

# 분자모사를 위한 그리드 컴퓨팅 시스템 개발

김동욱<sup>0</sup>, 정갑주<sup>1</sup>, 황선태<sup>2</sup>, 정선호<sup>3</sup>, 이종현<sup>1</sup>, 김상선<sup>1</sup>, 최영진<sup>3</sup>  
<sup>0</sup>건국대학교 인터넷미디어 공학부, <sup>1</sup>국민대학교 컴퓨터학부, <sup>2</sup>건국대학교 미생물공학과  
{dwkim<sup>0</sup>, jeongk<sup>1</sup>, lejoy<sup>1</sup>, ponyboy<sup>1</sup>}@konkuk.ac.kr  
sthwang<sup>2</sup>@kookmin.ac.kr  
{shjung<sup>3</sup>, ototo<sup>3</sup>}@konkuk.ac.kr

## Developing the Grid-Computing System for Molecular Simulation

Dongwook Kim<sup>0</sup>, Karpjoo Jeong<sup>1</sup>, Suntae Hwang<sup>2</sup>, Seunho Jung<sup>3</sup>, Jonghyun Lee<sup>1</sup>, Sangsun Kim<sup>1</sup>,  
Youngjin Choi<sup>3</sup>

<sup>1</sup>Dept. of Computer Science & Engineering, Konkuk University

<sup>2</sup>Department of Computer Science, Kookmin University

<sup>3</sup>Department of Microbial Engineering, Kookmin University

### 요 약

분자모사 시뮬레이션은 일반 컴퓨터로는 수행이 불가능한 대량의 연산을 요구하기 때문에, 현재까지는 적극적으로 활용되는 한계가 있다. 그리드 컴퓨팅을 이용하여 요구되어지는 대량 컴퓨팅 파워를 해결할 수 있지만, 응용 특성에 최적화되지 설계된 그리드 컴퓨팅 시스템의 부재하기 때문에 활용되어지기 어렵다. 본 논문에서는 이러한 문제점 해결을 목표로 하는 분자모사에 최적화된 그리드 시스템 (Molecular Grid System) 구축에 관한 내용을 기술하고 있다.

### 1. 서 론

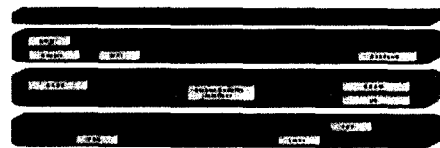
분자모사 시뮬레이션은 다양한 바이오·나노 물질 특성을 고비용·노동집약적 실험실내의 실험 없이 컴퓨터 시뮬레이션만을 통해 효율적으로 분석할 수 있는 기법이다.[4] 하지만, 이 시뮬레이션은 일반 컴퓨터로는 수행이 불가능한 대량의 연산을 요구하기 때문에, 현재까지는 초대형 슈퍼컴퓨터를 보유한 일부 연구기관에 속한 과학자들을 제외하고는 적극적으로 활용되고 있지 못하고 있다.

여기에 그리드 컴퓨팅은 이러한 대량 연산요구에 대한 근본적 해답이 될 수 있다.[3] 그러나 그리드 컴퓨팅을 분자모사 기반의 바이오·나노 연구에 적극적으로 활용하는 데는 심각한 문제가 존재한다. 가장 심각한 장애요인은 응용 특성에 최적화되지 설계된 그리드 컴퓨팅 시스템의 부재이다. 이와 반대로 기존의 응용에 최적화된 시스템들은 경직된 아키텍처 및 구현으로 새로운 그리드 환경에 통합이 매우 어렵다. 본 논문에서는 기존 분자모사 실험 소프트웨어 및 방식에 최적화되고 동시에 새로운 그리드 컴퓨팅 환경의 자원활용을 최적화할 수 있게 설계된 분자 그리드 시스템(MGrid)을 제안한다.

### 2. 관련 연구

그리드 관련 대표적인 미들웨어인 Globus[1]는 지리적으로 분산된 이종적인 컴퓨팅 자원들을 하나의 가상

컴퓨터처럼 사용 가능할 수 있도록 하는 소프트웨어 기반 구조를 제공한다. Globus의 특징은 객체 지향 모델과 같이 하나의 고정된 프로그래밍 패러다임을 취하는 대신 보안, 통신 및 자원 관리 등 기본적인 핵심 서비스만을 제공하고 응용 프로그램이 자신이 목적에 필요한 서비스를 선택, 조합함으로써 다양한 형태의 응용 프로그램 및 새로운 패러다임을 지원할 수 있다는 점이다. 현재 Globus 툴킷은 GRAM(자원 할당 및 프로세스 관리 서비스), NEXUS(통신 서비스), GSI(보안 관련 서비스), MDS(자원 관련 동적 정보 서비스), HBM(모니터링 서비스), GASS(원격지 데이터 액세스 서비스), GEM(실행 파일 관련 서비스), GARA(자원 예약 서비스) 로 이루어져 있다.



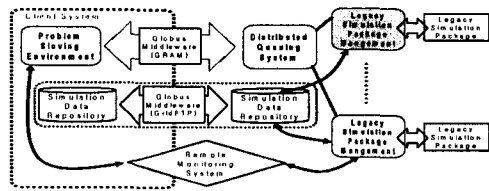
<그림 1> Globus Middleware Layer

### 3. MGrid 시스템

분자모사 시뮬레이션을 그리드 환경에 적용은 컴퓨팅 자원의 제공면에서 훌륭한 해결책이 되어질 수 있다.[3] 그러나 그리드 컴퓨팅 환경을 기반으로 기존 컴퓨팅 환경인 클라이언트·서버 모델을 구축할 수 없다. Globus와 같은 그리드 미들웨어를 통하여 모든 리소스에 접근

해야 하기 때문에 클라이언트·서버 모델을 적용할 수 없는 것이다. 이것은 그리드에서 사용자에게 편리한 환경을 제공하는데 제약으로 작용한다.

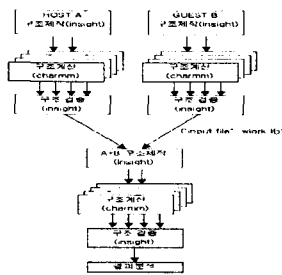
우리의 MGrid 시스템은 그리드 환경하의 사용자에게 로컬 환경과 동일한 작업환경을 제공하기 위해서 개발되어졌다. 아래 그림은 MGrid 시스템의 구조를 나타내고 있다. PSE(Problem Solving Environment)는 그리드 환경의 복잡성을 감추어서 사용자에게 편리한 실험 환경을 제공한다. PSE는 그리드의 대표적인 미들웨어인 Globus를 통하여 실제 자원에 접근한다. 그리고 사용자의 요청은 로컬 스케줄러에 의해서 분석되어 적당한 작업을 수행하고, 사용자는 Repository를 통하여 결과 파일에 접근하게 된다.



<그림 2> MGrid System Structure

3.1 PSE(Problem Solving Environment)

그리드 시스템을 효과적으로 이용하기 위해서는 바이오·나노 연구자 중심의 응용 플랫폼 환경이 요구된다. MGrid 시스템은 미생물학 분야에서 많이 수행되고 있는 단백질 분자 모사의 작업 수행과정을 대상으로 하여 분석하였다. 이 실험의 수행과정을 workflow 형태로 정의하고 각각의 작업 절차를 task로 구분하여 각각의 task 간 상호 의존성에 따른 작업 스케줄링 모델을 설계하였다.[6]



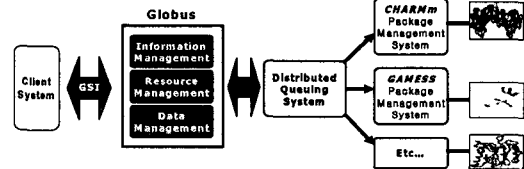
<그림 3> 단백질 3차 구조 생성을 위한 workflow

3.2 Legacy와 그리드 환경의 통합

현재 그리드에 관련된 대표적인 미들웨어인

Globus[1] 개발은 범용적인 툴의 형태로 진행되어지고 있다. 또한 바이오·나노 분자모사를 수행하는 위해 사용하는 전통적 분자모사 소프트웨어(Legacy Simulation Software)들(CHARMm, GAMESS, ..)은 Globus와 연동에 대해서 전혀 고려되지 않고 오랫동안 개발되어왔다. 그래서 Globus하에서 사용자에게 로컬머신상과 동일한 작업환경을 제공하는 것은 불가능하다.

이러한 이질적인 소프트웨어들을 연동시키기 위해서 MGrid 시스템은 Legacy 소프트웨어 관리 및 모니터링에 관련된 표준 통합 인터페이스를 정의하고 있다. 이 인터페이스를 통하여 새로운 Legacy 소프트웨어를 위한 모듈을 Plug-In 방식으로 연결할 있다. 따라서 새로운 실험 분야에 MGrid 시스템을 적용하기 위해서는 기존에 사용하고 있는 Legacy 소프트웨어를 지원하는 모듈을 새롭게 작성하여 MGrid 시스템에 접목하면 된다.

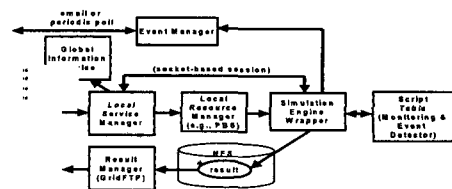


<그림 4> Globus환경하에 Legacy소프트웨어 통합

3.3 서비스 제어(Service Access)

그리드 환경에서 컴퓨팅을 위해서 제공되고 있는 자원들은 대부분은 클러스터와 같은 형태이다. 이러한 자원들은 중앙의 Master 노드를 통하여 접근하고, PBS[2] 또는 Condor 같은 로컬 스케줄러에 의해서 리소스에 접근한다. 모든 리소스의 접근이 로컬 스케줄러에 의해서 제어되기 때문에 이미 수행시킨 작업 다시 접근하는 것이 어렵다.

우리의 MGrid 시스템은 리소스들의 기존 로컬 정책을 유지하면서 이 문제를 해결할 수 있도록 구현되었다. 다음 그림과 같이 로컬 스케줄러를 이용하여 Legacy 소프트웨어를 제어할 수 있는 Engine을 기동시킨다. 그리고 이 Engine과 세션을 연결하고 Legacy 소프트웨어를 제어한다.



<그림 5> Resource Access & Notification

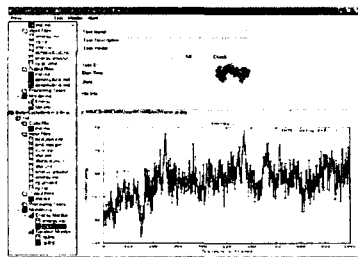
### 3.4 모니터링 및 통지(Monitoring and Notification)

일반적인 작업과 다르게 바이오·나노 시뮬레이션의 수행은 몇 일 또는 몇 주간의 작업을 수행한다. 사용자가 이 기간을 위해서 시스템에 접속하여 중간 상황을 계속해서 점검하는 것은 불가능한 일이다. 이를 위해서 MGrid 시스템에서는 사용자가 작업을 수행시키고, 이후 완료되는 시간까지 다른 작업을 수행할 수 있다. 작업이 완료되는 시점에 MGrid 시스템에서 사용자에게 E-mail 같은 도구를 이용하여 작업의 완료 또는 실패를 통지하여 사용자가 다음 작업을 수행할 수 있도록 하고 있다. <그림 5>는 Notification의 작업 흐름이 나타나있다.

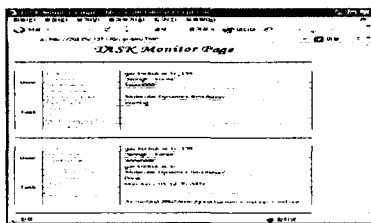
### 4. MGrid 시스템 구현 및 검증

현재 구축되어진 MGrid 시스템은 바이오·나노 분자모사 시뮬레이션을 위해 요구되는 많은 컴퓨팅 자원을 얻기 위해 분산된 자원을 그리드 미들웨어 톨인 Globus를 이용하여 분산된 컴퓨팅 자원들을 통합하고, 이를 기반으로 Legacy 분자모사 소프트웨어들은 연동시키기 위해서 Plug-in 방식의 표준 통합 인터페이스 개발 하였다.

사용자는 로컬 컴퓨터에서 원격 환경의 컴퓨터 자원에게 Legacy 분자모사 소프트웨어들을 관리 및 모니터링할 수 있도록 개발되었다.



<그림 6> PSE GUI 환경



<그림 7> 웹을 통한 모니터링 환경

### 5. 결 론

현재의 그리드 컴퓨팅에 관련된 기술들을 분자모사 기반의 바이오·나노 연구에 적극적으로 활용하는 데는 심각한 문제가 존재한다. 가장 심각한 장애요인은 응용 특성

에 최적화되게 설계된 그리드 컴퓨팅 시스템의 부재이다.

사용자에게 편리한 실험 환경을 제공하기 위해서는 바이오·나노 연구자 중심의 응용 플랫폼 환경이 요구된다. 우리는 이를 위해서 MGrid 시스템을 구축하고, 시스템 검증에 위하여 단백질 분자 모사를 수행과정을 분석하여 이를 workflow 형태로 정의하고, PSE 형태로 구축하였다.

Legacy 분자모사 소프트웨어 지원하기 위하여 Plug-in 방식의 표준 통합 인터페이스를 정의하여, 다양한 전통적 분자모사 소프트웨어(Legacy Simulation Software)들 통합할 수 있고, 이를 통하여 새로운 바이오·나노 실험에 MGrid 시스템을 적용할 수 있다. 그리고 원격 실시간 분자모사 작업 모니터링 및 Notification을 지원하여 사용자가 작업의 상태 정보를 확인할 수 있다.

현재 분자모사에 최적화된 응용 그리드 시스템(MGrid) 1차 프로토타입 개발하여 건국대 클러스터에 MGrid 시스템을 구축하였다. 구축된 시스템의 효용성 검증에 위하여 몬테카를로 (MC) 분자모사와 분자동역학 (MD) 분자모사를 결합해서 환형올리고당 Cyclodextrin (CD)의 광학선별성 연구를 수행중이며, 다양한 물질들에 대한 연구를 수행하여 결과를 데이터베이스로 구축하고 있다.

### 6. 참조 문서

[1] <http://www.globus.org/>  
 [2] <http://www.openpbs.org>  
 [3] I. Foster, C. Kesselman, S. Tuecke. International J. "The Anatomy of the Grid: Enabling Scalable Virtual Organizations." Supercomputer Applications, 15(3), 2001  
 [4] I. Foster, C. Kesselman. "Computational Grids." Chapter 2 of "The Grid: Blueprint for a New Computing Infrastructure", Morgan-Kaufman, 1999.  
 [5] I. Foster, C. Kesselman. "The Globus Project: A Status Report." Proc. IPPS/SPDP '98 Heterogeneous Computing Workshop, pp. 4-18, 1998.  
 [6] Doucet, J.P. and Weber, J. (1996) Computer-aided molecular design, Academic press, London.