



개의 미분방정식을 가진다. 상미분 방정식으로의 표현은 널리 사용되는 방법이지만, 생물학적 모델에서는 종종 대수적 제한 조건들을 가지므로 미분-대수 방정식이 필요하다.

- 미분-대수 방정식(differential algebraic equations) : 모델의 성분 농도에 관한 제한조건이 있을 때 사용한다. 상미분 방정식 체제를 포함하며, 대상 모델을 구축하는 연구자가 표현하고자 하는 구성을 지원할 수 있으므로 연구자들의 요구에 더 잘 들어맞는다. 미분-대수 방정식을 이용한 도구(solver)는 상미분 방정식을 이용한 것보다 구현하기가 더 힘들고, 따라서 많은 모사 도구들이 미분-대수 방정식을 제한적으로 지원한다.[5]
- 편미분 방정식(partial differential equations) : 시간과 공간 변수들을 모두 가지는 공간에서의 확산 모델처럼, 시스템에 두 개 이상의 독립 변수들이 있을 때 편미분 방정식을 이용한다. 편미분 방정식을 이용한 도구는 상미분이나 미분-대수 방정식을 이용하는 것보다 구현하거나 제대로 사용하기가 훨씬 더 힘들어서, 편미분 방정식의 틀을 사용하는 소프트웨어들은 매우 적다.

## 2.2. 확률적 접근

많은 연구들이 생물계에서의 확률적 변동과 잡음의 발생에 대해 보고하였다. 각 세포에서 유전자 발현을 관찰하면 전사의 확률적 본성을 잘 보여준다. 메신저 RNA는 비연속적이고 불특정한 펄스로 생산되며, 단백질은 연속적인 방법이 아니라 임의적인 시간간격으로 짧은 '폭발'을 통해 생산된다고 알려져 있다. 또한 조절 네트워크의 전개과정에서, 비슷한 초기 조건도 질적으로 다른 결과를 가져오게 된다.

세포 공정의 기능의 아주 작은 단위에서 DNA, mRNA, 단백질, 작은 분자들 간의 상호작용은 물리의 법칙을 따른다. 한편 이론통계물리에서  $\sqrt{n}$  법칙은, 시스템 크기의 지표로 여겨지는 입자 개수가 커질수록 어떤 시스템에서의 불특정성과 변동이 작아진다는 것을 말한다. 입자수가 작고 농도가 열으면, 대개 열적 진동이 상대적으로 높은 진동으로 나타나게 된다. 유전자 전사, 조절 작용, 신호 전달 등의 과정에 참여하는 생화학적 참여자들은 종종 그 수가 작으며, 그 결과, 기본적인 반응들이 상대적으로 넓은 반응시간대에 걸쳐 발생한다. 생화학적 상호작용의 고유한 성질 때문에 발생하는 이러한 확률적 효과는 종종 '내재적 잡음'으로 불린다.

유전자 발현에는 외재적 성분의 잡음도 존재한다. 이것들은 외부의 환경 조건 때문이다. 예를 들어, 주어진 어떤 유전자의 전사 요인은 종종 또 다른 유전자의 단백질 생산물이며, 따라서 그것의 생산 역시 임의적이다. 유전자가 확률적으로 야기한 단백질 생산이 또 다른 유전자의 교체로 이어지는 그러한 상황은 확률적 사건의 연속적 발생으로 나타난다. 여기서, 자극의 시간 간

격은 방대하게 다양한 결과를 나타내게 된다. 유전자의 발현에 정확한 자극 경계가 존재하는가, 유전자 발현에 따른 지역적 분화들에 어떤 전체적 기준이 있는가는 여전히 명확하지가 않다. 세포 공정에서 이 잡음이 암시하는 것은, 초기 조건이 동일할 때에도 시간이 지남에 따라 세포가 발달되는 경로가 달라져, 서로 다른 세포가 될 수 있다는 것이다. 어떤 한 종의 개체수를 측정해보면, 같은 유전 물질을 가진 세포에서도 같은 유전자에서의 발현 정도가 서로 심하게 달라진다는 것을 알 수 있다. 동질 유전자 개체 사이에서의 그러한 가변성의 근원은 대개 확률적 현상으로 설명된다.

확률성은 분자의 수가 작아질수록 더 현저하게 나타난다. 반응 물질의 농도가 증가할수록, 확률적 성질이 덜 두드러지게 나타나고 결정론적 해를 나타내는 경향이 있다. 이 사실은 계산의 복잡성을 생각할 때 중요한 고려사항이 되는데, 확률적인 모사를 하는 것이 결정론적 상미분 방정식의 모사보다 더 많은 시간이 걸리기 때문이다.

## 2.3 혼합형(hybrid) 알고리즘

혼합형 알고리즘(hybrid algorithm)은 시스템의 큰 관점에서의 결정론적 접근과, 미세한 관점에서의 확률적 접근을 모두 지원한다. 이러한 특징은 시스템에 급작스런 변화를 가져오는 불연속성을 가진 실제 생물 시스템의 거동을 포착하는 데 유용하다. 불연속적 변화로 연결된 세포주기모델, 불빛을 깜박이는 해저미생물인 비브리오 피케리(Vibrio fischeri)의 발광현상 등이 그 대표적인 예이다.

혼합형 알고리즘은 규모가 큰 문제를 모델링하고 모사하는데 드는 계산 비용을 상당히 줄일 수 있다. 그러나 알고리즘에 추가로 수적인 처리를 하게 되므로 더 많은 모수(parameter)들이 정의되어야 하고 해의 정확도가 모수에 크게 의존한다는 것이 중요한 단점이다. 따라서, 결과의 정확성, 시스템의 각 단계들 사이에서 모수의 항상성, 한 시스템을 하위 시스템의 불연속 집합으로 나누는 방법론과 그 하위 시스템들 간의 비선형 관계 등이 실제 적용 이전에 해결되어야 할 문제들이다.

## 3. 모델링 언어

### 3.1. SBML (Systems Biology Markup Language)

대용량 처리장치와 같은 고속 실험 도구의 발전으로 복잡하고 거대한 실험 데이터가 생성됨으로써 생물학자들이 컴퓨터모델링의 필요성을 인식하게 되었다. 복잡한 대상 모델을 서로 공유하기 위해서는 표준이 되는 형식이 절실하게 요구되었다. 2000년 일본의 게이오 대학의 키타노 교수 팀을 주축으로 소프트웨어 개발자와 모델을 구축하는 연구원들이 공동 연구하여, 대용량의

데이터를 빠르게 처리 할 수 있는 XML(Extension Markup Language)과 UML(Unified Modeling Language) 기반으로 Systems Biology Markup Language (SBML)을 개발하였다. 현재 85개의 프로그램에서 SBML을 사용하고 응용하고 있으며, 대표적인 예로 한국과학기술원 박선원 교수와 이상엽 교수가 개발한, 대사과정의 각 경로마다 생분자 유량을 계산해 내는 MetaFluxNet [12] 등이 있다.

기초적인 생화학적 네트워크 모델의 특성을 표현할 수 있는 Level 1에서, 세부적인 구조를 보강하고 동역학 방정식을 MathML로 표현하여, 복잡한 식을 손쉽게 사용자가 원하는 형태로 전환할 수 있게 하는 Level 2 를 발표 하였다. 그 후 계층적 모델과 네트워크 다이어그램을 보여 줄 수 있고 모델의 공간적인 특성을 잘 표현할 수 있는 Level 3 를 개발 중이다. 그리고 또한, 이를 기반으로 둔 Systems Biology Workbench (SBW)를 개발하여 아래와 같은 여러 언어와 호환성을 위한 도구들을 제공하고 있다.

표 1. SBML을 통해서 개발자가 응용할 수 있는 도구.

도 구	용 도
LibSBML	SBML 사용하기 위한 소프트웨어에 탑재 하여 SBML로 표현된 데이터를 이용할 수 있게 한다.
MathSBML	SBML으로 표현된 데이터를 Mathematica에서 처리 할 수 있도록 한다.
SBMLToolBox	SBML으로 표현된 데이터를 MATLAB에서 처리할 수 있도록 한다.
KEGG2SBML	KEGG 데이터베이스의 자료를 SBML 전환하여 유용하게 응용할 수 있도록 한다.
CellML2 SBML	CellML로 표현되어 있는 데이터를 SBML로 전환한다.

#### 4. 동적 모사 도구

##### 4.1. E-Cell

E-Cell은 일본 게이오 대학에서 개발한 가상세포이다. (<http://www.e-cell.org>) E-cell은 현재로서는 mRNA 전사, 단백질 해독, 해당 작용(glycolysis)에 의한 에너지 생성, 인지질 합성 등 세포의 생존에 필수적인 몇 가지 기능을 중심으로 구현하고 있다. 이 모형의 구축에 있어서 장애 요인은 각 대사 과정에 들어가는 모든 화학 반응식에 대한 자료가 불충분하다는 것이다. 예를 들면, 복잡한 화학반응 네트워크를 나타낼 때, 이것을 기술하기 위해서는 촉매로 매개되는 모든 화학 반응식을 수식화하여야 하고, 또한 이에 필요한 반응계수, 화학물질의 농도 등이 주어져야 한다. 현재로서는 일부분 알려진 자료를 이용하여 E-cell 모형을 구축하고, 시뮬레이션 분석 결과와 실제 실험

결과를 비교함으로써 모형에 사용하는 변수들을 고정시켜 나가는 작업이 진행 중이다.

최근 E-Cell 버전 3을 발표하면서 크고 복잡한 생물학적 세포 시스템의 분석과 모델링, 시뮬레이션을 위하여 객체 지향적으로 발전하였다. 또한, 결정론적 (ODE) 방법론과 확률적 (gillespie) 방법, 세포 자동자(cellular automata) 등을 이용하여 서로 다른 시간열(timescales)을 가지는 하부 시스템으로 구성된 세포 모형의 효율적인 모사를 다룰 수 있도록 설계되었으며, 이를 통하여 heat-shock model에 적용한 예를 발표하였다. XML을 기반으로 하는 E-cell Model description Language(EML)을 자체적으로 정의하였고, SBML도 지원한다. 공개 소스 프로젝트로서 누구나 재분배나 수정이 가능하다.

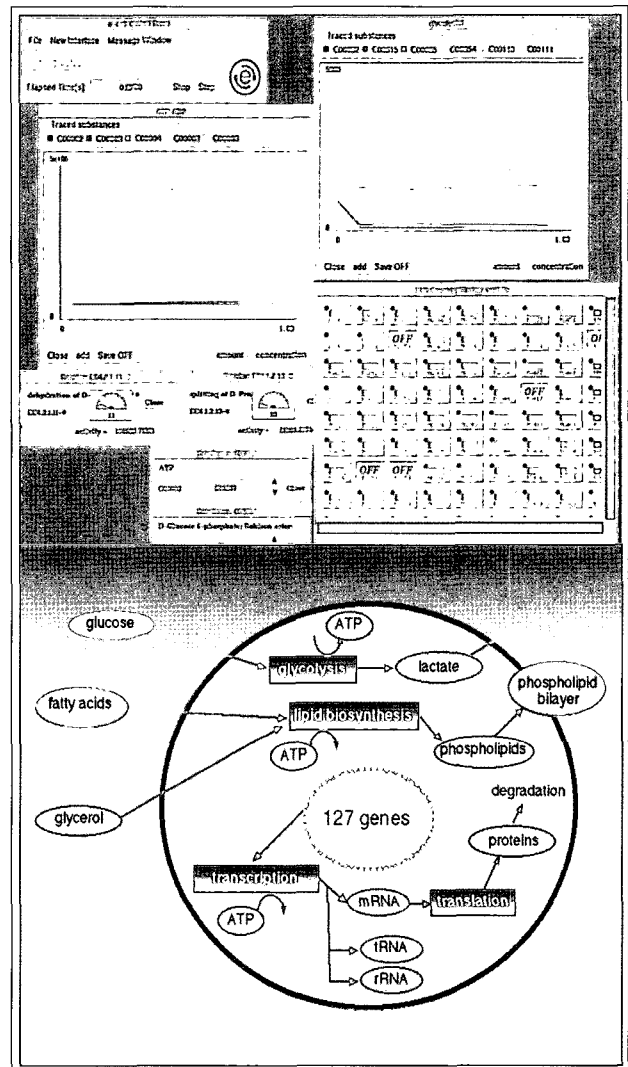


그림 2. M. genitalium 원핵세포 모사를 위한 모형과 E-Cell을 통한 동적모사.

#### 4.2. Gepasi/COPASI

세포 내 동적 대사과정을 중심으로 다루는 E-cell project와는 달리, Gepasi는 임의적 평면/입체 공간에서의 세포 간 또는 세포내 대류, 반응, 확산 현상을 모델링/모사하는데 목적을 둔다. 따라서 모델식들은 편미분 방정식과 상미분 방정식 등으로 표현된다.

버지니아 생물 정보학 연구소(<http://www.vbi.vt.edu>)는 실험 데이터에 부합되는 대사반응식 계수 추정을 위해 Gepasi(<http://www.gepasi.org/>)라는 프로그램을 개발하였다. Gepasi는 모델에 대한 최적화, 대사조절분석, 선형 안정성 분석 등을 제공하고, 모델에 대한 SBML(Systems Biology Makeup Language)을 통하여 다른 패키지와 연동이 가능하다.

하지만, 결정적인 결함의 발견으로 곧 이 프로그램의 후속편인 COPASI (Complex Pathway Simulator)를 공개 프로그램으로 배포할 예정이다. 버지니아 생물정보학 연구소와 독일의 유럽 미디어 연구소의 과학자들이 공동으로 참여하여 COPASI 프로젝트를 통해 모사뿐만 아니라 분석 도구를 함께 겸비한 소프트웨어를 개발하였다. 미국 측에서는 페드로 멘데스 박사와 독일 측에서는 우르술라 쿨머가 COPASI 연구를 주도 하였다.

COPASI는 추계적, 결정론적인 시간열(timescales) 분석이 가능하다. 또한, 정상상태 분석과 대사조절분석, 질량보존분석 등이 가능하고, 최근 발표한 버전에서는 분포(distribution)의 예측에 유용한 모수 탐색 기능을 추가 하였다. Gepasi와 COPASI는 각각 SBML Level 1, SBML Level 1,2 와 연동 가능하지만 자체적으로 XML을 기반으로 한 COPASI 파일 형식도 가지고 있다.

#### 4.3. INSILICO discovery

독일의 INSILICO 바이오테크놀로지(<http://www.insilico-biotechnology.com>)는 대사 경로 분석용 다양한 분석기법을 제공하고 사용자 환경이 뛰어나나 비교적 속도가 느리다. 특징으로는 표준 SQL과 XML을 이용하여 데이터베이스 통합을 용이하게 할 수 있는 인터페이스가 구성되어 사용자가 외부 데이터베이스의 적용과 응용이 가능하다. 대사흐름분석이나 동적 시뮬레이션 기능과 복잡한 생물학적 모델의 빠진 구성요소와 반응, 의미 없는 대사경로 탐색하여 대사 경로의 초고 생산 수율 및 다중 제저 변이와 신경 전달의 계층적 반응에 대한 예측 등이 가능하다.

#### 4.4 Virtual Cell

미국 NRCAM(The National Resource for Cell Analysis and Modeling, <http://www.ncam.uchc.edu>)의 Virtual Cell은 세포내의 일어나는 Ca<sup>++</sup>의 활동 분석에 매우 유용한 시뮬레이션 시스템으로서 자바 애플릿 환경으로 되어 있어 웹에

서 서비스한다는 강점이 있다. 자바 애플릿으로 구현되어 웹 환경에서 사용자가 응답 속도에 대한 불편함을 느낄 수 있다. 그러나 세포를 기하 이미지로 보여 줄 수 있으며 시공간적 시뮬레이션이 가능한 상미분 방정식과 편미분 방정식의 모사와 지역 모수에 대한 시스템 차원의 민감도 분석, 최적 모수를 찾을 수 있는 기능들이 가능하다. 최근 사용자들이 모델 데이터베이스 서버에 자신의 모델이나 정보를 저장하고 다른 사용자와 공유함으로써 인터넷 망을 이용한 이점을 살리고 있다.

표 2. 동적 모사 환경구축 연구 동향.

•\*: 웹기반, s: SBML

시스템	운영체제	Mathematical Framework	사이트
ECell 3 <sup>s</sup>	Linux	DAE, Hybrid, Stochastic(Gillespie, Gibson)	<a href="http://www.e-cell.org">http://www.e-cell.org</a>
Gepasi <sup>s</sup>	Linux, Windows	DAE	<a href="http://www.gepasi.org/">http://www.gepasi.org/</a>
SBW <sup>s</sup>	Windows	DAE, Hybrid, Stochastic(Gillespie, Gibson, Tau-Leap)	<a href="http://sbw.sourceforge.net/">http://sbw.sourceforge.net/</a>
Virtual Cell <sup>*s</sup>	Linux, Windows	DAE, PDE	<a href="http://www.ncam.uchc.edu">http://www.ncam.uchc.edu</a>
JWS <sup>**</sup>	Linux, Windows	DAE	<a href="http://ijj.biochem.sun.ac.za/index.html">http://ijj.biochem.sun.ac.za/index.html</a>
OBIYagns <sup>*</sup>	Linux, Windows	ODE, Genetic Algorithm	<a href="https://access.obigrid.org/yagns/">https://access.obigrid.org/yagns/</a>

### 5. 결론

복잡한 생물 시스템의 계산을 위한 모델링은 이용 가능한 대사 회로 정보를 조직하고 통합하며, 올바른 실험 계획을 하는 데 아주 중요하다. 그러나 아직 생물학적 모사의 예측 능력은 조절작용과 동역학에 대한 불충분한 지식과 정보에 의해 제한되고 있는데, 다중 시간열(multiple timescales), 다중 수준(multiple levels)의 문제 등 시스템 모델링과 모사 분야에서 가장 도전적인 문제들이다. [18]

세포 전체를 모사하는 가상세포가 완성되면, 유전공학적인 작업을 컴퓨터상에서 수행할 수 있게 되어, 독성 실험이나 약품 설계 등에 이용할 수 있다. 세포대사나 복잡한 네트워크의 특성을 파악하는 데 사용할 수 있고, 유전자 표현형의 예측 등을 더욱 정확하게 할 수 있다.

최근 우리 연구실에서는 국내의 인터넷 인프라를 최대한 이용하여서, 연구자들 간의 정보 공유를 통해 연구의 성과를 크게 향상시킬 수 있는, 웹 기반 가상세포의 개발에 힘쓰고 있다.

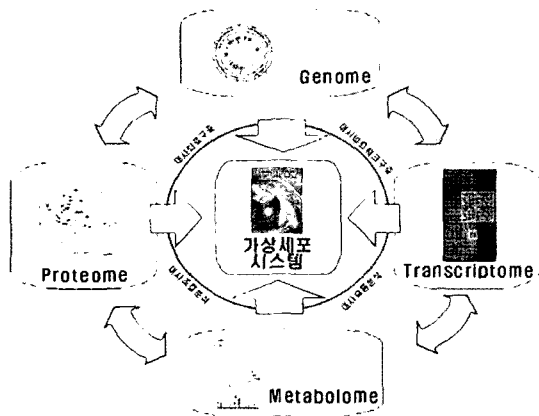


그림 3. 가상세포 시스템의 관련 연구 분야.

### 참고문헌

[1] O. Wolkenhauer, H. Kitano and K. -H. Cho, "Systems biology: Looking at opportunities and challenges in applying systems theory to molecular and cell biology", *IEEE Control Systems Magazine*, vol. 23, no. 4, pp. 38-43, 2003.

[2] S. Y. Lee, D. Y. Lee, and T. Y. Kim, "Systems biotechnology for strain improvement", *Trends in Biotechnology* (in press)

[3] I. Goryanin, T. C. Hodgman and E. Selkov, "Mathematical simulation and analysis of cellular metabolism and regulation", *Bioinformatics*, vol. 15, pp. 749-758, 1999.

[4] A. Pettinen, T. Aho, O. Smolander, T. Manninen, A. Saarinen, K. Taattola, O. Yli-Harja and M. Linne, "Simulation tools for biochemical networks: evaluation of performance and usability", *Bioinformatics*, vol. 21, pp. 357-363, 2005.

[5] U. M. Ascher and L. R. Petzold, "Computer methods for ordinary differential equations and differential-algebraic equations", *SIAM*, 1998.

[6] T. C. Meng, S. Somani and P. Dhar, "Modelling and Simulation of Biological Systems with Stochasticity", *In Silico Biology*, vol. 4, 0024, 2004.

[7] Li and H. Rabitz, "A general analysis of approximate lumping in chemical kinetics", *Chemical Engineering Science*, vol. 45, no. 4, pp. 977-1002, 1990.

[8] N. Novere and T. S. Shimizu, "Stochsim: Modeling

of Stochastic Biomolecular Processes," *Bioinformatics*, vol. 6, no. 6, pp. 575-576, June 2001.

[9] T. R. Kiehl, R. M. Mattheyses and M. K. Simmons, "Hybrid simulation of cellular behavior", *Bioinformatics*, vol. 20, no. 3, pp. 316-322, 2004.

[10] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, and H. Kitano, "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models", *Bioinformatics*, vol. 19, pp. 524-531, 2002.

[11] M. Hucka, A. Finney, B. J. Bornstein *et al.*, "Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project", *Systems Biology*, vol. 1, no. 1, 2004.

[12] D. Y. Lee, H. Yun, S. Park and S. Y. Lee, "MetaFluxNet: the management of metabolic reaction information and quantitative metabolic flux analysis", *Bioinformatics*, vol. 19, no. 16, pp. 2144-2146, 2003.

[13] M. Tomita, K. Hashimoto, K. Takahashi, T. S. Shimizu, Y. Matsuzaki, F. Miyoshi *et al.*, "E-CELL: software environment for whole-cell simulation", *Bioinformatics*, vol. 15, pp. 7284, 1999.

[14] P. Mendes, "Biochemistry by numbers: simulation of biochemical pathways with Gepasi3", *Trends Biochem. Sci.*, vol. 22, pp. 361-363, 1997.

[15] 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.

[16] J. L. Snoep and B. G. Oliver, "JWS Online Cellular Systems Modelling and Microbiology", *Microbiology Comment*, July 2003.

[17] S. Kimura, T. Kawasaki, M. Hatakeyama, T. Naka, F. Konishi and A. Konagaya, "OBIYagns: a grid-based biochemical simulator with a parameter estimator", *Bioinformatics*, vol. 20, no. 10, pp. 1646-1648, 2004.

[18] S. Okino, M. S. and Michael L. Mavrovouniotis, "Simplification of Mathematical Models of Chemical Reaction Systems", *Chemical Reviews*, vol. 98, no. 2, 1998.

..... 저자약력 .....



《박 신 원》

- 1970년 서울대학교 화학공학과 학사.
- 1983년 Oklahoma State University 화학공학과 석사.
- 1979년 University of Texas at Austin 화학공학과 박사.
- 1979년 University of Houston - Clear Lake (MBA).
- 1988년~현재 한국과학기술원 생명화학공학과 교수.
- 관심분야 : 공정제어, 공정설계, 공정최적화, SCM, LCA, 시스템생물학, 생물정보학.



《이 상 엽》

- 1986년 서울대학교 화학공학과 학사.
- 1987년 Northwestern University 화학공학과 석사.
- 1991년 Northwestern University 화학공학과 박사.
- 2002년~현재 한국과학기술원 생명화학공학과 교수.
- 관심분야 : 대사공학, 재조합 미생물공학, 생물정보학, 바이오칩 (DNA 칩 / 단백질 칩), in silico 바이오테크놀로지, 미생물 배양공학, 미생물 이용 일/이차 대사산물 생산, 생물 전환공학, 단백질 발현 및 생산 제반기술, 생물고분자.



《이 동 엽》

- 1998년 연세대학교 화학공학과 학사.
- 2000년 한국과학기술원 화학공학과 석사.
- 2004년 한국과학기술원 생명화학공학과 박사.
- 2005년~현재 싱가포르 국립대학교 조교수.
- 관심분야 : 시스템 생물학 / 생명공학, 신약 및 질병 모델링, 공급사슬망 최적화.



《윤 조 중 언》

- 2001년 영남대학교 화학공학과 학사.
- 2003년 영남대학교 화학공학과 석사.
- 2003년~현재 한국과학기술원 생명화학공학과 박사과정.
- 관심분야 : 공정시스템 및 시스템 생물학.



《김 영》

- 2003년 한국과학기술원 생명화학공학과 학사.
- 2003년~현재 한국과학기술원 생명화학공학과 박사과정.
- 관심분야 : 공정시스템 및 시스템 생물학.