

생체내 반응 시뮬레이션을 위한 신호전달 네트워크 브라우저 개발

Development of a browser for signal transduction network to simulate biochemical reaction in a cell

유석중, 이상주
한국과학기술정보연구원

Yu Seok Jong, Lee Sang Joo
Korea Institute of Science and Technology
Information

요약

분자상호작용에 대한 실험방법이 소개되면서 이 분야의 자료가 빠르게 증가되고 있으며, 이와 관련된 데이터베이스는 급격히 증가하고 있다. 하지만 이와 같은 정보를 기반으로 질병의 이해와 같은 분자수준의 반응 기작 연구의 활용에는 어려움이 있었다. 본 연구 우리는 다양한 형태의 분자상호작용 정보를 불러들이고 이를 가시화 할 수 있는 브라우저를 개발하였다. 이 프로그램은 복잡한 구조의 상호작용을 좀더 쉽게 이해시킬 수 있도록 3차원 형태의 네비게이션 기능을 제공하며, 이를 통해 연구자가 자신의 원하는 상호작용 모델을 생성할 수 있도록 해 준다. 또한 얻어진 모델을 이용하여 손쉽게 분자 시뮬레이션을 수행할 수 있도록 시뮬레이션 기능을 제공한다. 대장균의 주화성에 대한 신호전달기작을 대상으로 시뮬레이션 분석과정을 테스트해 보았다.

Abstract

After introducing some experiment methods including immunoprecipitation and yeast two-hybrid screening, the pool of molecular interaction data is growing fast and databases are produced dramatically. But it is difficult to apply the information to molecular kinetic studies for understanding disease. In this paper, we developed a program that can browse and visualize interactions of cellular molecules using importing heterogeneous external data file. This program support 3D view to navigate and understand more easily and making a signal transduction model that user wants and simulating function to research the model. It was tested for signal transduction of chemotaxis in bacteria.

I. 서론

인간 유전체가 밝혀진 이후 다양한 종류의 유전체 연구가 진행되고 있다. 특히 유전체 연구는 단순한 단일 유전자를 대상으로 하는 기존 연구에서 벗어나 전체 유전자를 연구하는 것으로 이를 위해 바이오인포매틱스 분야의 연구가 뒷받침되어야 한다. 더욱이 유전체 연구가 대량의 실험 결과를 분석하는 것 뿐만 아니라 분석된 정보를 이용하여 그 정보를 가공 분석하는 2차 데이터베이스분야의 연구도 활발히 진행되고 있다 [1].

유전체 연구가 진행되면서 많은 과학자들은 질병의 원인이 단일 유전자로 인해 발생하는 것이 아니라 세포내부의 복잡한 단백질 상호작용 네트워크의 오류로 인해 발생한다는 사실을 입증 하였으며, 이와 관련된 다양한 지식기반 데이터베이스가 공개되었다. 일본 교토대학은 미생물의 세포내의 반응 네트워크를 작성한 KEGG[2]데이터베이스를 처음으로 현재 EcoCyc[3], BIND[4], DIP[5], MINT[6]등 다양한 데이터베이스가 공개되어 제공되고 있다.

이 같은 지식기반의 2차 데이터베이스는 다양한 정보를 연구자에게 제공함으로써 세포내부에서 일어나는 복잡한 생체

반응 기작을 이해할 수 있도록 돕고, 외부 신호를 세포가 어떻게 반응하는지를 파악할 수 있는 중요한 정보를 제공한다. 특히 신호전달 분야는 암을 비롯한 다양한 질병과도 밀접하게 연관되어져 있으며, 생체내의 하나의 유전자 변형으로 세포내의 네트워크가 어떻게 혼선을 야기하는지를 연구하는 분야로 신약개발과도 밀접하게 연관되어져 있는 분야이다[7].

신호전달 과정의 연구는 다양한 단백질들 간의 정보교환을 어떻게 수행하는지를 밝혀내는 것으로 이를 위해서 연구자들은 세포내부의 반응 네트워크를 이해해야만 한다. 이 과정에서 단백질 상호작용에 대한 정보를 연구자가 빠르고 정확하게 파악하기 위한 도구의 필요성에 대한 요구가 증가하고 있는 실정이다.

현재 Cytoscape[8]와 VisANT[9]가 단백질 상호작용에 대한 가시화 도구로써 사용되고 있으나, 단순한 단백질 네트워크에 가시화를 목적으로 개발되어져 있는 프로그램들로써 신호전달 연구자들이 이해하기 쉽도록 가시화 하는데에는 불편한 점이 있는 상황이다.

본 연구는 신호전달 분야의 연구자들에게 현재 밝혀진 신호전달 네트워크 및 관련된 상호작용 단백질들을 가시화할 수

있는 도구를 제공함으로써 연구자들의 연구 효율을 향상시키고 이를 기반으로 신호전달 네트워크가 실제 생체 내에서 어떤 반응을 유발시키는지를 시뮬레이션 할 수 있도록 하였다.

II. 방 법

2.1 단백질 상호작용 데이터

현재 공개된 단백질 상호작용 데이터는 XML을 기반의 데이터형태가 주류를 이루고 있다. 단백질 상호작용에 대한 국제적인 표준 데이터 포맷에 대한 논의가 이루어 졌으며, 2002년 HUPO에서 PSI (Proteomics Standards Initiative)가 시작되면서 구체화 되었다. 현재 PSI에서 제안된 MI(Molecular Interaction)포맷이 국제적으로 이용되고 있다.

본 연구에서는 이와 같은 PSI-MI 포맷의 데이터를 불러들여 상호작용을 가시화 할 수 있도록 시스템을 설계하였으며, 이를 기준으로 기존의 DIP, BIND등에서 제공하는 XIN포맷의 자료와 Cytoscape에서 제공하는 텍스트 기반의 파일 포맷(tabular format)도 지원할 수 있도록 구현하였다. 또한 최근 신호전달 분야의 정보 교환을 목적으로 표준화 작업이 진행되고 있는 Biological Pathway Exchnage(BioPAX)에 대한 데이터 모델도 불러들일 수 있도록 설계하였다.

2.2 상호작용 데이터 모델

단백질 상호작용을 가시화하기 위해서 각 상호작용의 기본 단위인 단백질 및 작은 분자들의 정보를 xml파일로부터 불러들여 이를 객체화 시켰다.

이를 설계하기 위해 기존에 연구된 단백질 객체 모델을 참고하였으며, 각각의 단백질의 정보는 PSI-MI에 정의된 속성을 기준으로 삼았다. 또한 텍스트 형태의 자료를 불러올 경우를 처리하기 위해 각각의 텍스트 정보를 객체정보로 매핑하였다.

2.3 시스템 구현

신호전달 브라우저를 구현하기 위해 본 연구에서는 자바 버전 1.5를 기반으로 한 SWT(Standard Widget Toolkit)를 이용 하였다. 또한 2차원 신호전달 가시화를 위해서 오픈소스로 제공되고 있는 JPowerGraph (<https://sourceforge.net/projects/jpowergraph>)를 사용하였으며 3차원의 가상공간을 구현하기위해서 OpenGL을 사용하였다. 또한 OpenGL을 이용한 3차원 가시화 라이브러리로 Lightweight java game library(<http://lwjgl.org>)를 이용하였다.

2.4 주화성(Chemotaxis) 시뮬레이션

박테리아의 주화성에 대한 시뮬레이션 과정은 Smoldyn을 이용하였으며 주화성에 대한 신호전달 과정과 각 반응의 반응속도는 기존 논문의 참고하였다[11].

III. 결과 및 토론

3.1 단백질 상호작용 데이터와의 인터페이스

신호전달 경로에 대한 정보를 사용하기 쉽고 서로 교환할 수 있도록 하기 위해 몇 개의 XML형식이 제안 되었다. 특히 이중에서도 Systems Biology Markup Language (SBML)과 CellML등은 경로 정보에 대한 수학적 모델을 제시하고 있으며, 이를 통해 다양한 시뮬레이션 연구가 진행되고 있다. 또한 다른 한편으로는 Proteomics Standard Initiatives's Molecular interaction (PSI-MI) 형식과 Biological Pathway Exchnage(BioPAX) 형식이 있는데, 이들은 분자의 상호작용 정보를 제공하는데 있다[7].

본 연구에서는 아직까지 완전한 통합형 데이터구조가 없는 상황에서 가능한 단백질의 상호작용을 표현할 수 있는 데이터 포맷을 지원하고자 하였다. 현재 PSI-MI 포맷에 대한 상호작용 정보를 불러들일 수 있도록 하였으며, BioPAX에 대한 지원도 가능하다. 이를 통해 현재 많은 단백질 상호작용의 배포 형태를 지원할 수 있으며, 이를 확장할 수 있도록 고려하여 구현하였다.

3.2 신호전달 네비게이션용 프로그램 개발

다양한 종의 유전체 염기서열이 발표되고, 이를 기반으로 단백질간의 기능에 대한 연구가 활발해 지면서 세포내의 단백질간의 상호작용에 대한 다양한 정보를 표시해주는 다양한 도구가 개발되고 있다. 특히 단백질간의 상호작용뿐만 아니라 최근에는 단백질과 유전체와의 상호작용 단백질과 작은 분자와의 상호작용 등등 그 범위가 확대되고 있으며, 유전체와 유전체간의 상호작용인 유전자 조절 분야에서도 알려진 정보를 표시해 주는 것이 중요한 시점이 되었다.

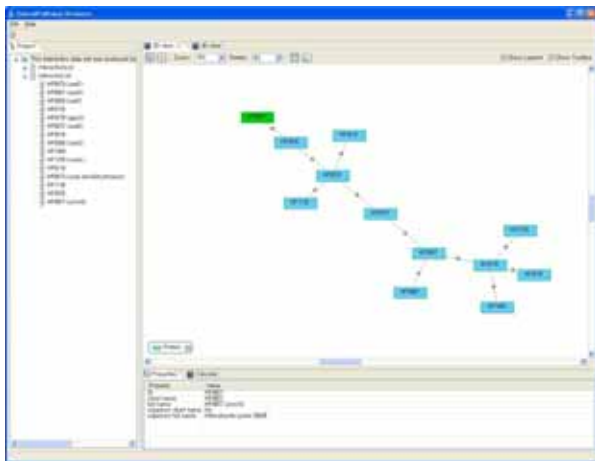
이 같은 이유는 상호작용의 대상이 광범위하고 복잡하기 때문이다. 특히 최근 경향을 반영한 다양한 형태의 상호작용을 모두 표시하는 것은 사용자 측면에서는 모든 내용이 표시되는 것이 오히려 복잡해 보이기 때문에 원하는 정보를 간결하면서 명확히 표현해 주는 것이 필요하다. 이러한 관점에서 현재 제공되고 있는 다양한 가시화도구는 그 한계점이 드러나고 있다.

현재 많이 사용되고 있는 단백질 상호작용 가시화 도구인 Cytoscape[8], VisANT[9]등등은 2차원을 기반으로 구현되어져 있으며, 단순한 상호작용의 가시화에는 적합하지만 대

량의 복잡한 구조의 네트워크의 표시에는 2차원의 한계점을 드러내고 있다. 사용자가 상호작용을 알아볼 수 없을 정도로 복잡하게 교차하는 선들은 상호작용의 내용을 파악하기 힘든 실정이다.

본 연구에서는 기존에 많이 쓰이고 있는 PSI-MI 형식을 불러들여 신호전달과정에 참여하는 단백질의 정보를 표시하고 이를 가시화 할 수 있는 네비게이션 프로그램을 개발하였다(그림 1). 구현된 2차원 가시화기능은 상호작용 종류별 레이어를 만들 수 있어 각 레이어별로 내용을 볼 수 있도록 하였다. 이 방법을 응용하면 다양한 신호전달의 교차를 중복적으로 표시하고, 이를 사용자가 원하는 신호전달 과정만 볼 수 있음으로써 복잡한 화면을 단순화 할 수 있을 것으로 기대된다.

특히 JPowerGraph라이브러리를 이용하여 사용자가 원하는 방향으로 단백질을 이동시키면 네트워크 내의 다른 단백질들이 같이 상호작용하면서 이동할 수 있도록 하여 보다 동적인 이미지를 제공할 수 있다.



▶▶ 그림 1. 신호전달 브라우저의 2차원뷰

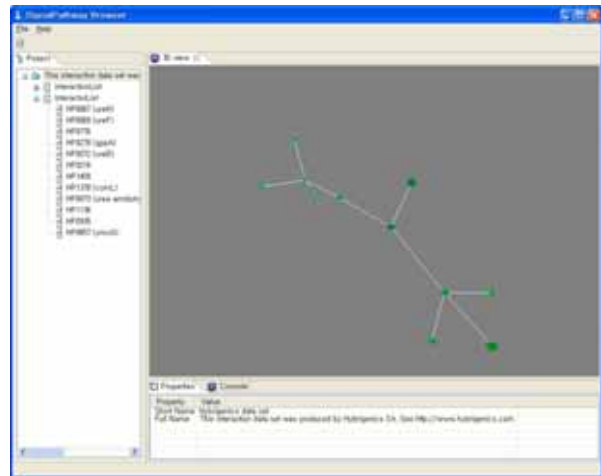
3.3 3차원 가시화

이미 알려진 가시화도구의 한계를 벗어나고자 3차원의 가시화 기법을 도입하였다. 3차원 가시화는 현재 2차원에 표시되어지는 상호작용을 3차원의 스프링 모델을 이용하여 3차원상의 최적 위치를 찾아갈 수 있도록 구현하였다.

어떤 복잡한 상호작용 구조도 3차원 내에 최적화하여 볼 수 있으며, 이 상호작용 구조를 3차원적으로 돌려 볼 수 있음으로써 다양한 시점에서 상호작용을 파악할 수 있도록 하였다(그림 2).

구현된 3차원 신호전달 가시화를 기반으로 향후 다양한 기능의 추가가 가능할 것으로 예상된다. 특히 구조 정보를 표시하여 신호전달 과정에 단백질간의 상호작용이 어떤 구

조에서 이루어지는지 연구자들이 쉽게 이해할 수 있을 것이다.



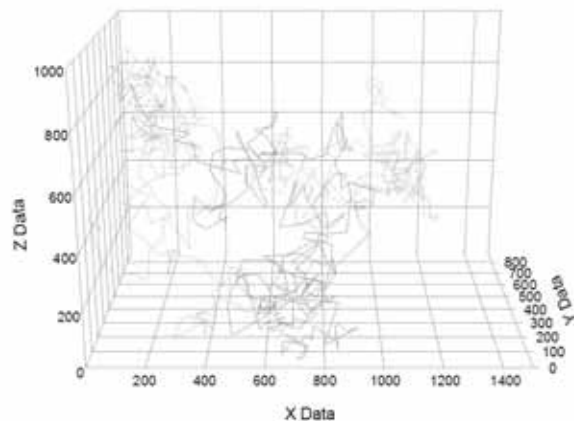
▶▶ 그림 2. 신호전달 브라우저의 3차원뷰

3.4 박테리아의 주화성(Chemotaxis) 시뮬레이션

미생물의 먹이를 향해 이동할 수 있는 신호전달 기작이 존재한다. 현재까지 연구결과 박테리아의 경우 세포질내의 다수의 단백질들(CheA, CheW, CheY, CheZ, CheR, CheB)의 상호작용이 확인되었으며, 외부 화학물질을 인지하는 수용체(Tsr, Tar, Tap)가 밝혀졌다[10]. 또한 이에 대한 신호전달과정 및 각 과정의 반응속도도 비교적 상세히 알려져 있다[11].

본 연구에서는 구성된 박테리아의 주화성 신호전달 경로를 신호전달 브라우저를 통해 형성하였으며, 이를 세포내 시뮬레이션에 이용하였다.

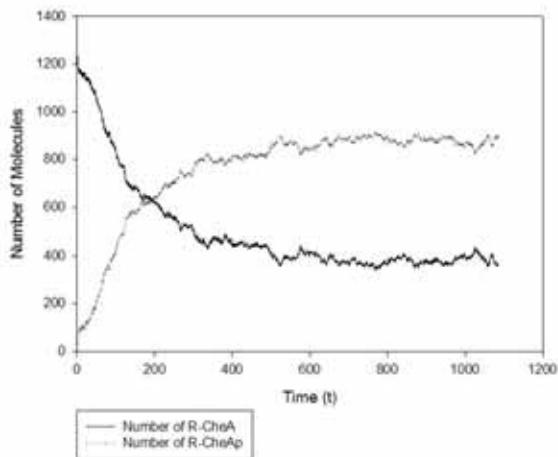
먼저 세포내의 반응을 시뮬레이션하기위한 세포의 물리적 환경을 구현하기위해 박테리아의 체적과 유사한 직사각형의 입방체(길이 2 μ m, 직경 0.84 μ m)를 정의하였으며, 각 반응의 반응속도를 입력하였다.



▶▶ 그림 3. 세포내의 CheY이동

시뮬레이션 결과로써 그림3은 주화성의 과정중 세포질내의 CheY가 세포막으로 이동하여 인산화된 이후 다시 세포질내로 이동하다가 섬모를 만나기까지의 이동과정을 추적한 그림이다. 이 과정에서 단백질은 브라운 운동을 하고 있음을 확인할 수 있다.

또한 그림 4는 시뮬레이션 결과를 통해 세포막의 수용체와 CheA가 화학물질을 감지하여 인산화되는 과정을 시간에 따라 추적한 것으로 이러한 시뮬레이션 결과를 통해 세포내의 화학반응을 통해 박테리아가 어떻게 먹이를 향해 이동하는지를 시뮬레이션 할 수 있으며, 더 나아가 각각의 단백질들의 변화에 따른 박테리아의 주화성 반응을 예상할 수 있는 기반을 마련해 줄 수 있다. 이를 확장한다면 인간의 세포내에서 발생하는 신호전달 과정을 이와 같은 형태로 시뮬레이션 해 볼 수 있으므로 신호전달 과정에 중요한 단백질을 찾아낼 수 있으며 더 나아가 향후 신약개발과 같은 고 부가가치 산업에 이용될 수 있을 것으로 기대된다. 특히 신호전달과정은 복잡다단한 단백질 상호작용 네트워크 속에 일부이다. 이러한 신호전달 과정의 최종단계는 세포의 핵내의 유전자 발현이므로 향후 유전자 발현 연구에도 응용될 수 있을 것이다.



▶▶ 그림 4. CheA의 인산화 반응 시뮬레이션

■ 참고 문헌 ■

- [1] Kanehisa, M. and Bork, P., "Bioinformatics in the post-sequence era", *Nature Genetics*, Vol 33, pp. 305-310
- [2] Ogata, H., Goto, S., Sato, K., Fujibuchi, W. Bono, H. and Kanehisa, M., "KEGG: Kyoto Encyclopedia of Genes and Genomes", *Nucleic Acids Res.*, Vol. 27, No. 1, pp. 29-34
- [3] Karp, P.D., Riley, M., Paley, S.M. and Pellegrini-Toole, A., "EcoCyc: an encyclopedia of Escherichia coli genes and metabolism", *Nucleic Acids Res.*, Vol. 24, No. 1, pp. 32-39
- [4] Bader, G.D., Betel, D. and Hogue, C.W.V. (2003) "BIND: the Biomolecular Interaction Network Database", *Nucleic Acids Res.*, Vol. 31, pp. 248-250
- [5] Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M. and Eisenberg, D. "DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions", *Nucleic Acids Res.*, Vol. 30, pp. 303-305
- [6] Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M. and Cesareni, G. "MINT: a Molecular INTERaction database", *FEBS Lett.*, Vol. 513, pp. 135-140
- [7] Helmreich, E.J.M., "The Biochemistry of Cell Signalling", pp. 3-23, Oxford, 2001.
- [8] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks", *Genome Res.*, Vol 13, pp 2498-2504
- [9] Hu, Z., Mellor, J., Wu, J. and DeLisi, C. "VisANT: an online visualization and analysis tool for biological interaction data", *BMC Bioinformatics*, Vol 5, pp 17-25
- [10] Parkinson, J. S., and Kofoid, E. C., "Communication modules in bacterial signaling proteins", *Annu. Rev. Genet.*, Vol 26, pp 71-112
- [11] Lipkow, K., "Changing cellular location of CheZ predicted by molecular simulations", *PLoS Computational Biology*, Vol 2, pp 301-310