

# 유전자 온톨로지와 연계한 단백질 상호작용 네트워크 시각화 시스템에 관한 연구

최윤규<sup>○</sup> 김석 이관수 박진아

한국정보통신대학교 공학부

[ckyun777@gmail.com](mailto:ckyun777@gmail.com), [gstone0103@icu.ac.kr](mailto:gstone0103@icu.ac.kr), [gsyi@icu.ac.kr](mailto:gsyi@icu.ac.kr), [jinah@icu.ac.kr](mailto:jinah@icu.ac.kr)

## Research on Protein Interaction Network Visualization Tool Combined with Gene Ontology

YunKyu Choi<sup>○</sup> Seok Kim Gwan-Su Yi Jinah Park

Information and Communications University

단백질 상호작용 네트워크는 어떤 단백질들 간에 상호 작용 관계가 있는지를 네트워크 형태로 나타낸 것이며 단백질 상호작용을 발견하거나 분석하는 것은 생명 공학에서 중요한 연구분야이다. 본 논문에서는 방대한 단백질 상호작용 데이터를 유전자 온톨로지와 연계한 시각화를 통하여 효과적으로 직관을 얻을 수 있는 효율적인 단백질 상호작용 네트워크 분석시스템 개발을 연구 하였다. 단백질 상호작용 네트워크를 시각화 하고 분석하는데 기본적인 문제점은 상호작용 네트워크 데이터양의 크기라고 할 수 있다. 이를 개선하기 위하여 단백질 네트워크 데이터를 동적이고 상호작용 가능한 FDP(Force-Directed Placement) Layout 알고리즘을 적용하였고, 관심 노드와 그 주변 노드를 표시하며 점진적으로 탐색할 수 있는 Context-based browsing 기법을 구현하였다. 또한 유전자 온톨로지를 각 노드에서의 네트워크 확장을 조합하고 동적으로 구성되는 그래프로 시각화하여 이와 연계하여 분석을 가능하게 함으로써 단백질 상호작용에 관한 분석을 용이하게 하고자 하였다. 이를 효과적으로 사용하게 하기 위하여 각각의 시각화 결과를 2차원 윈도우로 나란히 보여주는 인터페이스를 디자인 하였다.

### 1. 서 론

단백질 상호작용(Protein-Protein Interaction: PPI) Network는 각 단백질을 Node로 단백질 사이의 상호작용 유무를 Edge로써 표현한 Network이고, 이를 분석하는 일은 신약 개발 등에 이용될 수 있다.

유전자 온톨로지(Gene Ontology: 이하 GO)는 생물의 유전자의 정보를 나타내기 위한 전산화된 국제적인 데이터 모델로써, 생물학적 어휘나 개념, 그리고 이들 사이의 관계로 구성된다.

단백질 상호작용을 분석하고 시각화 해주는 시스템에 대한 연구는 국내외에서 꾸준히 연구되고 있으며 국내에서 연구된 시스템으로는 Protenica[1], PIVS[2], CPIN[3], BioINet[4][5]등이 있으며 국외의 사례로는 Osprey[6], Cytoscape, BiNGO[7]등이 있다. 이러한 시스템들은 각기 특징을 가지고 있으며 일부는 GO와 연계하여 PPI를 군집화 해주고 시각화와 분석 기능을 가지고 있지만, GO와 PPI를 동시에 그래프로 시각화 해서 보여주며 서로 쉽고 빠르게 연계하여 사용 할 수 있는 기능의 지원은 부족하다. 이러한 점들을 보완하여 GO와 PPI를 연계하여 단백질 상호작용에 대한 직관을 얻고 원하는 단백질 또는 단백질 상호작용을 찾는데 도움을 주고자 PINGO(Protein Interaction aNd Gene Ontology) 시스템을 개발하게 되었다. 이 논문에서는 GO와 PPI를 동시에 효율적으로 시각화 하고 연계하여 분석하기 위하여 효율적인 PPI 네트워크, GO의 시각화에 관한 연구와 둘을 연계하여 분석 가능하게 해주는 시스템 구현에 대한 내용을 다룬다.

### 2. 구현

본 연구에서 개발한 시스템인 PINGO (Protein Interaction aNd Gene Ontology)의 전체적인 구성은 그림 1과 같다. 그림 1에서와 같이 GO에 관한 부분을 왼쪽에 PPI에 관한 부분을 오른쪽에 배치하였고, GO와 PPI측의 동일/유사한 기능을 좌우로 나란히 배치하여 쉽게 이해하고 사용할 수 있도록 하였다. PINGO 시스템은 <http://cgv.icu.ac.kr/pingo/> 에서 다운로드 가능하다.

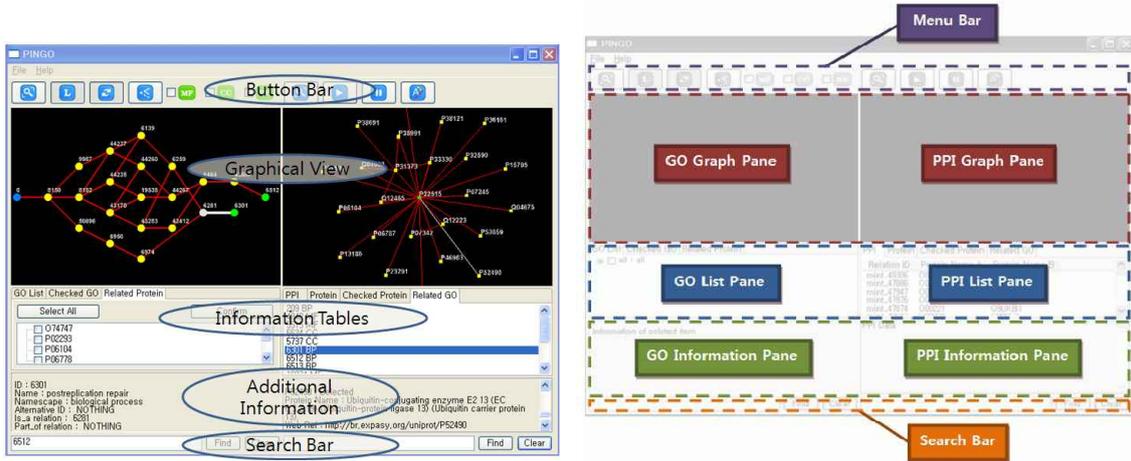


그림 1. PINGO 인터페이스

각 부분별 기능은 다음과 같다.

- (1) Button Bar: View 조작, 각종 기능 버튼 및 체크 박스
- (2) Graphical View: GO와 PPI를 각각 시각화된 그래프 형태로 표시해 주는 부분
- (3) Information Tables: GO와 PPI에 관한 Tree List, 또는 테이블 형태의 텍스트 데이터를 표시해 주는 부분
- (4) Additional Information: 선택된 노드의 정보와 프로그램 실행상태 관련 정보를 표시해주는 부분.
- (5) Search Bar: 각각 GO ID와 단백질을 검색해서 찾아주는 부분

시스템 구조는 GO측과 PPI측을 독립적 모듈로 구성하고 상위 GO-PPI 연동 계층에서 인터페이스를 통하여 양측을 모두 사용하는 기능을 수행하게 디자인 하였다. 이러한 연동은 GO Annotation (<http://www.geneontology.org>) 데이터를 통하여 하며, GO와 PPI간의 상호 참조를 사용함으로써 관심있는 기능이나 분류로부터 표시 할 단백질들을 선택하거나 표시된 PPI 네트워크의 의미를 분석하는 것을 도울 수 있다. 또한 효율적인 분석과 시각화를 위하여 GO측에서는 관심 GO용어들의 LCA (Least Common Ancestor)까지만 찾아서 표시하는 기능, PPI측에서는 Context-based browsing, 동적이며 상호작용 가능한 네트워크 등을 구현 하였다.

### Acknowledgement

이 논문은 한국과학재단 기초과학연구사업인 ‘유비쿼터 시스템 기반 단백질 조절 네트워크 지식 발굴 시스템’(R01-2005-000-10824-0)의 지원을 받았다.

### 5. 참고 문헌

- [1] Hee-Jeong Jin, Ji-Hyun Yoon, and Hwan-Gue Cho, "An Analysis System for Protein-Protein Interaction Data Based on Graph Theory", 한국정보과학회, 2006
- [2] Mi-Kyung Lee and Ki-Bong Kim, "A Visualization and Inference System for Protein-Protein Interaction", 한국정보과학회, 2004
- [3] Dong-Soo Han, Suk-Hoon Jung, Woo-Hyuk Jang, and Choon-Ho Lee, "Constraints Based Dynamic Protein Interaction Network", 한국정보과학회, 2005
- [4] <http://www.meta-biz.net/html/product03.htm>
- [5] SunLee Bang, JaeHun Choi, JongMin Park, SooJun Park, "단백질 상호작용 네트워크의 개념 분류 레이아웃", 한국정보과학회 학술발표논문집, pp. 61-63 (3 pages), 2006
- [6] Bobby-Joe Breitkreutz, Chris Stark, Mike Tyers, "Osprey: a network visualization system", Genome Biology, 2003
- [7] Steven Maere, Karel Heymans and Martin Kuiper, "BINGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks" Bioinformatics, 2005