

Systems biology and engineering (SBE)

임영일

CAPEC, Department of Chemical Engineering,
 Technical University of Denmark, DK-2800 Lyngby, Denmark
 Phone: +45 4525 2804, Fax: +45 4593 2906, Email: lim@kt.dtu.dk

3. 살아있는 세포 모사

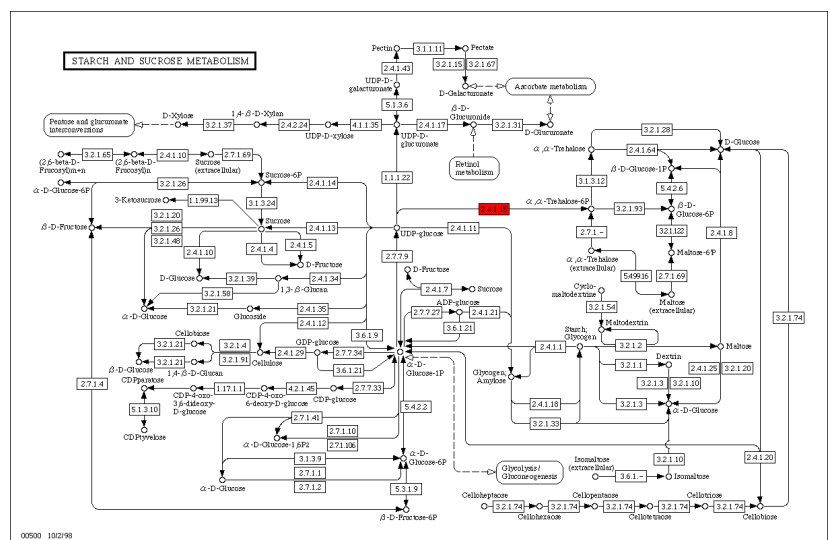
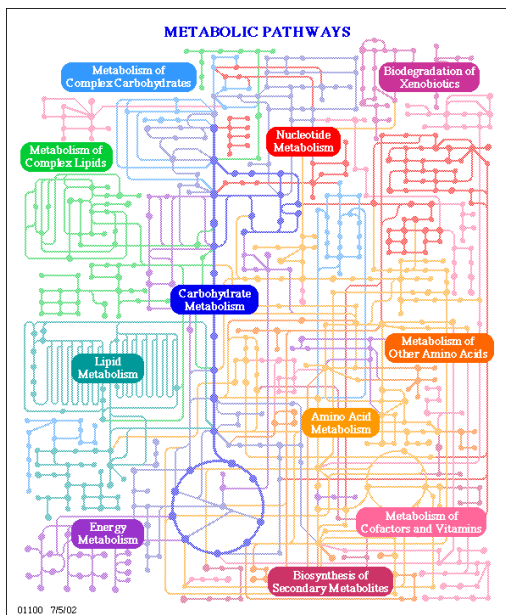
살아있는 세포/유기체의 정적/동적거동을 수학적 모델링을 통하여 체계적으로 이해하려고 하는 systems biology 는 genome (DNA) 으로부터 transcriptome (mRNA), proteome (enzymes and proteins), metabolome (metabolic pathway and kinetics) 그리고 fluxome (metabolic fluxes) 과 같은 세포내 분자간 상호관계에 대한 정보를 통합하여 세포거동을 예측하는 것에 목적을 둔다. 결여된 인간 유전자 (human orphan genes) 에 따른 생체내 기능변화에 대한 연구 (O'Donovan et al., 2001) 가 진행중인 현 시점에서, 세포 모사는, 장기적인 안목으로 볼 때, 기능 유전체학 (functional genomics) 에 큰 기여를 할 수 있을 것이다. 또한 세포모사는 생물공정의 동적거동을 유전자 그리고 분자 수준에서 이해할 수 있는 기법이 될 것이다.

대사경로 (metabolic pathway) 또는 이들의 집합체인 대사망 (metabolic network) 분석과 모델링은 유기체내의 복잡한 생물학적 기능 이해에 있어서 필수적이다 (Lee et al., 2003). 정상상태 가정하에서 이루어지는 이 분석을 세포의 정적모사라고 하고, 이들 대사경로상의 대사반응식과 유전자 조절, 신호전달을 포함하는 접근을 동적모사라고 한다. 본 장에서는 세포의 정적/동적 거동의 모델링과 모사에 관하여 설명한다.

3.1 정적 모사 (유전정보로부터 대사경로 모사, metabolic pathway/network analysis)

어떤 유기체의 대사 경로에 대한 연구는 오랜 기간에 걸쳐 하나씩 하나씩 발견되어 왔다. 또한 미생물의 유전자에 대응하는 대사경로분석 (genome-scale metabolic network) 은 미생물의 기능과 생성물 변화를 예측하는데 큰 역할을 담당한다 (Papin et al., 2003). 이 연구는 재조합 미생물의 생물학적 시스템 (engineered biological system) 의 설계와 가능성있는 대사경로의 발견으로 응용되어질 수 있다.

효모 (yeast) 로 알려진 *S. cerevisiae* 는 6300 여개의 유전자를 보유하고 있고, 이 유기체의 알려진 전체 대사경로는 그림 1 (a) 에서 보여준다 (<http://www.tigr.org/> 참조). 이 중 metabolism of complex carbohydrates 의 한 부분인 Starch and sucrose metabolism 은 그림 1. (b) 에서 보여준다.



(a) overall map of metabolic pathway

(b) Starch and sucrose metabolism in metabolism of complex carbohydrates

그림 1. Metabolic pathway of *S. cerevisiae* (<http://www.tigr.org/> 참조).

이러한 데이터베이스에서는, 대사경로분석에 관한 *in silico* experiment 를 위한 반응물, 생성물, 각 대사반응의 stoichiometry, 반응의 촉매역할을 하는 효소이름, 이들 효소합성 정보를 지닌 유전자, 그리고 EC (Enzyme Commission) number of each metabolic reaction 에 관한 정보를 얻을 수 있다.

708 개의 유전자로 부터 발현되는 1175 대사반응식/경로 (metabolic reaction/pathway) 과 584 대사물질 (metabolite) 을 포함하는 효모의 대사경로 모사는 J. Forster (PhD thesis, 2002) 의 박사논문에서 찾아볼 수 있다. 이러한 대사경로 모사에서는 유전자 변이에 따른 미생물의 생사, 생성물 변화 등을 예측할 수 있어 생물분자 개발에 이용할 수 있다. 이 연구에서는 반응상수값이나 유전자/생성물 조절과정 (regulatory pathway) 은 포함하지 않고, stoichiometric 모델식에 기반을 두고 있기에 정적모사라고 볼 수 있다. Metabolic pathway analysis 는 크게 다음 두분야에서 활발한 연구가 진행중에 있다.

Metabolic flux analysis (MFA)

대사 유량 분석은 정상 상태하에서의 대사물질 (metabolite) 에 관한 대사반응의 양론값 (stoichiometry), 질량수지에 바탕을 둔 대사경로분석을 말한다 (Lee et al., 2003). 모든 대사물질과 대사경로 그리고 각 경로에서의 stoichiometric matrix (S_{ij}^T , metabolite i in the j^{th} pathway) 가 알려져 있다면, 대사유량벡터 (v_j , flux of j^{th} pathway) 를 계산할 수 있는데, 시간에 따른 metabolite X 의 변화는 모든 대사반응의 유량의 합으로 나타낼 수 있다.

$$S^T v = \frac{dX}{dt} \tag{3-1}$$

만일 시간에 따라 X 의 변화량이 일정하다고 가정하면, 즉 정상상태 가정하에서, 이 식은 다음과 같이 정의 된다.

$$S^T v = 0 \tag{3-2}$$

앞서 설명한 효모의 대사경로 및 유량분석 (Forster, 2002) 에서는, 584 의 대사물질에 대하여, 1175 반응식이 관여한다 (I=584, J=1175). 또한 각 반응에 관여하는 708 개의 유전자가 명시되어 있어, 어떤 유전자의 삭제는 곧 이에 관련된 반응이 일어나지 않음을 의미한다.

하지만, 경로만이 알려져있고, 양론값 (stoichiometric value for each metabolite and pathway) 과 대사유량 (v_j) 이 부분적으로 알려진 경우가 많아 식 (3-2) 은 다음과 같이 확장된다.

$$S^T v \equiv S_m v_m + S_u v_u = 0 \tag{3-3}$$

이식에서, 실험적으로 알려진 양론값 ($S_m(I \times M)$, I=total metabolite number, M=total stoichiometrically-known reaction number) 과 유량 ($v_m(M \times 1)$) 의 내적으로 정의된 행렬과 알려지지 않은 양론값 ($S_u(I \times U)$) 과 유량 ($v_u(U \times 1)$) 의 곱으로 된 두 행렬로 분리된다. 만일 알려지지 않은 유량벡터 (S_u) 의 $rank(S_u)$ 가 U 보다 작으면 (즉, 방정식보다 변수의 수가 더 많으면), MFA 는 다음에 기술될 FBA 로 확장된다.

한국과학기술원의 이상엽교수가 이끄는 대사공학연구실 (Metabolic & Biomolecular Engineering National Research Laboratory, <http://mbel.kaist.ac.kr/>) 과 박선원교수가 이끄는 Process systems lab. (<http://love.kaist.ac.kr>) 에서는 대사과정에서 각 경로마다 생물분자 유량을 계산해 내는 MetaFluxNet v1.6 (Lee et al., 2003) 를 개발했고, 계속적으로 갱신하고 있다. 그림 2 는 이 프로그램에서 구한 대사유량 위상도를 볼 수 있다. 이 공개 프로그램은 <http://mbel.kaist.ac.kr/mfn/> 에서 내려받을 수 있으며, Linear programming (LP) optimization algorithm 을 이용하여 알려지지 않은 유량 ($v_u(U \times 1)$) 을 계산한다.

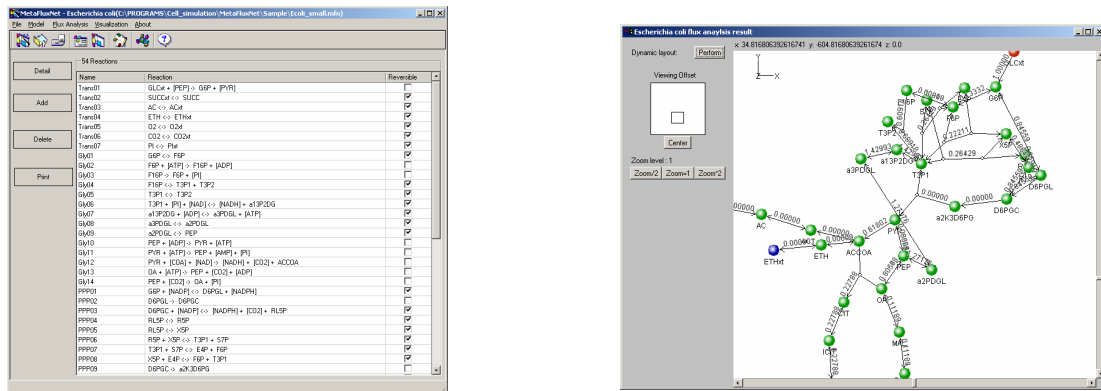
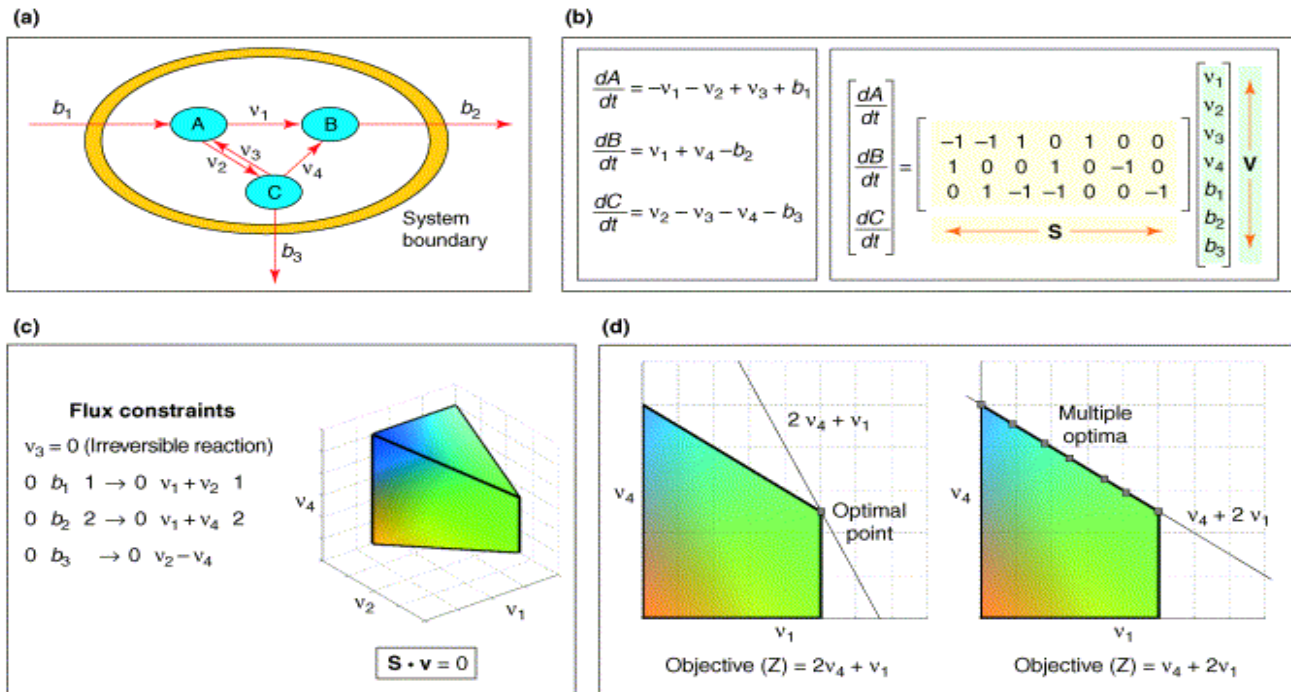


그림 2. MetaFluxNet (Lee et al., 2003) 에서 구현되는 FBA 를 이용한 MFA.

Flux balance analysis (FBA)

대사반응에 대한 반응상수값이 알려지지 않았다 하더라도, 대사경로상의 stoichiometry 에 의하여 제한되어지는 정상상태하에서의 flux balance analysis (FBA) 를 통하여 대사시스템의 여러 현상을 이론적으로 알아볼 수 있다. 양론학적 제한식 (3-3) 은 일반적으로 underdetermined system 으로 정의되는데, 특정 대사반응의 유량은 특정범위로 한정되어질 수 있으며, 여러 물리 화학적 제한식이 존재하므로, 유일한 최적 대사유량을 찾아낼수 있다.

University of California, San Diego (<http://www.ucsd.edu/>) 의 The Genetic Circuits Research Group (<http://gcrucsd.edu/>) 을 이끄는 Palsson 교수에 의해서 제안된 이 분석법은 그림 3 (Kauffman et al., 2003) 에서와 같이 단순한 한 예를 통하여 도식적으로 설명할 수 있다. 3 개의 metabolite (A, B, C), 4 개의 대사 반응에 관련된 유량 (v_1-v_4), 3 개의 외부계와의 교환유량 (b_1-b_3) 은 그림 3 (a) 에서 볼 수 있다. 따라서 3 개의 대사물질에 대한 양론식은 그림 3 (b) 에서와 같이 표현되고, 열역학적 (e.g., $v_3=0$) 또는 실험적 유량 제한식 (e.g., b_1, b_2, b_3 에 관한 최대/최소값) 이 그림 3 (c) 와 같이 존재할 수 있다. 특정 목적함수에 따른 최적 유량은 LP (linear programming) 를 이용하여 계산되어진다 (see 그림 3 (d)).



Current Opinion in Biotechnology

그림 3. FBA 에서의 최적 유량 계산을 위한 LP (Linear programming) 기법 (Kauffman et al., 2003): (a) system definition, (b) mass balance, (c) defining measurable fluxes and (d) optimization.

그림 3 (a) 에서, 대사경로는 유전자 서열정보로부터 1 차적으로 작성되어질 수 있다. 유전자에 의해서 합성되고, 특정 대사반응에 관여할 효소에 관한 대사 경로/반응은 알려져 있는 것도 있지만, 알려지지 않은 것도 있다. 이 단계에서는, 효소가 관여하는 대사반응과 대사경로 (i.e., v_i) 뿐만 아니라, 세포벽을 통해 들어오거나 나가는 경로들 (i.e., b_i) 도 명시된다. 방정식의 수 (i.e, 3 개의 metabolites) 보다 유량변수의 수 (i.e, 7 개 변수 in Fig. 3 (b)) 가 많기 때문에, 그림 3 (c) 에서와 같이 유량에 관한 몇개의 제한식들이 제시된다. 이들 제한식과 함께 어떤 목적함수를 최적화함으로써 모르는 유량값을 얻어 낸다. 즉, 수식으로 표현하면,

$$\begin{aligned} \text{Max } Z &= \sum c_i v_i \\ \text{s.t. } \quad S^T v &= 0 \text{ and } \alpha_{\min,i} \leq v_i \leq \alpha_{\max,i} \end{aligned} \quad (3-4)$$

이 식 (3-4) 은 목적함수와 제한식들이 모두 선형으로 정의되므로 LP algorithm 을 이용하여 해를 구할 수 있다. 일반적으로 목적함수는 성장속도의 최대화등으로 정의된다.

FBA 는 현재 Boolean logic approach 를 사용하면서, 대사과정 조절 (or gene expression regulation) 에 관한 제한식을 추가한 FBA 기법으로 발전해 가고 있다 (Covert et al., 2001).

3.2 동적 세포모사 (Whole cell simulation)

어떤 유기물의 유전 정보 (genome) 가 모두 해독되었다고 가정해 보자. 이들 유전자의 생성물 (e.g., mRNA, enzymes, proteins) 에 대한 구조와 기능이 또한 알려져 있고, 모든 metabolic pathway 지도가 완성되었다 하더라도, 주위 환경변화나 유전자의 이상기작등에 대한 세포의 동적거동은 예측할 수 없다.

살아있는 세포의 동적거동을 수학적 모델링하기 위해서는 대사반응경로 (metabolic pathway) 는 물론 경로상의 모든 반응상수 (kinetic constants) 와 regulatory pathway 및 signal transduction pathway 와 이들 반응식도 알려져야 한다. 현재로서는 실질적인 세포의 동적모사가 매우 힘들어 보이지만, 급속도로 증가하는 생물학 정보에 힘입어 언젠가는 결실을 맺을 수 있을 것으로 보인다. 세포 동적모사의 선두적 위치에 있는 E-cell project (Tomita et al., 1999; Tomita, 2001) 는 세포의 동적모사를 위한 환경을 구축하고 몇가지 좋은 결과를 내기도 했다.

E-cell platform (<http://www.e-cell.org>)

일본 게이오 대학의 Tomita 교수가 주축이 되어 개발되어 왔으며, 소스를 공개하고 있는 E-cell 은 살아있는 세포의 동적거동을 예측하고, 세포내 분자들의 상호관계를 모사하기 위해 개발되고 있다. 대사과정의 분석뿐만 아니라, 유전자 조절 과정, DNA 복사 등 세포내 여러 현상을 예측하기 위한 프로그래밍 환경을 제공한다. 따라서 multi time-scale simulation 을 위한 알고리즘이 내장되어 있다.

이 프로그램을 사용하면서, 가장 작은 유전자 (약 480 genes) 를 보유한 것으로 알려진 *M. genitalium* (<http://www.tigr.org/tdb/mdb/mdbcomplete.html>, Fraser et al., 1995) 원핵세포에 대하여, 생존가능한 최소 유전자 127 개를 바탕으로 동적모사를 하였다 (Tomita, 2001). 그림 4 는 생존 가능한 127 개 유전자에서 비롯되는 대사과정 모형과 동적모사 결과를 보여준다. 여기에서, 105 개의 단백질 생성 유전자, 22 개의 RNA 생성 유전자, 그리고 495 개의 대사반응식이 이용되었고, 조절과정이나 신호전달에 관한 식들은 사용되지 않았다.

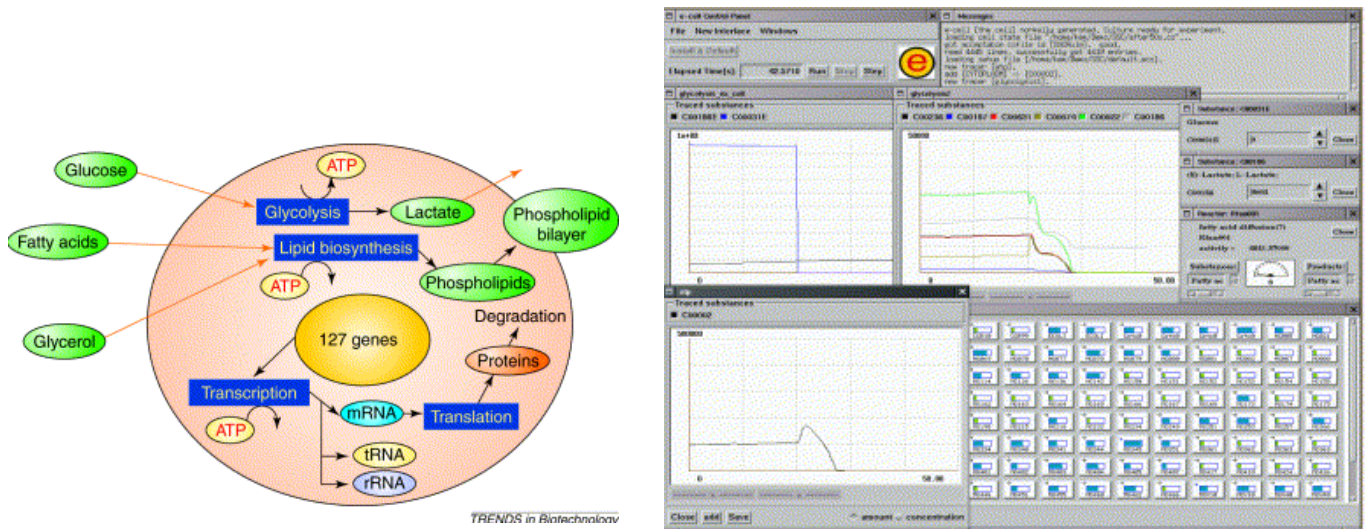


그림 4. *M. genitalium* 원핵세포 모사를 위한 모형과 E-cell 을 통한 동적모사 (Tomita, 2001).

TRENDS in Biotechnology

V-cell project (Loew & Schaff (2001))

University of Connecticut Health center 의 V-cell project (http://www.nrcam.uhc.edu/vcell_development/) 는 세포내 동적 대사과정을 중심으로 다루는 E-cell project 와는 달리, 임의적 평면/입체 공간에서의 세포간 또는 세포내 convection-reaction-diffusion 현상을 모델링/모사하는데 목적을 둔다. 따라서 모델식들은 편미방과 미방등으로 표현된다. 자세한 정보는 Loew & Schaff (2001) 를 참조하기 바란다.

Metabolic kinetic parameter estimations

Virginia bioinformatics institute (<http://www.vbi.vt.edu>) 은 실험 데이터에 부합되는 대사반응식 계수 추정을 위해 GEPASI (<http://www.gepasi.org/>) 라는 프로그램을 개발하였다 (Mendes and Kell, 1998). 하지만, 결정적인 bug 의 발견으로 곧 이 프로그램의 후속편인 COPASI (Complex Pathway Simulator) 를 공개프로그램으로 배포 예정에 있다. 그림 5 에서는 이프로그램을 이용하여 실험결과에 부합되는 반응상수식을 찾는 과정을 볼 수 있다.

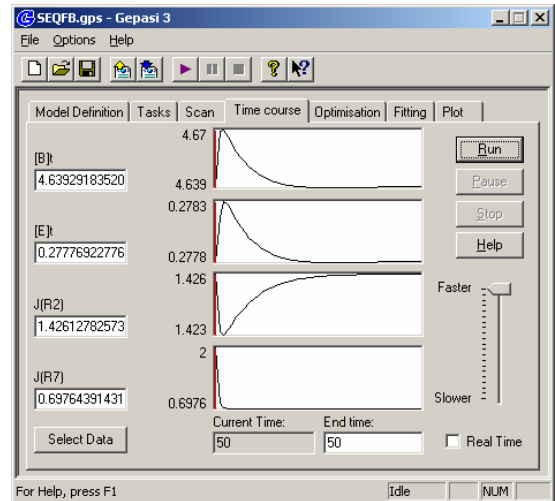


그림 5. 대사반응식 계수 추정 프로그램, GEPASI (<http://www.gepasi.org/>).

Modeling language

세포내 대사/조절/신호 경로 모델링에 있어서 범용의 재활용 가능한 모델링을 위한 프로그래밍 언어 개발은 필수적이다. 이 분야는 Software engineering 과 밀접한 관계를 갖고 있으며, SBML (System biology markup language, <http://www.sbml.org>, Hucka et al., 2003; Kinato, 2002) 은 이러한 목적으로 개발 중에 있으며, XML (extensible markup language) 과 UML (unified modeling language) 에 기반을 두고 있다.

일본 게이오 대학 Institute for Advanced Bioscience (Keio university, <http://bioinfo.sfc.keio.ac.jp/IAB/>, Japan) 과 미국의 California institute of technology 에서 함께 추진중인 The SBML project (Hucka et al., 2003) 은 생물학적 대사반응과 대사경로를 표현하기 위하여 개발되고 있고, 현재 30 여개의 프로그램 (e.g., E-cell, V-cell, GEPASI etc.) 에서 사용/지원되고 있다. 이언어는 대사 경로/반응에 대하여 compartment, species, reaction, parameter, unit definition, rule 등과 같은 성분으로 구분하고 있다. 자세한 내용은 Hucka et al. (2003) 을 참조하기 바란다.

이들 연구진들은 또한 SMBL 을 기반으로 둔 systems biology workbench (SBW, <http://www.sbw-sbml.org/>) 를 개발하고 있으며, SBW 에서는 여러 언어와의 호환성을 위한 여러 도구를 제공하고 있다.

3.3 모델링을 위한 생물학적 데이터 분석 (biological data analysis)

세포모사에 필요한 정보를 얻기위하여 genome, transcriptome, proteome, metabolome 등의 연구에서 생성되는 엄청난 양의 생물학적 데이터를 정리하고 분석하는 것 (or bioinformatics) 은 매우 중요한 학문으로 성장하였다. 한국 bioinformatics center (Center for computational biology and bioinformatics, CCBB, <http://www.cccb.re.kr>) 에서는 이러한 기능을 위한 여러 정보들을 제공하고 있다.

The University of Texas Health Science Center at Houston 의 Department of Neurobiology and Anatomy 에서는 BIO-SPIICE (<http://nba.uth.tmc.edu/darpa/>) 라는 프로그램 개발하였고, 생물학적 데이터 처리를 위하여, 1) databasing, 2) data mining and analysis, 3) model building, and 4) numerical simulation and analysis, 기능을 제공한다.

Databasing 에서는 유전자 서열, 분자구조에 대한 정보뿐 아니라, molecular profiling 과 image data 로부터 대사경로, 신호전달과정, 대사반응식 상수 등에 관한 정보를 체계적으로 저장한다. data mining and analysis 은 실험결과로부터 얻을 수 있는 여러 정보의 도출을 위한 여러 기능을 제공하는 것이다.

3.4 결론

Post genome-era 에 있어서 축적된 실험결과와 생물학적 정보기술의 비약적 발전에 힘입어 연구자들은 점차적으로 체계적 생물학 (systems biology) 에 관심을 갖어 가고 있다 (Kauffman et al., 2003). 유전자 서열 정보와 생물 분자간의 기능에 대한 정보를 바탕으로 세포/조직/기관/유기체 기능의 전체적인 모델링과 모사는 앞으로 다가올 수십년간의 과제로 남아 있다.

세포의 정적 모사 (steady-state simulation) 인 metabolic flux analysis (MFA) or flux balance analysis (FBA) 은 생체내 분자들간 상호관계의 동적모사를 연결하는 필수 단계이며, 여러 유전자 조합과 여러 세포 배양 조건에서의 대사 거동을 예측할 수 있다고 알려져 있다. 따라서 특정 유전자의 삽입과 삭제에 따라 나타나는 현상학적 거동 (phenotypic behavior) 을 살펴볼 수 있고, 미생물 유전자 설계에 도움이 될 것이다. 세포 정적모사는 현재 매우 활발한 연구가 진행되고 있고, 곧 의약품 생물분자 개발에 응용되어 상업적 가치를 보여줄 수 있을 것으로 기대된다.

대사경로 뿐아니라, 대사 반응식, 유전자 조절 과정, 신호전달등이 포함된 세포의 동적 모사 (dynamic simulation) 는 아직 구체적인 결과를 기대하기는 어려워 보인다. 이는 수만년에 걸쳐서 진화해온 유기체의 복잡한 대사경로나 조절과정에 대한 이해를 위해서는 아직도 많은 연구가 필요하기 때문이다. 즉, 아직까지 알려지지 않은 세포내 많은 반응과 경로로 인하여, 좀더 많은 실험결과와 자료의 축적이 요구된다.

하지만, 세포 정적모사 (steady-state simulation) 연구가 어느 궤도에 오르면서, systems biology 의 연구 방향은 세포 동적모사 (dynamic simulation) 로 옮겨갈 것은 자명한 사실이다. 따라서, 이 분야에 대한 꾸준한 기초연구가 필요하다고 말할 수 있다.

References

- Covert et al., 2001, Regulation of gene expression in flux balance models of metabolism, *J. Theor. Biol.*, 212, 73-88.
- Forster, J. et al., 2002, Genome-scale reconstruction of the *S. cerevisiae* metabolic network, PhD thesis, DTU, Denmark.
- Fraser et al., 1995, The minimal gene complement of *Mycoplasma genitalium*, *Science*, 270, 3(5235), 397-403.
- Hucka, M. et al. (2003), The systems biology markup language (SBML); a medium for representation and exchange of biochemical network models, *Bioinformatics*, 19(4), 524-531.
- Kauffman, K.J.; Prakash, P.; Edwards, J.S., 2003, Advances in flux balance analysis, *Current Opinion in Biotechnology*, 14(5), 491-496.
- Kitano, H. (2002), Systems biology: a brief overview, *Science*, 295, 1662-1664.
- Lee D.Y, Yun H.S, Lee S.Y. and Park S.W. (2003) MetaFluxNet: the management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics*, 19, 2144-2146.
- Loew, L. M. and J. C. Schaff (2001), The Virtual Cell: a software environment for computational cell biology, *Trends in Biotechnology*, 19(10), 401-406.
- Mendes, P. & Kell, D.B. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14**, 869-883.
- O'Donovan et al., 2001, The human proteomics initiative (HPI), *Trends Biotechnol.*, 19, 178-181.
- Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., and Palsson, B.O. (2003), "Metabolic Pathways in the Post-Genome Era," *Trends in Biochemical Sciences*, 28(5), 250-258.
- Tomita, M. (2001), Whole-cell simulation: a grand challenge of the 21st century, *TRENDS in Biotechnology*, 19(6), 205-210.
- Tomita, M. et al. (1999), E-CELL: software environment for whole-cell simulation, *Bioinformatics*, 15, 72-84.