

제약조건에 기반한 동적 단백질 상호작용 네트워크

한동수, 정석훈[○], 이춘오, 장우혁

한국정보통신대학교

{dshan, jsh[○], lcol, torajim}@icu.ac.kr

Constraints Based Dynamic Protein Interaction Network

Dong-Soo Han, Suk-Hoon Jung[○], Choon-Oh Lee, Woo-Hyuk Jang
Information and Communications University

요 약

본 논문에서는 단백질 상호작용 네트워크의 복잡성을 제어하고 생물학자가 자신이 설정한 조건을 만족시키는 환경에서 추가적인 다양한 제약 조건을 가하면서 원하는 상호작용 네트워크를 구성하고 조작할 수 있도록 지원하는 Constraints Based Dynamic Protein Interaction Network 이라는 새로운 개념의 단백질 상호작용 네트워크를 소개한다. 본 기법에서는 기존의 단백질 상호작용 네트워크에서 주로 사용하는 단백질 상호작용 정보뿐 아니라 단백질 상호작용에 영향을 미칠 수 있는 개개 단백질의 물리 화학적 특성 및 위치 정보와 상호작용의 환경 정보도 단백질 상호작용 네트워크 구성에 활용한다. 제안된 네트워크상에서 생물학자는 단백질 상호작용 네트워크 구성 조건을 변경하거나 얻어진 네트워크에 변경을 가하면서 점차 자신이 원하는 의미 있는 대사경로 모델을 찾거나, 제약조건의 다양한 조작을 통하여 생물학적 실험을 통하여 얻어진 대사모델의 유효성을 검증하는 것도 가능하다.

1. 서 론

단백질 상호작용 정보는 미지의 단백질 기능 추정에서부터 단백질 상호작용을 통해서 이루어지는 대사 경로 예측에 이르기까지 다양하게 활용될 수 있다. 따라서 그 동안 많은 단백질 상호작용 예측 실험을 통하여 다량의 단백질 상호작용 정보가 축적되었다. 또한 축적된 단백질 상호작용 데이터를 이용한 미지의 단백질 상호 작용 예측 기법도 활발히 개발되어 왔다.

단백질 상호작용 정보가 앞서 기술한 단백질 기능과 대사 경로 예측에 실질적인 도움을 주기 위해서는 단백질 상호작용 정보를 이용한 단백질 상호작용 네트워크 구성 및 시각화를 지원하는 도구의 지원이 절실하다. 일반적으로 생물학자 또는 단백질 상호작용 정보를 활용하는 연구자들은 단편적인 단백질 상호작용 정보로부터 특정 바이러스에 포함된 단백질들이 특정 장기 내에 있는 단백질 집단과 상호작용을 일으키는 패턴을 보여주는 단백질 상호작용 네트워크를 통해서 더 유용한 정보를 얻는 것이 보통이다[1],[2].

그 결과 많은 다양한 단백질 상호작용 네트워크 구성 기법과 그것으로부터 유용한 정보를 추출하는 기법들이 개발되어 왔다. 개발된 단백질 상호작용 네트워크 구성 틀들은 많은 경우 단백질 상호작용 정보의 시각화 관점에서 단백질 상호작용 네트워크를 단순화 계층화하는 데 집중하여

왔다. 하지만 이와 같은 기존의 단백질 상호작용 네트워크 구성 도구를 통해 생물학자가 원하는 유용한 정보를 추출하기는 쉽지 않다.

단백질 상호작용 네트워크로부터 유용한 정보를 얻고자 하는 생물학자들이 당연한 위와 같은 문제점을 해결하기 위하여 본 논문에서는 생물학자로부터 제공되는 정보를 조건의 형태로 받아들이고 제시된 조건에 부합하는 단백질 및 단백질 상호작용 네트워크를 구성하여 보여주는 단백질 상호작용 네트워크 구성 기법 Constraints Based Dynamic Protein Interaction Network를 소개한다.

생물학자들은 제안된 네트워크를 이용하여 필요한 조건을 제시하고 이에 따른 단백질 상호작용 네트워크 구성의 변화를 보면서 단백질 상호작용 네트워크로부터 얻고자 하는 최종적인 정보를 추출하게 된다. 이를 위해서 본 기법은 단백질 상호작용 네트워크를 구성하는 단백질을 단순히 시각화를 위한 하나의 노드로 간주하지 않고 단백질의 다양한 특성 정보를 포함하는 하나의 객체로 인식한다는 점에서 기존의 단백질 상호작용 네트워크 구성 방식과 구분된다. 네트워크를 구성하는 단백질을 하나의 객체로 인식하여 취급함으로써 얻을 수 있는 장점은 다양하지만 무엇보다도 해당 단백질 상호작용 네트워크를 사용자 요구에 따라 유연하게 확장하거나 적용시킬 수 있다는 강점이 있다.

2. Constraints Based Dynamic Protein Interaction Network

2.1 제안 배경

전통적인 단백질 상호작용 네트워크 구성 기법은 주어진 단백질 상호작용 정보를 네트워크 형태로 연결하여 시각화하고 해당 네트워크로부터 얻을 수 있는 다양한 정보를 추출하여 보여 주는 기능을 포함하고 있다. 전통적인 단백질 상호작용 네트워크 구성에 있어서 사용자의 역할은 다소 수동적인 입장에서 도구에서 제공하여 주는 정보를 받아서 필요한 정보를 추출하는 수준에 머물렀다. 그 결과, 생물학자들이 기존의 단백질 상호작용 네트워크 구성 도구를 사용하여 다양한 정보를 유연하고 효과적으로 추출하는 데 있어서 많은 어려움이 있었다. 이러한 어려움을 극복하기 위해서 본 논문에서는 Constraints Based Dynamic Protein Interaction Network을 제안한다.

2.2 특징 및 설계목표

제안된 네트워크 기법은 다음과 같은 특성과 설계 목표를 가지고 있다. 첫째, 사용자들과 네트워크의 적극적인 상호작용을 통하여 단백질 상호작용 네트워크를 구성하고 변경하는 절차를 반복하는 과정에서 궁극적으로 자신들의 원하는 정보를 추출할 수 있도록 한다. 그러기 위해서 본 네트워크 기법 상에서는 사용자가 자신이 가지고 있는 단백질 관련 정보 및 자신이 구성하고자 하는 단백질 상호작용 네트워크가 만족시켜야 하는 조건을 제공하고, 네트워크 틀은 사용자로부터 주어진 조건정보를 만족시키는 단백질 상호작용 네트워크를 구성한다. 사용자는 구성된 단백질 상호작용 네트워크를 보면서 필요한 정보를 추가로 제공하거나 조건을 변경시켜 가면서 궁극적으로 자신이 알고자 하는 단백질 상호작용 네트워크 또는 대사경로를 찾게 된다. 사용자가 제공하는 정보로는 특정 바이러스에 포함된 단백질 리스트, 상호 작용할 단백질들의 조직상 위치 정보 및 기타 단백질 상호작용을 일으키는 경우에 만족되어야 하는 환경 조건 등 다양하다. 표 1은 본 네트워크 기법에서 사용되는 제약 조건 및 틀이 제공해주는 정보를 나타내고 있다.

표 1 제약조건 및 일반정보

	제약조건	일반정보
단백질 관련	family, organism, location, tissue	cross ID, alias, description, gene, domain, motif, publication
상호작용 관련	temperature, pH, salt concentration, experimental evidence, experimental system,	binding site, bait condition, publication, description

둘째, 본 네트워크 기법에서는 실험을 통하여 얻어진 단백질 상호작용 정보뿐 아니라 예측을 통하여 얻어지는 단백질 상호작용 예측 정보도 적극적으로 수용하여 반영한다.

현재 본 기법은 Han's 연구 그룹에서 제안되었고 Yeast 에서 높은 예측 정확도를 보여주고 있는 도메인 조합 기반 단백질 상호작용 예측 기법을 통하여 제공되는 단백질 상호작용 정보를 적극 활용하고 있다[3],[4],[5]. 사용된 예측 기법은 도메인 조합 정보를 기반으로 하고 있어, 해당 단백질의 도메인 정보만으로 기존 단백질 및 새로이 발견된 단백질에 대하여 상호작용 예측 정보를 제공한다.

셋째, 단백질을 상호작용 네트워크를 구성하는 단순한 노드로 취급하지 않고 해당 단백질 관련 정보를 담고 있는 객체로 간주하고 새로운 정보가 얻어지는 경우에는 해당 객체의 속성 값을 추가함으로써 쉽게 반영할 수 있도록 하고 있다. 또한 만들어진 단백질 상호작용 네트워크에 새로운 단백질 또는 단백질 상호작용 정보를 손쉽게 추가하거나 삭제하여 반영할 수 있도록 하였다.

마지막으로 Constraints Based Dynamic Protein Interaction Network 는 구성된 단백질 상호작용 네트워크를 계층화 단순화하는 기능과 해당 상호작용 네트워크에서 Hub 단백질을 추출하는 기능 등도 갖추고 있다.

2.3 구현

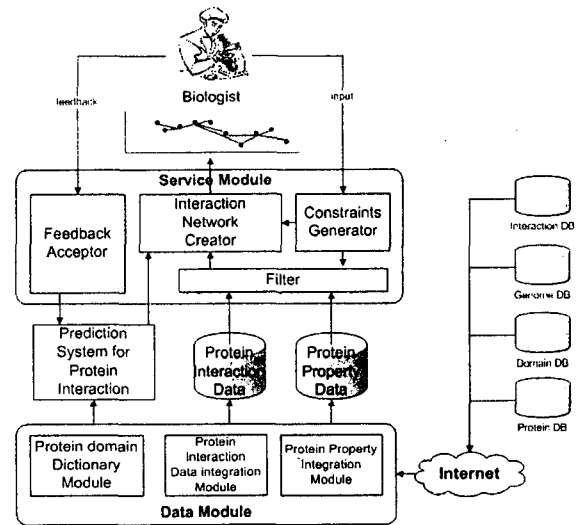


그림 1. Constraints Based Dynamic Protein Interaction Network의 구성

Constraints Based Dynamic Protein Interaction Network는 크게 데이터 모듈과 서비스 모듈로 구성되어 있다. 데이터 모듈은 인터넷 상에 산재된 관련 데이터들을 수집하고 통합하여 서비스 모듈에 기초 데이터를 제공한다. 인터넷 상에 분산된 데이터들은 그 특성상 데이터의 내용과 구조가 다양하기 때문에, 데이터 모듈은 이에 맞는 여러 기법들을 사용하여 정보의 수집 및 통합을 수행한다.

서비스 모듈은 데이터 모듈로부터 주어진 정보를 기초로 하여 사용자의 제한조건 설정에 따른 네트워크를 제공한다. 단백질 상호작용 예측 시스템에서 주어진 예측 정보 또한 네트워크 구성의 기초 데이터로 사용되며, 서비스 모듈은

새로운 단백질에 대한 상호작용 예측 정보를 예측 시스템에 추가적으로 문의할 수도 있다. 결과적으로 서비스 모듈은 사용자와의 동적인 피드백을 통하여 점진적으로 사용자가 원하는 네트워크 구성을 가능케 한다. 그림 1은 Constraints Based Dynamic Protein Interaction Network 구성의 개요를 보여주고 있다

그림 2는 제안된 네트워크 기법의 사용자 인터페이스 현황을 보여주고 있다. 사용자들은 좌측 상단의 패널과 검색 상자를 통하여 네트워크 구성 및 확장을 위한 특정 단백질, 또는 단백질 그룹을 추가할 수 있다. 좌측 하단의 패널은 선택된 단백질이나 단백질 상호작용의 정보를 보여준다. 중앙 패널은 생성된 네트워크를 보여주며, 사용자는 우측 상단 패널을 통하여 네트워크 구성에 사용된 제한 조건들을 열람하거나, 새로이 제한조건을 설정할 수 있다. 본 네트워크 기법에서는 사용자가 인식하기 용이한 네트워크의 시각화를 위하여 Fruchterman-Reingold Force 모델[6]을 사용하였다.

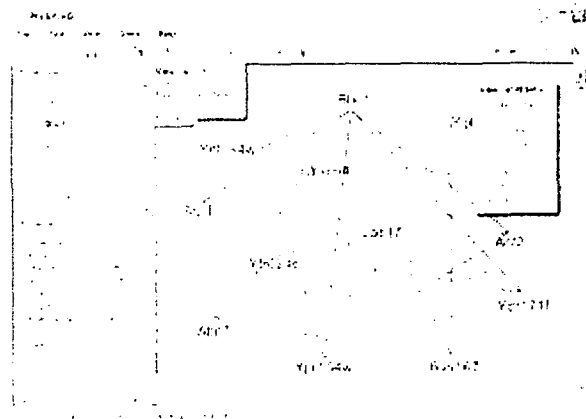


그림 2. 사용자 인터페이스

3. Summary

본 논문에서는 새로운 개념의 단백질 상호작용 네트워크 구성 기법인 Constraints Based Dynamic Protein Interaction Network 를 제안하였다. 제안된 네트워크상에서 생물학자들은 관심 있는 특정 단백질 집단을 중심으로 다양한 제한조건을 제시하면서 단백질 상호작용 네트워크를 구성하는 것이 가능하다. 본 기법은 단순히 주어진 단백질 상호작용 정보를 기반으로 네트워크를 구성하고 시각화하기 보다는 사용자와의 동적인 상호작용을 통하여 점진적으로 원하는 단백질 상호작용 네트워크를 구성할 수 있도록 도와준다는 점에서 기존의 단백질 상호작용 네트워크 구성 방식과는 구분된다. 또한 도메인 조합에 기반한 단백질 상호작용 예측 기법을 통하여 예측된 단백질 상호작용 데이터를 사용한다는 점에서도 기존의 방식과 구분되는 특징을 가지고 있다. 향후에는 본 논문에서 제안된 네트워크 기법을 사용하여 의미

있는 단백질 상호작용 네트워크를 구성하는 일반적인 절차를 고안하는 것이 필요하다. 또한 구체적인 바이러스 또는 질병을 일으키는 단백질 집단을 대상으로 고안된 절차를 적용하여 의미 있는 단백질 상호작용 네트워크 구성이 도출되는 지에 관한 검증도 필요하다.

Reference

- [1] On the structure of protein-protein interaction Networks A. Thomas, R. Cannings, N.A.M. Monk, and C. Cannings. Biochemical Society Transactions (2003) Volume 31, part 6.
- [2] The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes. Andreas Wagner; Mol. Biol. Evol. 18(7):1283-1292. 2001.
- [3] Dong-Soo Han and et al. Domain combination based probabilistic framework for protein-protein interaction prediction. Genome Informatics, 14:250-259, 2003.
- [4] Dong-Soo Han and et al. PreSPI: a domain combination based prediction system for protein-protein interaction. Nucleic Acids Research, 32(21), 2004.
- [5] Dong-Soo Han and et al. PreSPI: Design and implementation of protein-protein interaction prediction service system. Genome Informatics, 15(2):171-180, 2004.
- [6] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. Software -Practice and Experience, 21(11):1129-1164, 1991.