

## Bio-Grid

노 경 태 ktno@camd.ssu.ac.kr

숭실대학교 분자설계연구센터 소장

### I. 서론

#### II. GRID 개념을 이용한 데이터, 분석기술 및 연산능력의 통합



현재 우리나라에서도 생명과학 분야의 연구와 산업이 활성화됨에 따라 많은 정보들이 폭발적으로 생산되고 있으나 조직적으로 관리되지 못하고 있는 상태이다. 이렇게 양산되는 데이터의 공유는 작게는 한 기업 내에서, 크게는 한 국가가, 더 나아가서는 세계적으로 풀어야 할 중요한 과제이다.

### I. 서론

#### 1. 21세기의 과학과 기술을 이끌어갈 생명정보

생명현상에 관한 정보의 구축 및 이를 가공하고 분석하는 기술은 21세기의 과학과 기술분야에서 가장 중요한 위치를 차지하며 전 세계의 산업에 미치는 영향력은 가히 짐작하기 어려운 정도이다. 특히 20세기말의 정보기술(Information Technology : IT)의 발전은 생명현상에 관련된 수많은 정보들을 조직적으로 관리하고 이를 분석하여 학문적인 이용 및 산업적 응용을 가능케 해주었다.

생체체의 다양성과 각 생체체가 포함하고 있는 수많은 생명현상은 엄청난 양의 생물정보 (Bioinformatics)를 생산해 내고 있으며 그 속도는 점차 가속화되고 있다. 특히 최근 인간 유전체 염기서열을 밝히는 연구를 시작으로 인간 외의 무수한 생체체의 염기서열들이 알려지 DB화되고 있다. 또한 유전체 정보보다 훨씬 복잡한 단백질의 서열, 구조, 기능 등의 정보들을 비롯하여 화합물과의 상호작용 및 대사 그리고 세포, 장기 및 이 들이 결합된 상태의 생체체로부터 나오는

무수한 데이터들이 계속 축적되고 있다. 이 정보들은 인류의 복지 및 생존에 매우 귀중한 자료일 뿐 아니라 21세기 산업을 이끌어 갈 인류 공동의 재산이다.

#### 2. 생명정보 관련 DB가 세계적으로 산재

생명현상 관련 정보는 전 세계적으로 생산되고 있으며 특히 선진국 그 중에서도 미국이 반 이상의 정보를 소유하고 있다. 이 중 공공기관이 소유하고 있는 정보는 대부분 외부에 공개되어 있거나 일정한 조건 하에서 무료로 제공받을 수 있다. 또한 많은 대학이나 기업 연구소 등에서 보유한 데이터는 여러 이유로 인하여 접근하기 힘들다. 현재 우리나라에서도 생명과학 분야의 연구와 산업이 활성화됨에 따라 많은 정보들이 폭발적으로 생산되고 있으나 조직적으로 관리되지 못하고 있는 상태이다. 이렇게 양산되는 데이터의 공유는 작게는 한 기업 내에서, 크게는 한 국가가, 더 나아가서는 세계적으로 풀어야 할 중요한 과제이다.

#### 3. 생명정보를 해석하는 기술(업고리즘)의 세계적 산재

생명현상의 데이터를 조직화하고 공유하는 과제 못지 않게 중요한 기술이 바로 이 데이터를 분석 가공하여 생명체에 관한 유용한 정보를 도출해내는 기술이다. 바로 이 해석 기술이 생명정보에 가치를 부여하며 생명 관련 산업들의 바탕이 된다. 몇몇 유전체 관련 분석 기술들은 이미 여러 공공기관에서 공개적으로 제공하고 있으나 훨씬 많은 기술들은 개인이나 산업체가 소유하여 내가를 지불하고 사용할 수 있는 경우가 대부분이다. 또한 많은 경우 이런 기술이 개별적으로 사용되어 얻을 수 있는 정보는 대개 매우 제한적이며 큰 가치를 가질 수 없다.

#### 4. 생명정보의 해석 및 응용에 소요되는 엄청난 연산 능력

생명현상이 워낙 수많은 단위 현상들의 조합으로 일어나고 또한 다루는 물질들이 거대 분자이기 때문에 이를 모사하고 분석하는데 엄청난 계산 시간이 필요하여 슈퍼컴퓨터를 이용하는 경우에도 원하는 시간 내에 분석결과를 얻지 못하는 경우가 매우 많다. 현재 공간적으로 산재되어 있는 컴퓨터들을 통신망을 통하여 효율적으로 연계 사용할 수 있는 많은 방법들이 실용화되어 가고 있으며 또한 대형 컴퓨터 제작사들이 생물정보처리 에 효율적인 시스템들의 개발에 많은 재원을 투자하고 있다.

## II. GRID 개념을 이용한 데이터, 분석기술 및 연산능력의 통합

GRID의 기본 개념 및 관련 기술에 관해서는 다른 원고를 참고하기 바람에 여기에서는 생물학적 문제와 관련된 응용분야로서의 Bio-Grid에 초점을 맞추어 알아 보기도 한다.

### 1. Bio-GRID 유형

현재 GRID를 응용한 모델의 유형들은 명확하게 제시할 수 있을 정도로 응용분야는 다양한 형태의 유형이

가능하며 계속적으로 그 개념 및 범위를 넓혀가고 있다. 현재 실제 구축되었거나 구축하려는 Bio-GRID를 유형별로 보면 다음과 같다.

*Computing-Bio-GRID* : 주로 분산컴퓨팅(Distributed Computing) 기술을 기반으로 하여 연산 능력을 최대한 확보

*DATA-Bio-GRID* : 산재되어 있는 생물학적 DB들을 효과적으로 연계하여 통합적인 DB 시스템을 구축

*Analysis-Bio-GRID* : 산재되어 있는 서로 다른 생물정보 분석 방법, 주로 S/W들을 연계하여 생물정보의 분석에 사용

*Portal-Bio-GRID* : 생물학적인 특정 문제 해결을 위하여 가능한 컴퓨터, 데이터 및 분석 방법들을 유기적으로 연계하여 운영되는 복합적인 시스템

## 2. Bio-GRID 예

현재 구축되어 운영되고 있는 대표적인 Bio-GRID들(표 1)에 정리하였다.

몇 대표적인 Bio-GRID의 설립 목적 및 구축 시스템에 관하여 간단히 소개한다.

### 1) Virtual Laboratory Project

"The virtual laboratory project is engaged in research, design, and development of Grid technologies that help in solving large-scale compute and data intensive science applications in the area of molecular biology."

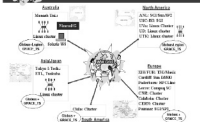
### 사용 S/W Resources

- Molecular Modeling Tools (DOCK)
- Parameter Modeling Tools (Nimrod/enFusion)

## Grid-응용

〈표 1〉 대표적인 Bio-GRID 프로젝트로, 시스템 또는 S/W를 정리

영칭/homepage	주관기관	내 용	구 분
Folded www.folded.org	UC Berkeley	단백질 구조 분석	계산
ImM-UD Cancer Research members.ucd.com/ projects/cancer/i ndex.htm	Intel, 미국암연 구재단, United- Devices, Oxford U., 일합회	단백질 상호작용 임 의한 선도물질 도출	계산
Virtual Lab. Project www.cse.mon ash.edu.au/~raj kumar/lab	호주 Monash 대학	POB의 단백질과 조합물론의 상호작용	계산데이터
Folding@home www.stanford.e du/group/pande group/CoSim/sd ence.html	Stanford 대학	단백질 구조 분석	계산
APBioNet www.apbionet.org	Nat. U. Singapore, APAN	Bioinformatics network	Portal
EURO Bio-GRID www.eurogrid.org	EURO GRID Bio W/G	여러 종류의 계산 package 계산자원	계산 Portal/거원
MOE www.chemcom p.com	Chem. comput group	컴퓨터를 이용한 화 학, 생물 과학 물질 설계, 정압 시스템	Portal
Genomics GRID	GFK Genomics W/G	유전체 응용 분야	Portal/거원
Postgenomics GRID	GFK Postgenomics W/G	post genomics 관련분야 응용	Portal/거원



〈그림 1〉 Virtual Laboratory Project의 computational resources 시스템

- Grid Resource Broker (Nimrod-G)
- Protein Data Bank (PDB) Management and Intelligent Access Tools
- Grid Middleware (Globus and GrACE)
- Grid Fabric Management (Fork/LSF/Condo/Codine/...)

## 2) APBioNet

"APBioNet is dedicated to the advancement of the field of bioinformatics, specifically, the development of the bioinformatics network infrastructure, the exchange of data and information, the development of training programs, workshops and symposia and encouragement of collaborations in the field of bioinformatics."

APBioNet는 다음의 생물정보관련 및 기술지원 W/G으로 구성되어 있다.

- Computational Structural Biology W/G
- Computational Immunology W/G
- Genome Analysis W/G
- Bio-informatics Directory W/G
- Advanced Networking W/G
- Integrated Resources Working Group.

## Resources

- DB : Australia Mirror, China Mirror, Japan Mirror, Malaysia Mirror, Singapore Mirror, USA Mirror
- Computation : Proposed Beowulf Linux Cluster Project

## 3) EUROGRID- BioGRID

"The Bio GRID will develop an access portal for biomolecular modeling resources. Bio GRID will develop interfaces to enable chemist and biologist

to be able to submit work to HPC facilities. This task will focus on development of various interfaces to biomolecular applications and databases.\*

#### Integrating S/W: UNICORE

#### Resources:

##### S/W

Quantum Chemistry:

Gaussian98 Gaussian94 Gamess TurboMole

CPMD ADF Molcas Molpro

Molecular Dynamics:

CPMD Amber Charmm Gromacs96

##### H/W

tsunami Cray T3E, tajfun Cray SV1, hydra Alpha Linux Cluster, Cray T3E-900, Cray T3E-1200, ZAMPano Intel Linux Cluster, AIX-Test Alpha (IBM), turing Cray T3E-1200, fermat SGI O2000 green SGI O3000, Cray T90, SGI O2000

##### DB

#### Literature databases

PubMed Literature database public access  
NIH (US)

MEDLINE and INSPEC Literature databases  
account required ICM

ERSCO Literature databases account required  
SWETS Literature databases account required

#### Structural databases

Protein Data Bank Biomolecular structural DB  
mirror at EBI (UK)

PDBLite PDB Search engine mirror at EBI  
(UK)

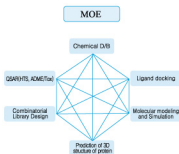
#### Sequence databases

CMS Molecular Biology Resource Biomolecular  
resources (US site)

Germany mirror (Univ. Bielefeld)

### 3. Bio-GRID의 형태를 위한 상업적 시스템 예

GRID의 개념을 생명정보관련 다양한 DB, 분석 S/W 및 장비의 통합에 중점을 두어 Molecular Operating Environment라는 시스템 통합 운영 시스템을 개발하여 운영.



〈그림 2〉 MOE를 통하여 연계되어 있는 DB 및 분석 subsystem들

### 4. 한국의 Bio-GRID Community (Working Group)

현재 국내에서는 Grid Forum Korea 소속의 두 개의 Bio-GRID 관련 W/G이 활동하고 있다.

- Genomics Bio-GRID: 주로 genomics 관련 DB 및 분석 기술 그리고 컴퓨터를 연계하는 portal 시스템을 구축할 예정임

- Postgenomics Bio-GRID: proteomics, metabolomics, physiomics, cheminformatics 관련 DB 및 분석 기술 그리고 컴퓨터를 연계하는 portal 시스템을 구축할 예정임

관련 내용은 GFK 홈페이지를 참조하십시오.