

## 그리드 기반의 가상 스크리닝 클라이언트(DrugScreenr-G)

### 설계 및 구현

이상도, 이세훈, 황순욱, 김진철  
{sdlee, sehooi, hwang}@kisti.re.kr  
한국과학기술정보연구원

## DrugScreenr-G: Towards Grid-enabled Large-Scale Virtual Screening Coming into Handy

Sangdo Lee, Sehoon Lee, Soonwook Hwang, Jinchel Kim  
KISTI

### 요약

그리드 인프라를 활용하여 대규모 계산 자원이 필요한 가상 스크리닝(Virtual Screening)에 관한 연구가 국제 WISDOM 프로젝트를 중심으로 활발하게 진행 중에 있다. 기존의 연구들은 싱글 클러스터나 개인의 컴퓨터를 중심으로 계산 자원을 활용하는 방식으로 연구가 진행이 되었다. WISDOM을 중심으로 대규모 말라리아(Malaria), 조류 독감(Avian Flu)에 대한 대규모 신약 후보 탐색에 관한 연구들이 진행이 되었으며 한국과학기술정보연구원(KISTI)의 그리드 개발팀은 2007년도부터 국제 공동 파트너로 연구에 참여하고 있다. 공동연구에 참여하면서 신약 연구 개발자들이 Docking 및 그리드 자원을 활용하기 위한 접근 방식에 있어서 기존 방식들을 활용하기 어려운 점과 검색된 신약 물질을 분석하는 도구의 활용이 어려움이 많아 그리드 기반의 가상 스크리닝 도구인 DrugScreenr-G(이하, DSG-G 클라이언트)를 개발하게 되었다. 본 논문은 개발된 DSG-G 클라이언트에 관한 개발 및 구조에 관한 내용이다.

### 1. 서론

2007년부터 국제 가상 스크리닝을 위한 WISDOM[1] 프로젝트에 한국과학기술정보연구원(KISTI)에서는 그리드 기반 가상 스크리닝 워크벤치 개발을 통해 KISTI 그리드 환경을 손쉽게 활용할 수 있는 e-Science 서비스 환경을 구축하고 그 기반위에서 DSG-G 가상 스크리닝 워크벤치 클라이언트의 개발을 진행하고 있다. DSG-G 클라이언트 개발을 위한 가상 신약 물질 탐색 프로젝트 관리 프레임워크를 디자인하고 구현함으로써 관련 연구 기관 및 연구자들이 신약 개발에 도움이 되고 서비스 분야의 대중화를 촉진시키는 기반을 구축하는데 있다.

본 논문의 2장에서 관련연구 및 EGEE glite에 관한 개략적인 내용에 대한 것이다. 3장에서는 DSG-G 구조 및 설계에 대해서 설명 및 기능에 관한 것이다. 결론에서는 적용 사례 및 향후 계획에 대한 것으로 논문으로 구성되어 있다.

### 2. 관련연구

#### 2.1 WISDOM 프로젝트

유럽 중심의 EGEE glite-그리드 인프라를 활용한 신약 개발 분야에서 가상 스크리닝에 그리드 컴퓨팅 자원을 적용한 최초의 프로젝트로 프랑스 PCSV(Corpuscular Physics for Life Sciences) 팀에서 국제 공동 협업을 통한 대규모 in-Silico 도킹 실험을 진행하고 있으며 매년 두 차례 정도의 실험을 진행하고 있다.

#### 2.2 가상 스크리닝

가상스크리닝(Virtual Screening)은 신약설계 연구에서 주어진 단백질에 대한 최선의 신약후보물질을 탐색하는 계산을 통한 방식이다. 스크리닝은 인비트로(in-vitro) 실험으로도 가능하지만 인비트로(in-vitro) 실험으로 수백만 개의 후보물질을 모두 시험하는 것은 비용 및 시간상으로 해결할 수 없기 때문에 인비트로 실험의 대상을 필터링 하기위한 용도로 활용되고 있다. 가상스크리닝은 주어진 단백질에 후보 물질이 적합하게 붙을 수 있는지를 시험(도킹, docking)하기 위해 AutoDock과 MD(molecular Docking)

엔진을 활용한다. 세계적으로 수백만 개의 후보 화학 물질들이 존재하기 때문에 가상스크리닝 역시 수많은 컴퓨팅 자원을 필요로 한다. 현재 EGEE와 같은 그리드 컴퓨팅 인프라는 8만개 이상의 컴퓨팅 CPU에 대한 접근을 제공함으로써 가상스크리닝과 같은 대용량 계산을 가능하도록 WISDOM 과 같은 프로젝트 들이 관련 연구를 진행하고 있다.

### 2.3 가상 스크닝 과정

가상스크리닝을 위한 단계별 진행단계는 WISDOM에서 진행했던 조류 독감 (H5N1)의 치료제를 예를 들어 설명하면, 주요 타깃 효소인 뉴라미니데이스에 활성 저해화합물을 스크리닝 하기위해 Autodock[FlexX,Gold]과 같은 가상 스크리닝 방식을 이용하여 단백질 활성부위에 결합하여 표적 단백질을 저해하는 화합물을 대규모 계산을 통해서 신속하게 후보 물질을 선택하는 것이다.

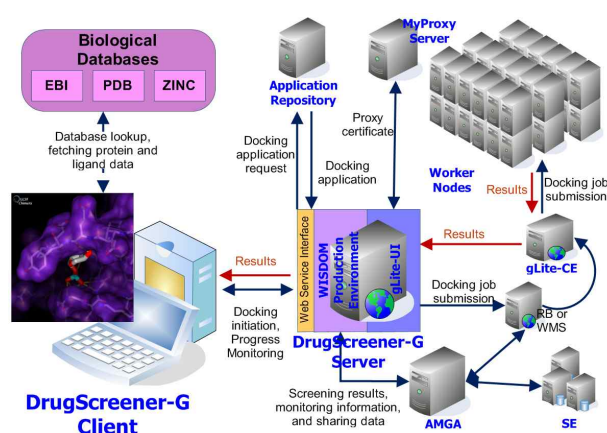
첫 번째, 도킹을 이용하여 단백질의 활성부위와 기질의 리간드 결합에너지를 계산하여 가장 안정한 형태의(에너지가 낮은) 리간드를 후보 물질로 추출한다. 두 번째, 후보 물질로 추출된 표적 단백질의 간단한 발현 시스템에 의한 최적 생산과 활성을 측정한다. 세 번째, 선발된 화합물 또는 선발 화합물에 표적 단백질을 혼합한 후 활성저해 정도를 측정하여 기존의 알려진 저해제와의 비교/평가하여 새로운 저해제(inhibitor)를 찾는 단계로 가상 스크리닝이 진행된다[2].

## 3. DSG-G 클라이언트 구성 및 설계

### 3.1 구성 및 동작 과정

DrugScreener-G는 크게 사용자 환경을 제공하는 클라이언트와 그리드를 이용한 가상 스크리닝을 수행하고 결과를 수집, 조직하고 사용자에게 제공하는 서버로 구성된다. 본 논문은 클라이언트에 관한 내용이다. 사용자는 클라이언트에서 수행할 가상 스크리닝 프로젝트에 필요한 생물 정보 데이터를 얻기 위해 클라이언트에 내장된 웹 브라우저를 이용하여 가상 스크리닝에 특화되어 조직된 형식으로 생물 정보 데이터베이스를 열람할 수 있고, 필요한 생물 정보 데이터를 획득할 수 있다. 클라이언트는 기본적으로 가상 스크리닝에 사용할 스크리닝 프로그램에 맞는 프로젝트 편집기와 도구들을 플러그인 형태로 제공하게 된다. 사용자가 사용할 스크리닝 프로그램을 위한 변수 설정, 생물 정보

데이터의 가공 및 전처리등의 과정이 프로젝트 편집기 및 데이터 전처리 도구들을 이용해 이루어진다.



〈그림1〉 DrugScreener-G와 그리드 인프라와의 관계 및 작업 수행 도식

사용자가 프로젝트 편집기 및 데이터 전처리 도구들을 이용해 가상 스크리닝 프로젝트의 준비를 마치면, 웹 서비스를 통해서 준비된 프로젝트의 정보와 데이터를 서버측에 전송하고 그리드 상의 메타데이터 카탈로그 서비스와 저장장치 서비스에 프로젝트 정보 및 데이터를 배치(deploy)하게 된다. 배치된 프로젝트에 저장된 정보를 바탕으로, DrugScreener-G 서버내의 태스크 관리자(task manager)에서 가상 스크리닝을 위한 작업 리스트 및 작업 기술 파일(job description file), 작업 제출을 위한 정보들을 생성하여 메타데이터 카탈로그 서비스에 저장하게 되고, 생성된 작업 정보를 이용하여 작업 관리자(job manager)를 통해 그리드에 작업을 제출하여 수행하게 한다. 현재 작업 관리자는 안정적인 작업 수행을 위해 작업 실행 에이전트(agent)가 그리드상의 CE(Computing Element)에 작업 노드(Worker Node)에 비례하는 수만큼 제출되어 각 작업 노드에서 실행되면서 메타데이터 카탈로그 서비스에 저장되어 있는 태스크 정보를 참조하여 작업을 끌어와(pull) 실행하도록 되어 있다. 그리드에 제출되어 수행되는 작업의 상태를 작업 실행 에이전트를 통해 작업 관리자(job manager)가 중간중간 점검하여 수행이 끝난 작업은 그 결과를 메타데이터 카탈로그 서비스에 저장하고, 중간에 실행에 실패하거나 오류가 생긴 작업들은 그 오류의 정도를 판단하여 다시 실행시키게 된다.

작업을 실행시키기 위해서는 그리드를 사용할 수 있도록 사용자의 인증서를 통해서 인증을 받아야 한다. 사용자의

인증서는 DrugScreener-G 클라이언트의 그리드 인증서 관리자(certificate manager)를 통해서 관리된다. 인증서 관리자를 통해서 사용자는 이미 발급받은 인증서를 DrugScreener-G 클라이언트에 배치시켜놓고 필요할 때마다 클라이언트가 자동으로 그리드에 인증을 받게 할 수 있다. 클라이언트에 배치된 인증서는 DrugScreener-G 서버를 통해 사용하는 그리드 사이트의 MyProxy 서버의 위치를 넘겨받아 MyProxy 서버에서 그리드 인증을 받고, 프락시 인증서를 생성, 저장하여 사용자가 생성하여 제출하는 작업들이 그리드의 CE에서 실행될 수 있도록 한다.

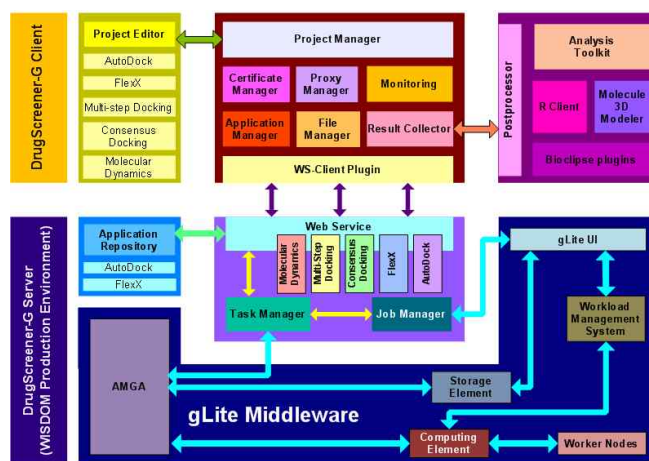
그리드에서 가상 스크리닝을 위한 작업 실행이 모두 완료되면, 모든 가상 스크리닝 응용프로그램 수행 결과는 메타데이터 카탈로그 서비스에 저장되게 된다. 사용자는 가상 스크리닝 작업 실행 진행상황을 모니터링하고, 작업이 모두 실행되면 다시 서버에 필요한 만큼의 결과 보고를 요청한다. 사용자로부터 결과 보고 요청을 받은 서버는 사용자의 결과 보고 요청 조건에 맞게끔 결과를 가공하여 사용자에게 반환하고, 반환된 가상 스크리닝 결과는 클라이언트에서 사용자의 요청에 적합한 형태로 보고되게 된다. 사용자는 반환된 결과를 열람하면서 추가의 분석이 필요한 경우, 클라이언트에서 제공되는 R 언어 기반의 분석 환경에서 추가의 분석을 수행하여 신약 물질 후보들을 가려내게 된다.

### 3.2 기능 및 설계 구조

DrugScreener-G 클라이언트의 기능은 크게 세 가지로 나눌 수 있다. 사용자와 직접 상호 작용하면서 가상 스크리닝 프로젝트를 편집하고 실행을 지시하는 프로젝트 에디터(project editor), 서버와 웹 서비스 통신을 하고, 프로젝트 관리 및 모니터링, 사용자 인증서 및 그리드 인증 관리, 응용프로그램 관리, 결과 수집 등의 핵심 기능을 하는 클라이언트 코어(Core), 그리드에서 수행된 가상 스크리닝에 대한 결과를 사용자가 열람할 수 있게 하고, 생물 정보 데이터의 가시화 및 모델링, 가상 스크리닝 결과의 후처리등의 기능을 제공하는 분석 및 모델링 도구 모듈이다. 위의 클라이언트 기능은 모두 플러그인 단위로 구성되며, Eclipse의 플러그인 프레임워크를 활용하여 모든 기능들이 Eclipse 플러그인으로 제작된다.

#### o 프로젝트 에디터

프로젝트 에디터는 가상 스크리닝 응용프로그램별로 각



〈그림 2〉 DrugScreener-G의 아키텍처

가상 스크리닝 응용프로그램에 맞는 사용자 인터페이스로 제작되며, 각 응용프로그램에 해당하는 프로젝트 에디터가 하나의 플러그인이 된다. 프로젝트 에디터에서 사용하는 공통의 자원들, 예를 들면 생물 정보 데이터나 프로젝트 속성의 등의 자원들은 DrugScreener-G 클라이언트의 핵심 플러그인과 Bioclipse 플러그인에서 제공하는 프레임워크를 활용하여 공통으로 정의되고, 각 응용프로그램 형태별 프로젝트 에디터에서 공통으로 활용될 수 있도록 한다.

#### o DrugScreener-G 클라이언트 코어(Core)

DrugScreener-G 클라이언트 코어(Core)는 DrugScreener-G의 핵심 기능을 제공하는 플러그인들로 구성되어 있다. 서버와의 웹 서비스 통신을 담당하며, 클라이언트 측에서 사용자가 설정한 프로젝트 속성 및 가상 스크리닝 응용프로그램 파라미터 정보들을 서버 측에 전달하고, 서버 측으로부터 클라이언트에 제공하는 정보를 전달하는 SOAP 메시지를 받아 해석하여 사용자에게 결과 정보를 전달하는 역할을 웹 서비스 클라이언트 플러그인이 담당한다. 사용자의 그리드 인증서 및 프락시 인증 관리를 위한 인증서 관리자 플러그인(certificate manager plugin)과 프락시 관리자 플러그인(proxy manager plugin), 프로젝트 진행 상황 모니터링을 위한 모니터링 플러그인(monitring plugin), 각 가상 스크리닝 응용프로그램에 특화된 정보를 관리하는 응용프로그램 관리자 플러그인(application manager plugin), 그리드에서의 파일 관리를 위한 파일 관리자 플러그인(file manager plugin), 서버에서 수집한 가상 스크리닝 결과를 서버로부터 받아 클라이언트에서 접근 및 가공가능한 형태로 보관하는 결과 수집기 플러그인(result

collector plugin)등이 DrugScreener-G 클라이언트 코어의 구성요소이다.

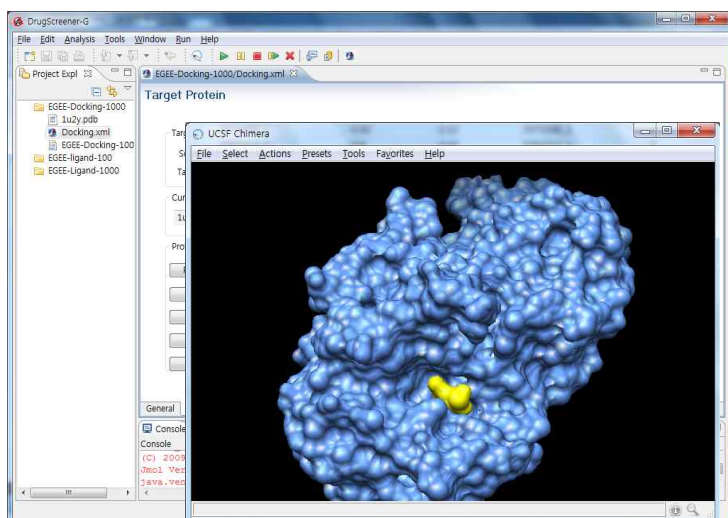
#### o 분석 및 모델링 도구

프로젝트 편집기 및 클라이언트 코어의 주요 구성요소 외에, 사용자의 가상 스크리닝 프로젝트 편집을 돕고, 결과의 후속 분석을 위한 분석 및 모델링 도구 모듈이 있다. 클라이언트 코어에서 서버를 통해 획득한 가상 스크리닝 결과는 결과 수집기를 통해 클라이언트에서 사용자가 열람하기 쉬운 형태로 변환된다. 이렇게 변환된 가상 스크리닝 결과 객체는 R 언어 등으로 정의된 후처리기(Postprocessor)의 후처리 모듈을 이용해 후속 분석 및 처리가 이루어진다.

후속 분석을 기술하기 위한 언어로 생물 정보학 및 생물 통계학등에서 널리 쓰이는 간결하고 강력한 함수 중심 언어(functional language)인 R 언어를 채택하였으며, R 언어 인터프리터 및 처리기가 프로토타입 후속 버전에서 통합될 예정이다.

플러그인 기반의 기능 확장 및 가상 스크리닝 방법 확장

DrugScreener-G는 현재 가상 스크리닝 기술이 빠르게 발전하고 다양한 가상 스크리닝 응용프로그램들이 개발되어 활용되고 있는 현 상황을 반영하여 새로운 가상 스크리닝 기술 및 방법이 쉽게 통합되어 활용될 수 있도록 플러그



인 기반의 확장 가능한 아키텍처를 기초로 설계되었다. DrugScreener-G 클라이언트는 Eclipse의 플러그인 프레임워크를 활용하여 플러그인 기능을 제공한다. DrugScreener-G 서버는 Eclipse에서 쓰이는 플러그인 프레임워크인 OSGi 플랫폼이 앞으로 통합되어 플러그인 확장 기능을 제공할 예정이다.

<그림3> DrugScreener-G 클라이언트 화면

#### 4. 결론 및 향후 과제

제2형 당뇨병 치료제 개발을 위한 대형 신약 후보물질탐색 데이터 챌린지 수행은 인간 체장에서 만들어지는 다당류 분해 효소인 알파 아밀라제(PDB: 1u2y)에 대해 31만개의 화합물을 대상으로 신약 후보 탐색 계산을 수행하여 전남대 김도만 교수 연구 그룹과 함께 테이블과 같이 EGEE 의 그리드 CE(Computing Element) (약 5000 CPU)를 활용하여 3.2 일에 걸쳐서 대규모 도킹 실험에 성공하여 테이블 <표1> 과 같이 기존의 알려진 Acarbose 와 유사한 수준의 후보 물질보다 도킹 에너지가 낮은(Lowest energy) 17, 18 번의 화합물을 찾아냈다. 현재 결과를 바탕으로 계

Compound No.	Lowest energy	M.W (g/mol)	clogP	Ki (μM)	IC50 (μM)	Type of inhibition
17	-16.43	473	3.04	19.8 ±1.2	58±4	competitive
18	-16.44	429	3.56	19.6±0.9	55±3	competitive
Acarbose	-12.62	645.605	-6.655	19.4	52±4	competitive

<표1>Autodock 을 이용한 실험 결과

속 연구를 진행하고 있다. 현재까지는 DSG-G 클라이언트가 베타 버전으로 개발되어 배포가 되지 않은 상태이고 몇몇 대학 연구실을 통해서 지속적으로 기능 및 도킹 계산 수행 결과에 대한 검증 작업을 진행하고 있다.

본 논문은 개발된 DSG-G 클라이언트를 통해서 신약개발에 도움이 될 수 있는 통합 환경에 대한 유저 클라이언트 소개 및 기능에 대해서 기술하였다. 베타버전으로 개발된 DSG-G 클라이언트를 이용하여 여러 대학교와 공동으로 검증작업을 진행 중에 있으며 향후 여러 개의 도킹을 방식을 동시에 지원 및 사용자의 보안 요구 사항을 반영하여 서비스를 제공할 수 있도록 진행할 계획이다.

#### 5. 참고 문헌

- [1]. WISDOM Project, <http://wisdom.eu-egge.fr>
- [2]. 김도만, "e-Science를 활용한 in silico 가상스크리닝 연구", 2007.1.1