

웹 서비스를 통한 분산 시스템 기반의 분자 모델링 도구의 실시간 상호작용

김보순¹, 박성준², 김지인³

드림위즈¹/건국대학교 컴퓨터공학과²/건국대학교 인터넷미디어공학과³
foxgap@dreamwiz.com¹, { hcipsj, jnkm }konkuk.ac.kr³

A Real-time Interaction of Molecular Modeling based on Distributed Processing System using a Web Service

Bosoon Kim¹, Sungjun Park², Jee-In Kim³

DreamWiz¹/

Department of Computer Science & Engineering, Konkuk University²/

Department of Internet & Multimedia, Konkuk University³

요약

분자 모델링 시뮬레이션은 신 물질, 신약 개발에 범용적으로 사용되는 중요한 컴퓨터 소프트웨어이다. 교육과 연구 등의 분야에서는 사용자가 직접 입력 도구를 사용하여 분자 모델을 시뮬레이션을 하는 경우가 존재한다. 사용자가 직접 시뮬레이션을 하기 위해서는 가상의 3 차원 시각화 환경은 물론 생화학적으로 안정성 여부 검증에 도움을 주기 위해 에너지 계산 결과를 제공해야 한다. 그러나 대다수의 분자 모델링 도구가 시각화 환경 제공을 위주로 개발되었으며 에너지 계산 수식이 복잡하여, 사용자가 시뮬레이션 하는 가운데, 실시간으로 에너지 계산을 제공하지 못한다. 이러한 단점을 극복하고자 본 논문에서는 어떠한 분자 모델링 도구라도 빠르게 에너지 계산을 반환 받을 수 있는 웹 서비스 기반의 분산 시스템 환경을 구현하였다. 또한 실시간으로 사용자가 시뮬레이션 할 수 있도록 작업 선별 처리 알고리즘(Job Skip Operation)을 개발, 적용하여 최신의 에너지 계산 요청에 대한 반환을 보장하였다. 본 연구는 사용자가 상호작용 기법을 통하여 가상의 분자 모델링 환경에서 화학적으로 안정된 분자 물질의 결합 위치를 빠르게 찾을 수 있도록 도와준다.

Keyword : HCI, VR, Bio Informatics, Web Service, Real-time Interaction

1 서론

분자 모델링은 분자의 3 차원 구조 해석을 기반으로 신 물질, 신약 개발에 활용되며, 이에 필요한 생화학적인 실험을 컴퓨터를 통하여 시뮬레이션 하도록 도와준다. 시뮬레이션에서 분자 물질의 결합은 3 차원으로 모델링 하여 그 모양이나, 특성, 안정성을 검사하여야 한다. 그러나 대부분의 분자 모델링은 3 차원의 분자 구조를 시각화하는 것에 중점이 맞추어져 있다.

실제로 신약 개발이나, 의학 실험에서 분자 모델링 시뮬레이션을 하기 위해서는 3 차원 분자 모델을 조작, 변형, 결합(Docking)해야 한다. 분자 모

델링에서 수용자(Receptor¹)의 특정 부위에 리간드(Ligand²)를 결합하는 작업은 실험실에서 많이 사용되며 이렇게 방법 등으로 결합된 분자 구조가 생화학적으로 의미가 있는지에 대하여 화학적 결합 알고리즘에 기반한 에너지 최소화 계산[1][2][3]을 사용하여 검증해야 한다.

그러나 이와 같이 분자 모델링을 통하여 사용자가 상호작용하는 동시에 에너지 계산 결과를 바로

¹ Receptor : 약물이 작용을 나타내려면 대부분의 경우 생체내의 어떤 세포의 특정부분 또는 성분과 결합하여야 하는데, 이때 약물 분자와 결합하는 특정 성분 또는 거대 분자를 말함.

² Ligand : 수용체에 결합하는 호르몬, 약제 등의 분자.

일기에는 에너지 계산 수식이 복잡하여 상대적으로 많은 시간이 소모된다. 또한 분자 물질의 효율적인 결합을 알기 위해 사용자는 지속적으로 분자 모델링과 상호작용하면서 분자 물질을 움직이게 되며, 이에 따라 실시간으로 에너지 계산 결과를 제공해줘야 하지만, 기존에 요청한 모든 에너지 계산 작업을 처리해야 하기 때문에 자연스럽게 에너지 계산 결과를 반환하는데 있어서 지연(delay) 문제가 발생한다.

본 논문에서는 사용자가 분자 모델을 시뮬레이션 하는 과정에서 에너지 계산 결과를 빠르게 제공하기 위해 웹 서비스(Web Service)를 이용한 분산처리시스템(Distributed Processing System)을 개발하였다. 또한 분산처리시스템에서 최신의 에너지 계산 요청을 지연 문제 없이 처리 할 수 있도록 하기 위하여 작업 선별 처리(Job Skip Operation) 알고리즘을 개발하여 적용하였다.

본 논문에서 제시하는 상호작용 환경이 어느 정도의 효율을 갖는지 검증하기 위하여 본 연구팀에서 개발한 가상현실 기반의 분자 모델링 시스템인 VRMMS[6] 분자 모델링 도구에 적용하였다. 그리하여 분산처리시스템의 처리 능력을 분자 구조의 크기와 처리 컴퓨터의 수를 달리 하면서 수행 속도를 비교하였고 작업 선별 처리 알고리즘의 성능을 분석하기 위하여 적용한 경우와 적용하지 않은 경우의 동일 시간의 위치에 따른 에너지 계산 반환 분석하였다.

2 관련 연구

RASMOL[4]은 단백질이나 핵산의 3 차원 구조에 대한 모델링 서비스를 제공해 주는 분자 모델링 도구이다. 범용적으로 사용되며, 빠른 3 차원의 분자 모델링 시각 환경을 제공한다. 마우스를 이용한 분자 모델의 회전이나 이동, 확대, 축소 등이 가능하다. 그러나 결합 등의 모의 실험 환경은 제공하지 않는다.

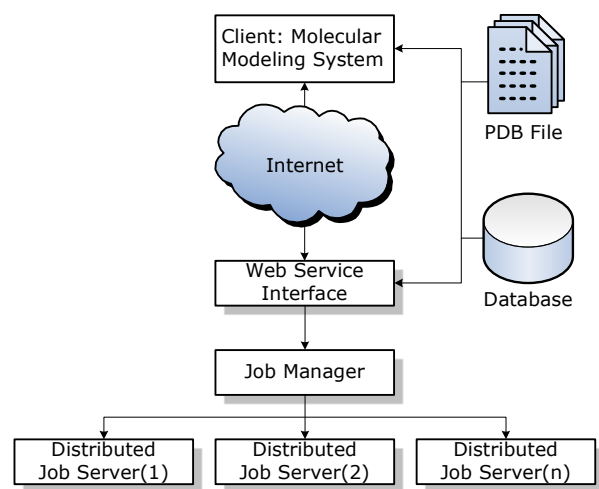
JMV(Java Molecular Viewer)[4]는 Java 와 Java 3D 로 구현된 분자 모델링 시각화 도구이다. 주로 시각화에 중점을 맞춰 분자 모델의 색상을 제어하거나 3 차원 모델링 출력에 대한 여러 가지 옵션을

지정할 수 있다. 스테레오스코픽(Stereoscopic)을 지원하여 가상의 3 차원 입체 영상을 제공한다. RCSB Protein Bank, BioCoRE 와 같은 분자 데이터뱅크 사이트와 연동하여 웹(Web)상에서 PDB(Protein Database)를 모델링 하여 출력한다. 그러나 결합 등의 시뮬레이션 환경은 제공하지 않는다.

WebLabViewer[5]는 Chime 로 유명한 MSI 사에서 만든 분자 모델링 도구로서 RASMOL 과 더불어 범용적으로 사용되는 도구이다. 무료로 사용할 수 있으며, RASMOL 보다 좋은 품질의 모델링을 제공한다. 그러나 RASMOL 과 기타 다른 분자 모델링 도구에 비하여 처리 속도가 전반적으로 느리며, 역시 시뮬레이션 환경은 제공하지 않는다.

3 웹 서비스 기반의 분산 시스템

3-1 시스템



[그림 1] 시스템 구성도

그림 1 은 웹 서비스의 분산 시스템을 이용한 분자 모델링 도구의 실시간 상호작용을 위한 시스템 구성도 있다. 시스템은 크게 클라이언트인 분자 모델링 시스템과 클라이언트가 요청한 에너지 계산을 처리하는 분산처리시스템으로 나눌 수 있다.

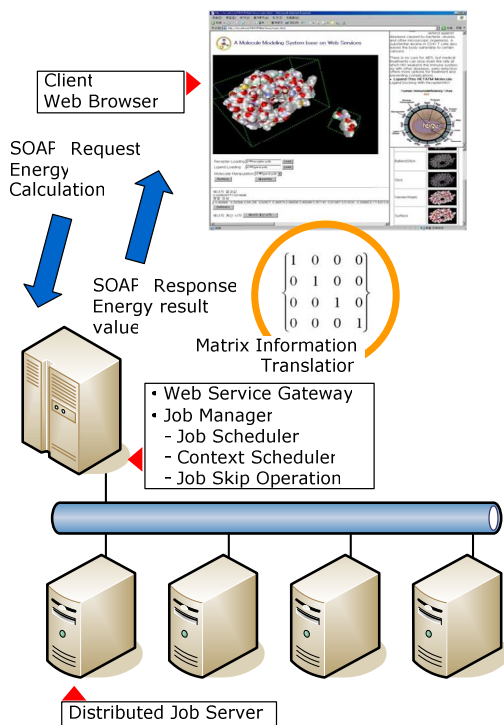
클라이언트는 사용자가 분자 구조 파일(PDB - Protein Data base) 또는 데이터 베이스(Database)로부터 분자 구조 정보를 얻어 3 차원의 분자 후보 물질에 대하여 시각화 하고 이를 조작, 변형, 결합

할 수 있는 환경을 제공한다.

분산처리시스템은 웹 서비스 인터페이스를 통하여 클라이언트가 요청하는 에너지 계산 정보를 받아 작업 관리자(Job Manager)를 통해 이를 스케줄링하여, 하위 다수의 작업 처리 서버(Distributed Job Server)에 요청하여 에너지 계산을 처리, 클라이언트에 반환한다.

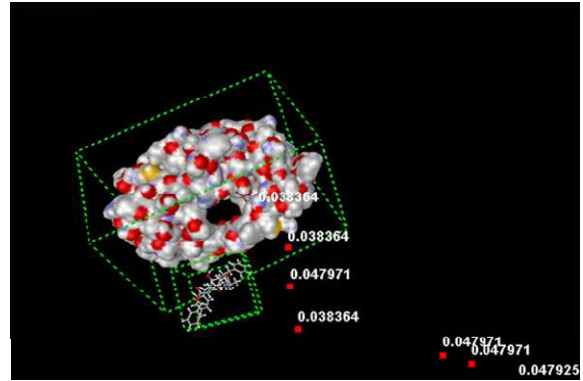
3-2 클라이언트 - 분자 모델링 시스템

그림 2 에서 보는 바와 같이 클라이언트는 3 차원 공간에 분자 모델을 시각화한다. 시각화한 분자 모델을 사용자가 조작, 변형, 결합 할 때 화학적 결합의 안정성을 판단하기 위해 인터넷 등의 네트워크를 통해 웹 서비스 인터페이스에 에너지 계산을 요청한다. 에너지 계산을 요청할 때 계산에 필요한 구성 요소로 분자 물질에 대한 이름과 위치 정보이다. 이러한 정보를 전달하면, 웹 서비스 인터페이스를 통해서 작업 관리자(Job Manager)에 에너지 계산을 요청한다. 요청한 작업은 내부에 분산된 작업 처리 서버(Distributed Job Server)를 통해 작업이 처리되어, 이러한 결과가 요청한 클라이언트에게 반환된다.



[그림 2] 클라이언트의 에너지 계산 요청과 응답 과정

그림 3 은 웹 서비스를 통한 분산 시스템에 본 연구에서 분자 모델링 도구로 적용한 VRMMS 의 분자모델링 도구를 연동한 모의 실험을 보여주고 있다. 사용자가 입력 도구로 분자 모델을 결합하는 과정에서 에너지 계산 결과를 제공한다.



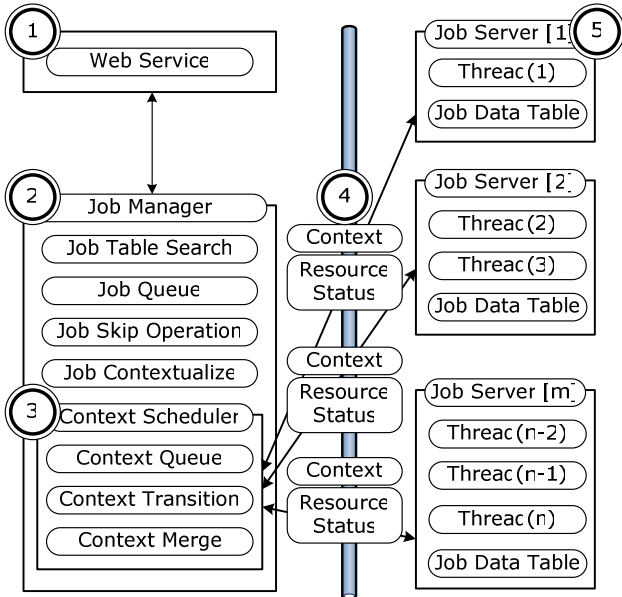
[그림 3] VRMMS 의 분자 모델 결합

3-3 분산 처리 시스템

분산처리시스템은 클라이언트(분자 모델링 시스템)의 에너지 계산 요청을 고속으로 수행하기 위한 분산 작업 처리 환경을 말한다[7].

분산처리시스템은 그림 4 과 같이 구성하였다. 웹 서비스(그림 4 의 ①)를 통하여 클라이언트는 에너지 계산을 요청하거나 계산된 에너지 결과를 반환 받는다. 요청한 에너지 계산은 작업 관리자(그림 4 의 ②)의 관리하에 분자 물질의 구성 정보를 탐색(Job Table Search)하게 되고 요청된 시간에 따라 작업 큐(Job Queue)에 쌓인다. 작업 관리자는 작업 선별 처리 알고리즘(Job Skip Operation)을 통하여 실시간 반환 보장을 하기 위한 작업(Job)을 선택하고 이를 분산 처리하기 위해 Context 화(Job Contextualize)한다. Context 에는 하나의 에너지 계산 작업을 완료하기 위해 분산 처리 서버의 스레드 하나당 처리되어야 할 분자 물질의 에너지 계산 구간과, 물질의 이름, 위치 정보가 담겨져 있다. Context 화가 완료되면 Context 스케줄러(그림 4 의 ③)에 의하여 Context Queue 에 들어간다. 작업 관리자가 원격지의 작업 처리 서버(그림 4 의 ⑤)에 휴먼(일을 마쳤거나, 작업이 할당되지 않은) 스레드가 존재하면 처리되지 않은 Context 를 보내어 이를 처리하고 반환 받는다. Context 의 전이 상태

를 주기적으로 검사(Context Transition)하여 작업에 대한 모든 Context 의 처리가 완료되면 이를 통합(Context Merge)하여 에너지 계산 결과를 반환한다. 작업 처리 서버의 상태(Resource Status)와 Context 의 처리 상태는 그림 4 의 ④와 같이 주기적으로 계속 작업 관리자에게 보고된다.



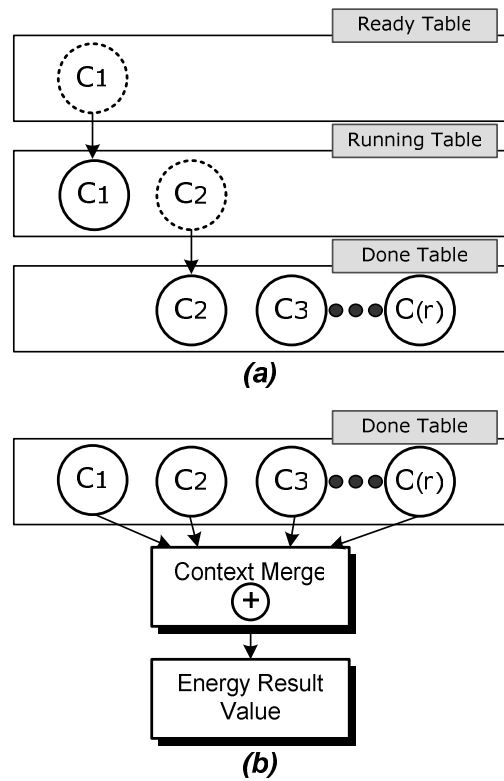
[그림 4] 분산 처리 시스템

3-4 Context 상태 전이

에너지 계산 요청 작업을 분산처리 하기 위하여 Context 화 하였을 때 이러한 Context 의 처리 완료 상태를 감지하기 위해서 Context 의 전이를 관리해야 한다.

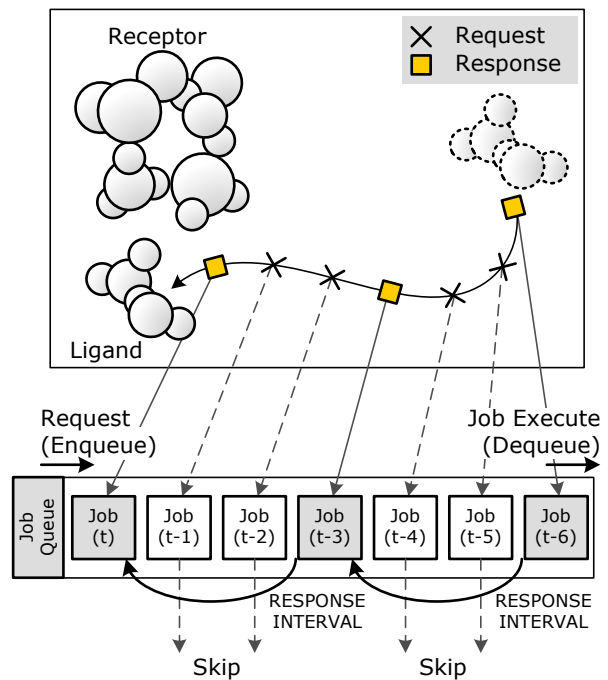
그림 5 의 (a)에서 보는 바와 같이 분산처리시스템은 Context 의 상태 전이를 관리 한다. 여러 개의 Context 는 초기에 준비 테이블(Ready Table)에 존재한다. 작업 관리자에 의하여 작업 처리 서버에 Context 가 할당되면 실행 테이블(Running Table)로 이동한다. 작업 처리 서버에서 처리가 완료되어 작업 관리자에 보고 되면, 해당 Context 는 완료 테이블(Done Table)로 이동한다.

그림 5 의 (b)에서 보는 바와 같이 하나의 작업에 대한 모든 Context 의 처리가 완료되어 완료 테이블에 존재하면 작업의 처리가 완료되었다고 판단하여 모든 Context 의 처리 결과를 합하여 에너지 계산 결과를 구한다.



[그림 5] (a) Context 상태 전이 (b) Context 의 에너지 계산 결과

3-5 작업 선별 처리



[그림 6] 작업 선별 처리(Job Skip Operation)

분산처리시스템은 에너지 계산 처리 요청을 빠르게 수행하는 장점을 제공하나 사용자가 증가하

거나 에너지 처리에 대한 소요 시간이 에너지 처리 요청이 쌓이는 시간에 비하여 긴 경우에는 보장된 시간 안에 에너지 계산 결과를 반환하지 못한다. 시간이 지날수록 사용자가 요청하는 최신의 에너지 계산 처리 요청에 응답하지 못하게 되면, 궁극적으로 사용자는 실시간으로 분자 물질의 결합에 대한 화학적 결합 안정성을 평가 하지 못하게 된다. 이러한 에너지 계산 처리에 대한 지연 문제를 해결하기 위하여 작업 선별 처리 알고리즘을 개발하여 분산처리시스템에 적용하였다.

그림 6 에서 보는 바와 같이 사용자가 분자 물질의 결합 하기 위하여 회전하거나 이동하는 사이에 작업이 일정한 시간 간격으로 분산처리시스템의 작업 큐(Job Queue)에 쌓인다. 처리 할 작업을 선별 하는 방법은 작업 큐에 쌓인 전체 작업에서 몇 개의 작업을 시간상 고르게 선택하는 것이다. 그리고 이러한 선택에서 제외된 작업은 무시(Skip)하여 큐에서 제거 한다. 작업을 처리할 때 마다 이런 방법을 사용하면, 작업 큐에 쌓인 작업이 아무리 많더라도, 최신의 에너지 계산 요청 작업을 처리하는데 지연 시간 문제가 발생하지 않는다.

4 실험

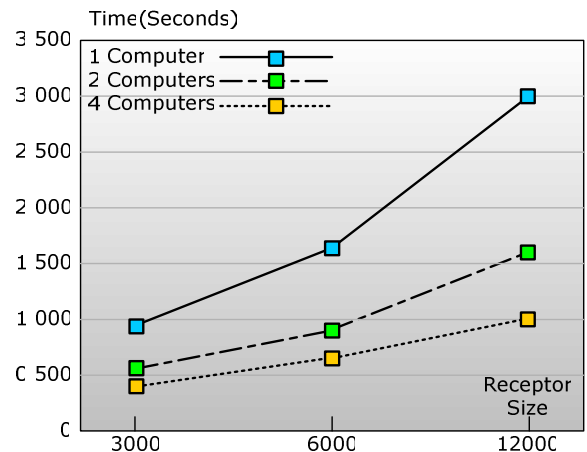
본 논문에서 개발한 웹 서비스의 분산 처리시스템을 이용한 분자 모델링 도구의 실시간 상호작용 환경의 성능을 평가하기 위하여 두 가지 측면으로 나누어 실험하였다. 첫째는 분산처리시스템 에너지 계산 처리의 성능평가이며, 둘째는 작업 선별 처리 알고리즘의 성능평가 있다.

실험에 사용된 수용자(Receptor) 물질은 HIV-1(Human Immunodeficiency virus)[8,9]을 사용하였으며 리간드(Ligand)는 HIV-1 의 재생 물질을 사용하여 결합 실험을 하였다.

4-1 분산처리시스템 성능 평가

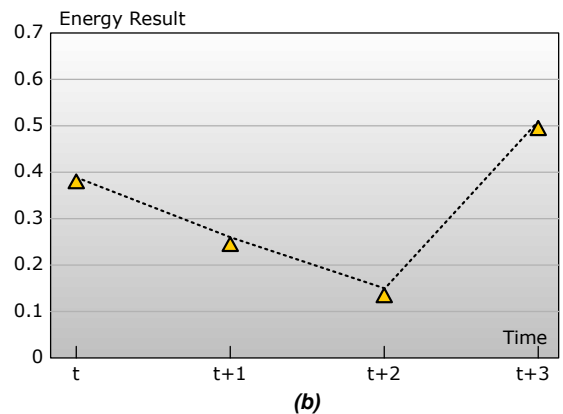
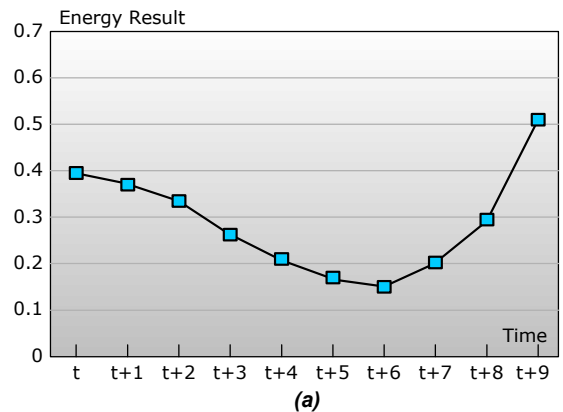
분산처리시스템의 성능 평가는 Receptor 의 분자 물질의 분자량을 3000 개, 6000 개 그리고 12000 개에 대하여 실험하였으며, 작업 처리 서버는 1 대, 2 대, 4 대 경우에 대하여 에너지 계산 결과 반환 시간을 측정하였다.

그림 7 에서 보는 바와 같이 분산처리시스템은 분자 물질의 분자량이 크면 클수록 처리 시간이 오래 걸렸다. 그러한 반면에 작업 처리 서버가 많으면 많을수록 에너지 계산 결과 반환 시간이 빨랐다. 이를 통하여 단일시스템 보다는 분산처리시스템이 동일한 분자량을 갖는 분자 물질에 대하여 더 빠르게 에너지 계산을 처리함을 알 수 있었다.



[그림 7] 분산처리시스템 성능 평가

4-2 작업 선별 처리 성능 평가



[그림 8] (a) 작업 선별 처리 알고리즘 미 적용 (b) 작업 선별 처리 알고리즘 적용

작업 선별 처리 알고리즘의 성능 평가는 수용자 (Receptor)와 리간드(Ligand)의 분자량을 일정하게 고정시킨 다음 리간드의 이동 경로 하나에 대하여 알고리즘을 적용한 경우와 적용하지 않은 경우의 에너지 계산 결과에 대한 동일 에너지 결과 패턴의 반환 시간을 비교하였다. 실험 결과 그림 8 과 같이 선별 처리 알고리즘을 적용하지 않은 경우 에너지 계산 결과가 모두 반환되어 조밀한 반면 적용한 경우에는 작업 선별 처리에 의하여 특정 작업은 무시(Skip)되었다. 이를 통해 작업 선별 처리의 적용이 사용자가 상호작용 하는 가운데 최신의 에너지 계산 요청을 빠르게 제공함을 알 수 있었다. 그러나 단위 시간당 에너지 계산 결과의 조밀도는 비록 최신의 에너지 계산 결과 반환은 느리나 적용하지 않은 경우가 더 높았다.

5 결론

웹 서비스를 통한 분산 시스템 기반의 분자 모델링 도구의 실시간 상호작용을 위하여 구현된 환경은 사용자가 분자 모델링 도구와 상호작용하는 가운데, 실시간으로 에너지 계산 결과를 제공하여 모의 실험을 할 수 있도록 해준다. 표준 웹 서비스를 통하여 어떤 분자 모델링 도구라도 분자 물질을 결합하는 모의 실험을 하는 가운데, 에너지 계산을 요청하여 반환 받을 수 있다. 작업 선별 처리 알고리즘의 적용에서 나타난 에너지 계산 요청 작업의 무시로 인한 낮은 조밀도는 사용자가 에너지 계산 결과를 보고 안정된 에너지 결과를 보인 위치에서 다시 실험을 진행하는 상호작용 방법을 통해 극복할 수 있다.

본 연구의 분자 모델링 도구의 실시간 상호작용 기법은 직접 사용자가 분자 모델링을 가상으로 실험하여 에너지 변화를 관찰하거나, 교육 목적 등의 모의 실험이 필요한 경우, 그리고 다양한 입력 도구를 이용한 분자 모델링의 조작에 큰 도움을 줄 것으로 예상된다.

참고 문헌

1. Ekachai Jenwitheesuk and Ram Samudrala, Improved prediction of HIV-1 protease-inhibitor binding energies by

molecular dynamics simulations, BMC Structured Biology, 3:2, (2003)

2. James C.Philips, Gengbin Zheng, Sameer Kumar and Laxmikant V.Kale, NAMD:Biomolecular Simulation on Thousands of Processors, IEEE, pp36, (2002)

3. B.R.Brooks, R.E.Brucoleri, B.D.Olafson, D. Vid J. States, S. Swaminathan, and M.Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J.Comp. Chem., 4:187-217, (1983)

4. <http://www.ks.uiuc.edu/Research/jmv>

5. <http://www.accelrys.com>

6. Jee-In Kim, SungJun Park, Jun Lee, Youngjin Choi, Seunho Jung, "Development of a Gesture-Based Molecular Visualization Tool based on Virtual Reality for Molecular Docking", Bulletin of the Korean Chemical Society, Vol 25, No 10, pp1571-1574, 2004(SCI)

7. SungJun Park, Bosoon Kim, Jee-In Kim, "A Web Service-based Molecular Modeling System using a Distributed Processing System", HS@I 2005, LNCS Vol. 3597, pp208-217, 2005(SCIE)

8. Earl Rutenber, Eric B.Fauman, Robert J.Keenan, Susan Fong, Paul S.Furth, Paul R.Ortiz de Montellano, Elaine Meng, Irwin D.Kuntz, Dianne L.DeCamp, Rafael Salto, Jason R.Rose, Charles S.Craik, and Robert M.Stroud, Structure of a Non-peptide Inhibitor Complexed with HIV-1 Protease, The Journal of Biological Chemistry, Vol. 268, No. 21, pp.15343-15346, 1993.

9. Junmei Wang, Paul Morin, Wei Wang, and Peter A. Kollman, Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA, Journal of American Chemical Society, Vol. 123, pp. 5221-5320, 2001.