

# 웹 기반 초고속 신약 후보물질 탐색-계산 시스템 구현 (Implementation of Web-based High-Throughput Screening Calculation System)

조 성 호 <sup>†</sup>      김 인 희 <sup>\*\*</sup>      공 지 원 <sup>\*\*\*</sup>      김 대 산 <sup>\*\*\*\*</sup>  
(Seong-ho Cho)      (Inhee Kim)      (Jiwon Kong)      (Daesan Kim)

권 남 훈 <sup>\*\*\*\*\*</sup>      로버트 E. 베리어 <sup>\*\*\*\*\*</sup>      김 성 훈 <sup>\*\*\*\*\*</sup>  
(Namhoon Kwon)      (Robert E. Burrier)      (Sunghoon Kim)

**요 약** 이 논문에서는 nanoluciferase 시스템의 생화학 반응 과정에서 도출한 데이터를 고속으로 계산하고 신약 후보물질을 신속히 찾아내도록 보조하는 웹 기반 초고속 계산-탐색 시스템(HTSCS) 개발 결과를 설명한다. 이 시스템에서는 단일 농도(Single-Dose)와 다중 농도(Multi-Dose)로 세포 실험체에 후보물질을 투여(*in-vitro*)한 대량의 실험 데이터를 계산하여, 차트 및 그래프 정보로 나타낸다. 따라서 HTSCS 시스템으로 신약 후보물질 발견에 필요한 시간과 노력을 줄일 수 있다.

**키워드:** 고속 신약 탐색, 발광 인자 실험, 신뢰 구간 반영 알고리즘, 비선형 회귀 분석, 데이터 정형화

**Abstract** In this paper, we describe the development of a web-based HTSCS (High Throughput Screening Calculation System), which is the process for assisting high-speed screening and calculation of drug candidates based on biochemical reactions from Nanoluciferase system. This system calculates and displays large amounts of experimental data with single-dose and multi-dose concentration *in-vitro* cell experiments through charts and graphs. Therefore, we can utilize the HTSCS to save time and effort required for drug screening.

**Keywords:** high-throughput screening, luciferase assay, trust region reflective algorithm, non-linear regression analysis, data structuration

- 
- 본 연구는 2017년 정부(과학기술정보통신부)의 재원으로 한국연구재단 글로벌 프린티어사업의 지원을 받아 수행한 연구임  
(한국연구재단-M3A6A4-2010-0029785, NRF-2015M3A6A4065724)
- 이 논문은 2018 한국컴퓨터종합학술대회에서 '웹 기반 초고속 신약 후보 물질 탐색-계산 시스템 구현'의 제목으로 발표된 논문을 확장한 것임
- <sup>†</sup> 정 회 원 : 서울대학교 의약바이오컨버전스연구단 정보관리팀 팀장  
(Seoul Nat'l Univ.)  
darkcircle.0426@biocon.snu.ac.kr  
(Corresponding author임)
- <sup>\*\*</sup> 비 회 원 : 서울대학교 의약바이오컨버전스연구단 Lead Validation Platform팀 연구원  
ihkim@biocon.snu.ac.kr
- <sup>\*\*\*</sup> 비 회 원 : 서울대학교 의약바이오컨버전스연구단 Drug Screening 팀 박사과정  
gong.jiwon@biocon.snu.ac.kr
- <sup>\*\*\*\*</sup> 비 회 원 : 서울대학교 의약바이오컨버전스연구단 Lead Validation Platform팀 팀장  
ichor@biocon.snu.ac.kr
- <sup>\*\*\*\*\*</sup> 비 회 원 : 서울대학교 의약바이오컨버전스연구단 실용화개발팀 교수/팀장  
nanaromi@biocon.snu.ac.kr
- \*\*\*\*\* 비 회 원 : 서울대학교 의약바이오컨버전스연구단 자문위원단 자문위원  
bburrier@excubio.com
- \*\*\*\*\* 비 회 원 : 서울대학교 의약바이오컨버전스연구단 교수/연구단장  
sungkim@biocon.snu.ac.kr
- 논문접수 : 2018년 8월 13일  
(Received 13 August 2018)  
논문수정 : 2018년 10월 30일  
(Revised 30 October 2018)  
심사완료 : 2018년 11월 20일  
(Accepted 20 November 2018)
- Copyright©2019 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.  
정보과학회논문지 제46권 제3호(2019. 3)

### 1. 서론

인류는 오랜 옛날부터 다양한 질병에 노출되어 왔고, 약물과 물리적 기술 방식을 활용하여 병마를 물리치는 방법을 끊임없이 연구하고 발전해왔다. 아직 인류는 암(癌)과 같은 완전히 정복하지 못한 몇가지 질병과 싸우고 있다. 현재 시판 중인 암 치료제보다 부작용이 적으며 효과가 뛰어난 다양한 암 치료제가 필요하며, 암 치료제 개발은 여전히 성공률이 낮고 오랜 개발 과정을 필요로 한다. 암 치료제의 경우 세포 독성을 가진 약물들이 많아 정상적인 세포에 대한 부작용이 있는 경우가 많다. 최근에는 질병에서 핵심 작용을 하는 단백질의 기능 조절을 통한 약물 개발이 활발히 진행중이다. 세포주 및 단백질 기반(in-vitro)의 약물 스크리닝 기법을 도입했고, 암 질병을 포함한 기타 다양한 질병에 대한 치료제용 화합물을 더욱 신속하게 찾아내기 위해 자동화 장비 및 정보처리 기술을 도입했다. 이 과정에서 나온 대량의 데이터에 대해, 생물 실험 과학자들은 스프레드시트 프로그램의 수식 매크로 기능을 주로 활용하여 계산하나, 개별 셀 반응 데이터를 직접 계산하는 과정은 많은 시간과 노력이 필요하다.

최근의 초고속 약물 탐색(HTS, High-Throughput Screening) 실험[1]에서는 목적에 맞는 스크리닝 기법(Assay) 구축, 화합물 정보 수집, 실험 명세 수립 과정을 거친 후, 일련의 실험 수행 과정의 일부 혹은 전체 자동화 기법을 활용한다. 실험 사례 데이터를 대량으로 축적하여 모델 기반 가상 실험을 진행하는 사례[2]도 있으나, 이 실험의 정확성 보장을 위해 실제 실험 데이터를 많이 축적해야 하는 조건의 한계가 있다.

유효물질(Hit), 선도물질(Lead), 후보물질(Candidate)로 이어지는 약물의 선별과정에는 몰 농도별 반응도, 특성, 독성 등에 대한 화합물 특성 규명이 필요하다. 초고속 약물 탐색 기법은 보통 96 well 플레이트 또는 384 well 플레이트에서 반응 실험을 진행하고 여기에 필요한 buffer, 세포분주 및 약물 처리를 위해 프로그래머블 리코드 컨트롤러를 활용한다. 플레이트의 well은 하나의 반응이 일어나는 시험관에 해당하며 이 논문에서는 well 개수가 96개인 플레이트를 다룬다. 플레이트 스캐너에서 각 well의 결과 상태를 정형 구조 수치 데이터를 획득하며, 규칙을 기반으로 데이터 계산을 자동화할 수 있다. 플레이트 실험 방식과 구역에 따라 계산 방식이 다르며, 플레이트에서 나온 데이터 96개를 사람이 직접 개별적으로 다루기에는 비효율적이므로 자동화 계산 방식이 필요하다.

이 논문에서는 후보 물질 스크리닝 실험을 수행한 후 데이터 계산 처리 과정, 후보 약물 탐색 시간을 단축하

고, 다수 연구자들의 체계적 연구 데이터 관리를 돕는 웹 기반 초고속 탐색-계산 시스템(HTSCS, HTS Calculation System)을 다룬다. HTSCS와 유사한 다양한 기존 구현체[3,4,5]는 다수의 사용자가 계산 시스템에 유연하게 동시 접근할 수 있고, 다수의 대상 화합물 군집 성향을 파악할 수 있다. 유사 구조 화합물 군집이 모여 있을 때 후보 물질 선별 대상을 더 빠르게 판단할 수 있다. 다만, 화합물 군집이 다수 실험 집단의 임의 순서 등록을 통해 섞였을 경우 특정 화합물에 대한 실험 명세, 후보 화합물의 기본 조건 만족 여부, 독성 경향을 바로 확인할 수 없는데 HTSCS에서는 필요한 정보를 필터링하여 보여주는 방법으로 문제를 해결했다. PHP, Python, PostgreSQL, RDKit[6]을 활용하여 만들어 운영 중인 Drug Library를 기반으로 scipy, numpy, matplotlib를 활용하여 만들었기 때문에, 기존의 화합물 정보를 보유한 상태에서 HTS 방식으로 실험하지 않은 실험 결과와 HTS 실험 결과를 서로 비교하여 HTS 실험의 일관성과 안정성, HTSCS 시스템의 처리 신속성을 확인할 수 있다.

### 2. HTSCS 처리 데이터 구조

이 논문에서는 96 well 플레이트에서 최대 80가지 화합물을 대상으로 실험하여 신약 후보 물질을 발굴하는 과정을 계산 처리, 차트 도출을 통해 가속하는 시스템을 다루며, 처리 대상 데이터는 각 화합물의 상태, 실험 조건, 실험 결과 수치 데이터이다. 단일 화합물 실험 조건/결과 데이터는 1차원으로 표현할 수 있으나, 이 데이터가 모이면 2차원 구조를 지닌다. 실험 데이터를 저장할 때 전통적인 테이블 구조를 활용할 경우 레코드 수가 기하급수적으로 늘어나는 문제가 있어, 특정 필드에 PostgreSQL 9.4부터 지원하는 JSON[7] 형식으로 데이터를 넣어[8] 효율성을 개선하였다.

#### 2.1 플레이트 구조

표 1은 plates 테이블 일부를 나타내며, 이 테이블에는 플레이트의 특성 및 well별 투여 화합물 정보가 들어간다. 플레이트를 실험에 직접 사용하지 않는 경우를 감안하여 실험 조건 및 결과 데이터는 별도의 테이블에 저장했다.

표 1 plates 테이블(일부)  
Table 1 plates table (partial)

Column	Type	Condition
plate_no	Integer	not null autoincrement
plate_id	varchar(10)	not null
plate_type	varchar(12)	not null
derived_plate_no	integer	not null default 0
plate_data	json	not null

표 2 화합물 물 농도 희석 전달 예제

Table 2 Example of chemical compound dilution and transfer

plate_no	plate_id	plate_type	derived_plate_no
1116	SP-00459	singledose	1115
1115	DP-00293	daughter	1112
1114	DP-00292	daughter	1113
1113	MP-00279	mother	0
1112	MP-00278	mother	0

plate\_no는 내부 일련번호이며, 실험 조건이 들어간 assay\_specs 테이블에서 참조한다. derived\_plate\_no는 플레이트 간 관계를 나타내는 참조 필드다. 표 2와 같이 derived\_plate\_no값으로 Mother plate → Daughter Plate → Single-Dose Plate로 이어지는 화합물 물 농도 희석 전달 과정을 표현할 수 있으며, 플레이트 단위 실험 이력 검색에 유용하다. derived\_plate\_no 값이 0인 경우는 Mother plate를 의미한다.

표 1의 plate\_data 컬럼에는 예 1의 정보를 json array 형식으로 저장한다. plate\_data 컬럼 데이터는 인덱싱 및 검색이 불필요하며 일괄 순차 조회하므로 plate\_data 컬럼 형식은 jsonb가 아닌 json이다.

예 1 plate\_data의 well 단위 정보(A01, Control1)

Example 1. Well information of plate\_data (A01, Control1)

```
{ "group":1, "idx":"A01", "value":0,
  "compound_id":"","compound_volume":-1,
  "concentration":-1,"type":"control"}
```

well 단위 정보의 type은 “control”과 “data”로 나뉜다. group은 type이 “control”일 경우 필요한 정보이며, type이 “data”일 경우 group 값은 0이다. idx는 well 위치 정보를 나타낸다. compound\_id는 화합물 ID, compound\_volume 값 단위는  $\mu$ l이며, concentration 값 단위는 g/mol이다. type이 “control”일 경우 compound\_volume과 concentration 값은 -1 (Not Available)이다.

## 2.2 실험 조건/실험 결과 데이터 구조

실험 조건 및 과정은 실험 프로토콜로 규정한다. 동일한 실험에 대해 화합물의 물 농도, 용량, 배치 위치 등의 실험 조건은 어떤 이유로든 달라질 수 있다. 실험의 무결성을 입증하려면 미 식약청 규정[9]의 최소 요건에 따라 이 데이터를 플레이트별로 표 3으로 나타낸 테이블에 저장해야 한다.

초기 설계 단계에서는 assay\_specs 테이블에 실험 정보만 넣었으나, 구현 편의상 실험 결과 데이터까지 들어가도록 assay\_result 컬럼을 추가하였다. assay\_spec\_no는 레코드 인덱스이며, 내부 인덱싱 목적 이외에는 사용

표 3 assay\_specs 테이블(일부)

Table 3 assay\_specs table (partial)

Column	Type	Condition
assay_spec_no	integer	not null autoincrement
related_plate_no	integer	not null
assay_spec	jsonb	not null
assay_result	jsonb	

하지 않는다. related\_plate\_no는 plates 테이블에 있는 실험 대상 플레이트 번호(plate\_no)가 들어간다. assay\_spec과 assay\_result 컬럼은 인덱싱 및 검색이 용이하도록 jsonb 형식으로 지정하였으며, 실제로는 단일 농도 실험과 다중 농도 실험시 각기 다른 json 형식의 데이터를 넣는다. 대체로, assay\_spec 컬럼에는 실험 프로토콜 정보와 계산 방식, 화합물 별 실험 횟수,  $\mu$ l단위 화합물 사용량, 실험시 임계 CV 값 정보, 실험 정상 수행 여부, 조건 설정용 실험의 여부 등의 정보가 들어감, assay\_result에는 Z'인자 값, CV 평가 값, 대상 실험 플레이트 정보, 계산일시, 계산 결과 데이터가 들어간다.

## 3. 약물 고속 스크리닝(HTS)을 위한 Assay 확립

암세포 성장에 중요하게 작용하는 두 단백질의 결합을 저해하는 화합물을 선별하기 위해 nano-luciferase system (Promega)을 스크리닝 시스템으로 도입했다[10]. 이 시스템에서는 두 단백질에 각각 nanoluciferase 조각을 붙여 CHO-K1 세포에 발현시킨다. 세포에서 발현된 두 단백질이 결합하면 각 단백질에 결합한 각각의 nanoluciferase 조각이 접근하여 원래의 nanoluciferase 구조를 재형성하여 luciferase 활성 상태를 나타낸다. 화합물이 두 단백질의 결합을 저해하면 그로 인해 nano-luciferase 조각도 분리되며 luciferase 활성도를 낮춘다.

본격적인 실험에서 두 단백질을 발현시킨 CHO-K1 세포를 96-well plate에 분주하고, 화합물을 4시간 처리한 뒤 nanoluciferase의 substrate를 처리, luciferase의 활성 값을 luminescence값으로 측정한다. 예비 실험을 통해 준비한 assay가 HTS에 적합함을 CV(Coefficient of Variation), Z'인자 값으로 확인했으며 FDA 승인 약물 테스트를 통해 본격적인 약물 탐색에 활용할 참고 물질을 확보했다.

## 4. HTSCS 계산 처리 과정

HTS 실험을 위해 화합물 정보를 미리 입력하는 조건으로 Mother Plate와 Daughter Plate에 배분한다. 화합물은 모두 DMSO(Dimethyl Sulfoxide) 용매에 용해하여 준비한다. Mother Plate 화합물 물 농도는 3~

10mM, Daughter Plate 화합물의 경우, 즉시 사용 시 3 $\mu$ M, 보관 화합물일 경우 수십~수백 $\mu$ M이다.

그림 1의 단일 농도 실험 과정에서는 두 단백질 결합의 상대억제를 유효 기준을 넘는 물질을 1차 유효 물질 (Primary Hit)로 선별한다. 우선 기준 물질과 DMSO 용매로 기준 물질을 용해한 대조군, 각각의 후보 화합물을 세포 실험 플레이트에 동일 몰 농도로 적량 투여한다. 일정 반응 시간 동안 대기 후, 플레이트 스캐너로 luminescence

값을 취한다. 이 결과 값을 HTSCS에 업로드한 후, CV 및 Z' 값을 계산하여 유효성을 확인한 뒤, 상대억제를 유효 기준을 넘는 1차 유효 물질을 찾아낸다. 유효 기준은 수식 모델 설명 부분에서 언급하였다.

그림 2의 다중 농도 실험 과정에서는 단일 농도 실험에서 선별(Cherry-picked)한 1차 유효 물질에 대해 몰 농도별 상대억제를 반응 값을 측정하고, 이를 sigmoidal-curve fitting을 이용하여 유효 물질에 대한 IC50(The Half

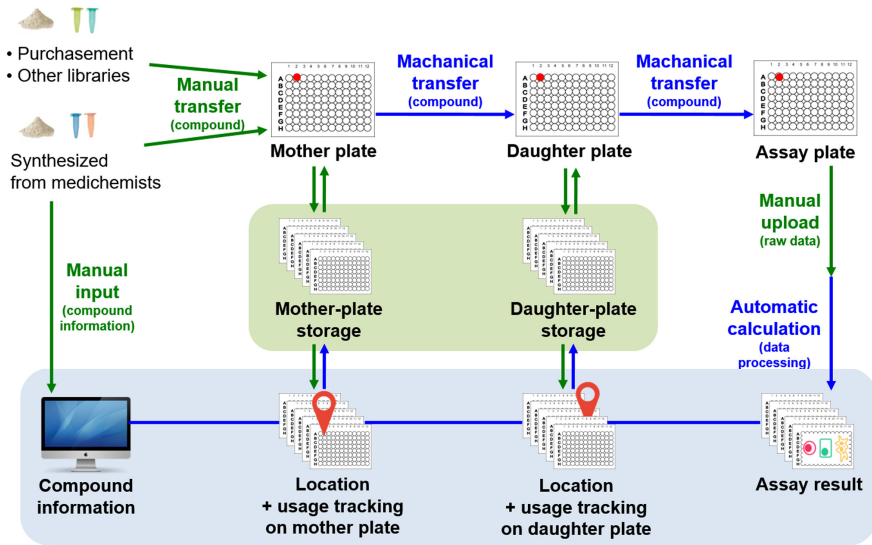


그림 1 단일 농도 실험  
Fig. 1 Single-Dose assay

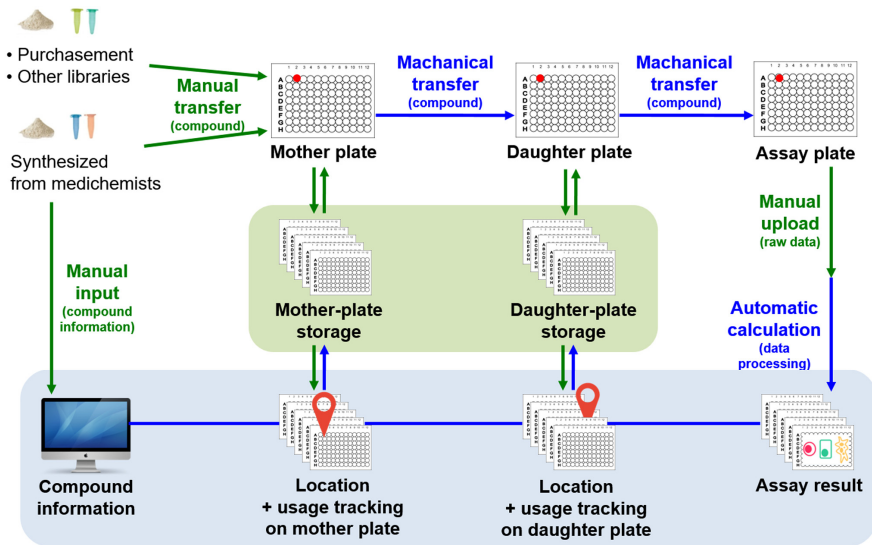


그림 2 다중 농도 실험  
Fig. 2 Multi-Dose assay

Maximal Inhibitory Concentration)값을 산출, 반응성 경향을 판단한다. 1차 유효 물질을 플레이트에 나란히 동일 물 농도 비율로 희석(Serially-diluted)하여 준비하고, 이를 DMSO용매로 기준 물질을 용해한 대조군과 함께 세포 실험 플레이트에 적량 투여하여, 반응 과정을 거친 후, 플레이트 스캐너로 luminescence 값을 취한다. 이 결과 값을 HTSCS에 업로드한 후, HTSCS를 통해 IC50 값을 계산하고, 반응 추세를 Z형 sigmoidal-curve로 나타낸다. 다중 농도 실험에서는 기준 물질을 1차 유효 물질과 함께 투여하여 실험하므로 CV값만 계산하며, 실험의 유효성을 나타내는 Z' 값은 취하지 않는다.

5. 계산 모델

5.1 플레이트 모델

HTSCS의 플레이트 공통 구성 형식은 그림 3과 같다. 이 논문의 시스템에서 처리할 실험에서 Control 1, Control 4에는 DMSO 화합물(negative control)을 넣고, Control 2, Control 3에는 비교 기준 화합물(positive control)을 넣는다. 다중 농도 실험에서는 타 화합물과의 비교를 위해 2~11번 컬럼 영역에 비교 기준 화합물을 직접 투여하는 경우, Control 2, Control 3을 사용하지 않는다. 플레이트의 조건 무결성 판단은 식 (1)[11]을 활용한다.

$$CV1 = 100 \times \frac{\sigma( Ctrl1, Ctrl4 )}{\mu( Ctrl1, Ctrl4 )}$$

$$CV2 = 100 \times \frac{\sigma( Ctrl2, Ctrl3 )}{\mu( Ctrl2, Ctrl3 )}$$

$$Z' = 1 - \left( 3 \times \frac{\sigma( Ctrl1, Ctrl4 ) + \sigma( Ctrl2, Ctrl3 )}{|\mu( Ctrl1, Ctrl4 ) - \mu( Ctrl2, Ctrl3 )|} \right)$$

식 (1) 플레이트 검증 수식  
Equation 1. Plate validation

$\sigma, \mu$ 는 괄호 안 구역의 화합물이 나타내는 luminescence 측정 값의 편차와 평균값을 나타낸다. 여기서 CV1과 CV2의 상한 값은 실험에 따라 별도로 정한다. Z' 값은 이상적인 경우 1이 나오며, 1을 초과할 수 없다. Z' 값이 0.5 이상인 경우(0.5 < Z' ≤ 1), 훌륭한 실험 구

성이라고 판단하며, 그 이하인 경우(0 < Z' ≤ 0.5), 실험을 진행할 수는 있지만 신뢰도가 낮은 실험으로 판단한다. Z' 값이 음수인 경우, HTS 방식으로 실험을 진행하기에 부적합한 실험으로 판단한다.

5.2 Single-Dose 수식 모델

단일 농도 실험에서 상대 반응억제율을 구하는 수식은 다음과 같다. 화합물을 well 하나에만 넣은 경우는 식 (2)와 같으며, 화합물을 가로/세로 방향으로 N번 반복 투여했을 경우 상대 반응억제율을 구하는 수식은 식 3 과 같다. HTS 시스템을 이용한 특정 화합물 반응에 대한 실험의 경우, 사람이 직접 화합물을 투여하여 실험하는 경우와 비교하여 개별 실험에 따른 오차가 크게 줄기 때문에, 동일 물농도 조건 단일 투여시와 다중 투여시의 상대 반응억제율이 큰 차이가 나지 않음을 확인할 수 있었다(Error rate in Single pipetting: 7.91%, HTS with auto multi-pipetting: 5.35%) [12,13].

$$\%RI = 100 \times \frac{w}{\mu( Ctrl1, Ctrl4 )}$$

식 (2) 단일 농도 단일 투여 화합물 상대 억제율  
Equation 2. Single-Dose compound percentage of relative inhibition

$$\%RI = 100 \times \frac{\sum_{n=1}^N w_n}{\mu( Ctrl1, Ctrl4 )}$$

식 (3) 단일 농도 다중 투여 화합물 상대 억제율  
Equation 3. Single-Dose compound multiplication percentage of relative inhibition

HTSCS에서 상대억제율 유효 기준(Primary Hit)의 기본 값은 70%로 정했다. 하지만 연구 분야별로 화합물의 세포 실험체에 대한 반응 성격이 다르므로 유효 기준을 다르게 설정하도록 했다. 이 모델을 이용하여 실험 데이터로부터 산출한 결과 예시는 그림 4와 같다.

5.3 Multi-Dose 수식 모델

Multi-Dose 계산 모델은 eccpy[14]와 GraphPad Prism

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B	Control 1											Control 2
C												
D												
E	Control 3											Control 4
F												
G												
H												

그림 3 HTSCS 플레이트 템플릿  
Fig. 3 HTSCS plate template

	1	2	3	4	5	6	7	8	9	10	11	12
A	Control 1 603700	BC-EXI-00002 667500 2.00	BC-EXI-00010 481300 29.34	BC-EXI-00018 196300 71.18	BC-EXI-00026 598500 12.13	BC-EXI-00034 685700 -0.67	BC-EXI-00042 666300 2.18	BC-EXI-00050 654600 3.89	BC-EXI-00058 677400 0.55	BC-EXI-00066 683000 -0.28	BC-EXI-00074 739800 -8.61	Control 2 188400
B	Control 1 621300	BC-EXI-00003 717600 -5.36	BC-EXI-00011 639700 6.08	BC-EXI-00019 305400 55.16	BC-EXI-00027 646200 5.13	BC-EXI-00035 637300 6.43	BC-EXI-00043 614300 9.81	BC-EXI-00051 616800 9.47	BC-EXI-00059 655500 3.76	BC-EXI-00067 712100 -4.55	BC-EXI-00075 733400 -7.67	Control 2 180900
C	Control 1 651700	BC-EXI-00004 545200 19.96	BC-EXI-00012 738500 -8.42	BC-EXI-00020 540800 20.60	BC-EXI-00028 466200 31.55	BC-EXI-00036 715700 -5.08	BC-EXI-00044 741900 -8.92	BC-EXI-00052 639600 6.10	BC-EXI-00060 816500 19.88	BC-EXI-00068 703500 -3.29	BC-EXI-00076 715300 -5.02	Control 2 182600
D	Control 1 639700	BC-EXI-00005 430400 36.81	BC-EXI-00013 524800 22.95	BC-EXI-00021 32360 95.25	BC-EXI-00029 767300 -12.65	BC-EXI-00037 664600 2.43	BC-EXI-00045 677100 0.59	BC-EXI-00053 702200 -3.09	BC-EXI-00061 670600 1.55	BC-EXI-00069 697600 -2.42	BC-EXI-00077 743200 -9.11	Control 2 186600
E	Control 3 174500	BC-EXI-00006 544400 20.07	BC-EXI-00014 7356 98.92	BC-EXI-00022 65.13 7.05	BC-EXI-00030 237500 6.17	BC-EXI-00038 632300 -4.16	BC-EXI-00046 652800 -13.53	BC-EXI-00054 773300 -12.70	BC-EXI-00062 715400 -5.03	BC-EXI-00070 665600 2.28	BC-EXI-00078 663100 2.66	Control 4 813200
F	Control 3 162800	BC-EXI-00007 466300 31.54	BC-EXI-00015 729200 -7.06	BC-EXI-00023 633100 7.05	BC-EXI-00031 674200 1.02	BC-EXI-00039 761800 -11.84	BC-EXI-00047 775000 -13.78	BC-EXI-00055 810500 -18.99	BC-EXI-00063 682600 -0.22	BC-EXI-00071 715700 -5.08	BC-EXI-00079 766800 -12.58	Control 4 700800
G	Control 3 144600	BC-EXI-00008 704500 -3.43	BC-EXI-00016 582500 14.47	BC-EXI-00024 594700 12.69	BC-EXI-00032 655000 3.84	BC-EXI-00040 661700 2.85	BC-EXI-00048 717800 -5.38	BC-EXI-00056 712800 -4.65	BC-EXI-00064 668300 2.32	BC-EXI-00072 690700 -1.41	BC-EXI-00080 729200 -6.91	Control 4 725700
H	Control 3 145900	BC-EXI-00009 619100 9.11	BC-EXI-00017 34150 94.99	BC-EXI-00025 -1.67	BC-EXI-00033 540900 20.59	BC-EXI-00041 629100 7.64	BC-EXI-00049 719000 -5.56	BC-EXI-00057 729000 -7.03	BC-EXI-00065 648200 4.83	BC-EXI-00073 644900 5.32	BC-EXI-00081 712800 -4.65	Control 4 692900

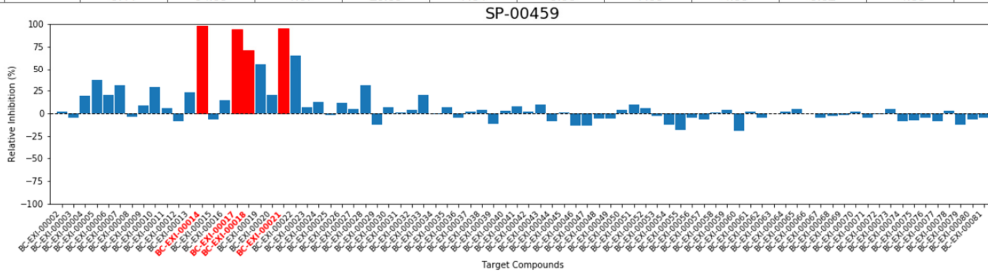


그림 4 단일 농도 실험 결과

Fig. 4 Single-Dose result

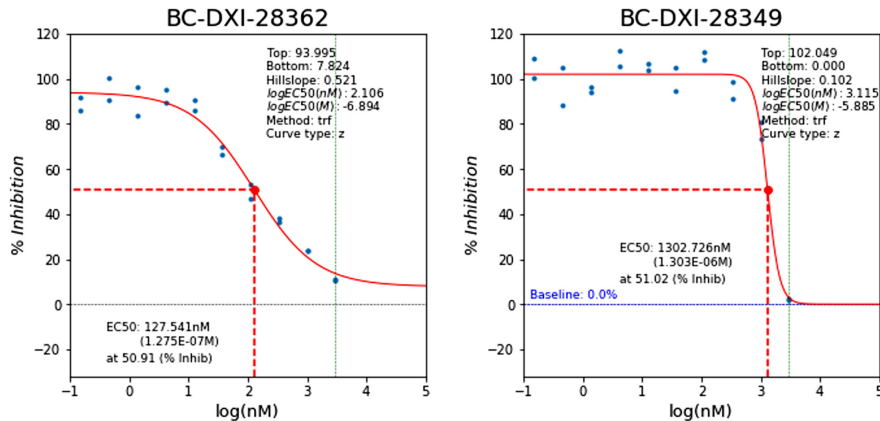


그림 5 다중 농도 실험 결과 분포 및 추세선 차트

Fig. 5 Multi-Dose result line-scatter chart

의  $\log(\text{inhibitor})$  모델[15]을 참고했다. ecppy에서는  $\log$  몰농도 값이 음수일 때 비정상 결과가 나타나므로,  $x$  값에 양수, 음수가 들어가는 각각의 경우에 대해 수식 모델을 분리했다. 실제 HTSCS에 반영한 수식은 다음과 같다. 그림 5에서의  $x$ 축 단위는 신약개발에서 주로 사용하는 범위에서의 편의성을 위해  $\log(nM)$ 을 사용했으며,  $x$ 축에서의 일부 데이터 값이 음의 영역으로 넘어가더라도 sigmoidal-curve를 올바르게 나타낼 수 있게 수식을 수정하였다.

$$y = lower + \frac{upper - lower}{1 + 10^{(x - IC50) / hillslope}}$$

식 (4) 다중 농도 반응 모델  
Equation 4. Multi-Dose reaction model

$x$ 는 해당 화합물의 몰 농도 값,  $y$ 는 상대 반응억제율(%),  $upper$ ,  $lower$ 는 각각 화합물의 최대, 최소 상대 반응 억제율(%)을 나타낸다. Trust Region Reflective 알고리즘[16]을 통해 위의 모델을 활용하여 최적의 피팅

곡선을 찾으며, 이 과정에서 *upper*, *lower*, *hillslope*, *IC50* 추산 값을 구한다. *hillslope*은 반응 억제율이 변화하는 정도를 나타내는 변수이며, *IC50*은 해당 화합물이 50%의 반응억제율을 나타내는 몰농도이다. *IC50* 반응 곡선 특성상 *lower* 추정치가 0 미만으로 나오는 경우, 실제 반응상의 특성을 고려하여 0으로 보정하여 재계산한다. 커브 피팅 모델에서 실험 몰 농도 값을 반영한 후, Brent's Method로 일 때의 조건에서 *IC50* 값을 도출한다. 과정상 나타나는 근소한 차이는 오차범위 내에서 유효하다.

## 6. 활용 및 개선 진행 현황

HTSCS는 실제 활용 중이며, 신규 개발 및 개선 역시 진행 중이다. 현재는 PHP언어와 Python언어를 혼용하여 개발한 초기 버전에서 PHP언어로 작성한 부분을 소프트웨어 공학 관점의 효율성을 추구하기 위해 Python언어로 전환하고 있다. 또한 전세계 신약 개발 과학자들이 공동으로 이용할 수 있도록 기능 개선 및 추가를 지속적으로 진행하고 있다.

2017년 8월부터 2017년 12월 31일까지 HTS 팀에서는 총 27676개 화합물에 대한 탐색 실험을 진행했고, 유효 물질 288개를 도출했다. 2018년 1월 이후로는 도출한 유효 물질로 추가 실험을 실시하여, 이 중 일부를 신약 후보 물질 대상으로 압축하였다. 의약 화학자들은 효능이 월등한 신약 후보 물질을 선정하기 위해 기존 물질에 구조적 변형 등 여러가지 화학적 기법을 적용하여 새로운 물질을 만들고 있으며, HTS 팀에서는 이 과정을 통해 만든 새로운 물질의 반응 우수성, 세포 독성, 의약 적합성을 검증하는 과정을 계속 진행하고 있다. 동시에 HTS 팀에서는 또 다른 질병에 대한 신약 후보 물질을 찾기 위해 신규 화합물을 반입하여 2017년도에 수행한 화합물 탐색 실험 과정을 반복하고 있다. HTSCS로 도출한 일련의 모든 HTS 실험 데이터는 후보 물질 라이브러리에 지속적, 체계적으로 저장, 관리하고 있다.

## References

[1] BROACH, James R., et al., High-throughput screening for drug discovery, *Nature*, 384.6604: 14-16, 1996.

[2] WANG, Lirong, et al. TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *The AAPS journal*, 15.2: 395-406, 2013.

[3] HUR, Junguk, et al., Tox21 Enricher: Web-based chemical and functional enrichment analysis tool for Tox21 toxicity screening platform. *The FASEB Journal*,

31.1\_supplement: 613.1-613.1, 2017.

[4] LIU, Rong, et al., HDAT: web-based high-throughput screening data analysis tools, *Computational Science & Discovery*, 6.1: 014006, 2013.

[5] PELZ, Oliver; GILSDORF, Moritz; BOUTROS, Michael, web cellHTS2: a web-application for the analysis of high-throughput screening data. *BMC bioinformatics*, 11.1: 185, 2010.

[6] RDKit, Open-Source Cheminformatics. <http://www.rdkit.org>.

[7] CROCKFORD, Douglas; BRAY, Tim. Rfc 7159: The javascript object notation (json) data interchange format, IETF, Mar. 2014.

[8] Appendix E. Release Notes, PostgreSQL 9.4 Documentation, <https://www.postgresql.org/docs/9.4/static/release-9-4.html>

[9] US FOOD AND DRUG ADMINISTRATION, et al. Part 11 Electronic Records; Electronic Signatures. CFR-Code of Federal Regulations Title, 1997, 21.

[10] HALL, Mary P., et al., Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS chemical biology*, 7.11: 1848-1857, 2012.

[11] ZHANG, Ji-Hu; CHUNG, Thomas DY; OLDENBURG, Kevin R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of biomolecular screening*, 4.2: 67-73, 1999.

[12] IVERSEN, Philip W., et al., HTS assay validation, 2012.

[13] DRAGIEV, Plamen; NADON, Robert; MAKARENKOV, Vladimir. Systematic error detection in experimental high-throughput screening, *BMC bioinformatics*, 12.1: 25, 2011.

[14] SCHANZENBACH, Christoph, et al., Identifying ionic interactions within a membrane using BLaTM, a genetic tool to measure homo- and heterotypic transmembrane helix-helix interactions. *Scientific reports*, 7: 43476, 2017.

[15] MOTULSKY, H. J. Prism 5 statistics guide, 2007. GraphPad Software, 31: 39-42, 2007.

[16] COLEMAN, Thomas F.; LI, Yuying. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, 6.2: 418-445, 1996.



조 성 호

2010년 순천향대학교 정보기술공학부 졸업(학사). 2015년 고려대학교 전기전자전파공학부 졸업(석사). 2016년~현재 (재)서울대학교 의약바이오컨버전스연구단



로버트 E. 베리어

1979년 메사추세츠 주립대학 생화학 학사 1983년 보스턴대학교 생화학 석박사. 1985년~1987년 보스턴 대학교 약학과 생화학 조교수. 1987년~1997년 웨팅-플라우 연구소. 2004년~2008년 Invitrogen 연구개발 부사장. 2011년~현재 Stemina Biomarker

Discovery 최고운영책임자, 연구개발 부사장. 2014년~현재 (재)서울대학교 의약바이오컨버전스연구단 자문위원



김 인 희

1999년 한림대학교 식품영양학과 졸업(학사, 석사). 1996년~1998년 매일유업. 2003년~2007년 한국해양연구원 극지연구소 2012년~2016년 농림수산식품부 산하 농림수산식품교육문화정보원. 2017년~현재 (재)서울대학교 의약바이오컨버전스연구단



김 성 훈

1981년 서울대학교 약학사. 1983년 KAIST 생명공학 석사. 1991년 브라운대학교 분자유전학 및 생화학 박사. 1994년~2001년 성균관대학교 생명과학부 부교수. 2001년~현재 서울대학교 약대학 교수. 2008년~현재 차세대융합기술연구원 바이오융합연

연구소장. 2010년~현재 (재)서울대학교 의약바이오컨버전스연구단장



공 지 원

2010년 서울대학교 자연과학대학 화학부 졸업(학사). 2012년 서울대학교 생물물리 및 화학생물학과 졸업(석사). 2013년~현재 서울대학교 약대학 약학과 박사과정 2013년~현재 (재)서울대학교 의약바이오컨버전스연구단



김 대 산

2011년 연세대학교 의용전자공학부 졸업(학사). 2017년 서울대학교 생물물리 및 화학생물학과 졸업(석박통합). 2017년~현재 (재)서울대학교 의약바이오컨버전스연구단 2018년~현재 주식회사 메타포어



권 남 훈

1997년 서울대학교 자연과학대학 미생물학과 졸업(학사). 2001년 서울대학교 수의과대학 졸업(학사). 2008년 서울대학교 수의과대학 수의미생물학 전공 졸업(석박통합). 2005년~2007년 University of Idaho 방문연구원. 2008년~2011년 서울대학교 약대학. 2011년~현재 (재)서울대학교 의약바이오컨버전스연구단

연구단