 생명현상 전산모델 연구 (E-cell project) 등 시스템즈 생물학 연구에 박차를 가하고 있다 (Kosen, 2005).

생체 네트워크 및 정보시스템 데이터베이스 구축의 주요한 목적은 시스템공학 기술을 이용하여 생물시스템의 정성적 및 정량적 네트워크 모델을 축진하는 것이다. 그러나 아직 이러한 네트워크 모델로부터 생물현상, 특히 생명현상에 대하여는 네트워크 모델링의 실제적인 활용은 현재 진행형이라고 할 수 있다. 따라서 현재 진행 중인 데이터로부터 다양한 학제들 간의 공동연구가 필요하며 본 기고에서는 생물 시스템 모델링에서 가장 중요한 상호작용 네트워크 모델과 시스템즈 생물학에 대하여 개략적으로 소개하였다.

..... 저자약력



《이 민 영》

- 2003년 포항공대 화학공학과 학사.
- 2005년 포항공대 화학공학과 석사.
- 현재 지능자동화연구센터 연구원.
- 관심분야 : 시스템즈 생물학, 생물정보학.
- 연락처 : E-mail: orange@postech.ac.kr



《유 장 규》

- 1993년 연세대학교 화학공학과 학사.
- 1995년 포항공대 화학공학과 석사.
- 1998년 두산기술원 전임연구원.
- 2002년 포항공대 화학공학과 박사.
- 2002년 ~ 2004년 벨기에 겐트대학교 BIOMATH학과 Post-doc.
- 현재 포항공대 환경공학부 연구조교수.
- 관심분야 : 시스템즈 생물학 (생물정보학), 공정시스템, 환경시스템공학.
- 연락처 : E-mail: ckyoo@postech.ac.kr



《이 인 범》

- 1977년 연세대학교 화학공학과 학사.
- 1979년 KAIST 화학공학과 석사.
- 1982년 한국과학기술연구원 연구원.
- 1987년 Purdue univ. 화학공학과 박사.
- 1988년-현재 포항공대 화학공학과 교수.
- 1998년-현재 지능자동화연구센터 소장.
- 2004년-현재 한국공학한림원 정회원.
- 연락처 : E-mail: iblee@postech.ac.kr