

계층적 유전자 조절 네트워크가 통합된 가상 미생물 시스템

이성근, 한상일, 황규석
부산대학교 화학공학과

1. 서론

가상 미생물 모델(*in silico* microbe model)은 세포 내에서 일어나는 모든 생리학적 현상을 컴퓨터상에서 모사하기 위한 모델이다. 최근 genomics, proteomics, transcriptomics, metabolomics, fluxomics에서 생성된 다량의 omics data를 총체적으로 통합하여 시스템 차원에서 해석하고자 하는 요구가 증가하고 있고, 이러한 다양한 omics data를 생리학적으로 해석하여 이용한 것이 바로 가상 미생물 모델(*in silico* microbe model)이다. 가상 미생물 모델을 구축하는 방법에는 상세한 속도식 모델(kinetic model), 대사 네트워크 모델(metabolic pathway model), 인공생체 모델(cybernetic model), 확률 모델(stochastic model) 등이 있고, 이 중에서 속도식 모델과 대사 네트워크 모델이 가장 많이 연구되고 있다.

1.1. 속도식 기반 가상 미생물 시스템

정량적 데이터로부터 만들어진 속도식을 이용한 가상 미생물 모델은 수년 전부터 미국, 일본, 유럽 등에서 많은 연구가 진행되고 있으며, 현재 개발된 시스템으로는 GEPASI, JARNAC, E-Cell, Virtual Cell 등이 있다. 특히 E-Cell은 게이오 대학에서 1996년에 Tomita교수를 필두로 연구가 시작되었고, 현재 가장 작은 지놈(580kb)과 유전자(~480개)를 가진 *Mycoplasma genitalium*을 대상으로 하여 처음으로 세포 전체를 모델링하는데 성공하였다. 그러나 속도식을 이용한 가상 미생물 시스템의 구축은 속도식에 필요한 모든 파라미터를 직접 실험으로 측정해야한다는 단점이 있다. 가장 적은 수의 지놈과 유전자를 가진 *Mycoplasma genitalium*는 적은 수의 파라미터와 속도식이 필요하기 때문에 전체 모델링이 가능했다.

만약 이 시스템을 4000개 이상의 유전자를 가진 *Escherichia coli*로 확대할 경우, 엄청난 수의 파라미터와 속도식들이 필요하며, 이를 실험에 의해서 모두 구하려면 많은 시간과 예산이 필요하다는 단점이 있다.

1.2. 대사 네트워크 기반 가상 미생물 시스템

상세한 속도식과 농도 데이터가 부족한 상황에서, 대사 네트워크를 이용한 흐름 평균 분석(flux balance analysis)은 가상 미생물 모델의 구축에 새로운 대안으로 떠오르고 있으며, 최근

주요 미생물에 관련된 genomic data의 급속한 증가와 많은 data를 정확하게 분석할 수 있는 전산학적 도구의 개발로 세포 전체의 대사 네트워크를 구축할 수 있게 됨으로써, 이 model의 적용 범위가 확대되고 있다.

흐름 평균 분석을 기반으로 하는 가상 미생물 시스템은 단지 대사경로의 양론계수, 대사 요구량, 그리고 기본적인 물리화학적 제약조건만으로, 세포의 거동, 대사 능력(metabolic capabilities), 그리고 원하는 성분을 생성시킬 수 있는 중요한 반응들을 정확하게 분석할 수 있다. 이 모델링에 관련된 연구는 캘리포니아 대학의 Palsson교수팀이 많은 연구결과를 발표하였으며, 1993년에 여러 가지 산소율(oxygenation rate)에서 대장균의 최적 흐름분배(optimal flux distribution)를 결정하였고, 이를 기초로 하여 시간에 따른 세포량, 기질의 농도, 대사 생성물의 농도변화를 예측하였다. 최근에는 흐름 평균 분석을 유전자 수준에서 염기 서열 수준으로, 미생물 대사 네트워크에서 인간 대사 네트워크로 그 영역을 확대하고 있다.

2. 흐름 평균 분석 기반 가상 미생물 모델

2.1. 대사 네트워크의 구축

흐름 평균 분석을 이용하여 가상 미생물 모델을 구축하기 위해서는 모든 대사 반응과 대사물질(metabolite)들이 포함된 대사 네트워크의 구축이 선행되어야 한다. 대사 네트워크는 완전히 서열화된 지놈으로부터 유전자를 찾고, 유사성 검색(homology search)을 통해서 각 유전자의 생산물들, 대사 효소(metabolic enzyme)들을 찾는다. 그런 다음, 이러한 대사 효소에 관련된

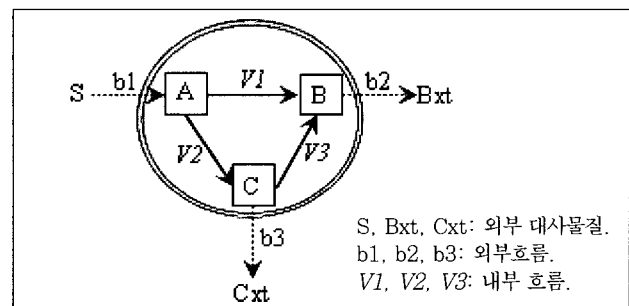


그림 1. 대사 네트워크.

대사 반응을 기능별로 분류하고 서로 연결시킴으로써, 대사 네트워크가 구축되어 진다(그림 1). 현재 주요 미생물의 대사 네트워크는 이미 KEGG와 MetaCyc에 구축되어 있다.

2.2. 양론과 흐름 행렬 구축

$$\begin{array}{l}
 \text{Balance Equation} \\
 \frac{dA}{dt} = -V1 - V2 + b1 \\
 \frac{dB}{dt} = V1 + V3 - b2 \\
 \frac{dC}{dt} = V2 - V3 - b3
 \end{array}
 \begin{array}{l}
 \left(\frac{dA}{dt} \right) \\
 \left(\frac{dB}{dt} \right) \\
 \left(\frac{dC}{dt} \right)
 \end{array}
 =
 \begin{array}{l}
 \text{In Matrix Form} \\
 \begin{pmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 \end{pmatrix}
 \cdot
 \begin{pmatrix} V1 \\ V2 \\ V3 \\ b1 \\ b2 \\ b3 \end{pmatrix}
 \end{array}$$

$$\frac{dA}{dt} = S \cdot V \xrightarrow{\text{steady state}} 0 = S \cdot V$$

그림 2. 물질수지식.

대사 네트워크의 구축으로 대사 반응과 대사물질이 정의 되면, 각 대사물질을 중심으로 물질 수지식(mass balance)을 세울 수 있다. 이때, 대사 체류시간(metabolic transients)이 환경변화와 성장율에 비해 매우 빠르기 때문에 정상상태로 가정할 수 있다. 이러한 정상상태 가정은 물질 수지식을 반응의 양론(stoichiometry)과 그 반응을 통해서 흐르는 흐름(flux)의 향으로 나타낼 수 있게 한다. 이 물질 수지식은 행렬표기법으로 일반화하여 나타낼 수 있으며(그림 2), S는 양론행렬(stoichiometry matrix), V는 흐름 행렬(flux matrix)을 의미한다.

가상 미생물 시스템에서 흐름 평균 분석의 주 목적은 대사 네트워크를 통해서 흐르는 최적의 흐름(optimal flux)을 예측하는 것이다. 그러나 일반적으로 미생물 시스템은 반응의 수가 대사 물질의 수 보다 많은 미결정적 시스템(underdetermined system)이기 때문에 추가적인 제한 조건들이 필요하다.

2.3. 추가적인 제한 조건 부가

가상 미생물 시스템은 일차적으로 정상상태의 물질 수지식들에 의해서 정의되고, 여기서 최적의 흐름을 예측하기 위해서 추가적인 제한 조건이 부가되어야 한다. 이러한 제한 조건에는 열역학적 조건과 이송능력(transporter capacity)이 있다. 열역학적 조건은 대사 네트워크 내에 있는 대사 반응의 가역 또는 비가역성 여부를 의미하고, 이송 능력은 외부 대사물질(extracellular metabolite)을 세포 내부로 흡수할 수 있는 양 또는 내부 대사물질(intracellular metabolite)을 세포 외부로 배출하는 양을 의미한다. 이러한 추가적인 제한 조건들은 미생물 대사에 관련된 지식과 실험에 의해서 결정되며, 정확한 수치로 정의되지 않고 대부분 허용할 수 있는 값의 범위로 주어진다.

2.4. 최적화

Range : $0 \leq b1 \leq 0.5, 0 \leq b2 \leq 1, 0 \leq b3 \leq \infty$
 Define an objective function : growth flux(V_{gro})
 e.g. maximize $V_{gro} = 3 \cdot V3 + V1$

Linear programming

$$V1 = 0, V2 = 0.5, V3 = 0.5$$

그림 3. 성장 흐름을 목적함수로한 최적화.

양론 행렬(S)과 흐름행렬(V)로 정의된 미생물 시스템은 흐름의 수가 더 많기 때문에, 이를 대수적으로 풀 수 없으며, 내부 흐름을 직접 실험으로 측정하는 것은 매우 힘들다. 직접 실험으로 측정하지 않고 이 흐름들을 계산할 수 있는 방법이 최적화(optimization)이다. 그러나 최적화를 실행하기 위해서는 대사 네트워크가 목적함수에 최적화된다는 가정이 필요하다. 만약 목적함수가 선형(linear)일 경우, 이 최적화문제는 선형계획법(linear programming)으로 해결된다(그림 3). 목적 함수가 최대 성장율일 경우, 식(1)과 식(2)의 형태로 나타낼 수 있다. 여기서 Z는 목적함수, V_{gro} 는 성장 흐름(growth flux)이고, M은 생합성 전구체(biosynthetic precursors), d_m 은 세포(biomass)를 생성하기 위한 M의 요구량을 각각 나타낸다.

$$\text{Minimize } Z = -V_{gro} \tag{1}$$

$$\sum dm \cdot M \xrightarrow{V_{gro}} \text{Biomass} \tag{2}$$

최대 성장율을 목적함수로 놓고 최적화할 경우, 최적 흐름 분배(optimal flux distribution)와 최대 성장율이 계산된다. 최적 흐름 분배는 물질 수지식과 추가적인 제약조건에 종속적이며, 최대 성장율은 시간에 따른 세포량, 기질농도, 대사 부산물의 농도를 예측하기 위해서 사용된다.

2.5. 성장속도식

$$X = X_0 \cdot e^{\mu \cdot \Delta t}$$

$$S_c = S_{c0} + \frac{S_{iu}}{\mu} (\mu_0 - \mu)$$

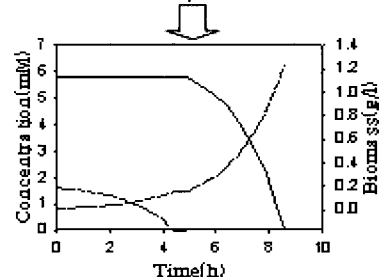


그림 4. 성장속도식에 의한 동적모사.

세포량, 기질농도, 대사 부산물의 농도를 예측하기 위해서, 그림 4의 성장속도식을 사용한다. 성장 속도식에서 초기 세포량 (X_0), 시간 구간(Δt), 기질 흡수율(S_u)은 정의된 값들이며, 성장률(μ)이 흐름 평균 분석에 의해서 계산된다. 이러한 데이터를 가지고 다음 시간 구간의 농도를 예측하기 위해서, 반복 알고리즘이 성장 속도식을 반복적으로 계산하게 된다. 최종적으로 작은 시간구간을 적분함으로써, 전 구간에 대해서 시간에 따른 농도 변화를 예측하게 된다. 이때, 시간구간 내에서는 흐름들(fluxes)이 동일하다고 가정하고 적분한다.

3. 계층적 유전자 조절 네트워크의 필요성

단지 흐름 평균 분석만을 이용한 가상 미생물 시스템은 흐름을 조절하는 다중 메카니즘(multiple mechanism)이 포함되지 않기 때문에, 실제 세포의 거동과는 상이한 결과를 생성시킬 수 있고, 이러한 문제점은 전역 조절(global regulation)이 발생할 경우 더욱더 커지게 된다. 대표적인 전역 조절로는 stimulon과 regulon이 있으며, 대사 네트워크의 내부 흐름에 영향을 주는 유전자를 조절하는 기능을 한다. 결국, 내부 흐름을 제어하는 유전자 조절 모델링은 가상 미생물과 실제 미생물 거동의 오차를 줄이는데 필수적인 요소이다.

시도하고 있다. 그러나 생성 규칙은 이진 시스템(binary system)이기 때문에, 복잡한 조절 단백질의 상호관계를 나타낼 수 없으며, 어떤 규칙이 적용되었는지 알 수 없기 때문에 유전자들의 조절 메카니즘을 알 수 없는 단점이 있다. 이러한 문제점들을 해결하기 위해서, (i) 조절 요소들(regulatory protein, genes, effectors etc.) 사이의 관계를 더욱더 유연하게 표현하기 위해서 결합의 정도를 나타내는 가중치 개념을 도입하고, (ii) 외부환경 변화에 따른 유전자들의 조절 메카니즘을 표현하기 위해서, 정의된 기호로 계층적 유전자 조절 네트워크(hierarchical regulatory network)를 모델링하는 것이 요구된다.

계층적 유전자 조절 네트워크는 그림 5에서 보는 바와 같이 stimulus, modulon, regulon과 같은 전역 조절이 모델링되어 있고, 이는 환경 변화에 따라서 내부 흐름을 제한하는 유전자의 활성/비활성의 메카니즘을 직관적으로 알 수 있게 한다. 만약 세포 외부 환경에 glucose와 fumarate가 함께 존재하는 경우, 정의된 기호와 가중치의 전파에 의해서 glucose가 먼저 대사된다. glucose가 대사되는 동안, fumarate를 대사시키는 효소 *fumB*와 *dcuB*가 비활성화되어 fumarate가 탄소원(carbon source)으로서 사용되지 않음을 알 수 있다. 만약 이러한 계층적 유전자 조절 네트워크를 포함하지 않는 가상 미생물 시스템은 glucose와 fumarate가 동시에 대사되어서 잘못된 예측결과를 생성시킨다.

이와 같이, 계층적 유전자 조절 네트워크는 첫째, 세포 외부 환경에 영향을 받는 유전자들이 어떻게 대사 흐름에 제한을 주는지를 알 수 있게 하고, 그러한 조절 제약 조건을 시스템에 부가함으로써, 기존의 시스템에서 발생 할 수 있는 대사 흐름분석의 오차를 바로잡을 수 있게 한다. 둘째, 환경 변화에 따른 대사 네트워크와 관련된 유전자들의 발현 패턴과 유전자 조절 메카니즘을 직관적으로 알 수 있게 한다. 셋째, 연속 논리 규칙(continuous logic rule)이며, 정의된 기호들로 모델링하였기 때문에, 예측 결과를 시간적으로 표현할 수 있게 한다.

4. 결론

포스트 지놈시대에 개발된 다양한 omics기술들은 세포 구성 요소들의 개별적 연구 관점에서 총체적 연구 관점으로 전환하게 하였다. 관점의 전환은 복잡한 생물학적 과정을 시스템 수준에서 분석할 수 있는 가상 미생물 모델을 필요하게 하였고, 이러한 상황에서 흐름 평균 분석을 이용한 가상 미생물 모델은 다른 가상 미생물 모델에 비해서 많은 장점을 가진다. 특히 구축 용이성 측면에서 보면, 단지 대사 반응 양론, 대사 네트워크, 종 특성 파라미터(strain specific parameter)와 같은 정보

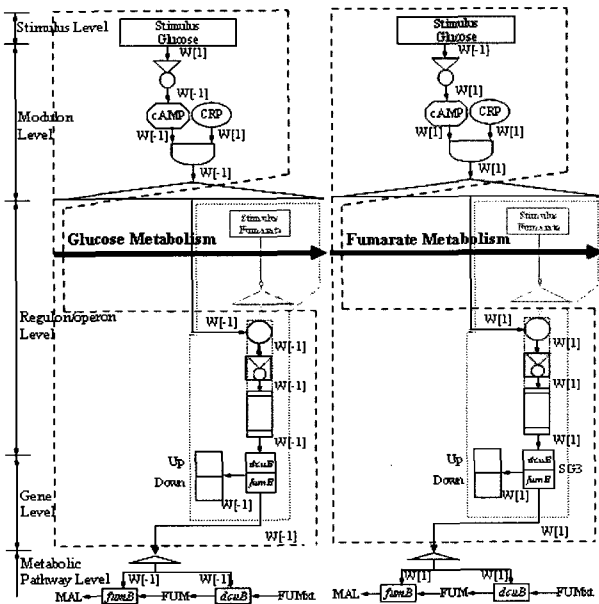


그림 5. 계층적 유전자 조절 네트워크.

최근에는 이러한 조절 제약조건(regulatory constraints)을 생성규칙(production rule-If then rule)과 불리언 논리(boolean logic)로 표현하여 흐름 평균 분석에 통합시키려고

만을 필요로 하기 때문에, 상대적으로 쉽게 구축과 확장(scale up)할 수 있다. 현재 흐름 평균 분석 기반 가상 미생물 시스템은, 산업적으로 유용한 미생물이고 생명현상 연구를 위한 모델 생명체인 대장균을 중심으로 연구가 활발히 진행되고 있다. 가까운 미래에 이 가상 미생물 시스템은 가상 모델생명체 역할을 수행함으로써, 산업균주의 개량, 새로운 생명현상의 규명, 유용한 유전자 확보뿐만 아니라, 나아가 신약 개발과 타겟 발굴연구로 그 활용범위가 확대될 것으로 기대된다.

참고문헌

- [1] J. S. Edwards, M. Covert, and B. Palsson, "Metabolic modelling of microbes: the flux-balance approach", vol. 4, no. 3, pp. 133-140, 2002.
- [2] K. J. Kauffman, P. Prakash and J. S. Edwards, "Advances in flux balance analysis", vol. 14, pp. 491-496, 2003.
- [3] S. G. Lee, "A study on *in silico* simulation of various carbon sources-grown *Escherichia coli* based on a hierarchical regulatory network and flux balance analysis", Doctoral Thesis, Pusan National University, 2005.
- [4] S. G. Lee, C. M. Kim, and K. S. Hwang, "Development of a software tool for *in silico* simulation of *Escherichia coli* using a visual programming environment", *Journal of Biotechnology*, vol. 119, issue 1, pp. 87-92, 2005.
- [5] S. G. Lee, K. S. Hwang, and C. M. Kim, "Dynamic behavior of regulatory elements in the hierarchical regulatory network of various carbon sources-grown *Escherichia coli*", *Journal of Microbiology and Biotechnology*, vol. 15, no. 3, pp. 551-559, 2005.

저자약력



《이성근》

- 1996년 부경대학교 화학공학과 학사.
- 1998년 부산대학교 화학공학과 석사.
- 2005년 부산대학교 화학공학과 박사.
- 관심분야 : 생물정보학, 시스템 미생물학.

• 연락처 : Tel: +82-51-510-3192

E-mail: lee_73@pusan.ac.kr



《한상일》

- 2003년 부산대학교 화학공학과 학사.
- 2005년 부산대학교 화학공학과.
- 관심분야 : 생물정보학, 시스템 미생물학.



《황규석》

- 1982년 부산대학교 화학공학과 학사.
- 1985년 일본 동경공업대학 화학공학과 석사.
- 1988년 일본 동경공업대학 화학공학과 박사.
- 관심분야 : 화학공장 자동화, 화학공정 전문가 시스템, 화학공정안전, 생물정보학.

• 연락처 : Tel: +82-51-510-2400

Fax: +82-51-512-8563

E-mail: kshwang@pusan.ac.kr