

시스템생물학분야 데이터 수집 및 시뮬레이션 동향

한국과학기술정보연구원 | 유석종
충북대학교 | 유재수*

1. 서론

생명현상에 대한 연구는 생명체를 구성하는 기본 물질과 생명체를 유지하는 다양한 기작들에 대한 연구를 기반으로 하고 있다. 특히 생명체의 구성 물질들이 규명된 이후로 유전자와 단백질 그리고 다양한 지질과 탄수화물 및 이온 물질들이 관여하는 생체내의 조절기작에 대한 연구가 활발히 진행되고 있다. 특히 유전현상에 대한 연구가 활발하게 진행되면서 유전체와 유전자의 역할과 기작을 이해하게 되었고, 유전자를 구성하는 핵산의 서열을 분석할 수 있는 방법이 개발되면서 유전자 분석이 본격적으로 수행되어 인간의 유전체를 분석할 수 있게 되었다. 이후 대규모의 분석 실험장비들이 등장하게 되면서 생물학 연구에도 많은 변화가 일어났다. 가장 큰 패러다임의 변화는 기존에 생명과학자들이 연구범위가 하나의 유전자나 단백질에 초점을 맞추어 연구하던 방식에서 생명현상에 관여하는 모든 유전자와 단백질 혹은 화합물들을 총체적으로 분석하려는 움직임이 등장하였다. 이러한 연구가 바로 오믹스 연구이다. 오믹스 연구는 다양한 분야에 적용되어 진행되고 있는데 가장 활발한 연구분야는 유전체와 단백질 그리고 대사체를 들 수 있다[1].

유전체 연구는 90년대 중반 미생물과 선충류의 유전체를 해독하면서 관련된 유전자의 염기서열 분석을 중심으로 빠르게 발전하였으며, 이후 인간유전체가 해독되고 이를 기반으로 기능유전체연구가 가속화 되면서 폭 넓은 분야로 확대되고 있는 실정이다. 특히 최근에는 기능유전체, 약물유전체, 구조유전체, 후성유전체, 메타게놈 유전체 등 다양한 분야로 세분화 되어 연구

되고 있는 실정이다. 단백질연구 또한 단백질 개개의 기능연구에서 벗어나 다양한 단백질간의 연관된 기능과 작용기작을 이해하는데 연구가 활발히 진행되고 있다. 또한 대사체 연구는 생체를 유지하는데 필요한 필수적인 대사과정에 대한 통합적인 분석이 진행되고 있다. NMR장비의 정밀화와 GC-MS/LC-MS 등 측정장비를 통해 생체내의 대사물질을 분석하고 이를 기반으로 다양한 질병진단 및 산업적 미생물개발 등 다양한 연구에 활용되고 있다.

자동차가 달리는 원리를 파악하기 위해서는 자동차를 구성하는 다양한 부품을 파악하고 각 부품이 역할과 동작원리를 해석해야 하듯이 생명현상의 근본기작을 이해하기 위해서는 단일 유전자나 단백질을 연구로는 그 한계가 존재하게 된다. 이러한 관점에서 시작된 학문이 시스템생물학으로 세포내의 각 구성물들을 동시에 파악하고 분석하여야만 정확한 생명현상의 기작을 해석할 수 있을 것으로 기대하고 있다. 일례로 단일 유전자가 질병을 일으키는 경우보다 여러 개의 유전자가 질병을 유발하는 경우가 많이 존재하게 되는데 이러한 측면에서 시스템생물학적인 관점에서 질병을 해석하는 것이 중요한 부분으로 제안되고 있다.

생명체의 시스템 수준의 이해를 위한 4가지 속성을 살펴보면, 첫째 시스템의 구조이다. 즉 유전자 상호작용과 생화학적 경로(pathway)의 네트워크 및 세포내 및 다세포간 구조의 물리적 속성을 조절하는 메커니즘을 포함한다. 둘째 시스템 동력학(Dynamics), 이는 어떻게 시스템이 다양한 조건하에서 시간에 따라 행동하는지를 연구하는 것으로 다양한 대사분석, 동력학분석 및 민감성분석 등이 여기에 속한다. 셋째는 조절방법이다. 세포의 상태를 시스템적으로 조절하는 메커니즘을 통해서 질병의 치료 타겟 선정이 가능하다. 넷째는 디자인방법으로 한정적 디자인원리와 시뮬레이션을 기반으로 원하는 속성을 지닌 생물학적 시스템을 설계하는 전

* 종신회원

† 본 연구는 한국과학기술정보연구원의 지원(과제 : K-13-L01-C02-S04)을 받아 수행되었습니다.

락을 말하고 있다. 이러한 연구를 위해서는 컴퓨터과학, 유전체학, 생물학적 측정기술 및 다양한 지식통합이 필요한 실정이다[2].

1995년에 시작된 일본 교토대학의 e-Cell 프로젝트를 시작으로 다양한 시뮬레이션 방법과 도구가 소개되고 있으며, 과거에는 생물학적 현상들을 연구하고자 할 때 직접 실험하는 방법이 주로 이용되었었지만, 이제는 컴퓨터상에서의 시뮬레이션을 통하여 원하는 정보를 얻는 것이 가능해졌다[3]. 이후 CellDesigner[4] 등 다양한 시스템생물학 분석도구가 개발되어졌다. 최근에는 이를 더욱 발전시켜 대규모 오믹스 정보 및 대용량 실험결과를 활용하여 시스템생물학 분석모델링을 수행할 수 있는 시스템의 개발이 요구되어지고 있는 시점이다. 최근 클라우드 시스템이 산업 및 연구에 활용되면서 유전체 분석에 널리 활용되고 있다. 이렇듯 대용량의 시스템의 필요성은 날로 증가하고 있으며, 이에 따라 시스템생물학 분야에서 고성능 컴퓨팅환경에 적합한 분석 프레임워크의 개발의 중요성이 날로 증가하고 있는 실정이다[5].

본 논문에서는 다양한 실험과 분석을 통해 생명체 내의 상호작용정보를 수집하는 방법과 이와 관련된 데이터베이스를 설명하고, 생성된 생물학적 네트워크를 기반으로 생물학적 모델을 수립하고 이를 시뮬레이션하는 다양한 방법을 소개하고자 한다. 결론에서는 향후 시스템생물학분야를 활용방향과 향후 연구에 대해 논하고자 한다.

2. 단백질-단백질 상호작용(PPI) 데이터 수집 및 데이터베이스

시스템생물학 분석을 위해서는 기본적으로 실험 데이터로부터 생물학적 상호작용을 추출하고 이를 분석할 수 있는 형태의 데이터로 저장하고 있어야 한다. 본 장에서는 생물학적 상호작용의 정보를 수집하기 위하여 현재 많이 수행되고 있는 실험방법과 이를 기반으로 하는 데이터베이스에 대해 살펴보기로 한다.

2.1 단백질-단백질 상호작용(PPI) 데이터 생성 방법

생명현상의 기본 단위는 화학반응이며, 이러한 화학반응이 실온에서 빠르게 일어나기 위해서 생체는 단백질을 이용한다. 또한 외부의 자극에 대처하고 이에 대한 새로운 화학반응을 만들기 위해서는 신호전달과정을 통하여 이루어진다. 즉 근본적으로 생명현상의 기본은 단백질과 단백질의 상호작용이 주된 역할을 수행하게 된다. 이러한 단백질간의 상호작용을 파악하기 위

서는 먼저 단백질 간에 물리적 결합이 동반되는지를 파악하여야 한다. 기존의 연구에서는 각각의 단백질을 정제하고 이들 간의 결합력을 실험적으로 검증함으로써 데이터를 얻는 과정을 거쳤다. 하지만 이러한 과정은 매우 비효율적이며 얻어지는 정보조차 미미한 상태이다. 이런 문제점을 해결하고자 Yeast two-hybrid 실험 방법과 면역침강법등이 고안되어졌다. 근본적으로 대규모의 실험을 통한 대량의 단백질 상호작용 정보를 얻으려는 시도이다. 각각의 방법에 대해 살펴보기로 하겠다.

• Yeast two-hybrid 실험

Yeast two-hybrid 스크린 실험은 유전자 발현을 유발하는 GAL4라는 유전자의 2개 도메인을 이용하는 방법이다. 먼저 DNA-결합 도메인과 활성도메인을 분리하고 여기에 외부 단백질을 붙여 발현시킴으로써 각각의 도메인에 결합되어 발현된 단백질간의 상호작용이 있는 경우에는 GAL4의 기능이 정상적으로 작동하여 레포트 유전자를 발현시키게 되고 그 결과 연구자가 단백질 상호작용을 하고 있는 미생물을 선별할 수 있도록 유도하는 방법이다[6,7].

즉 미생물의 색상변화를 통해 단백질-단백질 상호작용을 하는 조합을 수많은 조합에서 선별할 수 있게 되는 방식으로 초기 단백질-단백질 정보를 얻는데 가장 활발하게 활용되는 방법이었다.

• 면역침강법

면역 침강법은 항체가 특정단백질과 결합한다는 원리를 이용하는 방법으로 상호작용정보를 수집하고자 하는 단백질을 정제하여 특이적인 항체를 생성한다. 이후 목적한 단백질과 세포내의 다양한 종류의 단백질 용액과 반응을 시키면 특이적으로 결합하는 다양한 형태의 단백질의 복합체가 생성된다. 이후 생성된 항체를 이용하여 목적 단백질과 결합시킨 후 무게차이를 이용하여 항체만을 수거하게 되면 항체와 결합한 다양한 종류의 복합체를 얻을 수 있으며, 이렇게 얻어진 단백질들을 분석하게 되면 목적 단백질이 어떤 종류의 단백질과 상호작용하는지 알아낼 수 있게 된다[7].

• 그 밖에 방법들

최근에는 보다 다양한 방식의 단백질-단백질 상호작용 정보를 확보할 수 있는 방법들이 소개되었는데, 그 예로는 Biomolecular fluorescence complementation(BiFC)이 있다. 이 방법은 형광단백질을 2 도메인으로 분리하여 여기에 각기 다른 단백질을 발현시키게 되며 단백질 간 상호작용이 있는 조합만이 형광을 생성할 수 있는 특징을 이용하는 것이다. 이 방법은 yeast 2-hybrid

방법과 마찬가지로 해당 미생물을 형광을 통해 쉽게 스크린 할 수 있는 장점이 있다[8].

또 다른 방법으로는 Affinity electrophoresis 방법으로 이것은 단백질 간 상호작용이 발생하게 되면 그 전기적 특성도 달라져 전기영동 시 그 이동도가 달라진다는 특징을 이용하여 단백질 상호작용을 추적하는 방법이다. 앞으로 단백질 상호작용에 대한 대량의 스크린 방법은 지속적으로 등장할 것으로 예상되며, 이미 고효율의 스크린 방법들이 소개되고 있다. 이러한 장비들은 대규모의 단백질 상호작용 정보들을 생성해 낼 수 있을 것으로 기대된다.

2.2 관련 데이터베이스

단백질-단백질 상호작용(PPI) 데이터베이스는 생물학적 상호작용에서 가장 핵심적인 정보로 앞장에서 소개된 다양한 방법에 의해 생산된 정보를 통합하여 연구자들에게 제공되고 있다. 현재 서비스되고 있는 중요 단백질-단백질 상호작용(PPI) 데이터베이스를 살펴보기로 한다.

2.2.1 신호전달/대사네트워크

• BioModels

BioModels 데이터베이스는 영국의 EMBL-EBI와 미국의 SBML 팀, 일본의 시스템생물학연구소, 남아프리카공화국의 JWS online간의 협력을 통해 시작되었으며, EBI에서 구축 운영하고 있는 데이터베이스로 현재 전 세계에서 가장 많은 400개 이상의 생물학적 반응 네트워크를 보유하고 있다[9]. BioModels 데이터베이스는 각각의 연구자가 자신이 모델링한 생물학적 네트워크를 논문 게재이후 등록할 수 있으며, 등록된 생물학적 모델은 정량적 모델을 기준으로 직접 큐레이션하여 전 세계 연구자들에게 제공되고 있다.

큐레이션 작업은 BioModels 데이터베이스의 가장 큰 장점으로 다양한 프로그램에서 생성된 SBML 파일의 각각의 엔티티의 확인 작업을 통해 다양한 생물학적 반응 네트워크를 분석할 수 있는 다양한 소프트웨어에서 사용할 수 있다. 또한 SBML 표준의 레벨2 버전1부터 버전4까지 다양한 포맷을 지원함으로써 다양한 분석 도구에서 사용될 수 있도록 한다. 또한 Gene Ontology (GO) 용어나 질병 코드를 사용하여 생물학적 반응 네트워크에 주석을 부여한다. 이를 통해 연구자가 질병 이름이나 GO의 용어를 기준으로 데이터베이스를 검색할 수 있다.

• KEGG

일본 교토대학의 KEGG(Kyoto Encyclopedia of Genes and Genomes)는 대사 경로 데이터베이스로, 대사 네

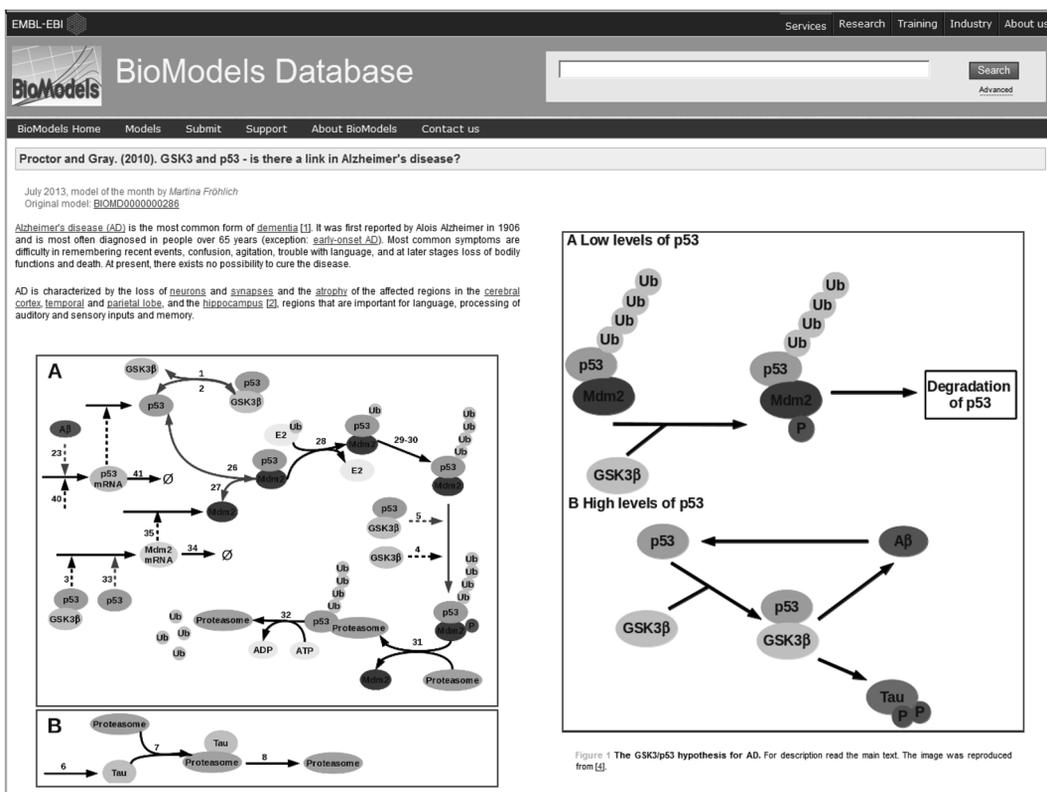


그림 1 BioModels Database[9]

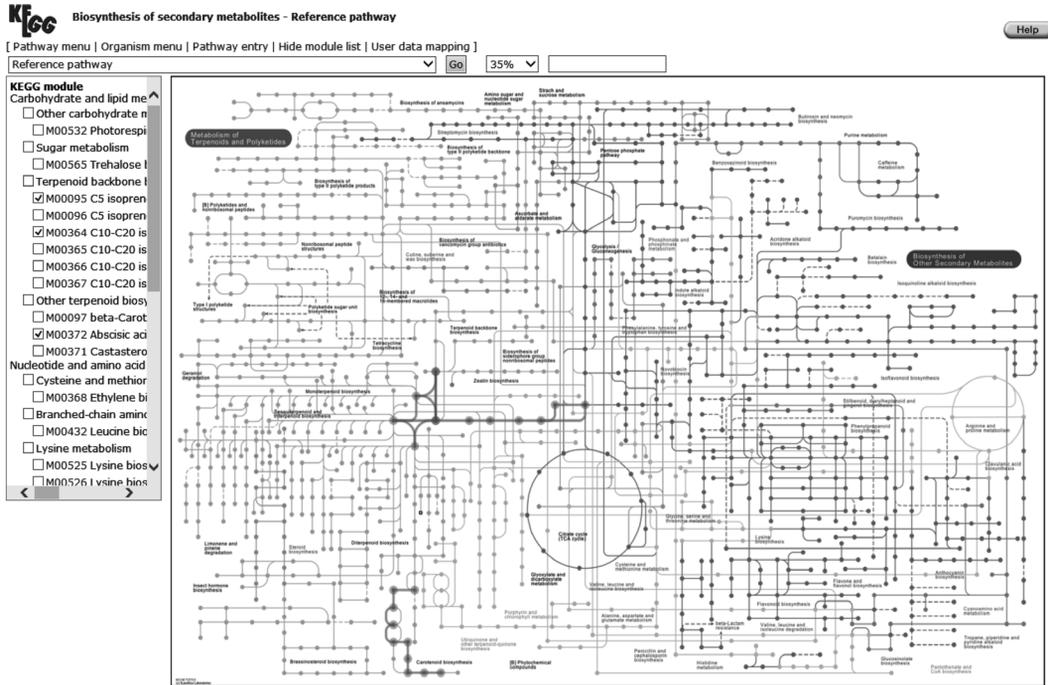


그림 2 KEGG Database[10]

트위크 지도를 제공하고 각 네트워크의 구성 원소들에 대해 여러 다양한 데이터베이스와 연결된 통합 데이터베이스 형태로 구성되어 있다[10]. 그 중에서 KEGG PATHWAY 데이터베이스에는 물질대사, 막 수송, 신호 전달, 세포 주기와 같은 세포 내 경로에 대한 정보가 저장되어 있는데, 여기에서 얻은 경로와 기능, 염기 서열과 같은 정보들을 유기적으로 연결해서 대사 기능에 대한 시스템적인 분석을 가능하게 한다. 그림 2와 같이 현재 KEGG 경로 내에 구축되어 있는 경로들로 목적에 알맞은 데이터를 선택하여 신호경로 매핑을 수행할 수 있으며, 이를 바탕으로 관찰하고자 하는 대사 경로를 분석할 수 있다.

• BioCyc

BioCyc는 현재 1,129개의 경로와 유전체 데이터베이스의 집합체이며, 각 데이터베이스는 생물의 유전체와 대사 경로 정보를 포함하고 있다. 이용되는 정도와 지속적인 생신 여부 등에 따라 데이터의 질을 1~3 단계로 구분한다[11]. 1단계는 논문에 게재되어 유효성이 검증된 정확성 높은 EcoCyc(*Escherichia coli*), MetaCyc(다양한 생물들의 대사 경로와 효소 데이터베이스), AraCyc(*Arabidopsis thaliana*), YeastCyc(*Saccharomyces cerevisiae*) 데이터베이스이며, 2단계와 3단계는 예측된 경로 정보를 갖고 있는 데이터베이스로 2단계에는 33개의 데이터베이스가, 3단계에는 968개의 데이터베이스가 존재한다.

MetaCyc에는 미생물, 식물, 동물 등 다양한 생물들의 유효성이 검증된 대사 회로와 효소에 대한 정보가 있고, 이와 같은 데이터베이스를 이용해서 분석이 가능하다. BioCyc에서는 데이터베이스로부터 얻은 정보를 바탕으로 대사 네트워크를 예측하는 경로 분석 소프트웨어인 PathoLogic을 이용할 수 있다. PathoLogic으로 예측된 경로들은 BioCyc 내 PGDB(Pathway/Genome DataBase)에 구성되어 있다.

2.2.2 단백질-단백질 상호작용 데이터베이스

• BIND

BIND는 단백질간의 상호작용을 수집한 데이터베이스로서 주로 단백질의 상호작용과 단백질 복합체 정보 그리고 경로정보를 저장하고 있다[12]. 여기에는 상호작용의 주체로서 단백질 핵산 그리고 작은 분자들이 있으며, 이 같은 물질들 간의 상호작용을 기술하고 있는 것이다. 또한 화학작용, 광화학작용, 구조변경 등의 정보도 기술하고 있다. 작은 분자에서 신호전달까지의 모든 것들이 그래프 이론을 이용하여 요약되어 있으며, 상호작용 네트워크 및 반응속도 시뮬레이션 연구에 이용되어지고 있다.

• DIP

DIP은 단백질간의 상호작용에 초점을 맞춘 상호작용 데이터베이스이다. 여기에 수집된 자료는 모두 실험적인 근거를 기반으로 구축되었으며, 574개의 종으로부터 25,628개의 단백질과 약 75,420개의 상호작용을

기술 하였다. 현재 지속적으로 자료 갱신이 이루어지고 있으며, 인간을 비롯하여 초파리, 효모 등 다양한 종의 자료 수집을 하고 있다[13].

또한 JDIP이라는 가시화 도구를 제공함으로써 사용자가 손쉽게 상호 작용하는 단백질을 시각적으로 볼 수 있으며, 해당 단백질의 상세정보도 같이 볼 수 있는 기능을 제공하고 있다. 또한 각 단백질의 정보를 Swiss-Prot, PDB, RefSeq, PIR, NCBI의 데이터베이스와 연결하여 관련 정보의 링크를 제공한다.

• BioGRID

물리적, 유전적, 기능적 상호작용은 단백질학 및 기능 유전체학에서 관심있어 하는 사항이다. BioGRID는 이런 항목들에 대한 분석 및 관계형 데이터베이스를 구축함으로써 사용자가 관심있어 하는 단백질에 대한 검색을 제공한다[14].

현재 약 48만개의 상호작용을 기술하고 있다. 웹을 통한 검색을 제공하고 있으며, 상호작용에 대한 그래프 표시는 “Osprey”라는 가시화 도구를 사용하여 볼 수 있다. 정밀한 검색을 원하는 사용자를 위해 기본 검색뿐만 아니라 고급검색 제공하고 있다. 검색결과에서는 GO의 분류를 보여주며, 관련논문 및 관련데이터베이스와의 링크를 제공한다.

• HPRD

HPRD는 2003년부터 존홉킨스 대학에서 서비스되고 있는 데이터베이스로 인간의 모든 단백질정보를 통합하고 제공하고 있다. 또한 다양한 실험정보를 통합하여, 단백질 도메인, 인산화정보, 아세틸화 등의 정보를 제공하고 있다. 또한 NetPath라는 암 및 면역의 신호전달 네트워크에 대한 데이터베이스를 구축하고 서비스하고 있다. 특히 HPRD에서는 단백질-단백질 상호작용에 대한 정보도 통합하여 다양한 단백질 상호작용 네트워크를 파악할 수 있다. 현재 약 41,327개의 단백질-단백질 상호작용 정보를 제공하고 있다. 또한 신호전달 네트워크도 32개의 확인된 신호전달경로를 제공하고 있다[15].

2.2.3 데이터 표준화

앞에서 기술된 바와 같이 현재 다양한 단백질-단백질 상호작용 데이터베이스 및 신호전달/대사네트워크 데이터베이스가 존재하고 있으며 이들 간의 데이터교환에는 많은 제약사항이 존재한다. 단백질-단백질 상호작용 데이터베이스만 하더라도 실험방법이 다양함으로 인해 서로 다른 방식으로 데이터베이스가 설계되어 있다. 초기에 이 같은 문제를 해결하고자 IMEx(international molecular Exchange) 콘소시움이 발족하여 데

이터간의 교환이 가능하도록 장려하고 있다[16]. 이러한 노력의 결과로 XML형식의 PSI-MI(Proteomics standards initiative-molecular interaction) 포맷이 개발되어졌다. 이밖에도 신호전달과정 등에 보다 초점을 맞춘 BioPAX(<http://www.biopax.org>)도 개발되어졌다[17]. 또한, 대사네트워크의 데이터를 교환하기 위한 방법으로는 SBML(Systems Biology Markup Language)가 개발되어졌으며, 대사 및 신호전달 네트워크를 모델링 및 시뮬레이션 시 활용하기 위해서 데이터 교환용 포맷으로 표준화 되었다[18].

2.3 시스템생물학 시뮬레이션 도구

일반적으로 생물체 내의 대사 시스템의 동적 거동을 설명하기 위하여 대사 반응들을 시간에 따른 각 대사물질의 농도변화를 나타내는 미분 방정식으로 표현하고, 이 반응 속도 식들의 해(solution)로부터 계산한 시간에 따른 농도 프로파일을 이용한다. 세포내의 물질 반응을 분석할 수 있는 전통적인 수리생물학적 방법은 생체 분자들 사이의 변환과 반응과정을 결합된 상미분방정식계로 기술하는 소위 연속 모델링(continuous modeling) 방법이다. 이러한 접근 방법의 한 예로 최근 효모의 세포 주기 과정에 관여하는 40여 개의 인자들의 농도 변화에 대한 속도식(rate equation)을 세워 이를 수치적으로 해석하는 연구가 행해졌다[19]. 이를 위해 다양한 분자생물학적 지식을 참고하고 새로운 실험을 수행하여 어떤 인자들이 각각의 반응 과정에 참여하고 또 반응 계수가 어떻게 되는지 최대한 실제에 가깝게 밝혀내는 과정이 필요하다. 이를 통해 얻어진 게놈 규모의 세포 주기 수치 모델은 실제 야생형(wild-type)의 효모의 세포 주기를 잘 시뮬레이션 하는 것으로 확인되었다. 최근에는 더욱 정교한 대사 기전을 시뮬레이션 할 수 있는 보다 폭넓은 연구에 활용되고 있다.

2.3.1 상미분방정식 분석

일반적으로 생물체 내의 대사 시스템의 동적 거동을 설명하기 위하여 대사 반응들을 시간에 따른 각 대사물질의 농도 변화를 나타내는 미분 방정식으로 표현하고, 이 반응 속도 식들의 해로부터 계산한 시간에 따른 농도 프로파일을 이용한다[20]. 세포 내의 물질반응을 분석할 수 있는 전통적인 수리 생물학적 분석 방법은 생체 분자들 사이의 변환과 반응 과정을 결합된 상미분 방정식계로 기술하는 소위 연속 모델링(continuous modeling) 방법이다. 이러한 접근 방법의 한 예로 최근 효모의 세포 주기 과정에 관여하는 40여 개의 인자들의 농도 변화에 대한 속도식(rate equation)을 기술하고 이를 수치적으로 해석하는 연구가 진행되었다

[19]. 이를 위해 다양한 분자 생물학적 지식을 참고하고 새로운 실험을 수행하여 어떤 인자들이 각각의 반응 과정에 참여한다. 또한, 반응 계수가 어떻게 되는지 최대한 실제에 가깝게 밝혀내는 과정이 필요하다. 이를 통해 얻어진 계수 규모의 세포 주기 수치 모델은 실제 야생형(wild-type)의 효모의 세포 주기를 잘 표현하는 것으로 확인되었다. 100 가지가 넘는 돌연변이 형질에 대해 그 영향을 잘 예측하는 것으로 판명되어, 앞으로 새로운 유전적 이상 현상에 대한 이해와 치료에 이용될 수 있을 것이다. 최근에는 더욱 정교한 대사 기전을 시뮬레이션 할 수 있는 보다 폭넓은 연구들이 진행되고 있어, 실제 효모의 전체 대사 네트워크도 구축되어 시뮬레이션에 활용되고 있다[21].

현재 몇 가지의 ODE분석 도구가 개발되어 사용되어지고 있다. Gepasi는 임의적 평면 또는 입체 공간에서의 세포 간, 세포 내 대류, 반응 및 확산 현상을 모델링하고 이를 시뮬레이션 할 목적으로 모델식을 편미분 방정식과 상미분 방정식 등으로 표현하였다[22]. 또한 버지니아 생물정보학 연구소와 독일의 유럽 미디어 연구소의 과학자들이 공동으로 참여하여 Gepasi의 모델에 대한 최적화, 대사 조절 분석, 선형 안정성 분석 등의 기능에 결정론적인 시간열(time scale) 분석, 정상 상태 분석, 대사 조절 분석, 질량 보존 분석, 모수 탐색 기능을 추가 개발한 COPASI(Complex Pathway Simulator)가 있다[23].

2.3.2 대사흐름분석

대사흐름분석(Metabolic Flux analysis, MFA)은 대사 네트워크에 대해서 대사물질의 흐름을 분석하는 수학적 접근방법이다. 특히 대사공학 분야에서 많이 사용되어지고 있는 방법으로 특정 산물 혹은 세포특성을 직접적으로 개선하는 것이 목적이다. 이를 위해서는 특정 생화학반응의 변화나 새로운 유전자를 도입함으로써 가능해 지는데, 단순한 실험방법으로는 다양한 조합의 유전자변형이나 유전자도입을 수행해야 하므로 비효율적이게 된다. 대사흐름분석은 이러한 목적을 빠르게 달성하기 위해 대사네트워크를 모델링하고 특정 산물을 얻기 위한 다양한 가상실험을 통해 원하는 최적의 해답을 찾고자 한다. 특히 대사흐름은 단순한 생화학적 반응에 따른 대사물질의 흐름을 파악하는 것이지만 그 이면에는 유전체, 전사체, 단백질체, 대사체 등 다양한 비선형 상호작용의 총합이라고 해야 할 것이다. 이같은 다양한 수준의 생물학정보가 통합되어 분석된다면 명확한 대사조절의 과정을 이해할 수 있을 것이다. 현재 대사흐름분석은 수 십년 간 유전체 수준의 대

사네트워크의 재구축 등 생화학적 네트워크연구에 활발히 이용되어졌다. 이러한 네트워크재구축은 다양한 종류의 대사물질과 반응을 포함하고 있고, 대사흐름분석을 통해서 이같은 대사네트워크를 통한 대사물질의 흐름을 계산할 수 있다. 그러한 활용의 사례로 현재 약 35개의 생명체에 대해서 생체 성장률을 예측하는 것이 가능해 졌다. 최근에는 $^{13}\text{C}/^{14}\text{C}$ 방사성동위원소를 이용한 고성능실험기술과 연계하여 보다 정확한 예측이 가능해 지고 있다.

대사흐름분석에 대한 다양한 분석도구가 현재 개발되어 제공되고 있다. COBRA는 대표적인 분석도구로 사용되어지고 있으며, 이를 오픈소스로 개발하여 제공하고 있는 OpenCOBRA가 최근에 공개되었다. 이와 더불어 OptFlux도 꾸준한 기능개선으로 연구에 활용되어지고 있는 실정이다[24].

2.3.3 불리언 네트워크 분석

위에서 언급된 분석방법은 수치분석을 통한 분석을 기반한 연구이다. 하지만 실제적으로 세포내의 반응속도상수를 얻는 일은 많은 시간과 노력을 필요로 한다. 최근 마이크로어레이 실험 기술이 발달하고 세포 내의 전체적인 유전자 발현 양상에 대한 데이터가 축적되면서 이러한 발현 데이터를 기반으로 네트워크를 생성 분석할 수 있는 기법들이 연구되고 있다. 이러한 네트워크의 특징은 정확한 속도 상수를 예상할 수는 없지만 유전자간의 활성화 및 비활성화에 대한 조절 형태는 알 수 있게 된다. 이렇게 구성된 네트워크를 이용하여 단순한 활성화/비활성화 상태를 활용하여 전체 네트워크의 특성을 분석할 수 있다.

불리언 네트워크 분석은 연구자가 문헌이나 실험을 통해 얻은 생물학적 상호작용의 정보를 활용하여 새로운 생물학적 모델을 생성한다. 이때 각 노드간의 연결은 활성화와 억제 2가지 특성으로 기술할 수 있으며, 하위의 노드는 상위의 노드의 신호에 대한 정보를 받아들이어 자신이 활성화될 것인지 비활성화될 것인지를 결정한다. 이때 각 노드(단백질)는 불리언 룰을 정의하여 자신의 전이를 정의할 수 있다. 이렇게 정의된 각각의 노드는 외부 자극에 의해 시작되는 신호가 어떻게 전달되는지를 분석할 수 있다[25].

2.4 세포반응 모델링 도구

2.4.1 CellDesigner

CellDesigner는 세포내의 다양한 신호 전달 과정을 디자인하기 위해 개발되어진 도구이다[4]. 그림 3은 CellDesigner를 이용하여 신호 전달 과정을 모델링하는

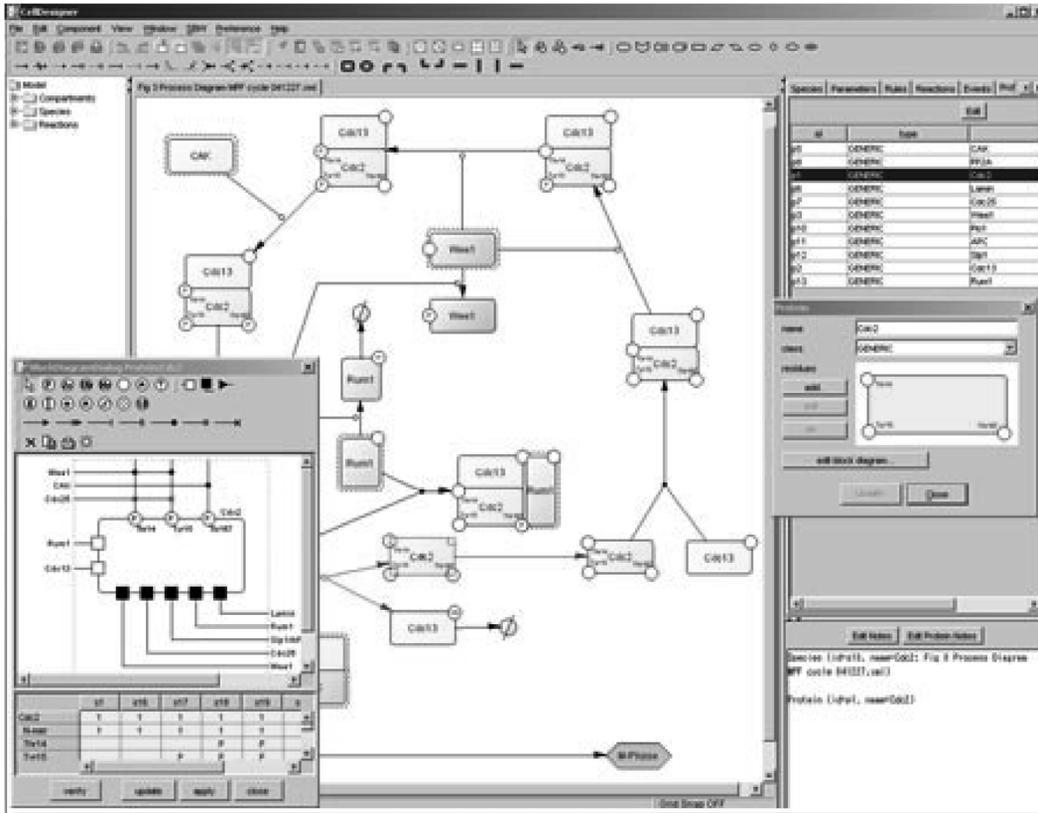


그림 3 CellDesigner 실행화면[4]

화면이다. 대표적인 특징은 생화학적 의미를 표현하고, 단백질의 상태를 상세히 기술한다. 신호 전달 과정을 표현하기 위해 SBML(Systems Biology Markup Language) 자료를 이용한다. 세포 내의 신호 전달 과정에 참여하는 다양한 종류의 단백질들을 서로 다른 모양으로 세분화하여 표시한다. 연구자들이 직관적으로 이해할 수 있는 다이어그램을 제공하는 것이 목적이다. 간단한 분석을 위해 구현된 모델을 상미분방정식을 이용하여 분석할 수 있는 기능을 제공한다.

2.4.2 Virtual Cell

Virtual Cell은 사용자가 세포 구조, 분자, 반응 파라미터 등에 대한 정보를 입력하면, 모델링 언어를 사용하여 생화학 반응을 상미분 방정식 또는 편미분방정식의 수학적 시스템으로 자동적으로 변환한다[26]. Swiss-Prot, KEGG 데이터베이스로부터 얻은 단백질, 화합물, 효소에 대한 정보를 모델과 연결하거나, KEGG 데이터베이스의 반응 정보 등 외부의 대사 경로 데이터베이스를 생화학 모델 구축에 사용할 수 있다. 또한, Bio-Models 데이터베이스에 있는 생물학 모델들도 Virtual Cell을 이용한 분석이 가능하다. 생리학적 모델 구성을 위한 입력 데이터로는 분자(유전자, 단백질, 화합물, ATP 등), 반응(효소 동역학, 수용체 결합, 세포막 유

동 등), 구조(소포체, 세포기질 등)에 대한 정보와 관련된 데이터들이 이용된다.

Virtual Cell을 통해 시간열(time scale), 민감도(sensitivity), 정상 상태(steady state)에 대한 정보를 얻을 수 있으며, 신경세포의 기하학 이미지와 함께 분자의 발현 농도 차이를 색의 변화로 관찰할 수 있다. 시뮬레이션을 실행한 결과, 시간과 위치에 따른 농도 변화를 표나 이미지로 얻을 수 있는데, 실험 데이터의 포맷과 같기 때문에 직접적인 비교 분석이 가능하다.

2.4.3 ezBioNet

ezBioNet은 신호전달/대사 네트워크 등 다양한 생물학적 반응네트워크를 가시적으로 모델링할 수 있으며 생성된 모델을 원격의 분석시스템을 이용하여 시뮬레이션 할 수 있는 분석도구이다[27]. 이 시스템은 클라이언트-서버구조로 개발되어졌으며, ezBioNet클라이언트는 이클립스 RCP프레임워크를 기반으로 개발되어졌다. 가장 큰 특징은 하나의 모델링도구를 활용하여 상미분방정식 분석과 블리언 네트워크분석을 선택적으로 실행할 수 있다. 또한 서버는 KISTI에 구축된 클러스터를 활용하여 원격 분석이 가능하며, 클라이언트와 서버간의 통신은 OPAL시스템을 활용하고 있다[28]. 그림 4는 ezBioNet 클라이언트 실행화면으로 분석결과

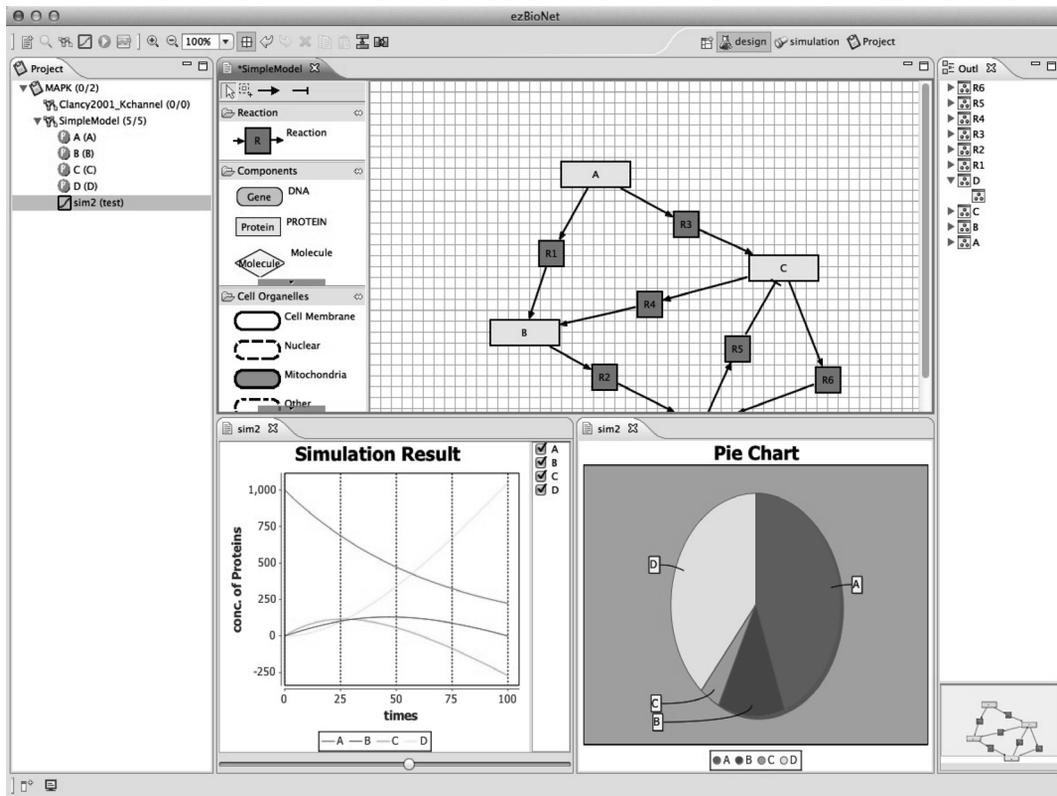


그림 4 ezBioNet 실행화면

에 대한 시계열 그래프를 제공하여 분석결과를 해석할 수 있도록 지원하고 있다.

수 있을 것으로 기대된다.

3. 결론

최근 생물학 관련 실험방식이 대량화하고 정밀해짐에 따라 생성되는 대량 실험 자료를 분석하고 이를 기반으로 세포의 기능분석을 수행할 수 있는 시스템생물학 기술의 요구가 증가하고 있다. 특히 단백질간의 상호작용을 관측할 수 있는 다양한 실험방법과 이에 대한 데이터가 계속적으로 증가하고 있으며 이와 관련된 데이터베이스도 속속 등장하고 있는 실정이다. 현재까지 다양한 표준화의 노력이 진행 중에 있다. 향후 단백질-단백질 상호작용 정보를 기반으로 대량의 생물학적 네트워크를 구성하고 이를 시뮬레이션 할 수 있는 시스템이 등장할 것이다. 특히 시스템 생물학 분석을 위해서 생물학적 모델로 생성하고 이를 테스트하여 실제 생명현상과 비교해 볼 수 있는 다양한 모델링 시스템과 시뮬레이션 시스템이 제공될 것으로 예상된다. 또한 최근 합성생물학의 등장과 더불어 산업적으로 유용성이 높은 미생물의 개발 등 산업적인 활용에 초점을 맞춘 다양한 연구가 진행되면서 다양한 오믹스 정보를 통합하여 다차원의 생물학적 현상을 통합적으로 시뮬레이션 할

참고문헌

- [1] M. Kanehisa and P. Bork, "Bioinformatics in the post-sequence era," Nat. Genet., Vol. 33, pp. 305-310, Mar. 2003.
- [2] H. Kitano, "Systems biology: a brief overview," Science, Vol. 295, No. 5560, pp. 1662-1664, Mar. 2002.
- [3] M. Tomita, K. Hashimoto, K. Takahashi, T. S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C. Venter, and C. A. Hutchison, "E-CELL: software environment for whole-cell simulation," Bioinformatics, Vol. 15, No. 1, pp. 72-84, Jan. 1999.
- [4] A. Funahashi, Y. Matsuoka, A. Jouraku, M. Morohashi, N. Kikuchi, and H. Kitano, "CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks," presented at the Proceedings of the IEEE, 2008, Vol. 96, No. 8, pp. 1254-1265.
- [5] D. P. Wall, P. Kudtarkar, V. A. Fusaro, R. Pivovarov, P. Patil, and P. J. Tonellato, "Cloud computing for comparative genomics," BMC Bioinformatics, Vol. 11, p. 259, 2010.
- [6] S. Fields and O. Song, "A novel genetic system to de-

- tect protein-protein interactions,” *Nature*, Vol. 340, No. 6230, pp. 245-246, Jul. 1989.
- [7] E. M. E. Phizicky and S. S. Fields, “Protein-protein interactions: methods for detection and analysis,” *Microbiol Rev*, Vol. 59, No. 1, pp. 94-123, Feb. 1995.
- [8] T. K. Kerppola, “Design and implementation of bimolecular fluorescence complementation(BiFC) assays for the visualization of protein interactions in living cells,” *Nat Protoc*, Vol. 1, No. 3, pp. 1278-1286, Oct. 2006.
- [9] N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, and M. Hucka, “BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems,” *Nucleic Acids Research*, Vol. 34, No. Database issue, pp. D689-91, Jan. 2006.
- [10] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Research*, Vol. 27, No. 1, pp. 29-34, Jan. 1999.
- [11] M. Krummenacker, S. Paley, L. Mueller, T. Yan, and P. D. Karp, “Querying and computing with BioCyc databases,” *Bioinformatics*, Vol. 21, No. 16, pp. 3454-3455, Aug. 2005.
- [12] G. D. Bader, I. Donaldson, C. Wolting, B. F. Ouellette, T. Pawson, and C. W. Hogue, “BIND-The Biomolecular Interaction Network Database,” *Nucleic Acids Research*, Vol. 29, No. 1, pp. 242-245, Jan. 2001.
- [13] I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte, and D. Eisenberg, “DIP: the database of interacting proteins,” *Nucleic Acids Research*, Vol. 28, No. 1, pp. 289-291, Jan. 2000.
- [14] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, “BioGRID: a general repository for interaction datasets,” *Nucleic Acids Research*, Vol. 34, No. Database issue, pp. D535-9, Jan. 2006.
- [15] T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, and A. Pandey, “Human Protein Reference Database-2009 update,” *Nucleic Acids Research*, Vol. 37, No. Database issue, pp. D767-72, Jan. 2009.
- [16] S. Orchard, S. Kerrien, S. Abbani, B. Aranda, J. Bhate, S. Bidwell, A. Bridge, L. Briganti, F. S. L. Brinkman, F. Brinkman, G. Cesareni, A. Chatr-aryamontri, E. Chautard, C. Chen, M. Dumousseau, J. Goll, R. E. W. Hancock, R. Hancock, L. I. Hannick, I. Jurisica, J. Khadake, D. J. Lynn, U. Mahadevan, L. Perfetto, A. Raghunath, S. Ricard-Blum, B. Roechert, L. Salwinski, V. Stümpflen, M. Tyers, P. Uetz, I. Xenarios, and H. Hermjakob, “Protein interaction data curation: the International Molecular Exchange(IMEx) consortium,” *Nat. Methods*, Vol. 9, No. 4, pp. 345-350, Apr. 2012.
- [17] E. Demir, M. P. Cary, S. Paley, K. Fukuda, C. Lemer, I. Vastrik, G. Wu, P. D'Eustachio, C. Schaefer, J. Luciano, F. Schacherer, I. Martinez-Flores, Z. Hu, V. Jimenez-Jacinto, G. Joshi-Tope, K. Kandasamy, A. C. Lopez-Fuentes, H. Mi, E. Pichler, I. Rodchenkov, A. Splendiani, S. Tkachev, J. Zucker, G. Gopinath, H. Rajasimha, R. Ramakrishnan, I. Shah, M. Syed, N. Anwar, O. Babur, M. Blinov, E. Brauner, D. Corwin, S. Donaldson, F. Gibbons, R. Goldberg, P. Hornbeck, A. Luna, P. Murray-Rust, E. Neumann, O. Reubenacker, M. Samwald, M. van Iersel, S. Wimalaratne, K. Allen, B. Braun, M. Whirl-Carrillo, K.-H. Cheung, K. Dahlquist, A. Finney, M. Gillespie, E. Glass, L. Gong, R. Haw, M. Honig, O. Hubaut, D. Kane, S. Krupa, M. Kutmon, J. Leonard, D. Marks, D. Merberg, V. Petri, A. Pico, D. Ravenscroft, L. Ren, N. Shah, M. Sunshine, R. Tang, R. Whaley, S. Letovksy, K. H. Buetow, A. Rzhetsky, V. Schachter, B. S. Sobral, U. Dogrusoz, S. McWeeney, M. Aladjem, E. Birney, J. Collado-Vides, S. Goto, M. Hucka, N. Le Novère, N. Maltsev, A. Pandey, P. Thomas, E. Wingender, P. D. Karp, C. Sander, and G. D. Bader, “The BioPAX community standard for pathway data sharing,” *Nat. Biotechnol.*, Vol. 28, No. 9, pp. 935-942, Sep. 2010.
- [18] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J.

- Wagner, J. Wang, SBML Forum, "The systems biology markup language(SBML): a medium for representation and exchange of biochemical network models," *Bioinformatics*, Vol. 19, No. 4, pp. 524-531, Mar. 2003.
- [19] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson, "Integrative analysis of cell cycle control in budding yeast," *Mol. Biol. Cell*, Vol. 15, No. 8, pp. 3841-3862, Aug. 2004.
- [20] U. M. Ascher and L. R. Petzold, *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM: Society for Industrial and Applied Mathematics, 1998.
- [21] B. D. Heavner, K. Smallbone, B. Barker, P. Mendes, and L. P. Walker, "Yeast 5-an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network," *BMC Syst Biol*, Vol. 6, p. 55, 2012.
- [22] P. Mendes, "GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems," *Comput. Appl. Biosci.*, Vol. 9, No. 5, pp. 563-571, Oct. 1993.
- [23] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer, "COPASI-a COMplex PATHway Simulator," *Bioinformatics*, Vol. 22, No. 24, pp. 3067-3074, Dec. 2006.
- [24] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, S. Soares, J. P. Pinto, J. Nielsen, K. R. Patil, E. C. Ferreira, and M. Rocha, "OptFlux: an open-source software platform for in silico metabolic engineering," *BMC Syst Biol*, Vol. 4, p. 45, 2010.
- [25] I. Albert, J. Thakar, S. Li, R. Zhang, and R. Albert, "Boolean network simulations for life scientists," *Source Code Biol Med*, Vol. 3, No. 1, p. 16, 2008.
- [26] L. M. Loew and J. C. Schaff, "The Virtual Cell: a software environment for computational cell biology," *Trends in Biotechnology*, Vol. 19, No. 10, pp. 401-406, Oct. 2001.
- [27] S. J. Yu, T. Q. Tung, J. Park, J. Lim, and J. Yoo, "ez-BioNet: A modeling and simulation system for analyzing biological reaction networks" *Journal of the Korean Physical Society*, 2012.
- [28] J. Ren, N. Williams, L. Clementi, S. Krishnan, and W. W. Li, "Opal web services for biomedical applications," *Nucleic Acids Research*, Vol. 38, No. WebServer, pp. W724-W731, Jun. 2010

약 력



유 석 종

1995 충북대학교 생화학 (학사)
 1997 충북대학교 생화학 (석사)
 2013 충북대학교 생물정보학 (박사)
 2003~현재 한국과학기술정보연구원 책임연구원
 관심분야: 생물정보학, 시스템생물학, 분산시스템
 E-mail : codegen@kisti.re.kr



유 재 수

1989 전북대학교 컴퓨터공학과 (공학사)
 1991 KAIST 전산학과 (공학석사)
 1995 KAIST 전산학과 (공학박사)
 1996~현재 충북대학교 정보통신공학부 및 컴퓨터
 정보통신연구소 교수
 관심분야: 데이터베이스시스템, 생물정보, 센서
 네트워크, 멀티미디어 데이터베이스, 클라우드컴퓨팅, 빅데이터
 E-mail : yjs@chungbuk.ac.kr