

# 단백질 허브 네트워크에서 도메인분석을 통한 단백질 기능발견 시스템

## Protein Function Finding Systems through Domain Analysis on Protein Hub Network

강태호\*, 류제운\*\*, 유재수\*, 김학용\*\*  
충북대학교 전기전자컴퓨터공학부\*, 충북대학교 생화학과\*\*

Tae-Ho Kang(thkang@netdb.cbnu.ac.kr)\*, Jea-Woon Ryu(aries84@naver.com)\*\*,  
Jae-Soo Yoo(yjs@chungbuk.ac.kr)\*, Hak-Yong Kim(hykim@chungbuk.ac.kr)\*\*

### 요약

본 논문에서는 단백질-단백질 상호작용과 도메인 분석을 통해 기능이 알려지지 않은 미지 단백질의 기능을 예측할 수 있는 알고리즘을 제안한다. 먼저 MIPS 데이터베이스로부터 효모에 대한 단백질-단백질 상호작용(PPI) 네트워크를 구축한다. 구축된 PPI 네트워크는(단백질 3,637개, 상호작용 10,391개) 많은 상호작용을 갖는 소수의 단백질들을 갖고면서 단백질 클러스터의 고유한 모듈성을 보이는 스케일 프리 네트워크와 계층적 네트워크의 특성을 보인다. 단백질-단백질 상호작용 데이터베이스는 Y2H(Yeast Two Hybrid) 실험 등으로 얻어졌기 때문에 부정확한 데이터를 포함하고 있다. 따라서 본 논문에서는 세포상의 localization을 고려하여 부정확한 데이터를 정제하여 PPI 네트워크를 재구축한다. 그리고 허브 단백질과 네트워크 구조를 분석하여 네트워크로부터 구조적 모듈을 발견하고 이를 정의한다. 또한 이러한 구조적 모듈로부터 단백질의 도메인을 분석하여 기능적 모듈을 밝히고, 높은 확실성을 가지는 기능적 모듈을 기반으로 미지 단백질에 대한 기능을 예측한다.

■ 중심어 : | 단백질-단백질 상호작용 | 단백질 기능발견 | 단백질 도메인 | 바이오인포매틱스 |

### Abstract

We propose a protein function finding algorithm that is able to predict specific molecular function for unannotated proteins through domain analysis from protein-protein network. To do this, we first construct protein-protein interaction(PPI) network in *Saccharomyces cerevisiae* from MIPS databases. The PPI network(proteins; 3,637, interactions; 10,391) shows the characteristics of a scale-free network and a hierarchical network that proteins with a number of interactions occur in small and the inherent modularity of protein clusters. Protein-protein interaction databases obtained from a Y2H(Yeast Two Hybrid) screen or a composite data set include random false positives. To filter the database, we reconstruct the PPI networks based on the cellular localization. And then we analyze Hub proteins and the network structure in the reconstructed network and define structural modules from the network. We analyze protein domains from the structural modules and derive functional modules from them. From the derived functional modules with high certainty, we find tentative functions for unannotated proteins.

■ keyword : | Protein-Protein Interaction | Functional Prediction | Protein Domain | Bioinformatics |

\* "본 논문은 2006년도 충북대학교 학술연구지원사업의 연구비 지원과 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구입니다." (지방연구중심대학육성사업/충북BIT연구중심대학육성사업단)

접수번호 : #071002-001

심사완료일 : 2008년 01월 08일

접수일자 : 2007년 10월 02일

교신저자 : 유재수, e-mail : yjs@chungbuk.ac.kr

## I. 서론

급속한 분자 생물학의 발달은 엄청난 생물학 데이터를 생산하였고, 방대한 생물학 데이터의 정리 및 해석을 주목적으로 하는 생물정보학이 생물학의 한 분야로 자리 잡게 되었다. 최근 생물정보학 분야의 성장에 따라 생물학적 정보의 양은 급속도로 증가하게 되었고 방대한 생물학적 정보들을 보다 빠르고 효과적으로 분석하기 위한 생물정보학 기술들이 요구되고 있다. 생물정보학에 의한 유전자 기능분석 및 예측은 기존의 생물학적 기법에서 요구되었던 엄청난 비용을 절감하고 또한 실험적 검증에 소요되는 많은 시간을 단축시켜준다.

2002년 인간 유전체 사업단 (Human Genome Project)이 인간 유전체에 대한 연구 결과를 발표한 후, 생명과학 분야에서는 새로운 패러다임으로의 전환을 요구하고 있다. 유전체 사업을 통해 밝혀진 많은 생물종의 유전체를 바탕으로, 유전체의 기능을 밝히는 연구, 그 기능을 수행하는 단백질 간의 관계를 밝히는 단백질체 (proteome) 수준에서의 연구를 비롯해, 세포 내에서 발생하는 모든 대사 과정과 신호 전이 경로에 관여하는 모든 단백질과 기질 사이의 상호관계를 밝히고자 하는 대사체 (metabolome) 수준에서의 연구 등이 이뤄지고 있다 [1]. 이러한 총체적인 수준의 연구는 DNA 칩, 단백질 칩, 단백질체학 (proteomics) 등과 같은 기술의 발달과 함께 유전자나 단백질에 대한 정보를 대량으로, 고효율로 얻을 수 있게 되었다.

현재까지는 단백질의 기능을 밝혀내기 위한 방법으로 생물학적 실험에 의한 방법을 주로 의존하고 있다. 하지만 이러한 실험을 위해서는 많은 비용과 시간이 요구된다. 최근에는 불필요한 실험을 막고 막대한 시간과 비용을 절약하고자 정보기술을 활용하는 노력들이 많이 시도되고 있다. 예를 들어 이미 많이 알려진 단백질들의 정보를 분석하여 생물학적 실험 범위를 줄여 불필요한 실험을 막아주거나 더 나아가 분석만으로도 미지 단백질의 기능을 예측할 수 있다면 많은 시간과 비용을 절약할 수 있다.

생명체 내에서 일어나는 대부분의 생명현상은 여러 단백질들이 복합적으로 상호작용함으로써 발생된다.

단백질들은 서로 매우 복잡한 상호작용 관계를 형성하는데 이들 전체 단백질들의 상호작용 관계를 연결하면 하나의 거대한 네트워크를 형성한다. 단백질의 상호작용은 기능과 밀접한 관계가 있다. 따라서 이들 상호작용들을 분석하여 단백질들의 기능적 관계를 파악하거나 기능적 관계를 형성하는 네트워크 모듈을 추출할 수 있다. 그리고 특정 기능과 밀접한 관계를 형성하는 기능모듈을 추출할 수 있다면 모듈 내에 존재하는 미지 단백질의 기능을 발견할 수 있다. 전체 단백질-단백질 상호작용 네트워크에서 기능의 분포를 살펴보면 특정 기능을 가지는 단백질들이 서로 상호작용하는 기능모듈들을 확인할 수 있다. 하지만 이러한 단백질 상호작용 네트워크는 매우 방대하기 때문에 이러한 기능모듈을 찾아내는 것은 쉽지 않다. 또한 이러한 모듈들은 특정 기능만이 모여 있는 형태만 존재하는 것이 아니고 다양한 기능들이 매우 복잡하게 연결되어 있다. 따라서 네트워크를 분석하기 위해 기능적으로 밀접하게 관련 있는 모듈로 구분할 수 있는 기준이 필요하다. 하지만 아직까지는 특정 기능을 수행하는 모듈을 발견할 수 있는 뚜렷한 방법이 제시되지 않은 상태이다. 이를 위해 본 논문에서는 단백질 상호작용 네트워크를 분석하여 기능 모듈을 추출할 수 있는 방법을 제시하고 이를 통해 미지 단백질의 기능을 발견할 수 있는 기법을 제안하고자 한다. 먼저 복잡한 단백질들의 상호작용을 분석하고 이해하기 위해 효모에 대한 단백질-단백질 상호작용 데이터를 이용한다. 그리고 기능모듈을 추출하기 위한 방법으로 기능별 주요 도메인 정보를 이용한다. 최종적으로 추출된 기능 모듈로부터 미지 단백질에 대한 기능을 예측한다.

본 논문의 구성은 다음과 같다. 2장의 관련연구에서는 단백질 기능 관련 데이터베이스 및 기능 발견 연구들을 기술하고 분석한다. 3장에서는 새롭게 제안하는 미지 단백질 기능 발견 알고리즘에 대해 설명한다. 4장에서는 3장에서 제안하는 미지 단백질 기능 발견 기법을 통해 예측된 결과가 정확성이 높음을 보인다. 마지막으로 5장에서는 본 논문에서 제안하는 단백질 기능 발견 기법의 특징을 요약 및 정리하고 결론을 기술한다.

## II. 관련연구

### 1. Localization 정보를 이용한 단백질 상호작용 데이터 정제

유전체나 단백체를 다루는 많은 데이터베이스들은 Y2H(Yeast Two Hybrid)를 통해 밝혀진 많은 양의 단백질-단백질 상호작용 결과를 포함하고 있다. 그러나 Y2H 결과 중 실제 세포에서 단백질 상호작용이 일어나지 않는 경우가 있다. 이러한 오류(false positive)의 한 원인이 될 수 있는 것이 단백질의 세포 내 localization이다. Localization은 특정 단백질이 특정 기능 또는 역할을 수행하기 위해, 세포에서 생성되어 적재적소의 세포 내 소기관으로 이동되어 위치하는 것을 의미한다. 따라서 단백질이 세포 내 소기관의 어디에 위치하느냐에 따라 단백질 상호작용 여부가 달라진다. 예를 들면, 퍼옥시솜에 있는 단백질과 미토콘드리아에 있는 단백질은 상호작용이 불가능하다. 단백질 기능 예측의 정확성을 높이기 위해 localization을 고려하여 실제 세포에서 발생하기 어려운 단백질-단백질 상호작용 (PPI)을 제거한다. 이를 정제(purified) PPIDB 또는 정제 네트워크라고 한다[2]. [표 1]은 초기(primary) PPIDB를 localization을 이용해 정제하는 과정을 보여준다.

표 1. Localization을 이용한 상호작용 정제

단백질 1	Localization	단백질 2	Localization	상호작용 가능성
ABF2	Nucleus	GCC1	Mitochondria	No
BNH1	Bud	SLT2	Nucleus	No
JSN1	Unknown	APA1	Nucleus	No
ACE2	Bud	HDA2	Nucleus	No
LSM1	Nucleus	YOR320C	Vacuole	No
MCM2	Nucleus	INH1	Mitochondria	No
YEL023C	Unknown	ZDS2	Cytoplasm	No
ABF2	Mitochondria	GCC1	Mitochondria	Yes
BEM1	Bud	FAR1	Cytoplasm	Yes
CDC12	Cytoplasm	SLT2	Nucleus	Yes
CLB2	Cytoskeleton	SLT2	Cytoplasm	Yes
CLB2	Cytoskeleton	MRPL50	Nucleus	Yes
LSM1	Cytoplasm	SWI1	Nucleus	Yes
BIM1	Cytoskeleton	YBR095C	ER	Yes
MCM2	Cytoplasm	INH1	Mitochondria	Yes
MOB1	Cytoskeleton	DBF2	Bud	Yes
MPS3	Nucleus	HTZ1	Nucleus	Yes
...	...	...	...	...

### 2. 단백질 기능 발견을 위한 기존 방법론

단백질 기능발견을 위해 기존에 연구되었던 방향은 서열로부터의 단백질 기능 예측[3], 구조로부터의 단백

질 기능 예측[4], 계통발생학 프로파일, 유전자 상관관계 그리고 단백질 상호작용을 이용한 단백질 기능 예측 방법[5][6]등으로 크게 5가지로 구분해볼 수 있다. 그중 단백질 상호작용으로부터의 기능예측 방법은 크게 네트워크에서 직접적인 연결을 기반으로 단백질의 기능을 예측하는 직접(Direct) 방식과 먼저 관계가 있는 단백질들의 모듈을 식별하고 모듈 내 단백질들의 알려진 기능을 기반으로 각 모듈의 기능을 추측하는 모듈(Module-assisted) 방식으로 나눌 수 있다[7]. 이 중 본문에서 제안하는 방법과 관련이 있는 대표적인 기존 연구들에 관해 설명한다.

#### 2.1 직접 방식

기능예측을 위한 모든 직접 방식에 적용되는 공통적인 원리는 PPI 네트워크 안에서의 서로 근접해있는 단백질은 유사한 기능을 할 가능성이 높다는 것이다. 즉, 상호작용 네트워크 안에서 가까운 두 단백질은 그들의 기능이 보다 유사하다. 이는 그림 1의 네트워크 거리와 함수적 거리의 상관관계를 통해 설명할 수 있다.

대표적인 직접 방식은 이웃노드 카운트 (Neighborhood counting) 방식과 이를 변형한 방식들이 있다. 이웃노드 카운트 방식은 기능이 밝혀지지 않은 단백질과 직접 상호작용하는 단백질들의 이미 알려진 기능을 기반으로 단백질 기능을 예측한다[8]. 이 방식은 단백질 기능 예측을 위한 가장 간단하면서도 직접적인 방법이다. 하지만 기능예측을 위해 다른 의미 있는 값들에 대한 연관성이 전혀 부여되지 않으며 전체 네트워크 토폴로지를 고려하지 않는다는 문제점이 있다. 그리고 특정 기능 클래스가 전체 기능 클래스에 차지하는 크기를 무시함으로써 기능 예측에 대한 편차를 가져온다.

이웃노드 카운트 방식과 마찬가지로 네트워크의 토폴로지는 고려하지 않으나, 한 단백질의 k-이웃에 대해 카이제곱을 적용해 기능을 예측하는 카이제곱 (Chi-Square) 방식이 있다[9]. 카이제곱 방식은 n-이웃 내의 단백질들을 p로부터 서로 다른 거리임에도 불구하고 같은 방식으로 취급하는 문제점을 가지고 있다. 이밖에도 Chi-Square의 문제점을 극복하기 위해 네트워크 거리와 기능적 유사성 사이의 관계를 적용하여 1

차 이웃과 2차 이웃노드에 대해 목표 단백질로부터의 거리에 따라 가중치를 다르게 부여하는 기능 스코어가 제안되었다[14].

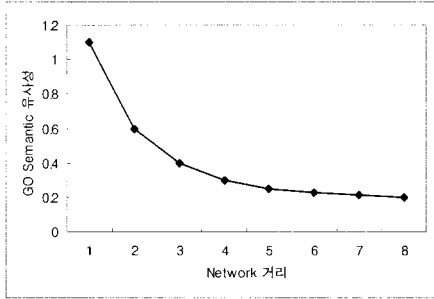


그림 1. 네트워크 거리와 함수적 거리의 상관관계

### 2.2 모듈 방식

분자 세포 생물학분야에서 모듈 이론이 널리 퍼짐에 따라 기능 예측에서 모듈의 사용이 확대되었다. 기능적 모듈은 일반적으로 세포를 구성하는 요소와 특정 생물학적 기능을 하도록 하는 상호작용들의 그룹을 의미한다. 개개의 유전자에 대한 기능을 예측하는 대신 모듈 방식은 먼저 유전자 그룹의 관련성을 찾고, 같은 그룹내의 유전자들에 기능을 부여한다. 모듈 방식들은 주로 모듈을 찾는 방식에서 그 차이가 있다. 모듈이 얻어지면 모듈내의 단백질의 기능을 부여하는 것은 비교적 간단하다. 예를 들어 모듈내의 유전자들이 대부분 모든 기능을 공유한다. 모듈 찾기 알고리즘은 네트워크 토폴로지 정보만을 사용하는 방식과 유전자 발현 데이터, 유전형질과 같은 부가적인 정보를 활용하는 방식으로 나뉜다. 알고리즘들은 중첩 모듈을 찾는 능력과 상호작용 신뢰성의 사용 등에 따라 여러 가지로 나뉜다. 각 알고리즘마다 기능적 모듈에 대한 정의 또한 다양하다. 어떤 알고리즘의 목적은 분자 복합체를 찾는 것이다. 또 다른 경우는 신호전이 경로 와 같이 최소 구조를 찾는 것을 목적으로 한다[11-13].

### III. 제안하는 단백질 기능발견 방법

이 장에서는 새롭게 제안하는 단백질 기능 발견 기법

을 설명한다. 먼저 기능발견을 위해 구성된 시스템 구조를 설명하고, 데이터베이스 구축 및 PPI 데이터 정제를 통해 네트워크를 구축하는 과정을 설명한다. 그리고 단백질 기능에 따른 도메인 연관성을 분석하여 PPI 네트워크로부터 기능 모듈을 추출하는 과정을 설명한다. 그리고 마지막으로 추출된 기능모듈에서의 미지 단백질에 대한 기능을 부여하기 위한 기능부여 기준을 제시하고 기능을 예측하는 과정을 설명한다.

#### 1. 단백질 기능 발견 기법

제안하는 단백질 기능 발견 기법의 전체적인 수행 과정은 그림 2와 같다. 먼저 PPI 데이터베이스를 구축하고 Localization에 의한 PPI 정제 및 중복 제거를 수행한다. 이후 각 기능별 도메인의 분포를 조사하여 기능과 특정 도메인 사이의 연관성을 분석하고 이를 네트워크에서의 상호작용 정제에 활용한다. 이를 위해 먼저 전체 PPI를 구성하는 단백질 리스트를 작성한다. 그리고 단백질 리스트의 각 단백질들을 중심으로 하여 2차 상호작용까지의 서브 네트워크를 추출한다. 추출된 각각 도메인 연관성에 의해 기능과 밀접한 도메인을 가지는 상호작용만을 남기는 정제작업을 수행한다. 마지막으로 각 서브 네트워크에서 특정 기능이 70%이상 발견되는 기능 모듈을 추출하고, 이들 기능모듈내의 미지 단백질의 기능을 부여한다.

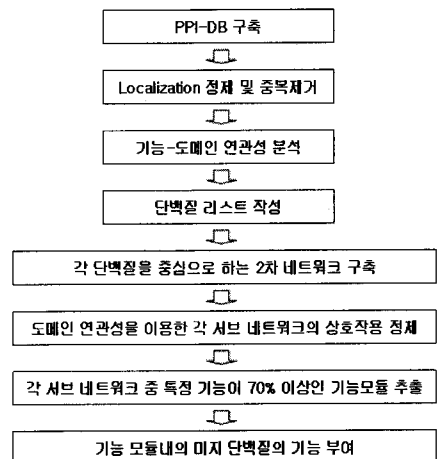


그림 2. 단백질 기능발견 과정

## 2. PPI 데이터베이스 구축

단백질-단백질 상호작용 데이터베이스는 Munich Information Center for protein Sequences[MIPS] 데이터베이스로부터 2006년 12월, 효모전체에 관련한 단백질 상호작용 목록, localization, 기능 카테고리에 대한 정보를 수집하였다. 수집된 효모의 단백질 상호작용 데이터베이스(PPIDB)로부터 한 단백질이 다른 단백질과 상호 작용하는 링크 수에 대해 특정 링크 수를 갖는 단백질 수를 계산한 결과, 그래프는 멱함수 법칙에 따른 분포 양상을 보였다(그림 3). 이것은 효모의 단백질 상호작용 네트워크가 스케일 프리 네트워크임을 나타낸다. 스케일 프리 네트워크는 노드의 대부분이 소수의 링크를 갖고, 많은 수의 노드들과 링크를 형성하고 있는 허브라고 불리는 소수의 노드로 이루어진 네트워크이다. 스케일 프리 네트워크는 어떤 노드가 제 기능을 못하게 되더라도 네트워크의 핵심적인 기능에 크게 영향을 미치지 않는 오류에 대한 내성이 있다. 비록 핵심 요소가 제 기능을 상실하더라도 그 영향은 국소적으로 나타나고, 네트워크의 전체적인 정보 전달 능력이 상실되는 경우는 거의 없다. 이러한 네트워크의 안정성과 견고함은 각각의 노드에 대한 평균적인 링크 수 없이 아주 적은 수의 링크를 갖는 노드에서부터 아주 많은 수의 링크를 갖는 노드에 이르기까지 노드에 대한 링크 수가 다양하기 때문이다. 특히, 다수의 링크를 갖는 소수의 허브들이 특정 노드의 기능 상실로 인한 전체 네트워크의 분절화를 최소화하고, 네트워크를 안정하고 견고하게 만든다[14]. 즉, 스케일 프리 네트워크를 구성한다는 것은 특정 유전자나 단백질의 손상에 대한 세포 내 작용에 있어 유연성 있게 대처하고 내외적인 요인에 의한 손상에 유연하게 대처할 수 있다.

## 3. 단백질 상호작용 데이터베이스의 localization을 이용한 정제

네트워크 구축에 앞서 MIPS로부터 수집한 단백질 상호작용 데이터에 대해 localization 정보를 기반으로 상호작용 정제를 수행하였다. 총 4387개의 단백질 노드와 12235개의 상호작용으로부터 정제 결과 3637개의

단백질 노드와 10391개의 상호작용을 추출하였다. 정제 이전의 상호작용 네트워크와 정제된 후의 상호작용 네트워크를 [그림 4]에서 보인다. 전체 네트워크는 허브 단백질들을 중심으로 구성 된다.

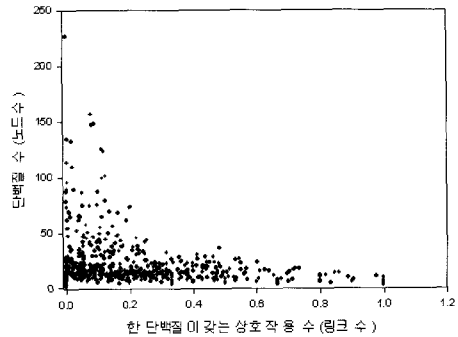


그림 3. 효모 네트워크의 멱함수 분포

여기에서의 허브란 여러 가지의 의미로 사용될 수 있지만 본 논문에서는 허브 단백질을 1차 상호작용의 수가 많아 상호작용 사이에서 중요한 역할을 수행하는 단백질들을 허브 단백질이라 정의하며 허브 단백질을 중심으로 1차 또는 2차까지의 상호작용 네트워크를 허브 네트워크 또는 네트워크 모듈이라 한다.

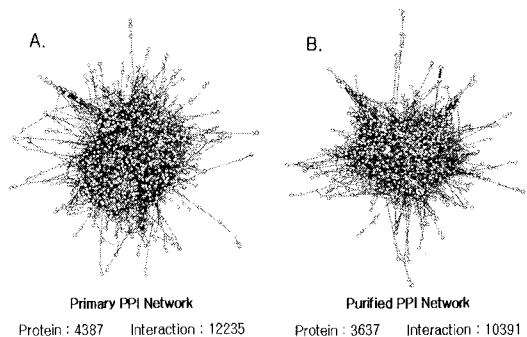


그림 4. 효모 단백질 전체 네트워크(A) 및 정제된 네트워크(B)

전체 단백질들의 기능 분포는 [그림 5]와 같다. 3637개의 단백질 중 기능이 알려지지 않은 단백질의 수는 499개로 전체의 14% 정도의 상당수의 단백질들의 기능

이 아직 밝혀지지 않았음을 알 수 있다.

표 3. MIPS 기능 카테고리

기능 번호	기능
01	METABOLISM
02	ENERGY
10	CELL CYCLE AND DNA PROCESSING
11	TRANSCRIPTION
12	PROTEIN SYNTHESIS
14	PROTEIN FATE (folding, modification, destination)
16	PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)
18	PROTEIN ACTIVITY REGULATION
20	CELLULAR TRANSPORT, TRANSPORT FACILITATION AND TRANSPORT ROUTES
30	CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM
32	CELL RESCUE, DEFENSE AND VIRULENCE
34	INTERACTION WITH THE CELLULAR ENVIRONMENT
38	TRANSPOSABLE ELEMENTS, VIRAL AND PLASMID PROTEINS
40	CELL FATE
41	DEVELOPMENT (Systemic)
42	BIOGENESIS OF CELLULAR COMPONENTS
43	CELL TYPE DIFFERENTIATION
99	UNCLASSIFIED PROTEINS

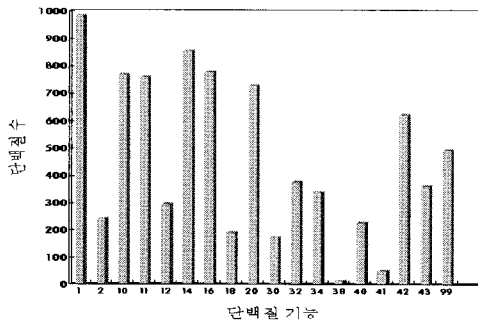


그림 5. MIPS 2006 단백질 데이터 기능 분포

5. 단백질 기능별 도메인 분포

본 논문에서는 특정 기능들 사이의 도메인 관계를 분석하여 단백질 기능을 유추하는데 이용하였다. 이를 위해 각 기능들이 발생하는 모든 단백질의 도메인 패턴을 분석하였다. 그 결과 특정 기능을 중심으로 발생하는 도메인 패턴들이 매우 다양함을 알 수 있었다.

먼저 각 기능들에 대해서 해당 기능을 가지는 단백질들을 모두 분류하고 각 기능별 단백질들이 포함하고 있는 도메인 패턴을 분석하였다. 분석결과 대부분은 [그림 6]의 A와 같이 특정 기능에 대하여 밀접하게 관련된 도메인들이 확연하게 드러나 보임을 알 수 있었으나 몇몇 기능에서는 [그림 6]의 B와 같이 도메인 분포 패턴에서 확연한 차이점을 찾을 수 없었다. 이러한 분석 결과는 각 기능에 따른 도메인 분포의 특성이 존재함을 말한다. 이러한 사실을 기반으로 본 논문에서는 이들 도메인 정보를 이용하여 각 허브 네트워크들에서 도메인 연관성이 큰 상호작용들만을 추출할 수 있었다. 그리고 도메인 연관성을 기반으로 추출된 허브 모듈들은 대부분 기능적으로 매우 유사한 기능 모듈을 구성하고 있음을 확인할 수 있었다(6절).

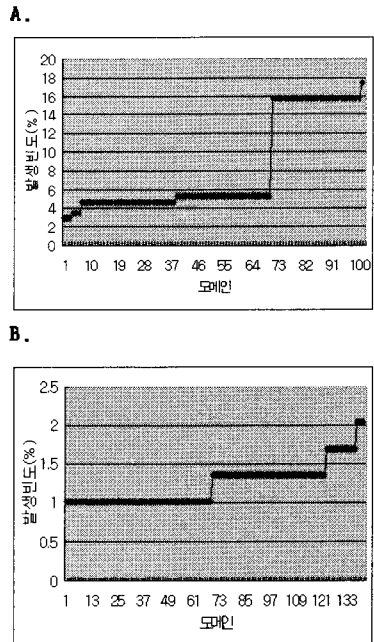


그림 6. 기능별 도메인 분포

[표 4]는 기능별 주요(빈도가 높은)도메인 정보이다. 이 중 굵은 글씨체는 빈도는 높으면서 여러 기능에서 공통적으로 나타나는 공통 도메인들을 그리고 나머지는 해당 기능에서만 빈도가 높은 특화 도메인을 나타낸다.

표 4. 기능별 주요 도메인 요약

기능	주요 도메인
1	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase
2	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, AAA+ATPase-core, ADH-N, ADH-Sf-Zn, ADH-Zn-bd, CBM-21, GAPDH, GAPDH-I, GAPDH-like, Myosin-head, PG-mutase, 6Pfruct-Kin, 6Phosfructo-Kin
10	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, AAA+ATPase-core
11	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, DEAE-like-N, Euk-arc-LSM-core, Fungi-Trscrp-N, Helicase-C, Helic-SF1/SF2-ATP-bd, Histone-fold, LSM-related-core, LSM-smRNP-core, RRM-RNP1, Znf-C2H2
12	DEAE/DEAH-N, DEAE-like-N, Helicase-C, Helic-SF1/SF2-ATP-bd, NA-bd-OB-fold, Transl-elong-init/rib-B-barrel, tRNA-ligase-II
14	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like
16	WD40, WD40-like, AAA-ATPase-core, AAA+ATPase-core, ABC-transp-like, ARM-like, DEAE-like-N, ENTH-VHS, Helicase-C, Helic SF1/SF2-ATP-bd, Nucleotide-bd, Prefoldin, RRM-RNP1, SH3, TRP-like-Helical
18	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, Cyclin, Cyclin-A-B-D-E, Cyclin-C, Cyclin-like, Cyclin-N, Cyclin-related, Cyclin-related-2, RabGAP-TEC, RhoGAP, Rho-GAP
20	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, WD40, WD40-like, AAA+ATPase-core, ARM-like, ENTH-VHS, GTPase-Rab, GTPase-Ras, GTPase-Rho, MFS, Mitoch-carrier, Ran-GTPase, Ras, Ras-trnsfrmg, Small-GTP-bd, Sub-transporter, Sug-transporter, Sugar-transpt, SNARE, t-snare
30	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, GTPase-Rab, GTPase-Ras, GTPase-Rho, M-pesterase, PH, Ras-trnsfrmg, RhoGAP, Rho-GAP, Small-GTP-bd, T-phatase-apaH, t-snare
32	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40-like, AAA+ATPase-core, CheY-like, M-pesterase, Response-reg, RRM-RNP1, SH3, Thiordxn-fold, Thiordxn-like-fd, TPR-1, TPR-like-helical, TRP-region, Znf-C2H2
34	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, AAA+ATPase-core, EF-hand-Ca-bd, MAP-kin, M-pesterase, Rho-GAP, SH3, Thiordxn-Fold, Thiordxn-like-fd, T-phatase-apaH, Wing-hlx-DNA-bd, Znf-C2H2
40	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, Cell-Div-GTP-bd, GTPase-Rab, GTPase-Ras, GTPase-Rho, PH, PH-type, Ras, Ras-trnsfrmg, Rho-GAP, SH3, Small-GTP-bd
41	Alpha-tubulin, Cell-Div-GTP-bd, DEAE-like-N, Gal-oxid-central, Helicase-C, Helic-SF1/SF2-ATP-bd, Kelch-1, Kelch-2, SH3, SMF2-related, Tubulin, Tubulin-Fts2, Tubulin-Fts2-C
42	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, AAA+ATPase-core, EF-hand-Type, GTPase-Rab, GTPase-Ras, GTPase-Rho, PH, PH-type, Prefoldin, Ras, Ras-trnsfrmg, SH3, Small-GTP-bd
43	pkinase-core, pkinase-like, Ser-thr-pkinase, Ser-thr-pkinase-AS, Tyr-pkinase, WD40, WD40-like, GTPase-Rab, GTPase-Ras, GTPase-Rho, PH, Ras, Ras-trnsfrmg, Rho-GAP, Small-GTP-bd, SH3

그리고 도메인을 이용한 상호작용 정제에서는 이들 공통도메인과 특화 도메인을 모두 사용한다. 그 이유는 공통도메인 중에서 일부는 몇몇 기능에서만 높은 빈도를 보이기 때문이다.

6. 기능 모듈 추출

기능 모듈을 추출하기 위해 먼저 전체 상호작용 네트워크에서 상호작용이 많은 순서로 HUB 단백질의 리스트를 작성한다. 그리고 리스트의 각 HUB 단백질을 중심으로 2차 상호작용까지의 네트워크 모듈을 추출하고 기능별 주요 도메인 정보를 이용하여 각 HUB 네트워크 모듈들을 추출하였다. 각 모듈에서는 상호작용 사이에 도메인 관계를 적용하기 위해 HUB 단백질의 기능들에서 보이는 도메인 패턴을 통해 필터링 작업을 수행하였다. 즉, 상호작용 하는 단백질의 기능들에서 보이는 도메인 패턴들 중 HUB 단백질의 기능에서 보이는 도메인 패턴에 있는 상호작용 단백질만 추출하였다. 추출되는 방식은 다음의 [그림 7]과 같다.

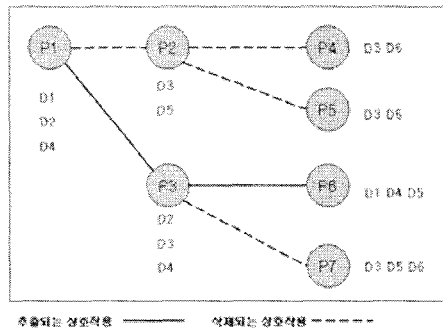


그림 7. 도메인 관계를 고려한 상호작용 추출

네트워크 모듈을 추출하고 각 모듈의 기능적 연관성을 알아보기 위해 각 모듈에서의 1차 상호작용, 2차 상호작용에 대한 기능 패턴을 분석하였다. 분석 결과 대부분의 추출된 네트워크 모듈에서의 기능 패턴이 상당히 유사함을 알 수 있었다.

먼저 1차 상호작용에서 특정 기능(주요 기능)이 60% 이상 발생하는 모듈을 추출 하였다. 모듈의 추출은 2차 상호작용까지 추출하였으며 이에 대한 예를 [표 5]에

나타내었다. [표 5]에서 1차와 2차 상호작용 단백질들의 대부분은 중심단백질의 기능인 30, 32, 34, 43 기능 중 하나이상을 포함하고 있음을 알 수 있다. 이 결과를 통해 도메인 연관성을 이용해 정제된 네트워크 모듈은 기능적으로 밀접하게 관계된 모듈임을 확인할 수 있다.

표 5. YER118c 중심의 네트워크 모듈

Hub 131	Hub 기능	1차 주요기능	2차 주요기능
YER118c	30:32:34:43	01:10:14:30:32:40:42:43:	01:11:14:32:34:40:42:43:
			01:11:30:32:42:
			01:10:11:14:16:20:32:34:40:42:
			01:10:14:30:32:34:40:42:43:
			10:11:30:32:40:
			01:10:14:30:32:34:40:43:
			01:10:12:16:30:40:42:
			01:10:14:20:30:42:
			01:10:14:30:32:40:43:
			01:16:30:40:42:43:
		88:	16:30:40:42:43:
			10:16:34:40:41:42:43:
			20:30:40:42:43:
			01:14:30:34:40:42:43:
			01:10:11:16:
			10:16:20:30:34:40:42:43:
			01:10:14:30:40:42:43:
			40:42:43:
		01:14:16:18:30:32:34:42:	01:14:18:30:32:34:
			01:11:14:30:32:34:
			01:14:30:32:34:43:
			01:14:18:30:32:34:42:
		01:14:	
		01:14:16:20:32:40:42:43:	18:30:34:
		01:14:30:32:34:42:	01:10:14:16:18:30:34:
			01:10:14:
			01:14:30:34:43:
			16:30:34:
			01:14:16:18:30:32:34:42:
			01:11:14:30:32:34:
			01:14:30:34:43:

도메인을 기반으로 정제되어 추출된 네트워크 모듈에서 1차 상호작용에서 특정 기능이 60%이상 발생하는 경우는 총 276개 모듈이다. 그 중 일부 단백질들의 1차 및 2차 상호작용에 대해 주요 기능 및 주요 기능 발생 빈도를 [표 6]으로 제시하였다. [표 6]에서 1차 상호작용에서의 기능 유사성의 평균(77.47)이 2차 상호작용의 평균(61.89)에 비해 높음을 확인할 수 있다. 따라서 본 논문에서는 1차 상호작용을 이용해 미지 단백질의 기능을 예측한다.

표 6. 1차, 2차 상호작용에서 주요 기능의 발생 빈도

HUB	1차 주요기능	주요기능 빈도(%)	2차 주요기능	주요기능 빈도(%)
hub_51	42	76.17	10	54.1
hub_54	11	94.03	11	72.31
hub_55	10	91.07	10	75
hub_56	14	78.05	10	50.82
hub_57	1	79.21	42	54.04
hub_58	10	61.67	10	71.19
hub_59	10	85.99	10	52.68
hub_60	20	95.83	14	59.09
hub_61	42	61.23	10	55.21
hub_62	42	61.12	10	56.22
hub_64	10	73.73	42	61.02
hub_65	42	64.12	42	54.35
hub_66	42	66.67	10	54.88
hub_69	20	78.26	14	60
hub_70	42	72.19	42	52.54
hub_71	10	78.75	10	75.95
hub_73	11	82.89	11	52.94
hub_74	10	56.36	10	52.73
hub_75	10	66.28	10	62.81
hub_76	11	89.89	11	64.77
hub_78	20	66.67	20	75
hub_80	42	75.8	42	53.66
hub_81	10	61.31	10	57.29
hub_83	11	100	14	54.03
hub_85	11	100	11	100
hub_87	42	80.95	10	50.34
hub_88	11	97.62	11	69.05
hub_90	10	62.5	10	54.74
hub_93	10	100	10	75
hub_96	10	70.91	10	70.91
hub_98	10	72.22	10	66.05
평균		77.47		61.89

실험 결과들로 볼 때 효모의 전체 네트워크는 소규모 기능 모듈들로 구성된다라는 것을 추측할 수 있다. 반대로 전체 네트워크를 소규모 기능 모듈들로 구분할 수 있다는 사실을 알 수 있다. 결과적으로 전체 상호작용



네트워크에서 기능적 관계가 밀접한 기능 모듈을 추출할 수 있다면 이들 기능 모듈내의 미지 단백질의 기능을 비교적 정확하게 예측할 수 있다.

다음으로 1차 상호작용에서 특정 기능이 80% 이상, 70% 이상 그리고 60% 이상 발생하는 모듈들만을 추출하여 각 기능 모듈에 포함된 미지 단백질의 기능을 예측하였다. 미지 단백질을 포함하면서 기능 유사성 기준을 각각 80%, 70%, 60%이상으로 하여 선별된 기능 모듈은 [표 7]-[표 9]와 같다.

표 7. 주요 기능이 80% 이상 발생하는 기능 모듈

단백질	HUB기능	1차 주요기능	주요기능 빈도(%)
YLR166c	10:16:20:43;	20	80
YOR127w	18:30:34:40:42:43;	42	85.71
YPL240c	01:02:10:14:16:32:34:43;	14	88.89
YOR008c	16:30:32:34:42:43;	01	85.71
YPL203w	01:11:14:16:30:34:43;	11	83.33

표 8. 주요 기능이 70% 이상 발생하는 기능 모듈

단백질	HUB기능	1차 주요기능	주요기능 빈도(%)
YLR447c	14:20:34;	01	71.43
YHR152w	10:43;	01:10:14	75
YER125w	10:14:16:32:43;	16	70
YDR217c	10:11:16:32;	10	75
YLR166c	10:16:20:43;	20	80
YMR312w	11;	42	75
YOR035c	16:20:34:42;	42:43	75
YOR127w	18:30:34:40:42:43;	42	85.71
YPL240c	01:02:10:14:16:32:34:43;	14	88.89
YGR040w	01:14:30:34:43;	01:14:34	75
YGR241c	14:16:20;	14	71.43
YGR249w	11:32:43;	14	75
YOR008c	16:30:32:34:42:43;	01	85.71
YFR052w	10:14:32;	14	75
YOR319w	11:16;	11	75
YPL203w	01:11:14:16:30:34:43;	11	83.33

[표 8]에서는 [표 7]과 달리 1차 상호작용의 주요 기능이 여러 개인 것들을 확인할 수 있다. 이는 기능 모듈

에서 여러 개의 기능이 주요 기능으로 동시에 같이 발생함을 의미한다. 예를 들어 YHR152w의 주요기능 "01:10:14"의 경우는 YHR152w를 중심으로 1차 상호작용하는 단백질들 중에서 75%이상 발생하는 주요 기능은 01, 10, 14의 기능 쌍임을 말한다. 즉, 1차 상호작용하는 단백질들의 대부분은 01, 10, 14기능을 가지고 있음을 의미한다.

표 9. 주요 기능이 60% 이상 발생하는 기능 모듈

단백질	HUB기능	1차 주요기능	주요기능 빈도(%)
YER016w	10:42;	10	65.85
YNL271c	10:16:20:30:32:34:40:42:43;	42	63.83
YHR030c	01:10:14:30:32:34:40:42:43;	42	64.29
YLR447c	14:20:34;	1	71.43
YMR263w	10:11:14:34:40:43;	10	62.07
YMR307w	01:42;	1	65.22
YNL322c	01:42;	42	61.9
YBL007c	14:16:20:40:42:43;	42	65.22
YMR117c	10:42;	1	66.67
YLR113w	01:11:14:30:32:34;	1	61.54
YOL149w	01:11:16;	11	64.71
YDL226c	10:16:18:20;	01:-14;	60
YHR152w	10:43;	01:10:14;	75
YER125w	10:14:16:32:43;	16	70
YDR217c	10:11:16:32;	10	75
YLR166c	10:16:20:43;	20	80
YMR312w	11;	42	75
YJR070c	16:18;	14	60
YLR116w	11:16;	11	60
YOR035c	16:20:34:42;	42:43;	75
YOR127w	18:30:34:40:42:43;	42	85.71
YER019w	01:30:32:34;	14:16:42;	66.67
YJR093c	11:16;	11:16;	66.67
YPL240c	01:02:10:14:16:32:34:43;	14	88.89
YDR192c	14:16:20:42;	14:20;	66.67
YNL047c	16:42;	42	60
YGR040w	01:14:30:34:43;	01:14:34;	75
YGR241c	14:16:20;	14	71.43
YGR249w	11:32:43;	14	75
YOR008c	16:30:32:34:42:43;	1	85.71
YFR052w	10:14:32;	14	75
YOR319w	11:16;	11	75
YPL101w	11;	42	66.67
YPL203w	01:11:14:16:30:34:43;	11	83.33
YMR104c	01:10:14;	1	66.67

## 7. 단백질 기능 발견

### 7.1 기능모듈로 부터의 단백질 기능 발견

본 논문에서는 실험의 결과를 통해 1차 상호작용의 주요 기능이 70%이상인 경우를 기능 모듈이라고 정의한다. 이는 70% 이상의 주요 기능을 갖는 모듈의 경우 특정 기능이 명확하게 구분되어 있지만 60% 이상을 기능모듈로 가정하고 모듈을 추출하였을 때는 주요 기능이 하나의 특정 기능이라고 명확하게 구분할 수 없는 경우가 발생되기 때문이다. 예를 들어, [표 9]에서 YDL226c번의 경우 1차 상호작용의 주요 기능이 "01;-14"로 표기되었는데 이는 1차 상호작용에서 01과 14번 기능이 서로 각자 60% 이상 발생함을 말한다. 이는 특정 기능에 해당하는 기능모듈이라 할 수 없다.

최종적으로 추출된 단백질 기능 모듈은 1차 상호작용의 주요 기능이 70%이상을 차지하면서 기능이 밝혀지지 않은 미지 단백질을 포함하고 있는 모듈이다. 이때 추출된 기능 모듈에 포함되어 있는 미지 단백질의 기능은 기능모듈의 주요 기능을 따를 가능성이 크다.

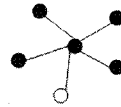
### 7.2 구조에 따른 기능부여 기준

추출된 기능모듈에서 포함하고 있는 미지 단백질의 기능을 예측할 때 정확하게 하나의 기능을 부여하기가 어려울 때가 있다. 다시 말해 몇몇 미지 단백질의 경우는 여러 개의 기능모듈에서 동시에 1차 상호작용으로 나타나기도 한다. 이러한 경우 미지 단백질은 여러 개의 모듈로부터 영향을 받아 여러 가지의 기능으로 예측되거나 판단을 내릴 수 없을 정도로 그 경계가 모호한 경우도 발생한다. 이와 같이 동시에 여러 개의 기능 모듈로부터 영향을 받는 미지 단백질은 어떤 기능 모듈을 영향을 더 받을지를 판단할 수 있어야 한다. 이를 위해 본 논문에서는 명확한 판단 기준을 제시하고자 추출된 기능 모듈들에서 상호작용이 서로 연결될 수 있는 구조를 확인하고 이들 사이에서 어느 쪽에 더 영향을 받을지에 대한 기준을 제시한다.

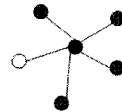
추출된 단백질 기능 모듈에서 모듈사이의 상호작용 관계가 형성될 경우 기능 부여를 위해 다음과 같은 4가지 기준을 제시한다. 첫째, 본 논문에서의 제안하는 방식으로 추출된 기능 모듈에서의 미지 단백질은 기능모

듈의 중심단백질로부터 1차 상호작용으로 나타난다. 둘째, 특정 기능으로 특성화된 클러스터 즉, 추출된 기능 모듈에 속한 미지 단백질은 그 기능 모듈의 주요 기능에 따른다. 기능 모듈의 중심 단백질과 기능과 중심 단백질에서의 1차 상호작용의 주요기능이 일치하는 기능이 존재하면 미지 단백질은 일치하는 기능을 따른다([그림 8] A). 셋째, 기능 모듈의 주요 기능에 따르되 중심단백질의 기능과 1차 상호작용의 주요 기능이 일치하지 않으면 미지 단백질의 기능은 1차 상호작용의 주요기능을 따른다([그림 8] B). 넷째, 미지 단백질이 두 개 이상의 기능 모듈을 연결하는 1차 상호작용인 경우 중심 단백질의 기능과 1차 상호작용 주요기능이 일치하는 기능 모듈의 기능을 따른다([그림 8] C). 여러 개가 존재할 경우 1차 상호작용 주요기능의 발생빈도가 높은 기능을 따른다([그림 8] D).

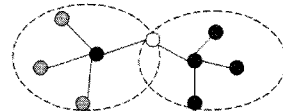
A. 기능모듈에 속한 미지 단백질은 그 기능 모듈의 주요 기능에 따른다.



B. 중심단백질의 기능과 1차 상호작용의 주요 단백질의 기능이 다를 경우 1차 상호작용의 주요 기능을 따른다.



C. 미지 단백질이 두 개 이상의 기능모듈을 연결할 경우 중심단백질과 1차 상호작용 주요기능이 일치하는 모듈의 기능을 따른다.



D. 미지 단백질이 두 개 이상의 기능모듈을 연결하고 모든 모듈의 중심 단백질과 상호작용의 주요 단백질 기능이 일치할 경우 상호작용이 많은 모듈의 기능을 따른다.

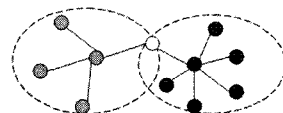


그림 8. 기능부여 기준

#### IV. 성능평가

본 논문에서 제안하는 단백질 기능 예측방법의 정확성을 측정하기 위해 성능평가를 수행하였다. 성능평가는 Intel(R) Xeon(TM) CPU 3.0GHz, 2.0GB RAM 시스템에서 Red hat Linux 3.2.3-10을 기반으로 MySQL 4.0.24 DBMS 와 gcc version 3.2.3의 구현 환경에서 수행되었다.

성능 평가를 위해 먼저 전체 단백질 상호작용 네트워크를 구축하였다. 그리고 기능이 알려진 단백질들 중 무작위로 100개를 추출하여 가상의 미지 단백질을 구성하였다. 그리고 기존의 기능 카테고리에서 미지 단백질을 나타내는 99번과 구별하기 위해 가상의 미지 단백질들에 대해서는 88번을 부여하여 이들을 구분하였다. 다시 말해 예측의 정확성을 측정하기 위해 무작위로 선별한 100개의 단백질을 기능이 밝혀지지 않은 미지 단백질인 것처럼 데이터를 조작하고 본 논문에서 제안하는 단백질 기능 예측 방법을 통해 기능을 예측해 보았다. 그리고 예측된 결과를 실제 알려진 기능과 대조하였다. 성능 평가를 위해 먼저 70% 이상의 기능모듈을 추출하고 예측 가능 단백질 수 및 정확히 예측된 단백질의 비율을 통해 예측의 정확성을 판단하였다.

그리고 기능 유사성을 60%, 70%, 80%로 변경하여 예측해보았다. 예측된 결과는 전체 가상의 미지 단백질 100개 중에서 기능 유사성이 60% 이상인 모듈에서 추출된 예측 가능한 가상의 미지 단백질의 수는 29개이다. 이중 예측된 기능이 실제 기능에 하나이상 포함된 경우, 즉 기능이 정확하게 예측된 경우는 총 18개로 62%의 예측 정확성을 말해주고 있다. 그리고 기능 유사성이 70%이상인 경우 예측된 단백질의 수는 감소하였지만 정확성은 더욱 높아졌음을 확인할 수 있다. 80% 이상인 경우에서는 는 예측된 미지 단백질의 수가 급격히 감소했음을 알 수 있다. 전체적인 예측 결과는 다음의 [표 10]과 같다.

결과적으로 도메인 연관성을 이용해 정제된 네트워크 모듈들에서 기능 유사성의 추출 기준을 낮출수록 예측 가능한 미지 단백질의 수는 많아지지만 정확성이 떨어지는 것에 반해 기능유사성의 추출 기준을 높일수록

예측 가능한 단백질의 수는 적어지지만 정확성을 높일 수 있다.

표 10. 기능 예측의 정확도

구분 기능유사성	예측된 단백질 수	정확히 예측된 단백질 수	정확성
60% 이상	29	18	62%
70% 이상	23	16	69.5%
80% 이상	10	8	80%

#### V. 결론

본 논문에서는 먼저 MIPS 데이터베이스로부터 효모에 대한 단백질-단백질 상호작용 네트워크를 구축하고 세포상의 localization을 고려하여 부정확한 데이터를 정제하였다. 그리고 네트워크의 구조적 모듈로부터 단백질의 도메인을 분석하여 높은 기능 유사성을 갖는 기능적 모듈을 추출하였다. 그리고 이를 통해 미지 단백질의 기능을 예측하는 방법을 제안하였다. 비록 제안하는 알고리즘을 통해 기능을 정확하게 예측하기에는 아직 그 한계가 있다. 하지만 본 논문에서는 정보기술을 이용해 실험의 타겟을 설정해주거나 범위를 줄여줌으로서 필요한 시간과 비용을 대폭 줄여 기존 생물학적 실험 방식의 비효율성을 개선할 수 있게 하였다. 향후에는 효모 단백질이 아닌 인간 단백질의 기능예측에 제안한 방법을 적용해볼 예정이다. 인간 단백질의 수는 효모 단백질에 비해 크기가 매우 크기 때문에 기능 관계 분석 및 도메인 연관성 분석을 보다 세밀하게 수행해야 할 것이라 생각된다. 보다 큰 규모의 데이터 분석을 통해 기능을 예측해보고 수정 및 보완을 통해 완성도를 높여 최종적으로 신뢰성 있는 인간 단백질 기능 예측 시스템으로 발전시킬 계획이다.

#### 참고 문헌

[1] E. Ravasz and A. L. Barabasi, "Hierarchical organization in complex networks," Phys. Rev.

E67, 026122, 2003.

[2] 강희영, *내피세포의 생존과 죽음 경로 및 효모의 단백질 네트워크 분석*, 충북대학교 대학원 석사학위 논문, 2005.

[3] G. Cynthia and J. Per, *Developing Bioinformatics Computer Skills*, O'reilly press, 2001(4).

[4] J. Jeffery, S. Fetrow, and K. Andrzej "Structural genomics and its importance for gene function analysis," *Nature Biotechnology*, Vol.18, No.18, pp.282-287, 2000.

[5] I. Takashi, *A comprehensive two-hybrid analysis to explore the yeast protein interactome*, Proceedings of the National Academy of Sciences of the United States of America, 2001.

[6] B. Schwikowski, P. Uetz and S. Fields, "A network of protein-protein interactions in yeast," *Nat Biotechnol*, pp.1257 - 1261, 2000.

[7] R. Sharan and R. Shamir, "Network-based prediction of protein function," *Mol. Sys. Biol.* pp.3-88, 2007.

[8] H. Hishigaki, K. Nakai, T. Ono, A. Tanigami, and T. Takagi, "Assessment of prediction accuracy of protein function from protein-protein interaction data," pp.523 - 531, 2001.

[9] H. N. Chua, W. K. Sung, and L. Wong, "Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions," *Bioinformatics*, Vol.22, pp.1623 - 1630, 2006.

[10] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani, "Global protein function prediction from protein-protein interaction networks," *Nat Biotechnol*, Vol.22 pp.697 - 700, 2003.

[11] A. U. Amin, M. Shin, Y. Mihara, K. Kurokawa, and S. Kanaya, "Development and implementation of an algorithm for detection of

protein complexes in large interaction networks," *BMC Bioinformatics*, 7: 207, 2006.

[12] R. Sharan and T. Ideker, "Modeling cellular machinery through biological network comparison," *Nat Biotechnol*, Vol.24, pp.427 - 433, 2006.

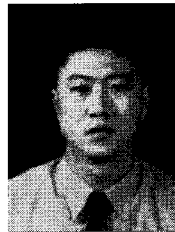
[13] V. Arnau, S. Mars, and I. Marin, "Iterative cluster analysis of protein interaction data," *Bioinformatics*, Vol.20, pp.364 - 378, 2005.

[14] A. L. Barabasi and E. Bonabeau, "Scale-Free Networks," *Scientific Mnerican*, Vol.288, pp.60-69, 2003.

### 저자 소개

#### 강 태 호(Tae-Ho Kang)

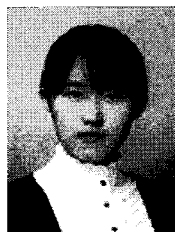
정회원



- 1999년 2월 : 호원대학교 정보통신공학과(공학사)
  - 2002년 8월 : 충북대학교 정보산업공학과(공학석사)
  - 2007년 8월 : 충북대학교 정보통신공학과(공학박사)
  - 2007년 9월 ~ 현재 : 충북대학교 전기전자컴퓨터공학부 Post-doc
- <관심분야> : 데이터베이스 시스템, 데이터 마이닝, 생물정보학, 시스템 바이오

#### 류 제 운(Jea-Woon Ryu)

준회원



- 2006년 2월 : 충북대학교 생화학 과(이학사)
  - 2006년 3월 ~ 현재 : 충북대학교 생화학과(석사과정)
- <관심분야> : 생물정보학, 신호전 이 네트워크, 시스템 바이오

유 재 수(Jae-Soo Yoo)

중신회원



- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : 한국과학기술연구원 전산학과(공학석사)
- 1995년 2월 : 한국과학기술연구원 전산학과(공학박사)

• 1996년 ~ 현재 : 충북대학교 전기전자컴퓨터공학부 및 컴퓨터정보통신연구소 교수

<관심분야> : 데이터베이스 시스템, 정보검색, 멀티미디어 데이터베이스, 분산객체 컴퓨팅, 생물정보학

김 학 용(Hak-Yong Kim)

정회원



- 1985년 2월 : 충북대학교 농화학과(농학사)
- 1987년 2월 : 충북대학교 화학과 (이학석사)
- 1994년 2월 : Connecticut University, Molecular Cell Biology(이학박사)

• 1998년 ~ 현재 : 충북대학교 생명과학부 교수

<관심분야> : 시스템 바이오, 신호 전이, 단백질 네트워크, 생체동역학