

# 인터넷 환경을 통한 생물학적 모델의 정량적 분석

## Quantitative Analysis of Biological Models under the Internet Environment

윤좌문, 이동엽, 조아연, 이상엽, 박선원\*  
 (Choamun Yun, Dong-Yup Lee, Ayoun Cho, Sang Yup Lee, and Sunwon Park)

**Abstract :** The computational modeling and simulation of complex biological systems are indispensable for new knowledge extraction from huge experimental data and ever growing vast amount of information in systems biology. Moreover, gathering and sharing of the existing information and newly-generated knowledge can speed up this research process. In this regard, several modeling projects have been undertaken for quantitatively analyzing the biological systems via the internet. They include Virtual Cell, JWS and OBIYagns. We also develop an integrated web-based environment, which facilitate investigation of dynamic behavior of cellular systems.

**Keywords :** systems biology, web-based application, kinetic modeling and simulation, cellular dynamics, SBML

### 1 서론

최근 생명공학 기술의 진보에 따라 사람의 유전체 서열이 완전히 해독되었고, 생명체의 염기서열과 유전자 해독이 이루어지고 있다. 이와 더불어 생물정보학의 발달로 가능해진 대용량 데이터 처리기술(high-throughput technology)의 발전에 따라 생명시스템에 대한 많은 양의 정보데이터가 쏟아져 나오고 있다. 그러나 여전히 유전자와 생물학적 기능에 대한 이해는 미흡한 것이 사실이다. 생명체와 같은 복잡한 시스템을 이해하고 응용하기 위해서는 각각의 구성요소의 개별적 기능뿐 아니라 그들의 상호작용에 대한 이해가 필요하다.

이러한 생명체의 상호작용에 대한 이해는 하나의 객체에 대한 구성요소를 보다는 전체 시스템의 표현에 대한 분석을 요구한다. 이는 몇몇 대사회로에 대해 정상상태 가정을 바탕으로 한 부분적 분석과 더불어, 전체적 시스템에 대한 동적 모사(dynamic simulation)에 의하여 비로소 가능해질 수 있다. 하지만, 이렇게 전체적인 생명체의 대사회로가 완성되었다 하더라도 주위의 환경이나 유전자의 이상기작 등에 대한 동적 거동을 예측하기 위해서는 더 많은 노력이 필요하게 현실이다[1].

이러한 노력의 일환으로 복잡한 생물학적 시스템의 전산 모사는 필수적이다. 그렇지만 현재의 모사 수준은 기반 지식의 부족으로 인하여 모델 분석 및 예측 능력이 어느 정도 제한되어 있는 것 또한 사실이다. 따라서 다음과 같은 지식 창출과정이 필요하다(그림 1)[2]. 원하는 정보를 얻기 위해, 먼저 가설을 세우고 실험을 설계한다. 실험을 통하여 데이터가 만들어지면 그것을 바탕으로 데이터 마이닝 혹은 데이터 클

러스터링 기법 등을 통하여 모델을 수립한다. 이어 컴퓨터상의 모사를 통해 수립된 모델의 타당성을 검증하여 가설을 보정한다. 이렇게 반복적이고 유기적인 과정을 통하여 새로운 지식을 창출함으로써 신약 개발이나 산업에 유용한 물질을 생산하는 데 이바지할 수 있게 된다.

이를 위한 모델링 기법으로는 결정론적, 확률적 접근법 및 혼합형 알고리즘 등이 있다. 결정론적 접근법은 생물체 내의 대사반응들을 시간에 따른 각 대사물질의 농도 변화를 나타내는 미분 방정식으로 나타내어 이 반응 속도식들의 집합으로 시스템의 동적 거동을 설명하는 모델을 구성하는 것이다 [3]. 민감도 분석, 평형 분석, 분기점 분석 및 파라미터 추정 도구들도 이를 위해 개발되었으며 대표적으로 상미분 방정식(ordinary differential equation), 미분-대수 방정식(differential algebraic equation), 편미분 방정식(partial differential equation) 등으로 표현된다[4].

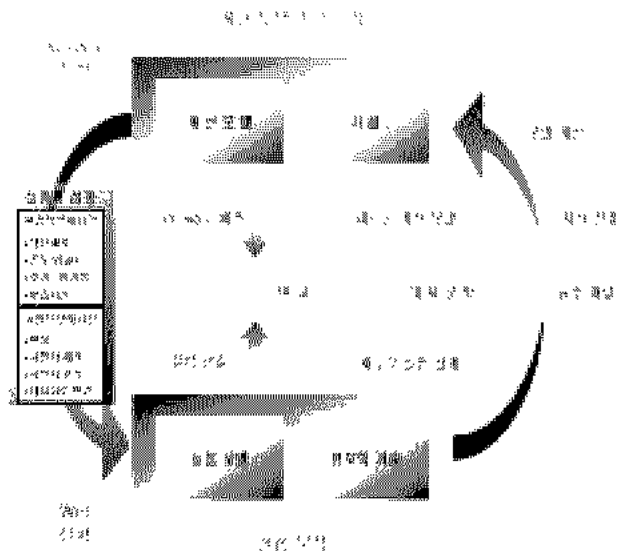


그림 1. 시스템 생물학 연구 과정.  
 Fig. 1. The research cycle of systems biology.

\* 책임저자(Corresponding Author)  
 논문접수 : 2005. 8. 25., 채택확정 : 2005. 9. 12.  
 윤좌문, 조아연, 이상엽, 박선원 : 한국과학기술원 생명화학공학과  
 (cmyun@pse.kaist.ac.kr/aycho@pse.kaist.ac.kr/sylee@kaist.ac.kr/sunwon@kaist.ac.kr)  
 이동엽 : 싱가포르 국립대학교(NUS) 생물공정기술연구소(BT)  
 (cheld@nus.edu.sg)  
 ※ 본 연구는 한국과학재단에서 시스템 생물학 사업과 BK21 프로젝트에 의하여 연구되었음.

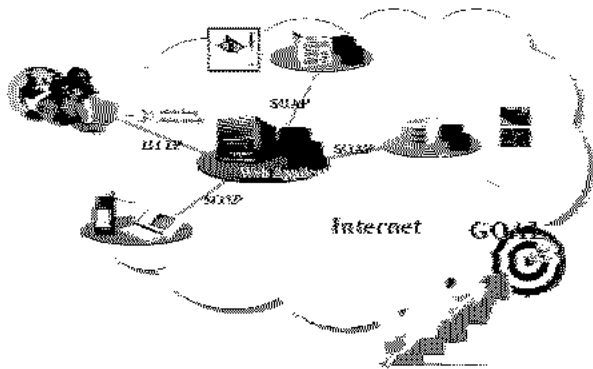


그림 2. 웹 기반의 서버-클라이언트 모델.  
Fig. 2. A web-based server-client model.

그러나 이러한 결정론적 접근법에서 시스템은 초기조건에 서부터 이미 결정된 경로를 따라 전개된다는 가정에 기초하므로 이벤트의 발생이 불연속적, 또는 확률적으로 이루어지는 경우, 그리고 비선형 방정식이나 다중 시간열을 가진 문제의 경우 결정론적 접근이 어려워진다. 따라서 확률적 변동과 잡음의 발생에 대해 초점을 두는 확률적 접근법과[5,6] 시스템에 급작스런 변화를 가져오는 불연속성을 가진 시스템의 거동을 포착하는데 유용한 혼합형 알고리즘을 채택할 수 있다[7].

**II. 웹 기반 환경**

웹 기반 환경은 지식 창출 과정의 효과를 극대화한다. stand-alone 방식[8-10]과는 달리 특별히 사용자가 이용하는 시스템마다 대상 프로그램을 설치해야 해야 하는 번거로움이 없으며, 서버-클라이언트 모델인 웹 환경에서의 모든 기능은 모두 서버에서 관리되므로 기능향상의 필요가 있을 때 즉각적으로 갱신되어 서비스를 수행할 수 있다. 즉, 버전업(version-up) 되었을 경우 프로그램을 따로 설치할 필요가 없이 유지보수가 손쉽게 이루어진다.

또한, 인터넷 환경의 이점을 최대한 이용함으로써, 사용자는 어디서나 기존의 모델 데이터베이스에 접근할 수 있을 뿐 아니라 본인이 디자인한 모델을 수정하고 분석할 수 있다. 개별 사용자들이 구축하여 공개한 모델이나 논문에 개제된 모델에 서버의 데이터베이스를 통해 접근할 수 있으므로, 빠르고 효과적인 연구 및 지식 축적이 가능할 것으로 기대된다.

**III. 생물학적 전산모사 도구**

**1. Virtual Cell[11]**

미국 NRCAM (the National Resource for Cell Analysis and Modeling, <http://www.nrcam.uchc.edu>)의 Virtual Cell은 세포내의 일어나는 Ca<sup>2+</sup>의 활동 분석에 매우 유용한 시뮬레이션 시스템으로서, 썬 마이크로시스템즈의 자바 애플릿 환경으로 되어 있어 웹에서 서비스한다는 강점이 있다. 서버에서는 마이크로소프트의 C++로 구현된 모사 서버에서 전산 모사를 수행하고, 데이터베이스나 모사의 제어와 그 결과는 자바 애플릿으로 표현되어, 웹에서 사용자가 응답 속도에 대한 불편함을 느낄 수 있다. 그러나 세포를 기하학적 이미지로 보여 줄 수 있으며, 시공간적 시뮬레이션이 가능한 상미분 방정식과

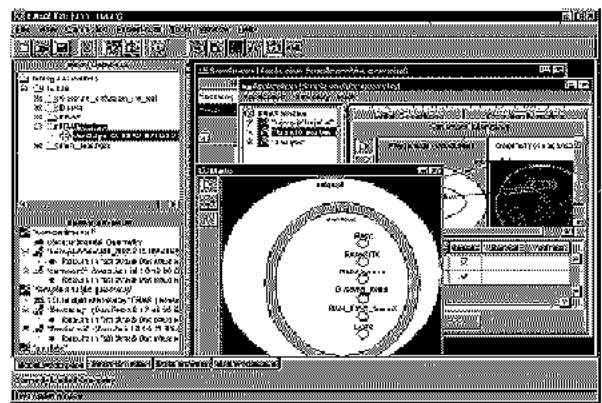


그림 3. NRCAM의 Virtual Cell. 모델 저장 데이터베이스와 분석 화면.  
Fig. 3. The screenshot of model repository and analysis modules of Virtual Cell.

편미분 방정식의 모사와, 지역 파라미터(local parameter)에 대한 시스템 차원의 민감도 분석, 최적 파라미터를 찾을 수 있는 기능들이 가능하다. 최근 더 많은 사용자들이 모델 데이터베이스 서버에 자신의 모델이나 정보를 저장하고 다른 사용자와 공유함으로써 인터넷 망을 이용한 이점을 살리고 있다.

**2. JW8[12]**

남아공 Stellenbosch 대학의 JWS는 사용자 중심의 생물학적 시스템 모델의 인터넷 동적 모사기이다. 자바 애플릿과 JLink interface를 통해 mathematica (wolfram research)의 연산 엔진과 통신 하도록 구성하여 결과에 대한 신뢰성을 제공하고 있다. 문헌조사를 바탕으로 구성된 동적 모델을 정리하여 'the silicon cell' 이라는 웹 페이지에 등재하여 사용자가 쉽게 모사할 수 있도록 하였다. 다른 모사기와의 연동을 위해 SBML (Systems Biology Makeup Language)[13-14] 과 PySoes (Python Simulation for cellular systems)[15]의 형식으로 다운 받아 사용할 수 있다. 또한 동적 모사 외에 정적 분석(steady-state analysis)와 대사 조절 분석(metabolic control analysis)등을 통해 생물학적 모델의 정량적이고 정성적인 분석이 가능하다. 하지만 생물학자나 시스템 과학자들이 실험을 통해 얻은 데이터를 바탕으로 모델을 형성하거나 분석할 수 없는 단점이 있으며, 자바 애플릿 환경으로 구현하여 Virtual Cell 과 같이 서비스 속도에 대한 문제점이 있다.

**3. CE Yagns[ 16]**

일본 RIKEN genomic science 센터의 OBIYagns (<http://access.obigid.org/yagns>). 기존의 생물학적 전산 모사기와는 달리 OBIGrid (Open Bioinformatics Grid)와 연계하여 grid 기술을 이용해 PC 클러스터의 병렬 계산을 수행 할 수 있다. 따라서, 고성능의 연산을 원하는 복잡하고 규모가 큰 생물학적 모델의 전산 모사나, 실험 데이터를 가지고 모델을 구성할 때 알지 못하는 파라미터에 대한 예측을 유전자 알고리즘(genetic algorithm)을 이용하여 해결할 수 있도록 구성하였다. 하지만 사용자 환경에 편이성이 부족하며, 다른 모사기와의 연동이 불가능하여 유용성이 부족하다. 또한 다른 웹 환경 모사 도구에 비하여 모델 저장 데이터베이스를 많이 가지고 있지 않다.

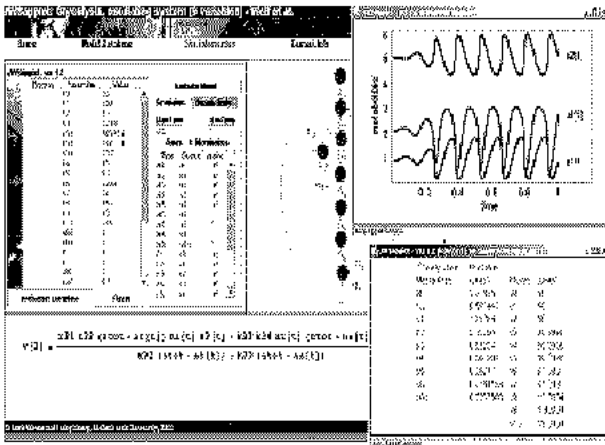


그림 4. JWS 애플릿: 네트워크 경로 시각화 화면 및 시뮬레이션 결과  
 Fig. 4. JWS applet: simulation results and a network pathway.

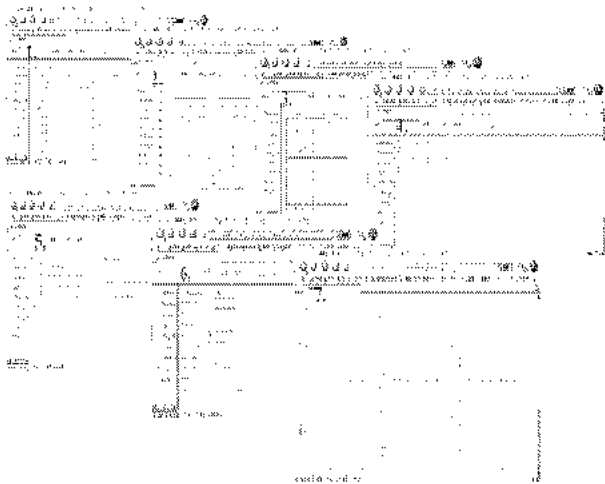


그림 5. OBIYagms: 모델 데이터 표 및 시뮬레이션 결과.  
 Fig. 5. OBIYagms: model data tables and simulation results.

**IV. WebCell**

WebCell[1]은 간편하면서도 이해하기 쉬운 모델링 환경을 제공한다. 사용자는 동적 반응(kinetic reaction) 모델을 새로 정의하거나 탑재(import) 할 수 있다. 이는 WebCell이 생화학 네트워크를 표현하는 모델링 언어인 SBML(Systems Biology Markup Language)을 지원함으로써 가능하며, 이렇게 만들어진 모델은 또다시 SBML level 1, 2나 MATLAB, CSV(Comma Separated Value)와 같은 다양한 형식으로 생성(export) 할 수 있다. SBML model repository, JWS online, BioModels database (<http://www.ebi.ac.uk/>)의 라이브러리를 모두 포함하며, 논문에 게재되어 유효성이 엄격하게 검증된 모든 모델들을 대사, 신호, 조절 경로로 분류하여 제공한다. 또한 구성된 대상 네트워크를 시각화 도구를 이용하여 효과적인 분석이 가능하다 (그림 6).

모델 분석 기법으로는 보존 분석(conservation analysis)과 함께 상미분 방정식, 미분-대수 방정식을 풀 수 있는 RK (Runge-Kutta), Euler, 암묵적 RK 등의 다양한 알고리즘을 포

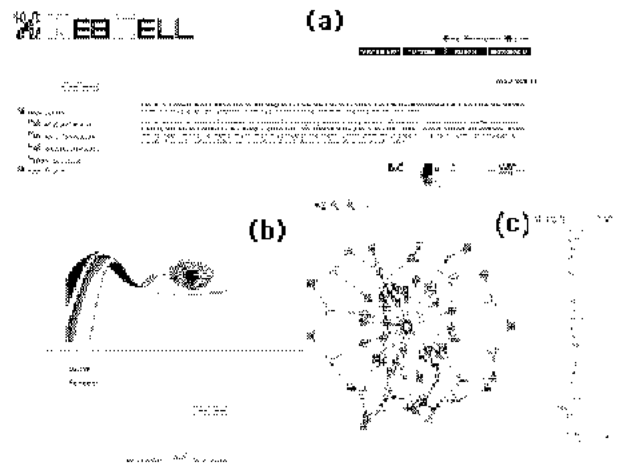


그림 6. (a) WebCell의 초기화면. (b) WebCell의 로그인 화면. (c) WebCell이 제공하는 시각화 기능.  
 Fig. 6. (a) The front page of WebCell. (b) Login page of WebCell. (c) Visualization provided by WebCell.

함한다. WebCell은 모델의 동적 파라미터(kinetic parameter)의 에너지 수치를 고려하여 모델 유효성 검사를 한다. 이는 WebCell만의 특징적인 기능으로서, 폐쇄된 반응 네트워크에서 순환경로(cyclic pathway)를 발견하고, 그 경로에 존재하는 동적 파라미터들이 열역학적 원리에 부합되는지 검사하여, 동적 모델의 유효성을 자동으로 검사하는 것이다. 이러한 특징은 확고하고 정확한 모델을 구축하는 데 있어 매우 중요한 역할을 한다.

또한 WebCell에서는 혼합형(hybrid) 모델의 구현이 가능하며, 연속적, 불연속적 성질을 모두 가진 생화학적 모델의 실제에 근접한 모사를 할 수 있다.

**1. 시스템 구조**

WebCell은 윈도우 2003 서버 운영체제 하에서 Apache's Jakarta Tomcat, MySQL, JAVA 2 Platform Standard Edition (J2SE) 5.0을 기반으로 구축되었으며, 주로 인터페이스와 엔진에 사용된 언어는 JSP와 JAVA이다. MySQL 데이터베이스는 기본적으로 모델 저장 데이터베이스를 제공하며, 사용자가 생성하는 사용자 모델을 저장할 수도 있다. WebCell의 몸체인 JSP 엔진은 사용자 환경을 담당하는 디스플레이 매니저, 사용자 계정을 담당하는 로그인 매니저, 모델의 파일 포맷 저장 및 내보내기를 담당하는 업로드 및 import/export 매니저, 데이터베이스를 담당하는 DB 매니저, 그리고 수치 해석 엔진을 담당하는 시뮬레이션 매니저 등으로 이루어진다.

디스플레이 매니저는 SOAP (Simple Object Access Protocol) 기술을 결합하여 사용자의 메시지를 받아들이고 분석하며, 네트워크 경로를 시각화 하기 위한 알고리즘과도 연결되어 있다. 시뮬레이션 매니저에서는 사용자가 모델을 생성하면 JSP와 JYTHON 스크립트에서 이를 처리하여 JAVA 파일을 동적으로 만들어낸다. 이렇게 만들어진 JAVA 파일은 혼합형 모델 및 상미분 방정식, 미분-대수 방정식 등을 수치적으로 풀어내는 엔진을 모듈로 하여 런타임으로 실행, 결과값을 얻게 한다. 결과값은 시뮬레이션 결과 그래프, SBML, MATLAB, Excel 또는 CSV 포맷으로 출력 가능하다.



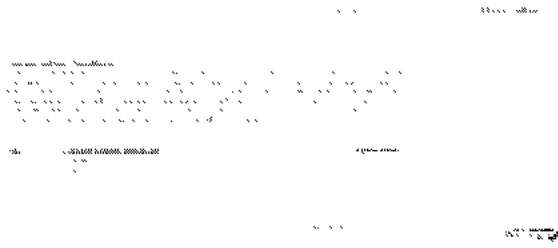


그림 12. 보존 관계 분석 결과.  
Fig. 12. The result of conservation relation analysis.

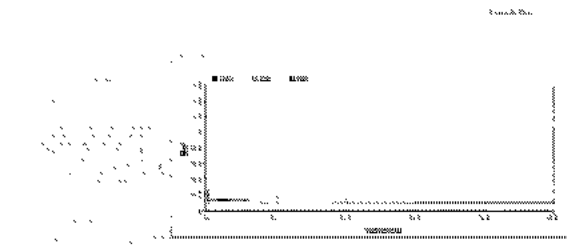


그림 13. Teusink model의 모사 결과.  
Fig. 13. Simulation result of Teusink model.

3.3 열역학적 에너지 수지식 (파라미터 검증) [18]

파라미터 검증을 위하여 열역학적 에너지 수지 관계를 이용할 수 있다. 생화학적 반응에서 깃스 자유에너지는 반응의 경로에 상관없이 변한다. 따라서, 순환 경로(cyclic pathway)에서 깃스 자유에너지의 합은 0이 된다. 즉, cyclic pathway를 이루는 각 반응물질의 표준 깃스 자유 에너지에 양론계수를 곱한 것의 합이 0이 되어야 한다.

$$\sum G_{\text{reaction}}^0 = \sum G_{\text{products}}^0 - \sum G_{\text{reactants}}^0 = \sum_i \nu_i G_i^0 = 0 \quad (1)$$

또한 평형 상태에서,

$$\Delta G = 0 = \Delta G^0 + RT \ln K_{eq} \quad (2)$$

위의 두 관계를 이용하여 동적 파라미터(kinetic parameter) 검증을 위한 열역학적 에너지 수지 관계를 다음과 같이 질량 작용식과 Michaelis-Menten식 각각에 대하여 얻을 수 있다.

$$\sum_{i=1}^2 \left( (\ln(k_{f,i}) - \ln(k_{r,i})) - \frac{\Delta G_3^0}{RT} - \frac{\Delta G_4^0}{RT} \right) = 0 \quad (3)$$

$$\sum_{i=1}^2 \left( \ln(V_{f,i}) + \ln(K_{mP,i}) - \ln(V_{r,i}) - \ln(K_{mS,i}) \right) - \frac{\Delta G_3^0}{RT} - \frac{\Delta G_4^0}{RT} = 0 \quad (4)$$

3.4 동적 모사

정의된 모델의 반응식을 풀어 반응의 유량 또는 구성성분의 농도를 계산한다. 사용자는 알고리즘과 시간열(timescale)을 조정하고 알고자 하는 물질을 선택하여 모사한다. 구성성분이나 반응식만 선택상자에 보여지도록 할 수 있으며, 동시에 여러 물질의 모사가 가능하다. 이렇게 나온 결과물은 CSV 포맷으로 내려 받을 수 있어 사용자의 임의대로 활용할 수 있게 한다.

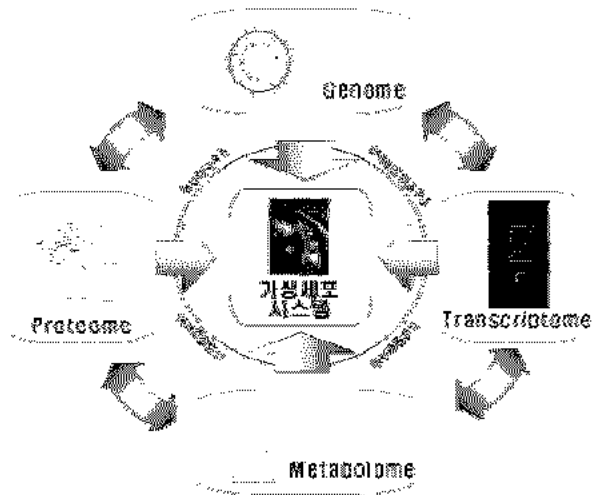


그림 14. 가상세포모델의 개념도.  
Fig. 14. The conceptual diagram of an *in silico* cell.

VI. 결론

본 논문에서는 생물학적 모델의 웹 기반 모사의 이점과 동향, 그리고 본 연구팀에서 개발한 WebCell에 대하여 소개하였다.

웹 환경은 기존의 stand-alone 방식에 비하여 언제 어디서나 접근 가능하고 유지보수가 뛰어난 이점을 지닌다. 통합 데이터베이스의 구축에 따른 정보의 축적, 지식의 공유에 따른 새로운 지식 창출이 가능해지는 것 또한 웹의 강력한 이점이다.

또한 본 논문에서는 웹 환경의 대사 네트워크 시뮬레이션 프로그램인 Virtual Cell, JWS, OBYagns 그리고 WebCell의 각각에 대한 특징을 살펴보았다. 여기에서 특히 WebCell은 웹 기반 대사 네트워크의 시뮬레이션 도구로서, 기존의 동적 모델 뿐 아니라, 네트워크의 정량적이고 동적인 분석을 위한 사용자 모델 생성을 돕는 포괄적이고 통합적인 시스템이다. 또한 현재 렘핑(lumping), 민감도 분석(sensitivity analysis)이나 시간 열 분석(time scale analysis) 등과 [19,20] 같은 기법 등을 추가로 적용 중이다. 이러한 방법에 의해 모델의 복잡도를 완화시켜 웹 환경에서 시스템 레벨의 이해가 가능해질 것이다.

참고문헌

- [1] D.-Y. Lee, C. Yun, B. K. Hou, S. Park and S. Y. Lee, "WebCell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks," *submitted*.
- [2] S. Y. Lee, D.-Y. Lee and T. Y. Kim, "Systems biotechnology for strain improvement," *Trends in Biotechnology*, vol. 23, pp. 349-358, 2005.
- [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] U. M. Ascher and L. R. Petzold, "Computer methods for ordinary differential equations and differential-algebraic equations," *SIAM*, 1998.
- [5] T. C. Meng, S. Somani and P. Dhar, "Modelling and simulation of biological systems with stochasticity," *In Silico Biology*, vol. 4, 0024, 2004.

- [6] N. Novere and T. S. Shimizu, "Stochsim: Modeling of Stochastic Biomolecular Processes," *Bioinformatics*, vol. 6, pp. 575-576, 2001.
- [7] T. R. Kiehl, R. M. Mattheyses and M. K. Simmons, "Hybrid simulation of cellular behavior," *Bioinformatics*, vol. 20, pp. 316-322, 2004.
- [8] 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, pp. 2144-2146, 2003.
- [9] 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.
- [10] P. Mendes, "Biochemistry by numbers: simulation of biochemical pathways with Gepasi3," *Trends Biochem. Sci.*, vol. 22, pp. 361-363, 1997
- [11] L. M. Loew and J. C. Schaff, "The Virtual Cell: A Software Environment for Computational Cell Biology," *Trends Biotechnol.*, vol. 19, pp. 401-406, 2001.
- [12] B. G. Olivier and J. L. Snoep, "Web-based kinetic modeling using JWS Online," *Bioinformatics*, vol. 20, pp. 2143-2144, 2004.
- [13] 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.

- [14] 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, pp. 41-53, 2004.
- [15] B. G. Oliver, J. M. Rohwer and J. S. Hofmeyr "Modeling cellular systems with PySCeS," *Bioinformatics*, vol. 21, pp. 560-561, 2005.
- [16] 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, pp. 1646-1648, 2004.
- [17] H. M. Sauro, B. Ingalls, "Consevation analysis in biochemical networks: computational issues for software writers," *Biophys. Chem. J.*, vol. 109, pp. 1-15, 2004.
- [18] D. A. Beard S. D. Liang and H. Qian, "Energy balance for analysis of complex metabolic networks," *Biophys. J.*, vol. 83, pp. 79-86, 2002.
- [19] S. Okino and M. L. Mavrouniotis, "Simplification of mathematical models of chemical reaction systems," *Chemical Reviews*, vol. 98, 1998.
- [20] M. Schauer, R. Heinrich, "Quasi-steady-state approximation in the mathematical modeling of biochemical reaction networks," *Mathematical Biosciences*, vol. 65, pp. 155-170, 1983.



#### 윤 과 문

1976년 11월 6일생. 2001년 영남대학교 화학공학과 학사. 2003년 영남대학교 화학공학과 석사. 2003년~현재 한국과학기술원 생명화학공학과 박사과정. 관심분야는 공정시스템 및 시스템 생물학.



#### 조 아 연

1982년 2월 15일생. 2005년 연세대학교 화학공학과 학사. 2003년~현재 한국과학기술원 생명화학공학과 석사과정. 관심분야는 공정시스템 및 시스템 생물학.



#### 박 선 원

1948년 1월 9일생. 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. 관심분야는 공정제어, 공정설계, 공정최적화, SCM, LCA, 시스템 생물학, 생물정보학.



#### 이 동 업

1973년 10월 4일생. 1998년 연세대학교 화학공학과 학사. 2000년 한국과학기술원 생명화학공학과 석사. 2004년 한국과학기술원 생명화학공학과 박사. 2005년~현재 싱가포르 국립대학교 조교수 및 생물기술연구소 연구원(겸임). 관심분야는 시스템 생물학/생명공학, 신약 및 질병 모델링, 공급사슬망 최적화.



#### 이 상 업

1986년 서울대학교 화학공학과 학사. 1987년 Northwestern University 화학공학과 석사. 1991년 Northwestern University 화학공학과 박사. 2002년~현재 한국과학기술원 생명화학공학과 교수. 관심분야는 대사공학, 재조합 미생물 공정, 생물정보학, 바이오칩 (DNA칩/단백질 칩), *in silico* 바이오테크놀로지, 나노 테크놀로지, 미생물 배양공정, 미생물 이용 일/이차 대사산물 생산, 생물전환공정, 단백질 발현 및 생산 제반 기술, 생물고분자.