

## 단백질 상호작용 네트워크를 이용한 기능성군 사이의 계층구조에 대한 연구

고영진 · 육순형 · 김엽

경희대학교 물리학과 및 기초과학연구소, 서울 130-701

(2007년 11월 21일 받음)

각 노드들이 연결된 커뮤니티들 사이에서 커뮤니티 구조와 계층구조를 검출하는 것은 복잡계 네트워크 연구 분야에서 가장 흥미로운 연구 주제 중 하나이다. 본 연구에서는 개별 단백질을 이웃하는 단백질과 상호작용하는 진동자로서 간주한다. 이러한 단백질들 사이의 상호작용은 복잡계 네트워크 (complex network)로 나타낼 수 있다. 본 논문에서 우리는 단백질 상호작용 네트워크 (protein interaction network)에서 기능성군 (functional class)들의 동기화에 대하여 알아보았다. 그리고 각각의 기능성군의 동기화를 바탕으로 기능성군 사이의 계층구조 (hierarchy)를 정의하기 위하여 매개변수  $r_{link}$ 를 도입하였다. 여기에서  $r_{link}$ 는 국소적 동기화의 정도를 나타내는 변수이다. 이를 이용하여 단백질 상호작용 네트워크의 성질과 기능성군의 계층적인 구조사이의 관련성에 대하여 논의했다.

PACS numbers: 87.16.Yc, 89.75.Fb, 05.45.Xt

Keywords: 복잡계 네트워크, 생체 네트워크

### I. 서론

단백질 상호작용은 많은 생물학적 기능에서 매우 중요한 역할을 한다. 예를 들어 한 개의 세포 외부로부터 나오는 신호는 세포 속에 있는 단백질 상호작용을 통해 조절된다. 이러한 단백질 상호작용에 관한 방대한 데이터는 yeast two-hybrid 실험 등을 통해서 얻을 수 있다 [1,2]. 생물학적 상호작용 네트워크의 전체적인 구조 (large-scale topology)는 다양한 동역학적 특성들과 생물학적 안정성 등에 영향을 받는다 [3,4]. 단백질 상호작용의 특성 중 하나는 같은 기능성군에 속한 단백질들이 서로 쉽게 상호작용을 하는 경향을 보인다는 것이다 [5]. 따라서 단백질 상호작용 네트워크 (protein-protein interaction network; PIN)상에서 같은 기능성군에 속한 단백질들끼리는 서로 뭉쳐져 있는 구조를 갖는다 [5]. 이러한 기능성군은 PIN에서의 modular structure를 유발한다. 많은 종류의 PIN 구조에서 다양한 방법에 의한 modular structure를 정의하려는 시도가 있어왔다 [6-9]. 특히 참고문헌 [6,7,9]에서는 coupled Kuramoto oscillator [10,11]의 동역학적 특성을 이용하여 modular structure 및 그들 간의 계층구조를 연구하였으며, 이는 몇몇 복잡계 네트워크의 구조 형성 과정이 phase oscillator의 동기화 과정으로 표현될 수 있는 동역학과 밀접한 연관이 있음을 보여준다. 본 연구에서는 PIN의 modular structure 사이의 관계를 이해하기 위하여 각각의 단백질을 하나의 phase oscillator로

간주한다. 이를 phase oscillator의 동기화 현상을 이용하여 각각의 기능성군 사이의 계층구조의 특성을 살펴볼 것이다.

우선 우리 주변에서 많이 볼 수 있는 동기화 현상에 대하여 살펴보도록 하자. 예를 들어 우리 몸 안에서 뛰는 맥박이나 심장 등의 장기들이 동기화를 이루며 작용하고 있다 [12]. 또한 자연에서 많이 관찰되는 귀뚜라미들의 울음소리가 일치하여 들린다던지 [13], 혹은 반딧불이들의 발광 주기가 일치하는 것 [14] 등이 대표적인 동기화 현상이다. 서로 상호 작용하는 진동자들은 얼핏 보면 매우 간단한 시스템인 것 같지만, 아주 다양하고 흥미로운 현상들을 나타낸다. 이러한 동기화 현상을 설명하는 가장 간단한 모형으로 Kuramoto model을 생각할 수 있다. Kuramoto model은 임의의 시간  $t$ 에서 진동자  $i$ 의 위상 (phase)  $\theta_i(t)$ 의 시간에 따른 변화가 다음과 같은 식으로 주어진다.

$$\frac{d\theta_i}{dt} = \omega_i + \lambda \sum_{j=1}^N A_{ij} \sin(\theta_j - \theta_i) \quad i = 1, 2, \dots, N \quad (1)$$

여기서  $\omega_i$ 는 진동자  $i$ 의 고유진동수를 나타내고  $[-\pi, \pi]$ 에서 uniform하게 분포되어 있다고 가정하였다.  $\lambda$ 는 진동자  $i$ 와 연결된 진동자들과의 결합계수 (coupling coefficient)를 나타낸다.  $A_{ij}$ 는 진동자  $i$ 와 진동자  $j$ 가 연결되어 있다면 1이고, 그렇지 않다면 0으로 준다. 그리고

Table 1. Identification (ID) number of protein functional classes and the number of proteins in each functional classes.

Functional class ID	Functional class	Number of proteins
1	metabolism	177
2	energy	32
3	cellcycle and DNA processing	261
4	transcription	277
5	protein synthesis	104
6	protein fate	251
7	protein with binding function or cofactor requirement	303
8	regulation of metabolism and protein function	78
9	cellular transport, transport facilities and transport routes	161
10	cellular communication / signal transduction mechanism	65
11	cell rescue, defence and virulence	92
12	interaction with enviroment	100
13	cell fate	99
14	development	17
15	biogenesis of cellular components	211
16	cell type differentiation	142
17	unclassified protein	8

$N$ 은 총 진동자의 개수를 나타낸다. Kuramoto model의 동기화 전이현상을 기술하기 위하여

$$re^{i\varphi} = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j} \quad (2)$$

로 정의되는 질서매트 변수 (order parameter)를 사용한다. 여기서  $r$ 은  $0 \leq r \leq 1$ 의 값을 가지며 위상의 결맞음을 측정하는 값이고,  $\varphi$ 는 평균 위상을 나타낸다.  $r=1$ 인 경우 모든 위상이 동기화 된 phase를 나타내며  $r=0$ 은 incoherent phase를 나타낸다. 앞에서 언급된 바와 같이 참고문헌 [6, 7]에서는 복잡계 네트워크에서의 modular structure 또는 각 module사이의 계층구조를 Kuramoto model의 동기화 현상을 이용하여 보일 수 있었다. 본 연구에서는 참고문헌 [15]에서 기재되었던 yeast PIN에서의 modular structure의 상호관계를 Kuramoto model을 사용하여 설명하고자 한다.

## II. 단백질 상호작용 네트워크

우리는 단백질 상호작용 네트워크에 대하여 연구하기 위하여 앞에서 언급했던 yeast PIN을 고려할 것이다 [15]. Kuramoto model의 동기화 전이 현상을 이용하기 위하여 PIN에서 가장 큰 클러스터 (largest cluster; LC)만 생각하였다. 그 이유는 동기화 현상이 일어나기 위해서 인접노드와 상호작용이 있어야 하지만 따로 분리된 클러스터들이 존재한다면 전체적인 동기화가 이루어질 수 없기 때문이다. 본 연구에서 사용된 PIN의 largest cluster의 크기  $N$ 은 778이다. 이 네트워크를 이루는 각각

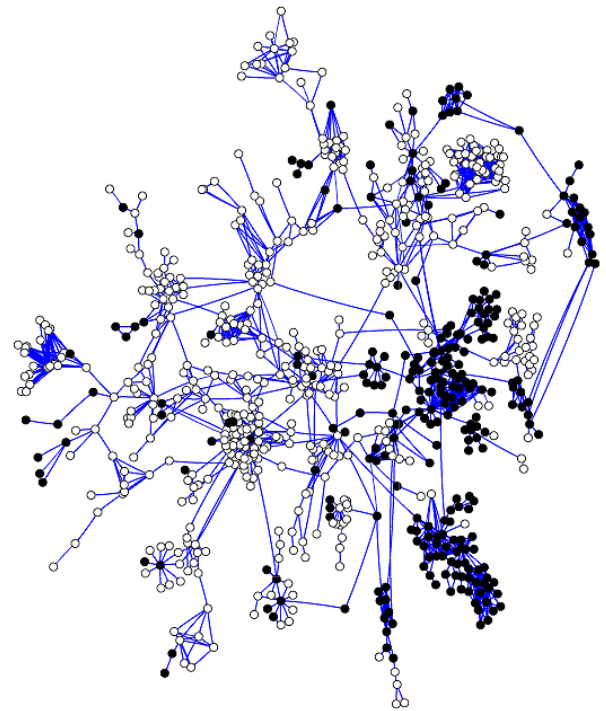


Fig. 1. Distribution of transcriptional proteins in PIN. Black circles represents the proteins belonging to the transcriptional class.

의 단백질은 참고문헌 [16]의 자료를 바탕으로 하여 17개의 기능성군으로 분류하였다 (Table 1 참조). 하나의 단백질은 한 가지의 기능성군에만 포함되는 것이 아니고 여러 가지 기능성군에 포함될 수 있다. 참고문헌 [5]에 의하면 PIN에서 동일한 기능성군에 속한 단백질들이 서로 뭉쳐있는 경향이 있음이 알려져 있다. Fig. 1은 이러한 경향을 보여주는 한 예이다. 여기에서 검은색 노드들은

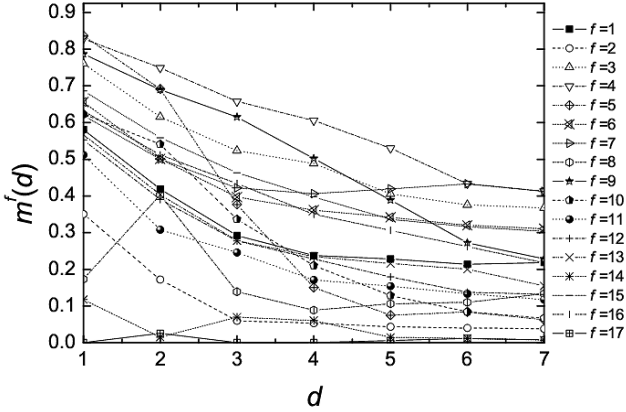


Fig. 2. Plot of  $m^f(d)$  against  $d$ .

전사군 (transcriptional class)에 해당하는 단백질을 나타낸다. 그림에서 볼 수 있는 바와 같이 검은 노드들은 검은 노드들 사이에서 더 많은 상호작용을 하고 있음을 볼 수 있다. 이와 같은 PIN에서의 현상은 기능성군에 따른 분리 (functional segregation) 현상으로 설명할 수 있고, 이것은 전사군 외의 다른 기능성군에서도 동일하게 관찰되었다 [5]. 이러한 기능성군에 따른 분리현상을 정량적으로 나타내기 위하여

$$m_i^f(d) = \frac{M_i^f(d)}{M_i(d)} \quad (3)$$

와 같이 segregation function을 정의하였다 [5]. 여기에서  $M_i^f(d)$ 는 단백질  $i$ 에서부터 거리  $d$  만큼 떨어진  $f$ 의 기능성군에 속해 있는 단백질의 개수를 나타내며  $M_i(d)$ 는 단백질  $i$ 에서부터 거리  $d$  만큼 떨어진 단백질의 개수를 나타낸다. 모든  $i$ 에 대한  $m_i^f(d)$ 의 평균  $m^f(d)$ 를 구함으로써 각각의 기능성군들이 어떻게  $d$ 에 따라 분포되어있는지를 살펴 볼 수 있다. 이때  $d$ 는 PIN에서 주어진 두 단백질 사이의 최소 연결선 수를 나타낸다. Fig. 2에서 보는 것과 같이 거리  $d$ 가 증가함에 따라 functional segregation 값이 감소하는 것을 볼 수 있다. 즉, 가까운 거리에서는 대부분 같은 기능성 군으로 이루어진 단백질들이 있다는 것이고, 이것으로 보아 같은 기능성군을 가지고 있는 단백질들끼리 서로 뭉쳐있다는 구조적인 특성을 정량적으로 확인할 수 있다.

### III. 동기화 클러스터

기존의 연구결과들은,  $A_{ij}$  행렬의 element들이 단순히 연결 유무만을 나타내지 않고 가중치를 가지게 되는 경우 전체 동기화가 가중치에 영향을 받음을 보였다 [17].

본 논문에서 우리는 PIN에서 상호작용을 하는 단백질들이 얼마나 동일한 기능을 많이 공유하느냐에 따라 가중치를 가정하였다. 이를 위해 다음과 같은 가중치 행렬  $W$ 를 도입하였다.

$$W_{ij} = \begin{cases} \frac{1}{2}(1 + n_{ij}) & \text{단백질 } i \text{와 } j \text{가 연결되어 있을 때} \\ 0 & \text{단백질 } i \text{와 } j \text{가 연결되어 있지 않을 때} \end{cases} \quad (4)$$

여기서  $n_{ij}$ 은 다음과 같이 정의되어 진다.

$$n_{ij} = \frac{1}{2} \left( \frac{O_{ij}}{n_i} + \frac{O_{ji}}{n_j} \right) \quad (5)$$

식 (5)에서  $O_{ij}(=O_{ji})$ 는 단백질  $i$ 와 단백질  $j$ 가 같이 포함하고 있는 기능성군의 개수이며  $n_i, n_j$ 는 각각 단백질  $i$ 와 단백질  $j$ 가 포함하고 있는 총 기능성군의 개수이다. 즉, 단백질  $i$ 의 기능성군이 이웃하는 단백질  $j$ 의 기능성군과 얼마나 동일 하느냐에 따라  $n_{ij}$  값이 결정되고 완전히 동일하다면 1, 그렇지 않다면 0의 값을 갖는다. 따라서  $W_{ij}$ 는 PIN에서  $i$ 와  $j$ 가 연결은 되어있으나 공통된 기능성군을 가지지 않는 경우 ( $O_{ij} = O_{ji} = 0$ )  $W_{ij}=1/2$  이 되고  $O_{ij}$ 가 증가함에 따라  $W_{ij}$ 는 1에 가까워지게 된다. 또한 정의 (5)에 의하면  $n_{ij} = n_{ji}$ 가 되므로  $W_{ij} = W_{ji}$ 를 만족한다. Fig. 3에서는 서로 연결된 (상호작용을 하는) 단백질  $i$ 와 단백질  $j$ 가 두개의 공통된 기능성군을 가지고 있을 때 ( $O_{ij} = O_{ji} = 2$ )  $A_{ij}$  값을 정의하는 예를 보여준다. 여기에서  $n_i=5$ 이고  $n_j=3$ 이라하면 이로부터  $A_{ij}=0.7655$ 를 얻을 수 있다. 식 (4)를 이용하여 각 연결선의 가중치가 다른 네트워크에서의 Kuramoto model은

$$\frac{d\theta_i}{dt} = \omega_i + \lambda \sum_{j=1}^N A_{ij} W_{ji} \sin(\theta_j - \theta_i) \quad (6)$$

로 표현될 수 있다. 본 연구에서는 식 (6)을 풀기 위해서 4차 Runge-Kutta method를 사용하였다. 이제 각각의 기능성군에 속해있는 진동자의 국소적 동기화를 정량적으로 나타내기 위하여 새로운 변수  $r_{link}$ 를 도입하였다 [18].

$$r_{link} = \frac{1}{2L} \sum_{i,j=1}^N \Gamma_{ij} \left| \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{t_r}^{t_r + \Delta t} e^{i[\theta_i(t) - \theta_j(t)]} dt \right| \quad (7)$$

여기서  $L$ 은 총 연결선 수를 나타낸다. 행렬  $D$ 의 element를 이용하여 동기화된 클러스터를 찾아내기 위해서 다음과 같이  $D_{ij}$ 를 정의하였다.

$$D_{ij} = A_{ij} \left| \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{t_r}^{t_r + \Delta t} e^{i[\theta_i(t) - \theta_j(t)]} dt \right| \quad (8)$$

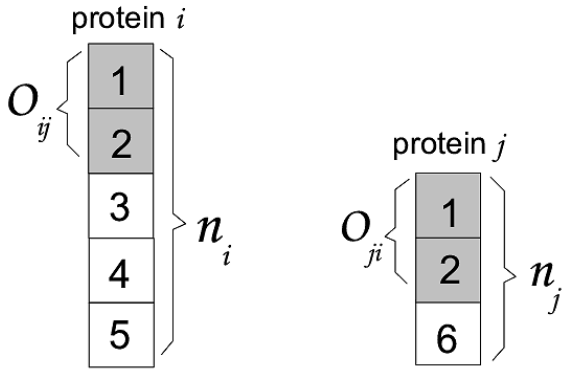


Fig. 3. An example to assign weights  $\omega_i$  and  $\omega_j$  when proteins  $i$  and  $j$  are connected.

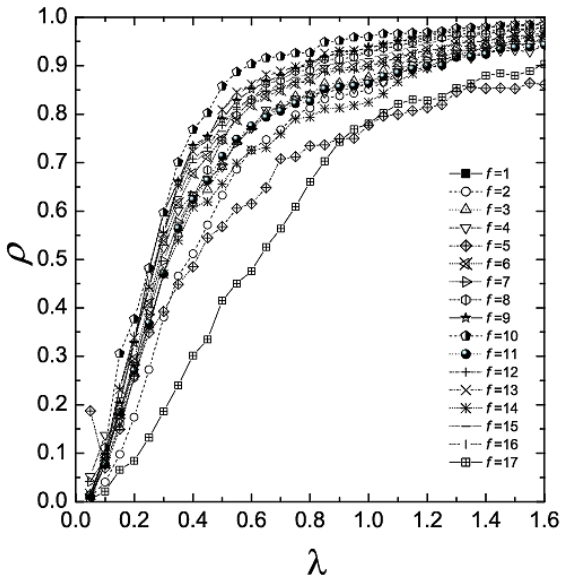


Fig. 4. Plot of  $\rho(\lambda)$  against  $\lambda$  for 17 functional classes.

여기에서 우리는 0과 1사이의 임의의 값을 갖는 threshold  $T_r$ 을 도입하여  $D_{ij} > T_r$ 이면  $i$ 와  $j$ 는 서로 동기화된 진동자로 간주하고 진동자  $i$ 와 진동자  $j$ 를 연결한다. 이와 같이 연결된 동기화된 진동자 쌍의 개수의 비율이  $r_{link}$ 와 동일하도록  $T_r$ 을 조절함으로써  $D_{ij}$ 를 filtering 할 수 있다. 이를 통하여 서로 동기화된 진동자끼리 묶여있는 동기화 클러스터를 정의할 수 있다 [18]. PIN의 경우에는 같은 클러스터에 속해있는 단백질들은 어떤 임의의 생물학적 기작을 위해 협동하고 있다고 가정한다. 따라서 동기화된 클러스터들 중 가장 큰 클러스터 안에 각각의 기능성군이 얼마나 포함되어 있는지를 조사해 보면 어떠한 기능성군들이 주어진 생물학적 기작을 위해 협동하는지에 대한 단서를 알 수 있으리라 예상된다.  $\lambda$ 값을 변화시키며 따라 동기화 클러스터의 크기가 다르게 나타나고 각각의  $\lambda$ 값에 따른 동기화 클러스

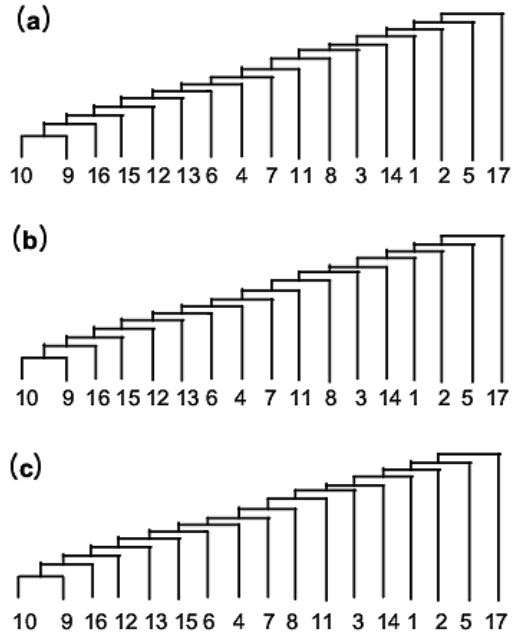


Fig. 5. Hierarchical structure between functional classes based on the order to reach a specific value of  $\rho$ . (a)  $\rho = 0.4$ , (b)  $\rho = 0.5$ , (c)  $\rho = 0.6$ .

터가 생성된다. 그리고 PIN에서 나타나는 동기화된 클러스터들 중에서 가장 큰 클러스터 (largest cluster)를 고려한다. 우선  $f$ 기능성군에 속해있는 단백질 전체의 개수  $n_{fT}$ 과 이중에서 largest cluster에 속해있는 개수  $n_{fL}$ 의 비율  $\rho = n_{fL}/n_{fT}$ 를  $\lambda$ 를 변화시키며 측정 하였다. Fig. 4에서 볼 수 있듯이 임의의 주어진  $\rho$ 에 도달하게 되는  $\lambda$ 는 기능성군  $f$ 에 따라 다르게 나타난다.

Fig. 5는 임의의 일정한  $\rho$ 에 도달하는 가장 작은  $\lambda$ 를 가지는 기능성군부터 차례로 묶은 결과를 보여준다. 그림에서 볼 수 있듯이 임의의 고정된  $\rho$ 에 도달하는 순서는  $\rho$ 에 상관없이 거의 동일한 경향을 보인다. 이는 주어진 임의의 생물학적 기작을 수행하기 위하여 일정한 비율의 단백질들이 서로 상호작용을 하고 있음을 보여준다. 또한 단백질간의 상호작용은 phase oscillator들의 상호작용을 근사시킬 수 있음을 보여준다.

단백질의 기능성군 사이의 계층구조를 조사하기 위하여, 임의의 주어진  $\rho = const.$ 에 도달하는  $\lambda$ 값의 차이가 작은 기능성군들은 서로 쉽게 협동을 할 수 있다고 가정 하였다. 이로부터 기능성군들 사이의 거리  $\ell$ 을 다음과 같이 정의한다.

$$\ell = |\lambda_i(\rho) - \lambda_j(\rho)| \quad (9)$$

이 때  $\lambda_i(\rho)$ 는  $i$ 기능성군이 LC에  $\rho$ 만큼 포함되었을 때의  $\lambda$ 값을 의미한다. 참고문헌 [5]에서는 다른 방법으로 정

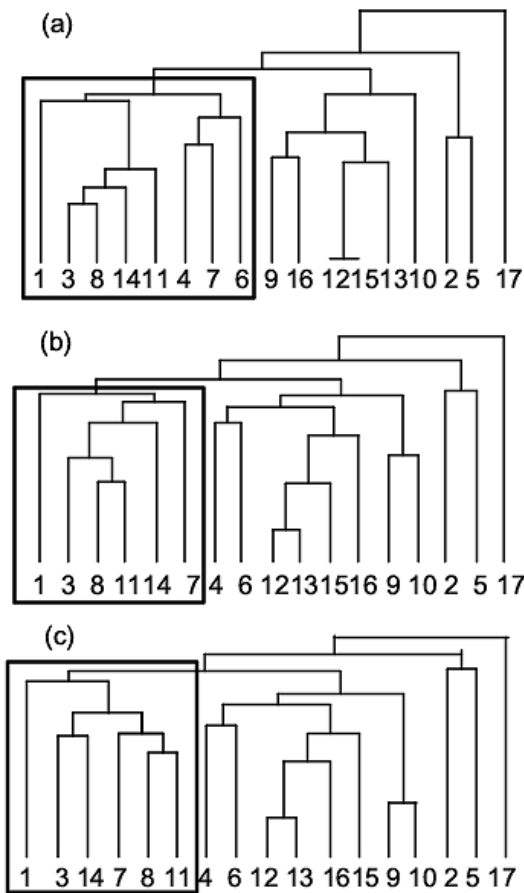


Fig. 6. Hierarchical structure between functional classes based on Eq. (9). (a)  $\rho = 0.4$ , (b)  $\rho = 0.5$ , (c)  $\rho = 0.6$ .

의된  $l$ 을 이용하여 나타낸 계층구조에 대하여 언급하고 있다. Fig. 6은 각각의 고정된  $\rho$ 에서 식 (9)를 이용하여 얻은 기능성군 사이의 계층구조를 보여준다. 이러한 계층구조를 살펴보면 모든  $\rho$ 값에 대하여 동일한 가지 안에 비슷한 기능성군들이 나타나는 것을 볼 수 있다. 예를 들어 사각형으로 표시된 첫 번째 가지에서 모든  $\rho$ 에 대하여 공통적으로 1, 3, 7, 8, 11, 14 기능성군이 자리하고 있음을 확인할 수 있다. 이러한 고정적인 패턴들이 단백질 상호작용 네트워크의 구조적 특성을 Kuramoto oscillator로 근사되어질 수 있는 단백질들 사이의 상호작용과 밀접한 관계가 있음을 보여주는 단서가 된다.

#### IV. 결 론

단백질 상호작용 네트워크의 기능성군 사이의 특성과 동기화 모델을 사용하여 나타낸 계층구조는 임의의  $\rho$ 값에 관계없이 나타나는 것을 알 수 있었다. 이는 단백질 상

호작용 네트워크 (PIN) 안에서 동일한 기능성군에 속한 단백질들 사이의 상호작용이 서로 다른 기능성군에 속한 단백질 사이의 상호작용보다 많이 일어남을 segregation function,  $m_i^f(d)$ ,를 통하여 확인 하였다. 이러한 구조적 특성은 PIN이 생물학적 기능성군들에게로 네트워크의 modular structure를 가지고 있음을 정량적으로 보여주고 있다. 참고문헌 [19]에 따르면 coupled phase oscillator 네트워크에서 modular structure 내부의 연결 상태에 따라 전역적 동기화가 일어나지 않더라도 각 module들 내부에서의 국소적 동기화가 발생됨을 보였다. 이러한 사실을 이용하여 본 논문에서는 PIN을 구성하는 각 단백질들을 phase oscillator (Kuramoto model)로 근사시키는 모형을 고려하였다. 이와 같이 근사되어진 phase oscillator들의 동기화 패턴으로부터 각 기능성군 사이의 일관된 계층구조를 발견하였다. 이는 PIN을 구성하는 근본 기작 (mechanism)이 coupled phase oscillator의 동기화 과정으로 설명되어 질 수 있다는 것을 뜻한다. 이러한 연구의 다양한 분야에서의 응용으로 인터넷 커뮤니티의 생성과정과 그 계층적 구조를 정량적으로 이해할 수 있는 방법 역시 제공할 수 있으리라 예상된다. 이 외에도 먹이 사슬 네트워크에서 볼 수 있는 구조들 또는 비즈니스 네트워크와 같은 경제계 등의 광범위한 네트워크에서 볼 수 있는 다양한 구조의 형성 메커니즘과 그 계층 구조연구에도 쉽게 적용될 수 있을 것으로 기대된다.

#### 감사의 글

이 논문은 2006년도 및 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원(R01-2007-000-10910-0, R01-2006-000-10470-0, F01-2006-000-10093-0)을 받아 수행된 연구임.

#### 참 고 문 헌

- [1] S. Fields, R. Sternglanz, The two-hybrid system: an assay for protein-protein interactions, Trends in Genetics **10**, 286 (1994).
- [2] Two-hybrid analysis of genetic regulatory networks, <http://cmmg.biosci.wayne.edu/finlab>.
- [3] S. Maslov, K. Sneppen and U. Alon, Science **296**, 5569, 910 (2002).
- [4] H. Jeong, S. P. Mason, A.-L. Barabaši and Z. N. Oltvai, Nature **411**, 41 (2001).

- [5] S. H. Yook, Z. N. Oltvai and A.-L. Barabaši, *Proteomics* **4**, 928 (2004).
- [6] S. Boccaletti and M. Ivanchenko, *Phys. Rev. E* **75**, 045102 (2007).
- [7] A. Arenas and A. D. Guilerá, *Phys. Rev. Lett.* **96**, 114102 (2006).
- [8] M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **103**, 8577 (2006).
- [9] J. Ren, H. Yang and Y.-C. Zhang, *Cond-mat/0703232v2* (2007).
- [10] Y. Kuramoto, in *International Symposium on Mathematical Problems in Theoretical Physics Lecture Notes in Physics* Vol. 30, edited by H. Araki (Springer, New York, 420 (1975)).
- [11] Y. Kuramoto, *Chemical Oscillation, Waves, and Turbulence* (Springer-Verlag, New York, 1984).
- [12] C. S. Peskin. *Mathematical Aspects of Heart Physiology* (Courant Institute of Mathematical Science Publication, New York, 1975), P. 268.
- [13] T. J. Walker. *Science* **166**, 891 (1969).
- [14] J. Buck. *Quart. rev. Biol.* **63**, 265 (1988).
- [15] J. D. Han, N. Rertin. *Nature* **430**, 88 (2004).
- [16] Munich information center for protein sequences - online protocol: <http://mips.gsf.de>.
- [17] A. E. Motter, C. Zhou and Kurths, *Phys. Rev. E.* **71**, 016116 (2005).
- [18] J. Gómez-Gardeñes, Y. Moreno and A. Arenas, *Phys. Rev. Lett.* **98**, 034101 (2007).
- [19] M. Zhao, T. Zhou, and B.-H. Wang, arxiv:0711.0530 [physics.data-an].

## Study of the Hierarchy Among the Functional Classes in Protein-protein Interaction Networks

Young-Jin KO, Soon-Hyung YOOK and Yup KIM

*Department of Physics and Research Institute for Basic Science, Kyung Hee University, Seoul 130-701*

(Received 21 November 2007)

Detecting community structures and hierarchy among communities has been one of the most attractive research topics in complex network studies. In this study, we regard each protein as an oscillator that interacts with its neighboring proteins. The interaction between the proteins produces a complex network. First, we study the synchronizability of the proteins in the same functional classes of the protein-protein interaction network (PIN). In order to define the hierarchy among the functional classes based on the synchronizability of each functional class, we introduce a parameter  $r_{link}$ . Here,  $r_{link}$  represents the fraction of all possible links whose end nodes are synchronized. Some possible relationships between the observed hierarchical structure of the functional classes and the properties of the PIN are also discussed.

PACS numbers: 87.16.Yc, 89.75.Fb, 05.45.Xt

Keywords: Complex networks, Biological network