

e-Organ Simulator 를 위한 Kepler 기반 협업 시스템

박상수**, 한영주*, 김동현*, 윤찬현**, 심은보***, 노경태****, 남기엽*****

*한국과학기술원 정보통신공학과

**한국과학기술원 전기및전자공학과

***강원대학교 기계메카트로닉스공학부

****연세대학교 생명공학과

*****분자설계연구소

e-mail : {sangsu, y.han, epick, chyoun}@kaist.ac.kr,
ebshim@kangwon.ac.kr, ktno@yonsei.ac.kr, kyn@bmdrc.org

Kepler based Collaborative System for e-Organ Simulator

Sangsu Park**, Youngjoo Han*, Dong-Hyun Kim*, Chan-Hyun Youn**,
Shim, Eun Bo***, Kyoung Tai No****, Ky-Youb Nam*****

*Dept. of Information and Communication Engineering., KAIST

**Dept. of Electrical Engineering, KAIST

*** Dept. of Mechanical & Biomedical Engineering, Kangwon National University

**** Dept. of Biotechnology, Yonsei University

***** Bioinformatics & Molecular Design Research Center

요 약

계산 집약적인 다양한 응용 프로그램들이 서로 유기적으로 통합되어 결과를 도출하는 컴퓨팅 환경에서는 응용들을 관리하기 위한 효과적인 워크플로우 기법과 컴퓨팅 자원관리 기술이 필요하다. 본 논문에서는 인체 생리기능 단위들을 모사한 신약개발 시뮬레이션을 지원하기 위한 OSICC (e-Organ Simulator-Integrated Cyber Computing) 시스템의 협업 기능을 평가한다. 제안하는 시스템은 워크플로우 관리자를 이용하여 시뮬레이션의 객체간 자원관리를 최적화하고 있으며, 프로토타입으로 구현되어 협업 컴퓨팅을 지원하는 OSICC 시스템에서 효율성을 확인하였다.

1. 서론

최근 들어 바이오 응용 기술을 개발함에 있어서 경제성, 사회적, 기술적 측면에서 효과적인 연구 및 개발을 위해 IT 융합기술에 대한 필요성이 증가하고 있다. 컴퓨터 모델을 이용한 시뮬레이션의 활용은 신약개발 기간을 단축할 수 있으며 50% 이상 신약개발 비용의 획기적 절감에 이바지할 수 있을 것으로 예상되어 신약개발에 있어서 IT 기술의 역할이 계속 증대될 것으로 전망하고 있다[1].

그러나, 시뮬레이션 결과를 컴퓨터상에서 파악할 수 있고 고성능 컴퓨팅 환경을 이용하여 신약 개발에 소요되는 시간, 비용을 획기적으로 줄이기 위한 가상적인 세포/장기의 컴퓨터 모델을 위해서는 크게 다음과 같은 2 가지 문제가 선결되어야 한다. 첫째, 다양한 응용 프로그램들이 서로 유기적으로 통합되어 결과를 도출하는 컴퓨터 모델 환경에서는 응용들의 상호관계를 관리하면서 수행을 조절하는 자율적이고 지능적이며, 동시에 유저에게는 투명한 워크플로우 관리기술이 필요하다. 기존 비즈니스 워크플로우는 복잡한 과학적 문제를 풀기 위해서 그리드 워크플로우와 같이 계산 집약적이고 데이터 분석에 초점을 맞춘 다른 속성을 가진 워크플로우에 적합하지 않다. 또한,

사용자가 요구하는 단위시간 내에 시뮬레이션 결과를 제공하기 위해서는 고성능 컴퓨팅을 제공하는 분산환경 인프라를 위한 관리 기술이 필요하다. 즉, 비용효율적인 대용량 컴퓨팅 제공을 위해 지리적으로 분산되어있는 컴퓨팅 자원을 유기적으로 결합하여 작업을 할당하고 관리하는 분산 자원관리 기술이 필수적이다. 기 개발된 PQRM 은 그리드 환경에서 사용자의 QoS 제약조건을 만족하면서 성능 면에서 유용함을 보였다 [3,4].

이에 본 논문에서는 세포-조직-장기(e-Organ)에 이르는 인체 생리기능 단위들을 모사한 신약개발 시뮬레이션을 지원하기 위해 상기에 제시한 2 가지 문제를 해결하는 OSICC (e-Organ Simulator-Integrated Cyber Computing) 시스템을 제안한다. 본 제안하는 시스템은 e-Organ 으로 구성되는 복잡한 약물독성 예측 시뮬레이션을 위해 워크플로우관리자를 이용해 시뮬레이션의 구성과 조합을 효율화하였으며, 분산된 자원을 효율적으로 이용하여 제한된 전체 시간 안에 시뮬레이션을 처리하기 위해 하부의 자원관리자를 연구하였다. 제안한 시스템은 실제 테스트베드에서 프로토타입으로 구현되어 신약개발 시뮬레이션을 위한 효율적인 컴퓨팅 환경 제공이 가능함을 확인하였다.

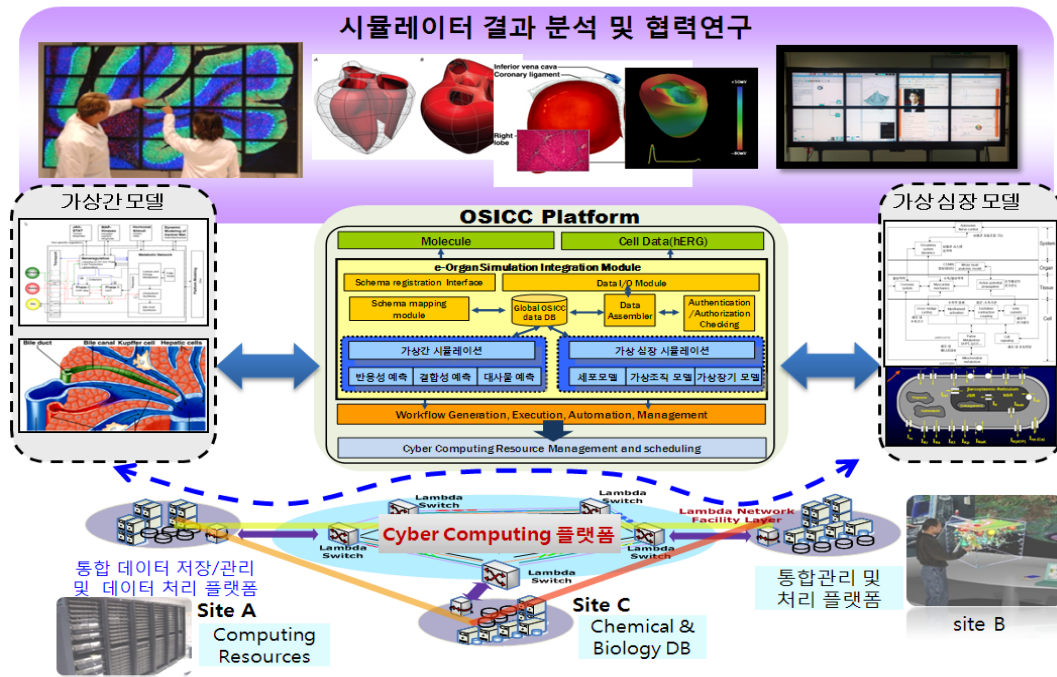
2. OSICC 시스템 모델

일반적으로 신약개발을 위한 약물독성 예측은 세포-조직-장기에 이르는 인체 생리기능 단위들을 모사한 시뮬레이션을 기반으로 약물 스크리닝 및 가상 임상 시험의 과정을 거치게 된다. 약물독성 예측 시뮬레이션은 크게 다음과 같은 2 가지 특성으로 규정할 수 있다. 첫째, 약물 독성 시뮬레이션은 최종결과를 얻기 위해 많은 프로그램의 조합으로 구성된다. 즉, 인체의 대사 반응을 모사하기 위해 복잡한 컴포넌트들의 연속된 수행으로 정의할 수 있다. 둘째, 시뮬레이션을 구성하는 각 개별 컴포넌트들은 계산 집중한 특성을 가진다. 따라서, 많은 컴퓨팅 파워를 이용하며 비교적 긴 실행시간이 요구된다.

이를 위해 본 논문에서는 상기에 제시된 2 가지 관점에서 분산된 사이버컴퓨팅을 이용하여 약물독성 예

에게 제공하기로 보장하는 서비스의 품질을 의미한다. 본 연구에서는 사용자의 요구사항을 취득하기 위해 SLA 를 사용하며 이것은 자원관리자가 최적 자원 선정 시 사용된다.

본 OSICC 시스템에 수행 요청이 이루어진 워크플로우는 일련의 과정을 거쳐 워크플로우관리자에게 전달된다. 복잡한 작업으로 구성된 응용을 효율적으로 관리 및 통제하기 위한 워크플로우관리자는 상위의 응용계층에서 요청한 작업을 인지하고 이를 효율적으로 처리하기 위해 자원관리자와 연동하여 사용자의 작업 처리를 지원한다. 또한, 자원관리자는 실제 작업의 처리가 일어날 물리 계층의 자원 관리 및 최적 자원 선정을 담당한다. 자원관리자는 실시간 자원 상태의 모니터링을 통해 실제 작업의 처리가 일어날 물리 계층의 자원 상태를 인지하고 이를 기반으로 최적의 자원을 선정하여 제공한다.



(그림 1) 약물독성 예측을 위한 OSICC 플랫폼

측 시뮬레이션을 효율적으로 지원할 수 있는 시스템을 설계하였다. 그림 1 은 OSICC 시스템의 개념도를 나타낸다. 본 논문에서 제안하는 응용 관리시스템은 디자인된 워크플로우와 SLA(Service Level Agreement) 를 시뮬레이션 실행을 위한 입력으로 사용한다. 약물독성 시뮬레이션을 수행하기 위해 사용자는 먼저 작업실행을 위한 워크플로우를 생성하여야 한다. 일반적으로 워크플로우의 디자인은 워크플로우 기술언어 (Workflow Description Language)를 이용하여 작업의 순서 및 작업들의 연결관계를 정하고 작업들 사이에 흐르는 데이터를 제대로 주고 받을 수 있도록 데이터 타입을 정의하는 것들을 의미한다. 본 연구에서는 Kepler [5]를 이용하여 명세된 워크플로우를 인지하여 실행할 수 있도록 하였다. 또한, 응용서비스 제공자와 사용자 사이의 응용 SLA 는 서비스 제공자가 사용자

2.1 워크플로우관리자

그림 2 는 워크플로우 관리시스템의 구조도를 나타낸다. 그림 2 의 상단에 도시된 워크플로우관리자는 시뮬레이션의 실행 및 관리 기능을 제공한다. 사용자가 케플러로 작성한 워크플로우 상세를 기반으로 워크플로우관리자는 워크플로우의 각 Actor(atomic operation) 기반의 관리를 통해 사용자의 워크플로우 제어 명령(실행, 일시정지, 다시시작, 정지)을 탐지하고 워크플로우의 흐름을 제어한다. 작업의 진행에 따라 작업을 진행하기 전에 작업의 정보와 수행 정보를 유저에게 전송하고, 작업수행 도중 에러가 발생하면 이에 따른 처리도 진행한다. 작업이 완료되면 마지막으로 작업 수행정보를 다시 한번 유저에게 전송한다. 이를 위해 다음과 같은 핵심 기능을 제공한다.

먼저, Kepler Execution web service server 는 케플러로 디자인된 워크플로우 명세로부터 웹 서비스 명세로 변경하는 기능을 담당한다. 즉, 단일한 워크플로우 명세를 통해 사용자는 다른 원격지에 있는 웹 서비스를 호출하여 사용할 수 있으며 추가적인 작업 없이 다른 웹 서비스를 이용하여 실험을 수행할 수 있다. 또한, 사용자가 제출한 추상 워크플로우(abstract workflow)의 각 actor 를 실제 수행할 자원에 매핑하는 기능을 수행한다. 이때, 사용자로부터 제출된 SLA 와 함께 자원 관리자를 호출하여 최적 자원을 선정하고 원하는 작업을 선정된 자원에 할당할 수 있다. 확장형 케플러 프로파일(Extended Kepler Profile, EKP)는 워크플로우의 각 actor 의 분산 환경에서의 실행을 관리하기 위한 기능을 제공한다.

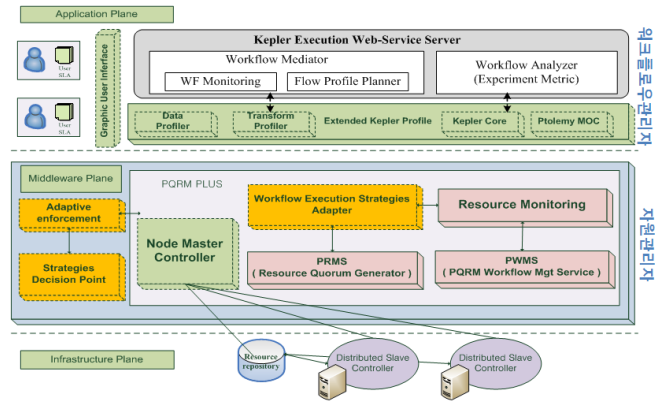
2.2 자원관리자

사용자는 워크플로우를 실행하는데 있어서 원하는 SLA 를 정의하여 워크플로우 관리시스템에 전달한다. 시스템은 사용자의 SLA 를 만족 시킬 수 있는 정책을 결정하고, 워크플로우가 실행될 때 결정된 정책이 사용자가 정의한 성능을 만족시키도록 한다. 전체 워크플로우 시스템은 사용자 응용 SLA 를 기반으로 자원 관리자를 통해 정책을 생성하고 워크플로우를 실행하여 사용자가 요청한 품질을 보장한다.

그림 2 의 하부에 도시된 자원관리자인 PQRM^{PLUS} 는 distributed controller 에 자원을 할당하기 위한 스케줄링 알고리즘을 제공하며 다음과 같은 하부 컴포넌트로 구성된다.

PQRM 워크플로우 관리서비스 (PWMS)는 PQRM 과 케플러 사이의 인터페이스를 담당하고 케플러의 추상 워크플로우에 대한 정보를 PQRM 에 전달한다. Strategy Decision Point 는 사용자의 SLA 와 정책의 조건(condition)을 비교하여 적절한 수행정책(policy)을 결정하며, 정책생성 시 Resource Quorum Generator 는 분산된 자원에서 워크플로우 실행을 위한 최적 자원 집합을 생성한다. 최적 실행 모듈(Adaptive Enforcement Module)은 PQRM^{PLUS} 에서 실제 실행(enforcement)이 일어나는 곳이다. Node Master Controller 는 분산된 자원을 명시하고 워크플로우를 실행하는 주된 모듈이며, 다른 원격 자원에서 실행되는 Distributed Slave Controller 를 통해 할당된 워크플로우를 실행하도록 한다. 마지막으로 자원 모니터링(Resource Monitoring)은 PQRM 이 최적 자원선정을 위해 필요한 자원상태

정보를 수집하는 시스템이며, 지원도구(Ganglia, IPerf)에 의해 수집된 자원 정보를 주기적으로 DB 화 하는 역할을 한다.



(그림 2) 워크플로우 관리시스템의 구조도

3. OSICC 프로토타입 및 성능 평가

OSICC 시스템은 IT 비전문가인 사용자가 약물독성 예측 시뮬레이션의 과정을 디자인하고 분산된 사이버컴퓨팅 인프라를 이용하여 워크플로우를 수행하고 시뮬레이션의 결과를 확인할 수 있는 통합서비스를 제공한다. 즉, 사용자는 작업수행 및 분산된 자원 사용에 대한 복잡한 작업 없이 사이버컴퓨팅 자원을 이용하여 쉽게 시뮬레이션을 실행하고 결과를 유지 관리할 수 있다. 그림 3 은 구현된 사이버컴퓨팅 기반 OSICC 시스템의 프로토타입을 나타낸다.

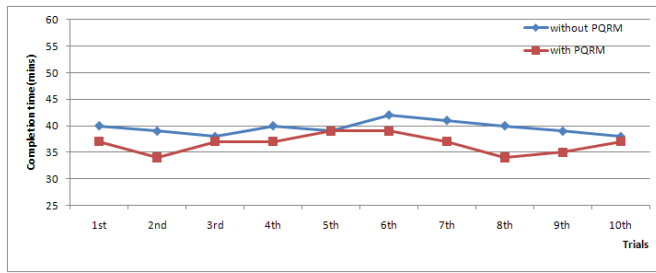
먼저, 그림 3 의 왼쪽은 사용자가 시뮬레이션하고자 하는 화합물을 upload 하는 과정이다. 오른쪽은 upload 된 신약 후보 화합물의 시뮬레이션 진행을 위한 파라미터 설정하는 화면을 나타낸 것이다. 또한, 자원관리자의 설정을 통해 사용자가 선택한 SLA 를 기반으로 최적자원을 설정할 수 있으며, 새로운 SLA 를 설정함으로써 자원 선정을 수정할 수 있다.

OSICC 프로토타입의 성능을 응용처리 요구 수준(QoS)에 대한 서비스 완료시간(Completion time)을 측정하여 OSICC 플랫폼의 최적 자원 관리 기능을 평가하였다. 그림 4 는 실험에 대한 결과로서, Lite 와 Premium 의 각 SLA level 에 대해서 실험을 하였다. PQRM 의 워크플로우 관리 서비스를 통해 워크플로우를 실행한 경우의 완료 시간이 그렇지 않은 경우에

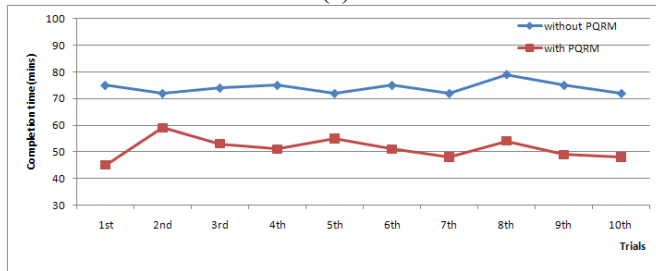


(그림 3) 사이버컴퓨팅 기반 OSICC 시스템

비해 단축 되는 것을 볼 수 있다.



(a)



(b)

(그림 4) PWMS 환경에서 완료 시간 비교
(a) SLA level in Lite(b) SLA level in Premium

4. 결론

본 논문에서는 계산 집약적인 작업으로 구성된 복잡한 시뮬레이션을 위해 최적 관리를 지원하는 OSICC를 실험적으로 구현하고 그 성능을 평가하였다. 제안하는 시스템은 워크플로우관리자를 통해 시뮬레이션의 구성과 실행을 최적화하였으며, 분산된 컴퓨팅자원을 효율적으로 관리하여 시뮬레이션을 실행하기 위한 자원관리자를 연구하였다. 제안하는 OSICC의 유효성을 평가하기 위하여 약물독성 예측 시뮬레이션을 지원하는 시스템 프로토타입을 구현하여 서비스 수준에 따른 그 효용성을 평가하였다.

Acknowledgement

본 연구는 지식경제부 및 한국산업기술평가관리원의 IT 산업원천기술개발사업의 일환으로 수행하였음. [2008-F-029-01, 사이버컴퓨팅 기반 e-Organ 시스템 개발]

참고문헌

- [1] PricewaterhouseCoopers report(www.pwc.com), 1999.
- [2] Workflow Management Coalition. <http://www.wfmc.org/>
- [3] C.-H. Youn, B. Kim, and Eun Bo Shim, "Resource reconfiguration scheme based on temporal quorum status estimation for grid management, IEICE Tran. on Comm., Vol.E88, no.11, pp.4378-4381, 2005.
- [4] C.-H. Youn et al., "Implementation and Evaluation of A PQRM-based PACS System," Proc. of 6th int'l special topic conf. on ITAB, 2007.
- [5] The Kepler Project. <https://kepler-project.org/>