

KCB newsletter

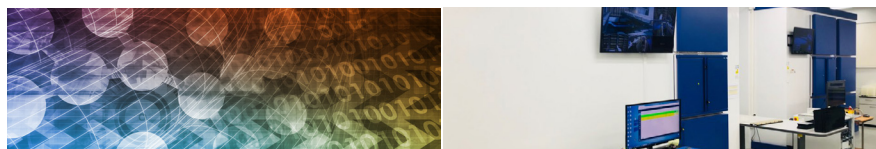
한국화학물은행 뉴스레터

2019년 제2호



CONTENTS

2019 Vol.2



[한국화학물은행 뉴스레터]

발행일 : 2019년 7월

발행인 : 김창균

편집인 : 이선경

발행처 : 한국화학물은행

한국화학물은행은 국가연구개발사업 수행을 통해 창출된 연구성과 중 화학물 및 관련 정보의 관리·유통 전담 기관으로 지정되어 있습니다.

[과학기술정보통신부 고시 제2017-7호]

- 03 한국화학물은행 뉴스
- 04 한국화학물은행 데이터 플랫폼 개발 현황
- 12 대외활동 · 주요방문인사
- 13 화합물 확보 현황
- 14 데이터 활용 현황
- 14 화합물 활용 현황
- 16 화합물 활용 논문 성과
- 18 인공지능 기반의 신약개발 연구현황
- 27 화합물 기탁 절차
- 28 한국화학물은행 보유 화합물 품질관리
- 29 데이터 활용 절차
- 30 화합물 활용 절차
- 31 한국화학물은행 제공 화합물의 활용결과(논문, 특허 등)에 대한 권리관계 규정
- 31 후속연구를 위한 화합물 추가합성 및 구매 진행 안내
- 32 한국화학물은행 제공 라이브러리 종류

한국화학물은행은 국내에서 합성되는 화합물(유기화합물 및 천연물) 및 관련정보를 범국가적 차원에서 통합 · 관리하고 공동으로 약효시험에 활용함으로써 국내 신약 개발연구 및 바이오 연구를 지원하기 위하여 지난 2000년에 설립된 이후로 국내 · 외에서 약 61만종의 화합물을 수집, 관리하고 있습니다.

한국화학물은행 뉴스



웹기반 신약소재 화합물 통합 데이터베이스 구축 협약

일시 : 2019. 02. 13(수) 오전 11:30

한국화학연구원 한국화학물은행은 인공지능 기반 신약개발에 필요한 화합물 정보 제공의 효율성 제고를 위해 (주)파로스아이비티와 함께 2월 13일 오전 11시 30분 한국화학연구원에서 “웹기반 신약소재 화합물 통합 데이터베이스 구축 협약식”을 개최했다.

한국화학물은행과 (주)파로스아이비티는 한국화학물은행에서 운영하고 있는 화합물의 구조와 약효 등의 화합물정보 데이터베이스 (Chemical Information Management System)와 화합물의 입고 (기탁), 출고(활용), 보유량 등의 정보를 관리하는 실험 종합 정보데이터베이스(Plate Information Management System), 그리고 연구자들의 연구데이터 및 해외 공공 화합물데이터를 모아 신약개발 분야의 국가 개방형 통합 데이터베이스를 구축할 계획이다. 본 협약을 통해 웹기반 데이터 플랫폼을 구축하여 향후 국내 신약개발 분야 연구자들에게 다양한 서비스를 제공할 것이며, 이는 인공지능 기반 신약개발에 필요한 시간과 비용의 감소 및 성공률 제고에 기여할 것이다.



글로벌 신약분야 발전을 위한 협약

일시 : 2019. 05. 24(금) 오전 11:30



한국화학연구원과 한국제약바이오협회, 한국보건산업진흥원 3개 기관이 5월 24일 한국화학연구원에서 신약개발을 위한 빅데이터 공유, 인공지능 인프라 구축, 공동연구개발 및 기술교류, 신약개발 및 의약생산에 대한 기술지원 등을 위한 업무협약식을 체결하였다.

한국보건산업진흥원과 제약바이오협회의 신약개발 관련 데이터, 한국화학연구원 한국화학물은행의 화합물 및 활용데이터 등을 통합해 신약개발 인공지능 인프라를 구축하고, 인공지능을 활용한 신약개발을 활성화한다는 구상이다.

이에 한국화학물은행은 인공지능 신약개발을 위한 공공 포털사이트를 구축하여 수요자 맞춤형 데이터 활용 시스템을 운영할 계획이다.

Focused 라이브러리 구성

한국화학물은행은 최근 신규 확보된 화합물로 구성된 focused 라이브러리 구축을 진행 중이다.

2015년 임상라이브러리를 구성한 이후 보유화합물 및 FDA 승인 약물이 증가함에 따라 임상라이브러리 재구성의 필요성이 증대되었다. 이에 한국화학물은행은 2018년에 임상화합물 200여종을 추가 확보하였으며 약 2,500여종 라이브러리 구축을 진행하고 있다.

또한 PPI 화합물 8,000여종을 신규 확보하여 올해 초 17,000여종의 PPI 라이브러리 구성을 완료하였다.

61만종의 골격분석과 안정성 및 약물성 검토를 통하여 선별된 2018 대표화합물 라이브러리 구성이 완료되었으며 7,000여종의 신대표라이브러리는 7월에 96-well 및 384-well plate 형태로 제공될 예정이다.

....

KCB 데이터 플랫폼 개발현황

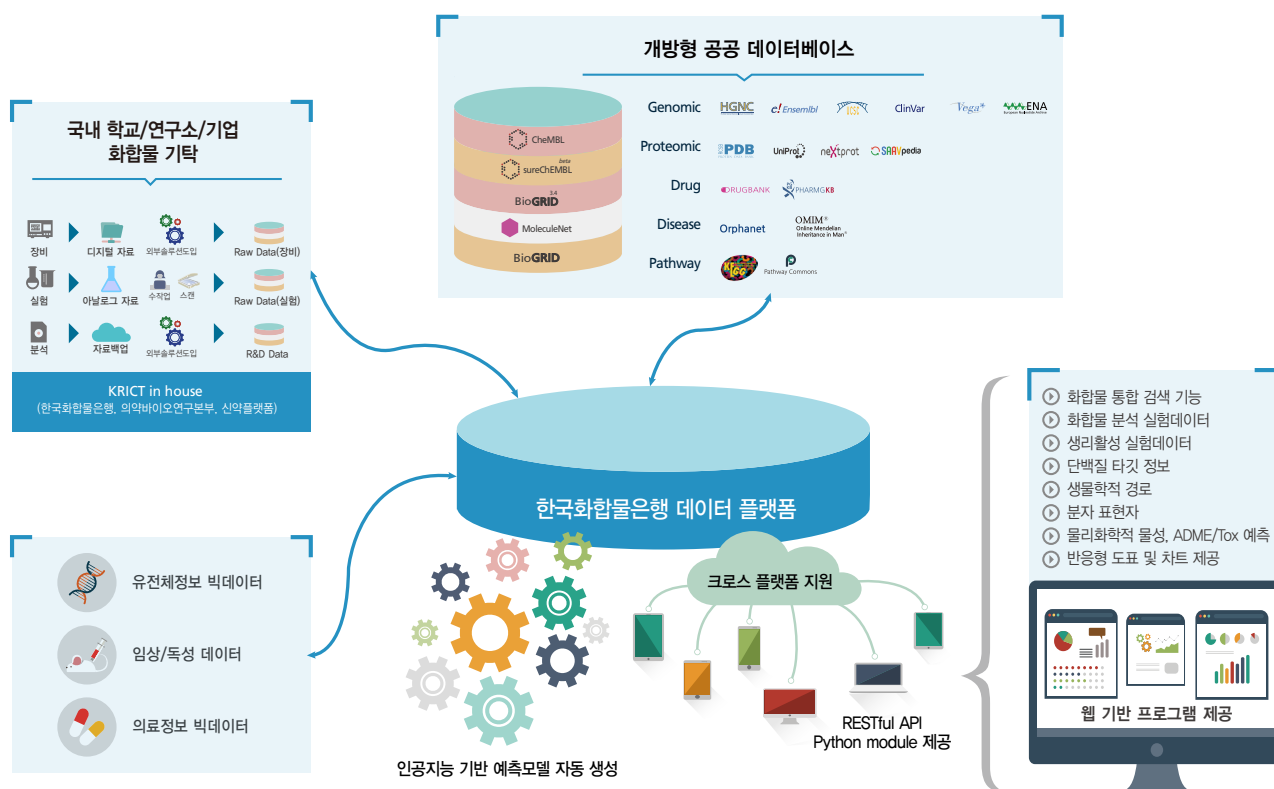
1

한국화합물은행 데이터 플랫폼

신약개발 분야에서는 이미 오래 전부터 컴퓨터를 이용한 화합물의 물리화학적 물성, 인체 내에서의 흡수/분포/대사/배설, 독성 예측 및 단백질과의 도킹 시뮬레이션에 의한 생물학적 활성 예측 및 약물 가상탐색과 같은 기술이 꾸준히 발전해 왔다. 최근 컴퓨터 성능향상과 함께 데이터를 활용한 인공지능 연구가 확대되고 있으며 한국화합물은행 또한 대한민국 신약개발 연구자들이 새로운 패러다임에 쉽고 빠르게 적응할 수 있는 신약개발 연구용 데이터 플랫폼을 제공하기 위해 노력하고 있다.

한국화합물은행 데이터 플랫폼 개발 계획

한국화합물은행은 2020년까지 자체 공개 가능한 보유 데이터 및 외부 공개 데이터를 수집/통합하여 포털 서비스 형태로 1) 화합물 통합 검색 기능, 2) LC-MS, NMR 등의 분석데이터, 3) 생리활성 실험 데이터 4) 단백질 타겟 정보, 5) 생물학적 경로, 6) 분자 표현자, 7) 물리화학적 물성, 흡수, 대사 및 독성 예측 정보를 웹 기반 데이터 플랫폼으로 제공할 계획이다. 2019년 현재, ChEMBL, Protein Data Bank(PDB), DrugBank 등의 외부 공공 데이터를 수집하여 한국화합물은행이 보유한 53만종의 공개 가능한 화합물과 통합 데이터베이스를 만들고 있으며, 2020년에는 생물학적 경로 및 질병 데이터, 장기적으로는 임상데이터 등도 수집 및 통합할 계획이다. 추가적으로 통합 수집된 데이터로부터 인공지능 기반 예측모델을 자동 생성하여 사용자에게 예측모델을 제공하는 것도 목표로 하고 있다.



< 한국화합물은행 데이터 플랫폼 시스템 >

한국화학물은행 뉴스



2

한국화학물은행 홈페이지 「www.chembank.org」 개편 계획

화합물 검색 시스템

The screenshot displays the Korea Chemical Bank website. The header includes the logo and navigation links. The main banner features a person holding a tablet with various chemical and industrial icons, accompanied by the text '공동활용을 통한 화합물의 가치 재창출' (Recreation of chemical value through joint use). Below the banner, there is a section titled '화합물 활용' (Chemical Use) which describes the bank's resources and services. To the right, a '반출 화합물 검색' (Export Chemical Search) interface is shown, featuring a chemical structure and a search bar. At the bottom, a blue banner promotes '신약개발의 첫 걸음을 시작하세요' (Start the first step of drug development) with contact information.

화합물 활용

매년 20-40만종의 한국화학물은행의 화합물이 국내 산·학·연에 무료로 제공되어 150-200여개 과제(신규착용점 60건이상) 대한 약효시험에 활용되고 있으며 현재까지 770건 이상의 다양한 작용점에 대한 약효검색에 활용되었습니다.

Hit 화합물에 대한 구조제공시 LC-MS 분석에 의한 순도 및 분자량 검증자료를 기본적으로 제공하고 있으며 Hit 화합물 관련정보 제공(유도체정보, 약물성 예측정보, off-target 정보)도 함께 제공하고 있습니다.

반출 화합물 검색

만들 가능한 화합물 구조를 검색해보세요.

신약개발의 첫 걸음을 시작하세요

신약개발연구에 필요한 화합물이 있다면 먼저 한국화학물은행에 문의주세요!

042-860-7190
chembank@krcit.re.kr

화합물 확보

< 새로운 한국화학물은행 홈페이지, <https://www.chembank.org/> >

2019년 9월 새롭게 개편되는 한국화학물은행 홈페이지 (<https://www.chembank.org/>)에서는 사용자가 분자 에디터를 통해 직접 분자를 그려 넣거나, 또는 화합물명, 화학식, CAS Number, SMILES 코드, InChI, InChIKey 등의 다양한 입력 값을 통해서 한국화학물은행이 보유하고 있는 화합물 정보를 통합 검색할 수 있을 것이다. 사용자는 웹 브라우저를 통해서 한국화학물은행의 보유 화합물 구조를 검색하거나, 간단히 Drug-likeness rules을 계산할 수 있게 된다.

위함울 소개 위함울 제품 위함울 서비스 위함울 파트너 위함울 채용 위함울 문의

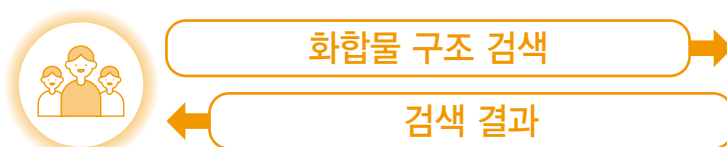
위함울을 통해 만

위함울 무료 만

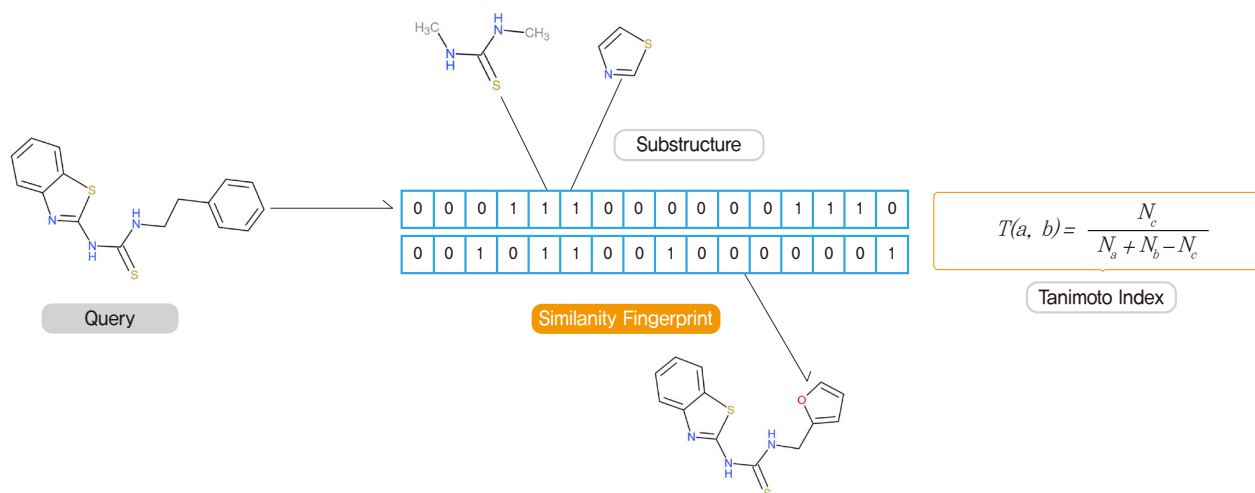
위함울을 함께

위함울의 첫 걸음을 시작하세요

위함울 확보

[illegible]

모든 화합물 데이터는 InChI, InChIKey, Canonical SMILES와 다양한 Fingerprints (Topological, MACCS Keys, Morgan 등)를 미리 생성하여 데이터베이스에 저장한다. InChI, InChIKey, Canonical SMILES는 입력 화합물 구조와 동일한 화합물 데이터를 검색하는데 사용된다. Fingerprints는 입력 화합물의 Sub-structure 검색 또는 유사도 함수를 이용하여 구조가 유사한 화합물을 찾을 때 사용되며, 유사도 함수는 기본적으로 Tanimoto coefficient가 적용된다.



< Chemical similarity search using 2D chemical fingerprints. >



기본적으로 사용되는 Tanimoto coefficient 외에도 fingerprints를 사용하는 다양한 유사도 함수(Dice, Cosine, Russel coefficient 등)를 적용하여 유사 화합물을 검색할 수도 있다. 약효시험을 통해 활성을 보인 화합물과 유사한 구조의 화합물을 한국화합물은행에서 확인하고 싶다면, 이제는 방문 검색을 하지 않고도 쉽게 홈페이지에 접속하여 유사 화합물 또는 sub-structure를 검색할 수 있다.

유사도 구분	수식	범위
Tanimoto coefficient	$c/(a+b-c)$	0 ~ 1
Euclidean distance	$\sqrt{(a+b-2c)}$	0 ~ N
Manhattan distance	$a+b-2c$	0 ~ N
Dice coefficient	$2c/(a+b)$	0 ~ 1
Cosine similarity	$c/\sqrt{ab}$	0 ~ 1
Russell-RAO coefficient	c/m	0 ~ 1
Forbes coefficient	cm/ab	0 ~ 1
Soergel distance	$(a+b-2c)/(a+b-c)$	0 ~ 1

< Fingerprints를 사용하는 유사도 리스트 >

* 두 개의 fingerprint A와 B가 있다면, m은 전체 fingerprint의 수, a는 fingerprint A의 수, b는 fingerprint B의 수, c는 fingerprint A와 B의 교집합 수이다.

분자 표현자 (Molecular descriptor) 계산 시스템

분자 표현자는 보통 신약개발 방법 중의 하나인 리간드 기반 가상탐색에서 단백질 타깃과의 상호작용을 예측할 수 있는 모델을 개발할 때 많이 사용된다. 리간드 기반 가상탐색 기술은 타깃 단백질 구조가 없어도 화합물의 구조와 생리활성 정보만을 이용하여 예측모델 개발이 가능하며, 일반적으로 Quantitative Structure-Activity Relationships (QSAR)로 알려져 많이 사용되고 있다. 성능이 우수한 QSAR 모델을 개발하기 위해서는 분자 표현자가 매우 중요한 요인으로 작용하기 때문에, 사용되는 모든 분자 표현자를 파악하고 확보하는 것이 매우 중요하다.

한국화학물은행 홈페이지에서는 화합물 통합 검색뿐만 아니라, 3,000개 이상의 1~3차원 Property type 분자 표현자와 다양한 Fingerprints 생성 기능도 제공한다. 사용자는 분자 표현자를 웹 브라우저를 통해 “<https://descriptor.chembank.org/>”에 접속하여 원하는 분자 표현자 선택 및 분자 구조를 입력하게 되면 얻을 수 있다. 여기에서 생성되는 분자 표현자는 다양한 화합물 물성을 예측하는데 입력 Features로 사용될 수 있다. 생성된 분자 표현자는 결과 데이터 테이블로 정리되어 보이며, 해당 데이터는 MDL SD 파일, 엑셀, CSV, PDF 파일로도 저장이 가능하도록 지원하고 있다.

< 분자 표현자 생성 페이지 >



Export SDF

Copy Excel CSV PDF

Show 10 entries

Search:

ID	Value
AlogP98_value	-0.099900
AMolRef	49.333400
Polarizability_Miller	19.185000
Polarizability_MPEOE	17.918008
SMolRef	49.381500
Solvation_Free_Energy	-14.940000**

Show 10 entries

Showing 1 to 6 of 6 entries

Previous 1 Next

< 분자 표현자 계산 결과 페이지 >

다수의 분자 구조로부터 빠르게 분자 표현자를 얻으려는 사용자는 Web API (<https://descriptor.chembank.org/api/>)를 이용하여 문제를 해결할 수 있다. 현재는 Python 사용자에게 한하여 MDL SD 파일을 입력 값으로 Web API를 이용하는 예제를 한국화학물은행 GitHub 저장소(<https://github.com/ChemBank/kcb-descriptor/>)에서 제공하고 있으며, 2019년까지 Java, R, Perl언어에 대한 활용 예제도 추가할 계획이다.

Korea Chemical Bank

<https://www.chembank.org>

Repositories Packages People Teams Projects Settings

Find a repository... Type: All Language: All Customize pins New

kcb-descriptor

KCB Molecular Descriptor Calculator

Apache-2.0 0 0 0 0 Updated Jun 19, 2019

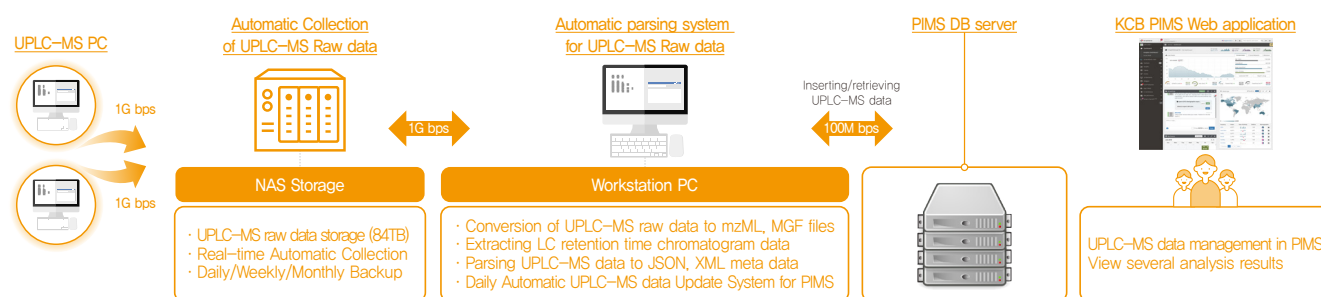
Top languages

Loading...

< 한국화학물은행 GitHub 저장소 - <https://github.com/ChemBank/> >

UPLC-MS 데이터 자동 수집/변환 관리 시스템

한국화합물은행에서는 Ultra performance liquid chromatography-mass spectrometry (UPLC-MS)를 이용하여 순도를 측정하여 80%이상의 화합물만 기탁 받거나, 반출해 주고 있다. 순도를 측정함에 따라서 UPLC-MS 분석 데이터가 매우 빠르게 증가하게 되었지만, 쌓여가는 데이터를 관리하고 한국화합물은행의 실험정보관리시스템 (Plate Information Management System, PIMS)의 데이터베이스로 저장하는 과정은 수작업을 통해 하나씩 입력할 수밖에 없었다. 따라서 질량분석기를 담당하는 연구원은 UPLC-MS 데이터의 입력에 많은 시간과 노력을 쓰지 않을 수 없었다. 이런 문제점을 해결하기 위해서 UPLC-MS 분석 데이터를 실시간으로 자동 수집 저장/변환하여 데이터베이스 서버에 저장하는 시스템을 구축했다.



<UPLC-MS 데이터 자동 수집/변환 관리 시스템 구성도 >

전체적인 시스템을 간략히 설명하면,

- 1) 모든 UPLC-MS Raw 데이터를 수집 저장하기 위해서 각 UPLC-MS 분석기기의 컴퓨터와 NAS 저장장치가 연결되었다.
- 2) NAS 저장장치는 실시간으로 연결된 UPLC-MS 분석기기 컴퓨터에서 Raw 데이터를 실시간 파악 후 수집하게 된다.
- 3) 수집된 UPLC-MS Raw 데이터를 보유한 NAS 저장장치는 고성능 Workstation 컴퓨터와 네트워크로 연결되어 실시간으로 Raw 데이터를 동기화 한다.
- 4) 동기화된 UPLC-MS Raw 데이터를 Workstation 컴퓨터는 ProteoWizard와 MassLynx 프로그램을 이용하여 질량분석 표준 파일 포맷인 mzML로 자동 변환을 수행한다.
- 5) 변환된 mzML파일은 파이썬 proteomics 모듈을 이용하여 JSON과 XML 파일로도 자동 변환을 수행한다.
- 6) Raw 데이터, mzML, JSON, XML 파일은 모두 데이터베이스 서버에 저장된다
- 7) 사용자는 데이터베이스에 저장된 UPLC-MS 데이터를 웹 프로그램을 통해서 이용할 수 있다.

대부분의 질량분석기기의 분석결과는 해당 기기를 제공해주는 Vendor의 특정 프로그램을 통해서만 데이터를 살펴볼 수 있다. 한국화합물은행이 보유한 질량분석기기의 결과 데이터 역시 MassLynx 프로그램을 이용하지 않으면 질량분석 결과를 살펴보기 힘들었다. 따라서 측정된 모든 UPLC-MS Raw 데이터는 ProteoWizard 프로그램을 통하여 mzML파일로 변환하는 작업이 필요했고, 해당 작업은 고성능 Workstation 컴퓨터에서 변환 작업을 담당하였다. 하지만, 여전히 질량분석 표준 파일 포맷으로 알려진 mzML도 이진(Binary) 파일이므로 사람이 해석할 수 있는 데이터 형태는 아니며 사용하기 위해서는 항상 mzML Parsing 프로그램이나 모듈을 이용해야 한다. 따라서 mzML 파일을 사람이 이해할 수 있으며, 많은 프로그래밍 언어에서 쉽고 빠르게 적용이 가능할 수 있도록 JSON과 XML형태로 변환을 수행하였다.

변환된 파일들은 PIMS 데이터베이스에 저장되며 한국화합물은행 관리자는 웹 브라우저를 통해서 질량분석을 수행한 화합물 리스트를 살펴볼 수 있게 된다.



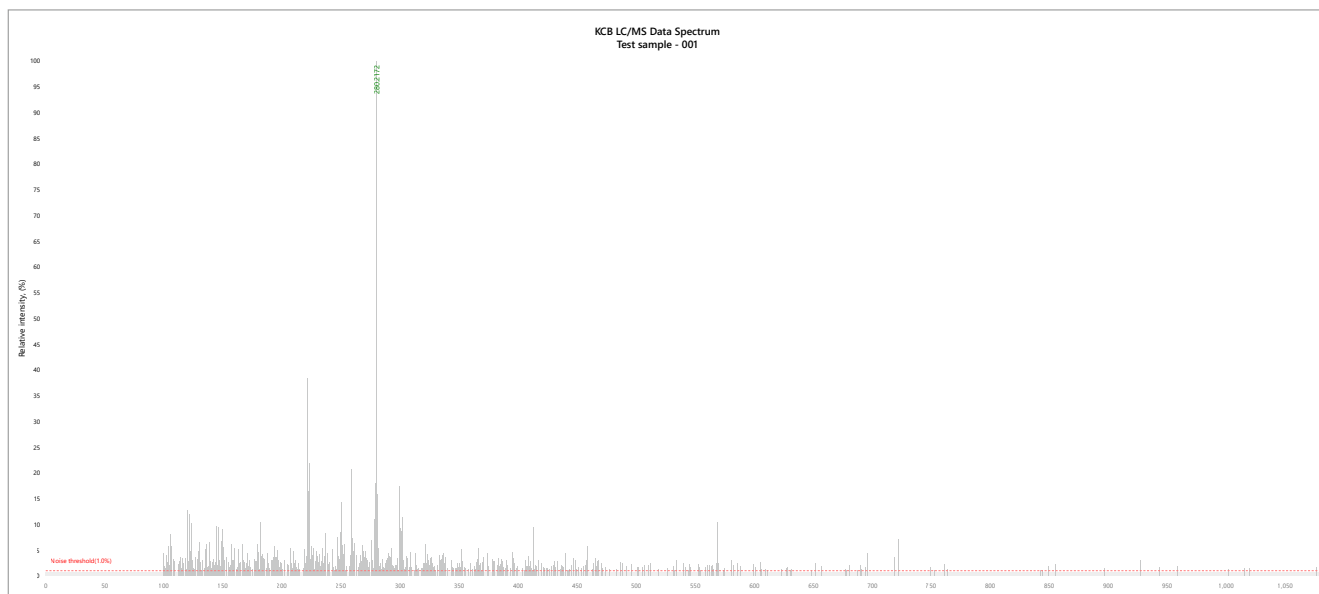
Search for something... Welcome to PIMS. 16 Log out

화합물 정보
Home / 기본정보관리 / 화합물 정보

Index	2D Structure	Formula	InChI	InChIKey	SMILES	LC/MS data
1		C11H15NO3S2	InChI=1S/C11H15NO3S2/c1-3-14-10	WIPAMIXVLMVRHF-MRVPVSSYSA-N	CCOC(=O)C(=C1SCCS1)C1CC(C)=NO1	O
2		C12H14ClNO2	InChI=1S/C12H14ClNO2/c1-12(2)-1	XZSIAPVWRQZGGZ-SNVBAGLBSA-N	CC1(C)COC(c2ccccc2C)NC1=O	O
3		C15H18N2O3	InChI=1S/C15H18N2O3/c1-9(2)15(3)	HBEIBGCAXFZAX-HNNXBMYSA-N	COC(=O)c1ccccc1C1=NC(C)(C)C(C)=C	O

< 질량분석 데이터를 가진 화합물 정보 리스트 >

질량분석 데이터는 웹 기반으로 자체 개발된 Mass Spectrum Viewer (MS Viewer)를 통해서 m/z peak 정보를 쉽게 확인할 수 있다. MS Viewer는 확대/축소, 특정 Peak의 색 압하기, 특정 Peak의 Tooltip 보여주기, 그림파일 저장 등의 다양한 기능들을 제공해 준다. UPLC-MS 데이터는 현재 한국화합물은행 내부 관리자만 사용이 가능한 PIMS에서 작동하고 있지만, 2019년 10월까지 한국화합물은행 홈페이지에서도 화합물의 질량분석 결과 정보도 함께 볼 수 있을 것이다



< Mass Spectrum Viewer >

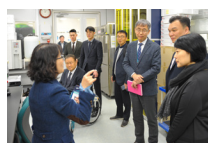
대외활동



한국화합물은행은 화합물의 기탁과 활용을 장려하기 위하여 다양한 방법으로 홍보활동을 진행하고 있습니다.

일정	내용	장소
2018.04.16	KCB 소개 발표 및 국제심포지움 참석	분당차병원
2019.04.17~19	대한화학회 춘계총회 및 학술발표회 홍보부스	수원 컨벤션센터
2019.06.02~05	생화학분자생물학회 KCB 특별세션 발표 및 홍보부스	제주 국제컨벤션센터
2019.06.27~28	대한화학회 의약화학분과회 하계워크숍 홍보원고 게재	거제마리나 대명리조트

주요방문인사



2019.04.03.
네이버랩 | 한성숙 대표
이상민 국회의원



2019.02.28.
부경대학교 | 최태진 교수



2019.03.19.
(주)파로스아이비티
남기엽 소장, 유응식 R&D총괄사장



2019.04.09.
녹십자 | 이지은 상무



2019.05.22.
기획재정부 | 양충모 국장

2019.02



2019.02.28.
과학기술정보통신부
김광수 국장, 한형주 과장,
박성국 서기관

2019.03



2019.03.20.
University of Pennsylvania
Molander 교수

2019.04



2019.04.18.
서울대학교 | 김흥기 교수

2019.05



2019.05.31.
티앤알바이오랩 | 김 현 이사



2019.04.22.
농업진흥청



2019.04.25.
파스퇴르 연구소



화합물 확보 현황

한국화학물은행은 2019년 현재 약 61.3만종의 화합물을 보유하고 있으며 다양성, 약물성, 특이성이 우수한 고수준 신약소재 화합물 라이브러리를 제공하기 위하여 다양한 방법으로 우수한 화합물을 확보하고 있다.

1 연구성과 기탁[법적의무 기탁]

「국가연구개발사업의 관리」 등에 관한 규정 제25조13항에 따라 국가연구개발사업 수행을 통해 창출된 화합물은 연구성과 관리·유통 전담기관인 한국화학연구원 한국화학물은행에 의무적으로 기탁하도록 되어 있다. 연구성과 기탁 효율제고를 위하여 2017년부터 국가연구과제 및 사업평가에 연구성과 기탁실적을 반영하고, 전담기관에 기탁된 성과만 인정 하도록 “국가연구개발 평가지침”이 개정되었다.

2 국내 전문가 공모 위탁합성

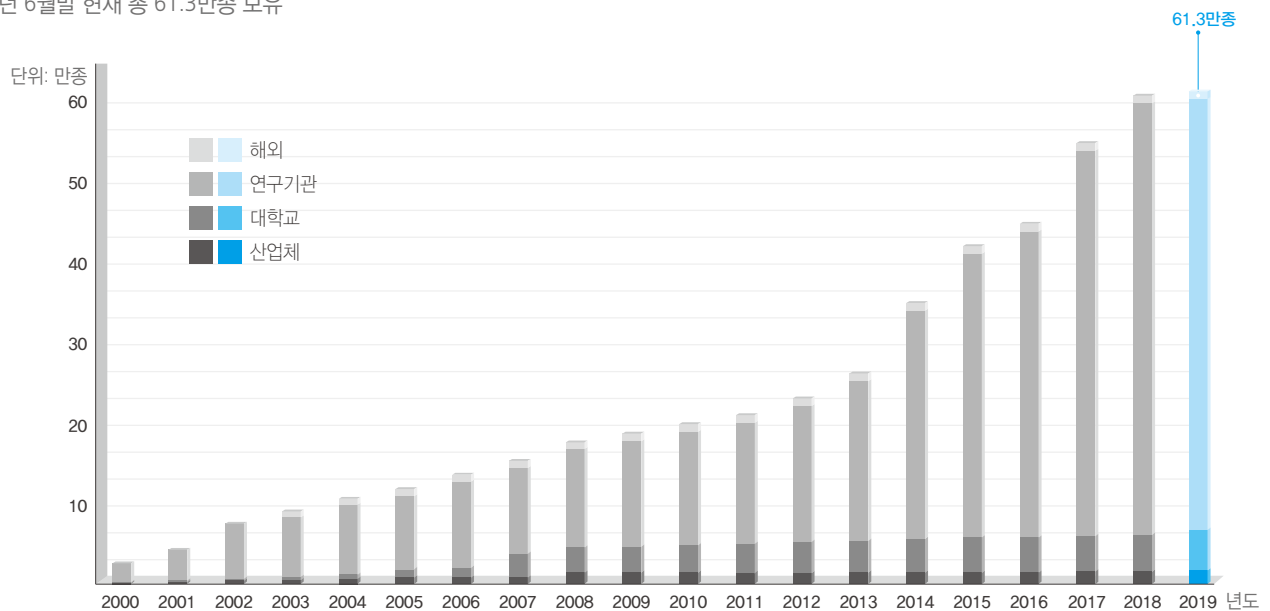
보유화합물의 구조 다양성 확대를 위하여 국내 합성 전문가를 대상으로 위탁 합성 과제를 공모, 선정을 통하여 화합물을 확보하고 있다. 전문가 위탁 합성은 2014년 부터 매년 15개 내외 과제를 선정하여 7,500여종의 화합물을 확보하고 있으며, 2018년에도 12개 과제에서 약 2,000여종을 확보하였다. 2019년에도 전문가 위탁합성공모를 통하여 14개 과제에서 2,500여종을 확보할 예정이다.

3 특수 골격화합물 외국 구매

단순 기탁만으로 충족하기 어려운 화합물 다양성을 신속하게 확보하기 위하여 의약 및 분자설계 전문가의 화학정보학, 분자 모델링 기술을 활용하여 약물성 및 다양성 극대화를 고려한 특수 골격 화합물을 외국 vendor로 부터 선별하여 구매, 확보하고 있으며 2019년에도 지속적으로 가치가 높은 라이브러리를 확보하여 제공할 계획이다.

화합물 확보 현황(누계)

2019년 6월말 현재 총 61.3만종 보유





데이터 활용 현황

한국화학물은행은 2018년부터 화합물 구조 및 공개 가능한 약효시험 데이터 정보를 제공하고 있다. 2019년 현재까지 26건의 데이터를 제공하였고, Hit 및 SAR분석 위한 구조정보도 84건 제공되었다.

정보제공 데이터 신청 내역별 지원 건수 현황 (2019.06. 누적)

분류	소분류	지원 건수
화합물 구조 정보	구매화합물 라이브러리	11건
	임상 화합물	1건
	Focused 라이브러리	1건
	GPCR 화합물	1건
	PPI 화합물	1건
	반출 가능한 라이브러리	3건
약효시험결과 정보	공개 가능한 정보	9건



화합물 활용 현황

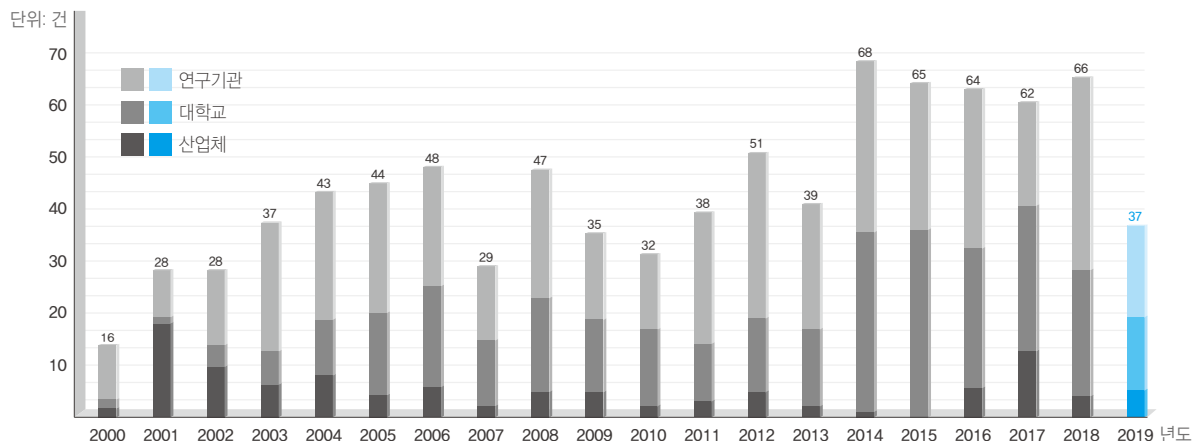
연 평균 60건 이상의 신규 작용점에 대하여 40만종 이상의 화합물이 국내 산·학·연에 제공되어 활용되고 있으며 현재까지 877건 이상의 특정 작용점에 대한 약효검색에 활용되었다. Hit 화합물에 대한 구조 정보 제공시 UPLC-MS 분석에 의한 순도 및 분자량 검증 자료를 기본적으로 제공하고 있으며 그 외 요청시 유도체 검색, 약물성 예측, off-target 정보 등도 함께 제공하고 있다.

최근 4년간 화합물 활용 현황

년도	활용과제 건수		반출 횟수		반출화합물 개수	
	전체과제	신규과제	전체과제	신규과제	전체과제	신규과제
2016	239건	64건	411회	184회	1,236,527종	1,148,136종
2017	195건	62건	371회	137회	361,561종	181,032종
2018	188건	66건	297회	117회	253,514종	191,810종
2019(상반기)	76건	37건	128회	55회	91,749종	63,520종

• 2019년(상반기): 128회에 걸쳐 총 91,749종 화합물 반출(평균 506개/일), 유도체 검색(의약화학)지원: 49회(평균 1.8건 / week지원)

연도별 신규 약효시험 활용 현황 (2000~2019.06.)



2019년 상반기 37건

약효시험명	구분	적용질환
*** 저해제 개발	연구기관	항생제
*** 치료 관련 화합물 탐색	대학교	기타
Identification of *** and *** inhibitors for novel antibiotics	대학교	항생제
*** 활성평가	연구기관	바이러스
*** 형성 저해	대학교	염증
*** 저해제 스크리닝	연구기관	항암제
*** 저해제 탐색	산업체	염증
*** 조절제를 찾기 위한 약물 스크리닝	대학교	기타
*** 항산화능 증진 물질 스크리닝	대학교	기타
*** 결합 유무 실험	산업체	대사성질환
*** 활용 - 세포주기 및 사멸 관련 약물 탐색	연구기관	항암제
*** 살균활성 리드화합물 발굴	연구기관	살균제
*** 세포의 ***에 대한 식작용과 분해	대학교	신경계
***와 ***의 결합 저해제 발굴	연구기관	항암제
*** 살균활성 리드화합물 발굴	연구기관	살균제
*** 유래의 *** 시스템을 기반으로 한 항생제 개발	대학교	결핵
*** 스크린을 사용한 *** inhibitor screen	대학교	결핵
*** 이용한 항바이러스 후보물질 선별	대학교	바이러스
*** 효소 활성	산업체	기타
*** 구조를 기반으로 작은 분자수준의 ligand 탐색	대학교	대사성질환
*** 유발하는 *** 활성억제 화합물 발굴	연구기관	신경계
*** 저해제	대학교	기타
*** 약효평가	연구기관	항암제
*** 저해제 후보물질 탐색 연구	연구기관	항암제
*** 사용한 *** 유효 약물 스크린	대학교	결핵
*** 활성 억제제 발굴	대학교	항암제
*** 활용 기술 개발	대학교	기타
*** 개선	대학교	염증
*** 활성 측정	산업체	기타
*** 작용물질 발굴	산업체	신경계
*** 억제 약물 발굴	대학교	항암제
*** 세포로의 분화 측정	연구기관	대사성질환
*** 능력 측정	연구기관	대사성질환
***의 효현 활성	대학교	대사성질환
*** 유효물질 도출	대학교	기타
*** 치료제 화합물 도출	대학교	기타
*** 환자 세포 유래 스크리닝 시스템	산업체	기타



화합물 활용 논문 성과

2019년 상반기 12건

한국화합물은행 제공 화합물 library를 활용하여 발표된 논문 목록

No	논문명	저자	저널
1	Psoralidin Stimulates Expression of Immediate-Early Genes and Synapse Development in Primary Cortical Neurons	Seojin Hwang, Seong-eun Lee, Sang-Gun Ahn, Gum Hwa Lee*	<i>Neurochemical Research</i> 2018 , 43(2), 2460-2472
2	Identification of a novel SIRT7 inhibitor as anticancer drug candidate	Ji-Hye Kim, Da hee Kim, Suk Joon Cho, Kwan-Young Jung, Jong-Hoon Kim, Jun Mi Lee, Hee Jung Jung, Kwang Rok Kim*	<i>Biochemical and Biophysical Research Communications</i> 2019 , 508(2), 451-457
3	Homoharringtonine Induces Apoptosis in Human Colorectal Carcinoma HCT116 Cells Via Downregulation of Wnt/ β -Catenin Signaling Cascade	Mikyung Park, Hojeong Kwon*, and Seong Hwan Kim*	<i>Bulletin of the Korean Chemical Society</i> 2019 , 40(2), 196-199
4	Stereoselective Synthesis of Highly Functionalized 5- and 6-Membered Aminocyclitols Starting with a Readily Available 2-Azetidinone	Raghavendra Achary, Hyeong Rae Kim, Hyeon-Kyu Lee*	<i>The Journal of Organic Chemistry</i> 2019 , 84(7), 4263-4272
5	DITMD-induced mitotic defects and apoptosis in tumor cells by blocking the polo-box domain-dependent functions of polo-like kinase 1	Ka-UI Kim, Ju Hee Lee, Mi Young Lee, Chong Hak Chae, Jeong Hyun Lee, Byung Ho Lee, Kwang-Seok Oh*	<i>European Journal of Pharmacology</i> 2019 , 847, 113-122
6	Small-molecule inhibitor of HlyU attenuates virulence of <i>Vibrio</i> species	Zee-Won Lee, Byoung Sik Kim, Kyung Ku Jang, Ye-Ji Bang, Suhyeon Kim, Nam-Chul Ha, Young Hyun Jung, Hyun Jik Lee, Ho Jae Han, Jong-Seo Kim, Jeessoo Kim, Pramod K. Sahu, Lak Shin Jeong, Myung Hee Kim & Sang Ho Choi*	<i>Scientific Reports</i> 2019 , 9, 436-449

No	논문명	저자	저널
7	A nano sensor for sensitive and selective detection of Cu ²⁺ based on fluorescein: Cell imaging and drinking water analysis	Prasad G.Mahajan, Nilam C.Dige, Balasaheb D.Vanjare, Seong-Hui Eo, Song Ja Kim, Ki Hwan Lee*	<i>Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy</i> 2019, 216 , 105-116
8	Discovery of Novel Pyruvate Dehydrogenase Kinase 4 Inhibitors for Potential Oral Treatment of Metabolic Diseases	Dahye Lee, Haushabhau S. Pagire, Suvama H. Pagire, Eun Jung Bae, Mahesh Dighe, Minhee Kim, Kyu Myung Lee, Yoon Kyung Jang, Ashok Kumar Jaladi, Kwan-Young Jung, Eun Kyung Yoo, Hee Eon Gim, Seungmi Lee, Won-Il Choi, Young-In Chi, Jin Sook Song, Myung Ae Bae, Yong Hyun Jeon, Ga-Hyun Lee, Kwang-Hyeon Liu, Taeho Lee, Sungmi Park, Jae-Han Jeon, In-Kyu Lee*, and Jin Hee Ahn*	<i>Journal of Medicinal Chemistry</i> 2019, 62(2) , 575-588
9	Synthesis and Biological Evaluation of PF-543 Derivative Containing Aliphatic Side Chain	Seon Woong Kim, Taeho Lee, Yoon Sin Oh, Sang Mi Shin, Joo-Youn Lee, Sanghee Kim, Dong Jae Baek*, and Eun-Young Park*	<i>Chemical and Pharmaceutical Bulletin</i> 2019, 67(6) , 599-603
10	A Small Molecule Targeting Mutagenic Translesion Synthesis Improves Chemotherapy	Jessica L. Wojtaszek, Nimrat Chatterjee, Javaria Najeeb, Azucena Ramos, Minhee Lee, Ke Bian, Jenny Y. Xue, Benjamin A. Fenton, Hyeri Park, Deyu Li, Michael T. Hemann,* Jiyong Hong,* Graham C. Walker,* and Pei Zhou*	<i>Cell</i> 2019, 178(1) , 152-159
11	In vitro activity of DNF-3 against drug resistant Mycobacterium tuberculosis	Md Imtiazu Islam, Choi Moon Han, Hoonhee Seo, Sukyung Kim, Hafij Al Mahmud, Kung-Woo Nam, Byung-Eui Lee, Venkata S. Sadu, Kee-In Lee, Ho-Yeon Song*	<i>International Journal of Antimicrobial Agents</i> 2019, 54(1) , 69-74
12	An insect serotonin receptor mediates cellular immune responses and its inhibition by phenylethylamide derivatives from bacterial secondary metabolites	Ariful Hasan, Hyun-Suk Yeom, Jaewook Ryu, Helge Bode, Yonggyun Kim*	<i>BioRxiv - The preprint server for biology</i> 2019, in press



인공지능 기반의 신약개발 연구현황

한국화학연구원 의약정보플랫폼센터 강영묵

평균 10년, 1조원의 비용

일반적으로 알려져 있는 신약개발에 필요한 기간과 비용이다. 이렇게 많은 시간과 비용이 드는 이유는 신약 후보 물질을 찾는 단계에서 상당한 시간과 비용이 소모되기 때문이다. 신약 개발에는 이러한 막대한 시간과 비용이 들어가지만 정작 성공 가능성은 매우 낮다. 신약 개발 과정은 질병을 일으키는 단백질을 찾아내고 해당 질병을 치료하기 위한 후보물질 탐색을 전임상, 안전성 및 약효를 평가하는 임상 1, 2, 3상 단계를 거친다. 이와 같은 과정에는 많은 시행착오가 생기며 이로 인해 막대한 연구개발 시간과 비용이 소요될 수 있으며, 가장 치명적인 점은 신약 개발과정은 비가역적이라 실패할 경우, 그 시점에서 개선하는 것이 아니라 처음부터 다시 시작해야 한다는 것이다. 물질에 문제가 있으면 초기 히트 탐색부터, 타깃에 문제가 있으면 타깃 검증부터 시작해야 하므로 연구자들과 소비자를 모두 지치게 한다.

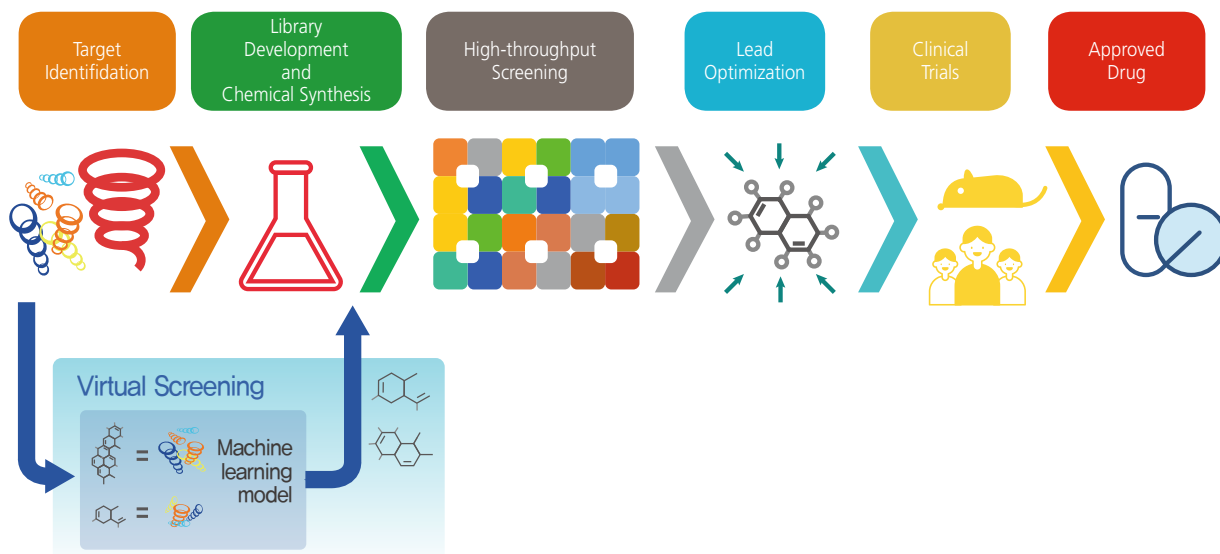
신약개발에 들어가는 시간과 비용을 줄이고, 성공 가능성을 높일 수 있는 방법은 없을까? 나아가 신약개발의 실패 결과가 다음 연구에 효율적으로 활용될 수는 없을까

신약 개발 업계 관련자들은 항상 이런 고민을 해왔을 것이며 이를 해결하기 위해 엄청난 노력을 꾸준히 하였다. 최근 4차 산업혁명이라는 이슈와 함께 이런 질문에 대한 대안으로 관심 받고 있는 기술이 있다. 바로 컴퓨터를 이용한 인공지능 기술이다. 인공지능은 하드웨어 발달 및 병렬처리 알고리즘 개발로 인한 컴퓨터 성능향상, 딥러닝을 기반으로 하는 획기적 데이터 분석 기술, 전세계를 연결하는 초고속 인터넷 등의 다양한 기술 발전의 결과물로서 이미 이미지, 동영상부터 음성인식까지 이미 오래 전부터 인간의 해석 능력을 초과하여 신약개발을 포함한 모든 과학분야의 영역으로 확대하고 있다.

신약개발 과정의 실패는 투자된 연구 비용의 회수가 어려운 것이 현실이다. 신약개발의 실패원인은 주로 효능에 문제가 발생한 경우 50%, 부작용으로 인한 실패가 25%라고 알려져 있다. 실패한 데이터를 활용하여 예측모델을 개발하면 개선된 가능성을 기대할 수 있을 것이다. 이런 문제에서 인공지능을 활용하게 되면 개선될 여지가 충분히 크기 때문에, 인공지능 기술은 기존 신약개발 연구개발 과정의 효율성을 높여 신약개발 성공률을 높일 수 있을 것으로 기대하고 있다.

신약개발 연구에서 사용되는 기존의 가상탐색 기술

컴퓨터를 이용한 신약개발 연구는 지난 수십 년 동안 화합물과 생물학적 타깃 간의 상호 작용을 컴퓨터, 물리, 화학, 통계 등의 융합 연구 결과물로서 가상탐색(Virtual screening) 기술을 개발하여 사용해 왔다. 가상탐색은 통계적 / 기계학습 모델들을 이용하여 약물-타깃 상호 작용을 빠르게 예측한다. 가상탐색 결과에서 약물-타깃 상호작용이 좋지 않게 평가된 결과를 제거한 후에 신약으로 가능성이 높은 활성 조합만을 다음 단계인 HTS 과정으로 보내어 신약개발 과정의 비용과 시간을 크게 줄일 수 있도록 도와주는 것이다.



< 신약개발 과정에서 가상 탐색이 사용되는 과정 >

기존 대부분의 가상탐색 기술들은 화합물의 생물학적, 물리화학적, 위상학적(Topological) 물성들과 단백질 타깃의 실험적으로 검증된 생리 활성 값을 이용하여 알려지지 않은 화합물의 생리활성을 예측하는 방법을 사용한다. 가상탐색 기술은 화합물과 단백질 타깃을 컴퓨터적으로 표현할 수 있는 분자 특징들을 정량적인 값으로 수치화 하여 화합물과 타깃 단백질 분자 사이의 상호작용을 모델링 하기 위한 입력 값으로 사용하는데, 이때 사용되는 입력 값들의 형태에 따라서 크게 1) 구조 기반 가상탐색, 2) 리간드 기반 가상탐색 방법의 2가지 형태의 가상탐색 기술로 분류할 수 있다.

구조 기반 가상탐색 기술은 단백질과 후보 화합물의 구조가 알려져 있는 경우에 사용이 가능한 방법으로 AutoDock[1], DOCK[2], Glide[3], GOLD[4], FlexX[5], Fred[6]와 같은 Docking 프로그램들을 이용하여 리간드 화합물 구조와 타깃 단백질의 결합 사이트(Binding site)의 상호작용을 평가하여 높은 결합 점수를 가지는 화합물과 타깃 단백질의 구조를 찾아내는 방법이다. 이 방법은 3차원 공간에서 리간드와 단백질의 원자 위치에 따라서 결합 점수가 달라지므로, 더 정확한 결합 점수를 평가하기 위해서는 리간드와 단백질의 구조변화를 다양하게 고려하여야 한다. 다양한 리간드 단백질의 구조 변화를 고려할 때, 회전가능한 분자 결합이 증가할수록 고려해야 하는 3차원 구조의 경우의 수가 기하급수적으로 증가하여 컴퓨터 계산 시간이 크게 증가할 수 있다는 점을 주의하여야 한다. 또한 결합 점수를 평가하는 에너지 함수에 Docking 결과가 굉장히 영향을 많이 받게 되는데 사용하는 에너지 함수를 제대로 이해하고 적용범위를 파악하는 것이 매우 중요하다.

리간드 기반 가상탐색은 화합물의 분자 표현자(Molecular descriptor)를 사용하여 단백질 타깃과의 상호작용을 예측할 수 있는 모델을 개발하여 접근하는 방법이다. 리간드 기반 가상탐색 기술은 타깃 단백질 구조가 없어도 화합물의 구조와 생리활성 정보만을 이용하여 예측모델 개발이 가능하며, 일반적으로 Quantitative Structure-Activity Relationships (QSAR)이라는 이름으로 알려져 많이 사용되어지고 있다. 성능이 우수한 QSAR 모델을 개발하기 위해서는 분자 표현자가 매우 중요한 요인으로 작용하기 때문에, 현재까지 개발되어서 사용되어지는 모든 분자표현자를 파악하고 확보하는 것이 중요할 것이다. 분자 표현자는 크게 Property type과 Fingerprint type으로 구분할 수 있으며, 분자 표현자가 계산되는데 사용되어지는 화합물 구조의 차원 정보에 따라서 1-3차원 분자 표현자로 구분하기도 한다.

구분	Property type	Fingerprint type
1차원 표현자	분자량, Constitutional descriptors (Atom number, Atom-type count, Other basic descriptors such as number of heavy atoms) Functional groups Reactive groups List of substructure fragments Substituent atoms	
2차원 표현자	Topological descriptors 2D Geometrical descriptors Graph-based substructures Connectivity bonds	Substructure keys based (MACCS, MNA, ALOGP) Path based (DayLight, FP2) Circular (ECFP, HOSE code)
3차원 표현자	3D geometrical descriptors Surface area, Volume 3D-based graph invariants 3D autocorrelation descriptors 3D WHIM, RDF, MoRSE descriptors	Geometrical (triangular descriptors) Pharmacophore
기타	SMILES Text-based molecular fingerprints ATC code annotations	

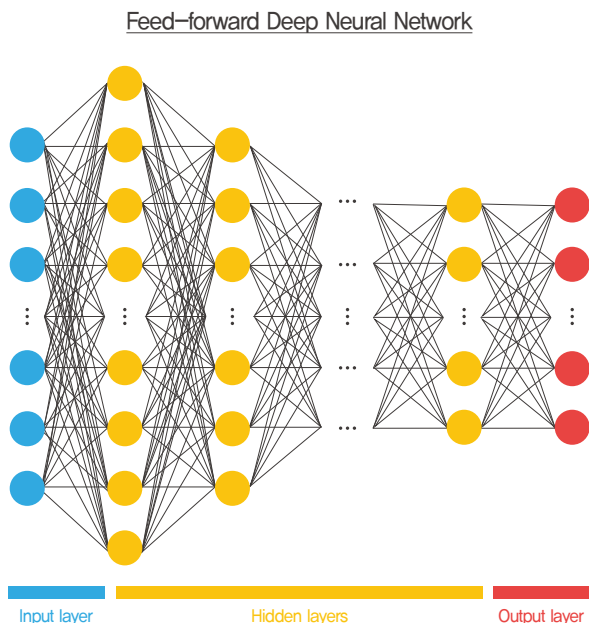
< 분자 표현자 분류표 >

현재까지 다양한 분자 표현자 계산 프로그램들이 존재하는데, 먼저 Property type 표현자들은 Dragon[7], RDKit[8], PreADME/T[9], PaDEL descriptor[10] 등의 프로그램을 통해서 적게는 수백 개에서 많게는 수천 개의 표현자를 얻을 수 있다. 또한 Fingerprint type의 표현자들은 RDKit[8], OpenBabel[11], DayLight Toolkit[12], Chemistry Development Toolkit[13], Open Eye Toolkit[14], ChemmineR[15], Indigo[16] 등의 다양한 프로그램들을 통하여 얻을 수 있다. 이런 다양한 프로그램을 통하여 분자 표현자를 생성한 것과 기존의 통계적/기계학습 모델들을 적용하여 예측모델을 개발하던 것이 기존의 리간드 기반 가상탐색 방법이라고 할 수 있다.

신약개발 분야의 새로운 가상탐색 방법론, 딥러닝

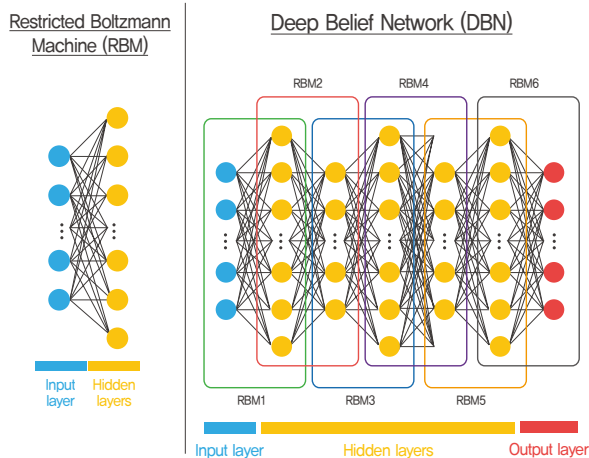
최근에는 정교한 기계 학습 방법 중 하나인 딥러닝 기술이 적용되어 개선된 예측 성능과 정확도를 가진 모델들이 꾸준히 보고 되고 있다. 초창기 컴퓨터 시각 및 자연 언어 처리분야에서 획기적인 발전을 가져온 후 인기가 높아진 딥러닝 기술은 현재 신약개발의 가상탐색에서까지 영역을 확장 시켜 나가고 있는 것이다. 현재까지도 매우 다양하고 새로운 딥러닝 방법들이 개발되어 보고 되고 있지만, 신약개발 과정의 가상탐색에서 가장 많이 사용되어지는 대표적인 딥러닝 방법들과 그 활용 사례들을 살펴보면 다음과 같다.

1 전방 전달 심층 신경망 (Feed-Forward Deep Neural Networks, FFDNN)



전방 전달 심층 신경망은 가장 기본적인 심층 신경망으로서 일반적으로 모든 신경망이 완전 연결된(Fully connected) 다중 은닉층들(Hidden layers)로 이루어져 있다. 다중 단백질 타겟의 약물-타겟 상호작용을 예측하는 모델에서 가장 많이 사용되어지는 인공신경망 구조 중의 하나이다. 심층 인공신경망의 성능이 전세계에 크게 알려지던 2014년에 Dahl은 다양한 분자표현자들을 이용하여 다중 단백질 타겟에 대하여, Unterthiner 연구팀은 ECFP12 fingerprint를 이용하여 각각 전방 전달 심층 신경망 모델들을 만들었다.[17] 이듬해 2015년에는 Ma 연구팀이 AP(Atom pairs)와 DP(donor-acceptor pair) 표현자와 전방 전달 심층 신경망을 이용한 QSAR (Quantitative Structure-Activity Relationship) 모델을 만들었다.[18] 그 외에도 Ramsundar 연구팀과 Koutsoukas 연구팀은 각각 2015년과 2017년에 ECFP4 fingerprint를 이용한 전방 전달 심층 신경망 모델들을 개발해 신약개발 연구에 적용하였다.[19,20]

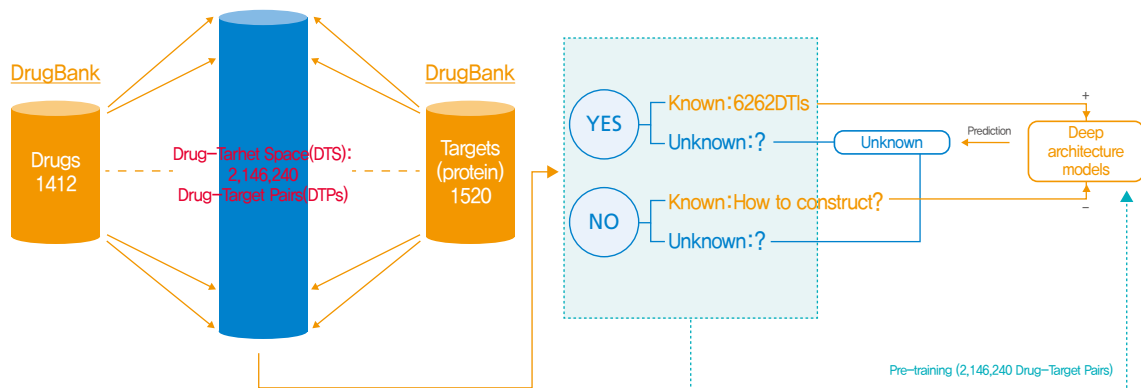
2 제한된 볼츠만 머신 (Restricted Boltzmann machine, RBM) / 심층 신뢰 신경망 (Deep belief networks, DBN)



제한된 볼츠만 머신은 입력층과 하나의 은닉층으로 구성된 비감독학습 (Unsupervised learning) 신경망 모델로서 차원 감소, 분류, 선형 회귀 분석, 특성 학습 등에 사용되고 있다. 제한된 볼츠만 머신은 2개의 층으로만 구성되어 있기 때문에 심층 신경망은 아니고 심층 신뢰 신경망을 구성하는 요소로 사용되기 때문에 여기에서 먼저 간략한 설명을 하고자 한다. 제한된 볼츠만 머신의 모든 입력 층 노드는 은닉 층 노드와 연결되어 있고, 모든 은닉 층 노드도 입력 층의 노드와 연결되어 있다. 그러나 같은 층에 있는 노드끼리 전혀 연결되어 있지 않다. 제한된 볼츠만 머신에서는 같은 층의 연결이 전혀 없는 이 구조 때문에 “제한된”이라는 수식어를 가지게 되었다. 제한된 볼츠만 머신은 다수의 출력 노드를 가지고 출력이 감독학습 신호를 받지 않으며 입력 층 노드에서 입력 받은 데이터를 은닉층에 전달할 때 확률에 따라서 전달 여부를 결정한다. 이것이 제한된 볼츠만 머신의 매우 큰 특징이라고 할 수 있다.

심층 신뢰 신경망은 앞서 설명한 제한된 볼츠만 머신을 여러 개의 층으로 쌓은 심층 신경망 구조이다. 심층 신경망에서는 학습 과정에서 출력층에서는 오차가 잘 교정되나 입력 층 쪽으로 이동함에 따라 오차가 적어져서 입력층에 대한 가중치 교정이 잘 이루어지지 않는 오차 소멸(Vanishing gradient) 문제가 발생한다. 이러한 문제를 해결하기 위해서 심층 신뢰 신경망은 입력층에 가까운 층부터 출력 층 쪽으로 순차적으로 선행 훈련을 수행한다. 입력층에 가까운 층부터 가중치를 먼저 학습한 후 이 가중치가 고정된 형태로 다음 층의 가중치를 학습하는 방법이다. 이 기술이 층별 선행훈련 (Layerwise pre-training)이라고 하며 딥러닝 학습에서 핵심 기술로 제안되었다. 심층 신뢰 신경망은 크게 입력과 같은 출력을 재생성 하는 오토인코더(Autoencoder) 방법과 분류기(Classifier)로 사용되고 있다. 오토인코더는 비감독 학습이며, 분류기는 감독학습에 해당된다. 일반적으로 심층 신뢰 신경망은 비감독학습 방법이지만, 분류 문제를 해결하기 위해서 비감독학습인 마지막

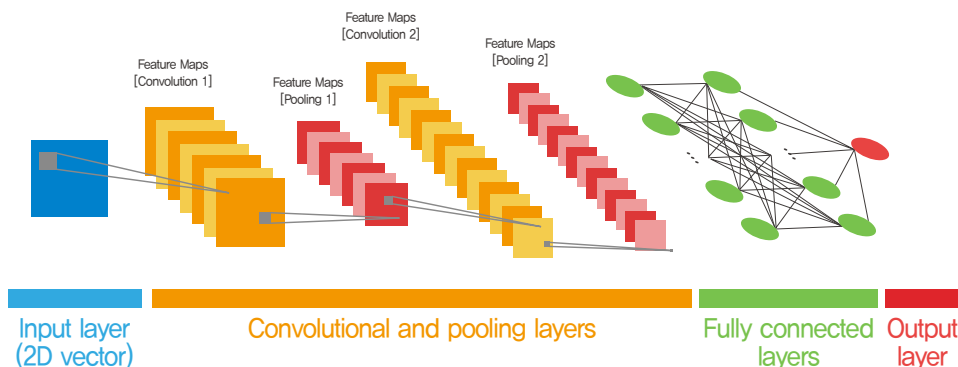
은닉층을 입력층으로 하는 출력층을 추가로 만들어 주고 난 뒤 감독학습을 수행하여 분류 문제 해결에 사용할 수 있다. 2017년 Wen 연구팀은 앞서 언급한 심층 신뢰 신경망을 분류 방식으로 사용한 것이 대표적인 심층 신뢰 신경망의 이용 사례이다. 해당 연구진은 1,412개 약물과 1,520개 단백질 타깃 간의 214만개 이상의 약물-타깃 쌍으로 선훈련을 수행한 후, 알려진 6,262 약물-타깃 상호작용을 훈련하여 나머지 알려지지 않은 것들에 대하여 예측결과를 보고하였다.[21]



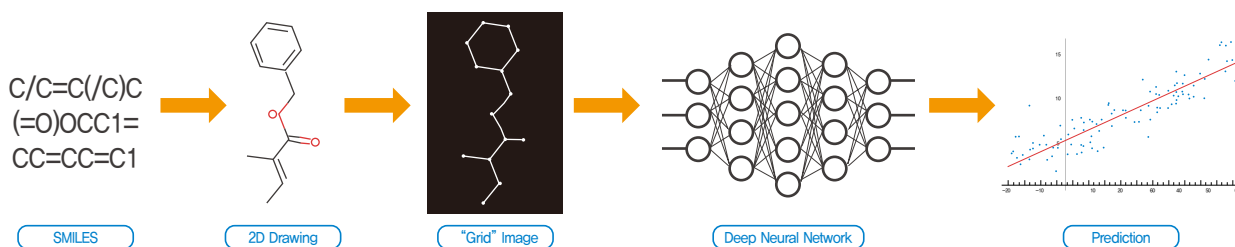
< Wen 연구팀이 개발한 DeepDTI 모델 순서도 >

3 합성곱 신경망 (Convolutional neural networks, CNN)

Convolutional Neural Network (CNN)

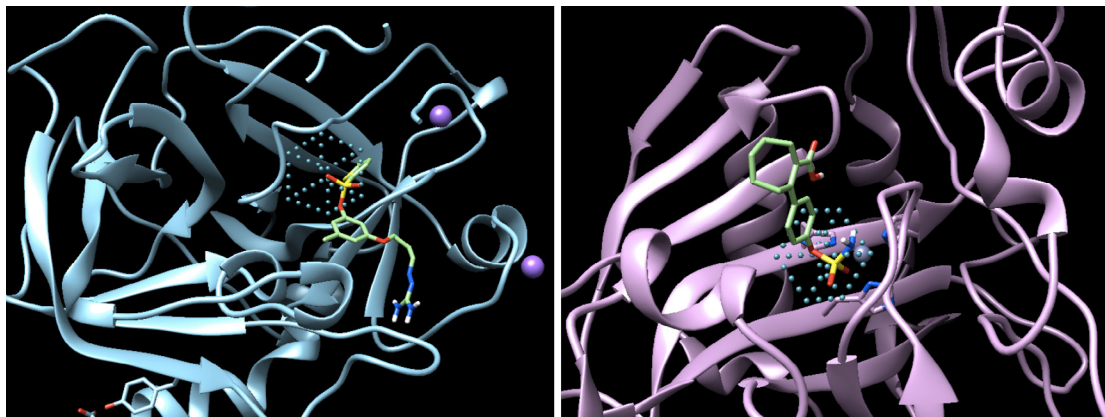


합성곱 신경망은 최소한의 전처리를 사용하도록 설계된 다층 신경망의 한 종류로서 하나 또는 다중 합성곱 계층과 그 뒤로 이어진 일반적인 인공신경망 계층들로 이루어져 있으며, 가중치와 통합 계층(Pooling layer)들을 추가하여 활용한다. 다른 심층 신경망 구조와 비교해 2차원 구조의 입력데이터를 처리하는 것에 큰 장점을 가지고 있어 이미지, 영상, 음성 분야 모두에서 좋은 성능을 보여주고 있다. 합성곱 신경망이 이미지 인식에서 매우 뛰어난 성능을 보여준다는 점에서 “화합물 구조 그림에서도 화합물의 물성을 예측할 수 있을까?”라는 호기심은 2017년 Goh 연구팀이 SMILES 코드로부터 생성된 2차원 화합물 이미지를 직접 합성곱 신경망에 적용하여 HIV 활성화도 예측모델 개발한 시도로 충분한 가능성을 보여주었다.[22]



< Goh 연구팀이 개발한 Chemception framework의 모식도 >

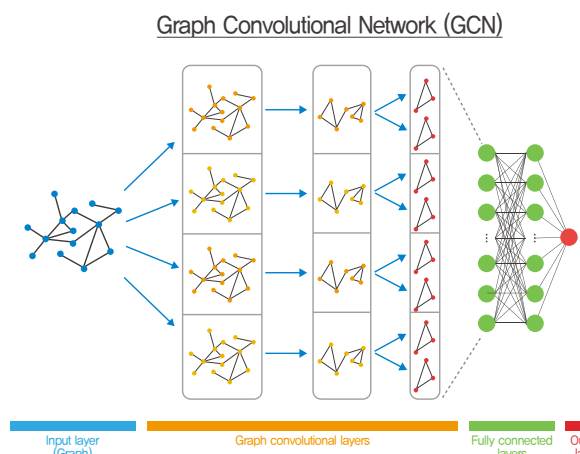
미국 실리콘밸리에 본사를 두고 있는 AtomWise 스타트업이 개발한 AtomNet은 분자 구조와 단백질 Binding site의 각 3차원 정보를 합성곱 신경망에 적용하여 정확도 증가한 결과와 약물-타겟의 결합을 시각적 결과와 함께 같이 보여줄 수 있었다.[23] 이를 계기로 AtomWise는 다국적 제약회사인 Merck와 말라리아 치료제 개발과 관련된 인공지능 신약개발 업무협약을 체결하였고, 현재 해당 인공지능 모델을 활용하여 에볼라에 효과가 있는 신약 후보물질 2개를 발견한 것으로 알려져 있다.



< AtomWise의 AtomNet을 이용한 가상탐색 결과 >

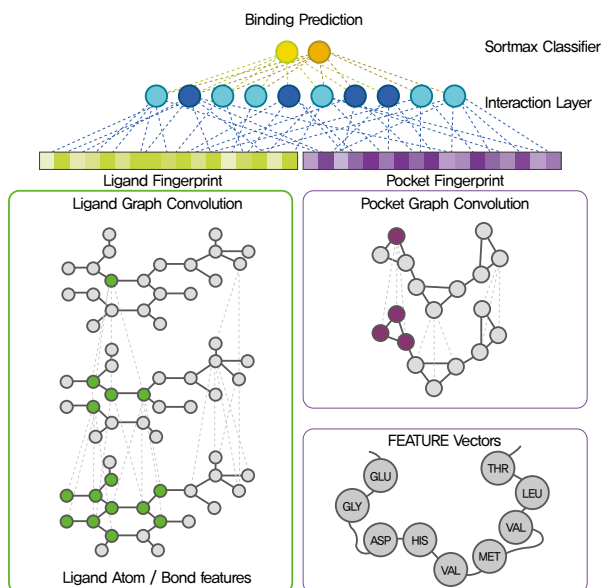
이렇듯 합성곱 신경망은 최근 신약개발 가상탐색 영역에서 가장 많이 사용되고 있는 심층 신경망 구조 중의 하나라고 할 수 있다. 또한 2차원 구조의 입력 값을 처리한다는 점에서 화합물 구조를 2차원 그래프로 표현한 입력 값으로 사용하는 그래프 합성곱 신경망 모델의 기초가 되는 구조라고 할 수 있다.

4 그래프 합성곱 신경망 (Graph convolutional networks, GCN)



그래프 합성곱 신경망은 그래프 표현 입력 값에 대해서 합성곱 신경망을 적용한 구조라고 할 수 있다. 해당 심층 신경망 구조는 넓게 보면 합성곱 신경망 카테고리에 속한다고 볼 수 있지만, 현재 합성곱 구조에 그래프를 입력 값으로 사용하는 연구 사례가 크게 증가하고 있으며, 표현 자유도가 높은 2차원 그래프 표현 방법이 학습 능력이 뛰어난 합성곱 신경망 구조와 결합하여 굉장히 좋은 예측 성능을 보여주는 사례가 증가하고 있다. 따라서 높은 자유도와 고성능 학습능력이 합쳐진 그래프 합성곱 신경망 구조는 인공지능 신약개발 분야에서 하나의 중요한 예측 모델 개발방법으로 자리 잡아 갈 것이 분명하다고 판단된다. 그래프 이론은 객체 간에 짝을 이루는 관계를 설명하기 위해서 사용되는 수학 구조인 그래프에 대한 연구를 다루는데, 여기서 그래프는 꼭지점(Vertex)과 변(Edge)으로 이루어져 있다. 이를 화합물 구조 표현에 사용한 것이 분자

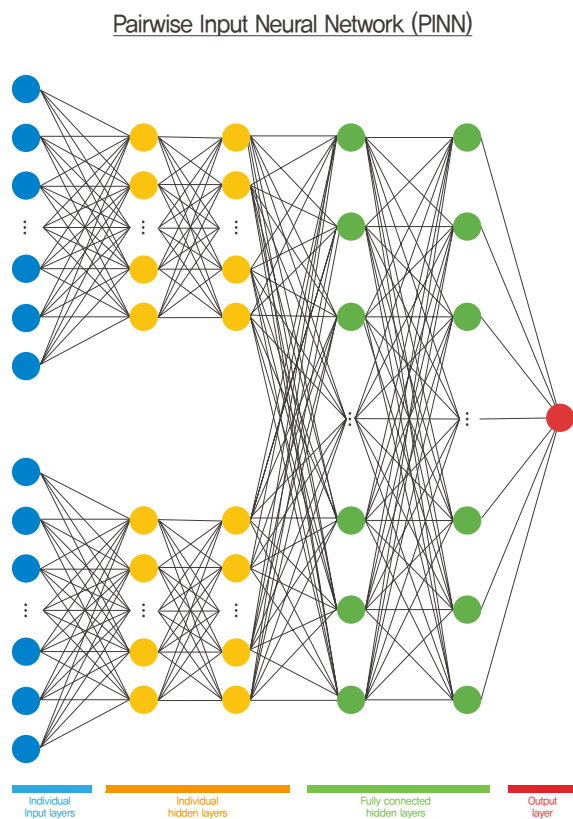
그래프(Molecular graph)라고 말하는데, 분자 그래프는 Arthur Cayley가 그래프라는 단어가 도입되기 전인 1874년에 이미 분자 그래프를 연구결과로 발표한 것으로 알려져 있다.[24] 이처럼 분자 구조를 그래프로 표현하는 것은 이미 오래전에 시작되어 왔지만, 딥러닝 기술이 개발되기 전까지는 표현 자유도 높은 분자 그래프와 화합물 물성의 관계를 설명하기에 좋은 성능을 보인 예측모델을 만들기 힘들었다.



< Torng 연구팀이 개발한 2단계 그래프 입력구조의
그래프 합성곱 신경망 모델 >

하지만, 딥러닝 기술이 개발되고 전세계로 확산되면서, 최근에는 벌써 화합물 구조를 그래프로 표현하여 입력 값으로 사용하는 인공지능망 모델들이 다양한 화합물 물성예측에서 굉장히 좋은 성능을 보여준다는 것이 다양한 연구진들을 통해 보고되었다. 이렇게 좋은 성능의 화합물 물성 예측의 그래프 신경망의 성공적인 연구결과들은 신약개발 분야의 약물-타겟 상호작용 모델 개발에도 크게 영향을 주게 되어 약물-타겟 상호작용 예측에 다양한 그래프 신경망 모델들이 보고되었다. 2017년 Gomes 연구팀은 원자 타입과 각 원자의 거리를 그래프 합성곱 신경망에 적용한 예측모델을 개발하였다.[25] 2018년 Feinberg 연구팀은 원자들의 유클리드 거리를 기반으로 그래프 합성곱 필터를 적용한 모델을 보고하였고,[26] Torng 연구팀은 약물-타겟 상호작용을 예측하기 위해 리간드와 단백질 포켓의 2단계 그래프 입력 구조를 가진 합성곱 구조를 이용하여 그래프 합성곱 신경망 모델들을 개발하였다.[27]

5 쌍방 입력 신경망 (Pairwise Input Neural Networks, PINN)

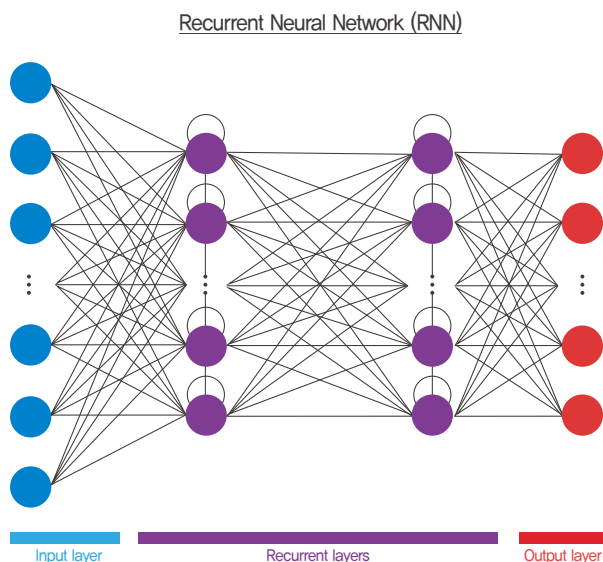


쌍방 입력 신경망은 2개의 다른 입력층과 다중 은닉층을 심층 인공지능망 앞부분에 가지고 있는 것이 전방 전달 심층 신경망과 다른 점이며, 마지막 출력 층(Output layer)에는 보통 하나의 출력 노드(Output neuron)를 구성하여 하나의 단백질 타겟에 대한 예측 모델을 만드는데 사용된다. 쌍방 입력 신경망은 약물-타겟 상호작용(Drug-Target Interaction, DTI)과 같은 양쪽방향 상호작용(Pairwise interaction)을 표현하는데 매우 좋은 학습방법으로 2개의 서로 다른 입력층에는 화합물과 단백질 타겟의 특징들이 분리되어 들어간다.

2014년 Wang 연구팀은 단백질 결합 사이트들(Binding sites)과 2차원 구조 Fingerprints를 이용하여 약물-표적 상호작용 모델을 개발했다.[28] 2016년에 Wan과 Zeng은 결합 사이트 대신에 단백질 아미노산 서열을 3개를 한쌍으로 읽어 들이는 방법으로 쌍방 입력 신경망에 적용했으며,[29] 2017년 Lenselink 연구팀은 리간드(Ligand)의 특징을 물리화학적 물성으로 단백질 타겟의 특징은 Morgan fingerprints로 사용하여 쌍방 입력 신경망 모델에 적용하여 인공지능망 모델들을 개발하였다.[30]

최근에 보고되는 심층 신경망 모델들을 살펴보면, 여기에서 분류해 놓은 단순한 기준의 한 가지에 속하지 않고 여러 신경망 구조들이 서로 복합적이고 융합적인 구조로 변형되어 사용되어가는 심층 신경망을 이용한 인공지능 모델들을 개발하려는 노력들이 많다는 점을 살펴볼 수 있다.

6 순환 신경망 (Recurrent Neural Networks, RNN)



순환 신경망은 유닛간의 연결이 순환적 구조를 가지는 것이 매우 큰 특징이다. 이러한 구조는 시계열 특징을 모델링 할 수 있도록 신경망 내부에 상태를 저장할 수 있게 해준다. 전방 전달 신경망과 달리 순환 신경망은 내부의 메모리를 이용해 서열 형태의 입력을 처리할 수 있다. 따라서 순환 신경망은 필기체, 음성 인식과 같은 시계열 특징을 가지는 데이터 처리에 사용되고 있다.

2017년 Goh 연구진들은 SMILES 문자열 서열을 순환 신경망에 적용하여 화합물 물성 예측모델을 만드는 방법을 소개한 바 있다.[31] 아직 신약개발 연구에서 단백질 서열에 대해서 적용되어 사용된 사례를 찾아보기 힘들었지만, 생명정보학이나 신약개발 분야에서는 유전자, 단백질 서열 등을 처리하는데 사용될 수 있을 것으로 기대되고 있다.

지금까지 신약개발연구에서 사용되어지고 있는 대표적인 딥러닝 기술과 사례들에 대해서 살펴보았다. 현재 딥러닝 기술은 우수한 학습 능력을 보여주면서, 기존의 기계학습 방법들보다 우월한 예측 성능을 보여주는 사례들이 발표되고 있다. 앞으로 신약개발 연구분야 문제해결을 위한 역할을 수행할 것으로 기대한다.

References

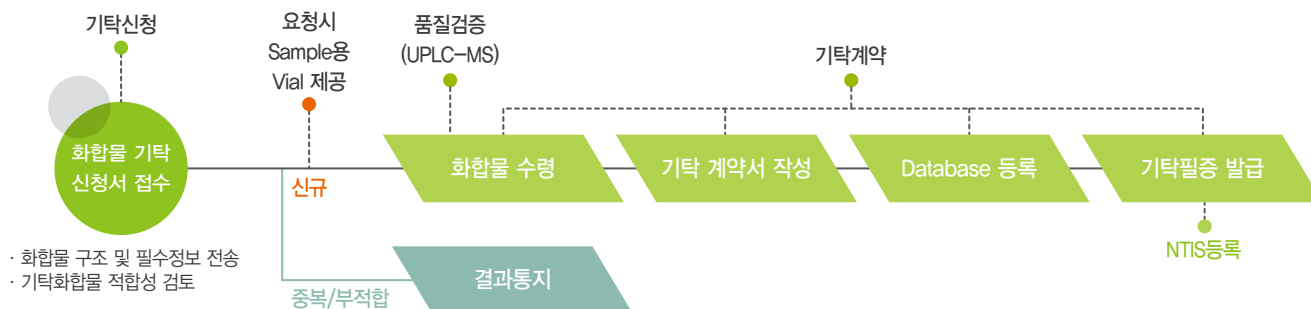
- [1] Morris GM, Goodsell DS, Halliday RS, et al. AutoDock-related material automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Comput Chem J Comput Chem* 1998;19(14):1639–62.
- [2] Ewing TJ, Makino S, Skillman AG, et al. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 2001;15: 411–28.
- [3] Friesner RA, Banks JL, Murphy RB, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;47(7): 1739–49.
- [4] Verdonk ML, Cole JC, Hartshorn MJ, et al. Improved protein-ligand docking using gold. *Proteins Struct Funct Bioinforma* 2003;623:609–23.
- [5] Kramer B, Rarey M, Lengauer T. Evaluation of the FlexX incremental construction algorithm for protein- ligand docking. *Proteins Struct Funct Genet* 1999;37:228–41.
- [6] McGann M. FRED pose prediction and virtual screening accuracy. *J Chem Inf Model* 2011;51(3):578–96.
- [7] Mauri a, Consonni V, Pavan M, et al. Dragon software: an easy approach to molecular descriptor calculations. *MATCH Commun Math Comput Chem* 2006;56:237–48.
- [8] RDKit: Open-Source Cheminformatics Software, 2019. <http://www.rdkit.org/> (30 June 2019, date last accessed).
- [9] PreADMET, 2019. <https://preadmet.bmdrc.kr/> (30 June 2019, date last accessed).
- [10] Yap, Chun Wei. "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints." *Journal of computational chemistry* 32 7 (2011): 1466–74 .
- [11] O'Boyle NM, Banck M, James CA, et al. Open Babel: an open chemical toolbox. *J Cheminform* 2011;3:33.

- [12] Daylight Toolkit, 2018. <http://www.daylight.com/products/toolkit.html> (30 June 2019, date last accessed).
- [13] The Chemistry Development Kit, 2018. <https://cdk.github.io/> (30 June 2019, date last accessed).
- [14] OpenEye Toolkit, 2018. <https://docs.eyesopen.com/toolkits/python/index.html> (30 June 2019, date last accessed).
- [15] Cao Y, Charisi A, Cheng LC, et al. ChemmineR: a compound mining framework for R. *Bioinformatics* 2008;24(15):1733–4.
- [16] Indigo Toolkit, 2018. <http://lifescience.opensource.epam.com/indigo/> (30 June 2019, date last accessed).
- [17] Unterthiner T, Mayr A, Klambauer G, et al. Deep learning as an opportunity in virtual screening. *Deep learn Represent Learn Work NIPS 2014*;2014:1–9.
- [18] Ma J, Sheridan RP, Liaw A, et al. Deep neural nets as a method for quantitative structure—activity relationships. *J Chem Inf Model* 2015;55(2):263–74.
- [19] Ramsundar B, Kearnes S, Edu K, et al. Massively Multitask Networks for Drug Discovery. *arXiv* 2015; arXiv: 1502.02072.
- [20] Koutsoukas A, Monaghan KJ, Li X, et al. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *J Cheminform* 2017;9:42.
- [21] Wen M, Zhang Z, Niu S, et al. Deep-learning-based drug-target interaction prediction. *J Proteome Res* 2017;16(4):1401–9.
- [22] Goh GB, Hodas N, Siegel C, et al. SMILES2vec: an interpretable general-purpose deep neural network for predicting chemical properties. *arXiv* 2017; arXiv:1712.02034.
- [23] Wallach I, Dzamba M, Heifets A. AtomNet: a deep convolutional neural network for bioactivity prediction in structurebased drug discovery. *arXiv* 2015;1510:1–11.
- [24] A. Cayley, *Phil. Mag.*, 1874, 47, 444–446, as quoted in N. L. Biggs, E. K. Lloyd and R. J. Wilson, "Graph Theory 1736–1936", Clarendon Press, Oxford, 1976; Oxford University Press, 1986, ISBN 0-19-853916-9
- [25] Gomes, J.; Ramsundar, B.; Feinberg, E. N.; Pande, V. S. *arXiv preprint arXiv:1703.10603* 2017
- [26] Feinberg, E. N.; Sur, D.; Wu, Z.; Husic, B. E.; Mai, H.; Li, Y.; Sun, S.; Yang, J.; Ramsundar, B.; Pande, V. S. *arXiv preprint arXiv:1803.04465* 2018,
- [27] Torng, W.; Altman, R. B. *bioRxiv preprint bioRxiv 473074* 2018,
- [28] Wang C, Liu J, Luo F, et al. Pairwise input neural network for target-ligand interaction prediction. In: *Proceedings of 2014 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2014, New Jersey: IEEE, 2014, 67–70.*
- [29] Wan F, Zeng J. Deep learning with feature embedding for compound-protein interaction prediction. *bioRxiv* 2016; doi:10.1101/086033.
- [30] Lenselink EB, Ten Dijke N, et al. Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set John Holliday. *J Cheminform* 2017;9:1–14.
- [31] Goh GB, Hodas N, Siegel C, et al. SMILES2vec: an interpretable general-purpose deep neural network for predicting chemical properties. *arXiv* 2017; arXiv:1712.02034.

화합물 기탁 절차

문의

E-mail. chembank@kriech.re.kr Tel. 042-860-7190 Web. http://www.chembank.org



유기합성 화합물 및 단일성분 천연물

- 범위 유기합성 화합물 및 단일성분 천연물

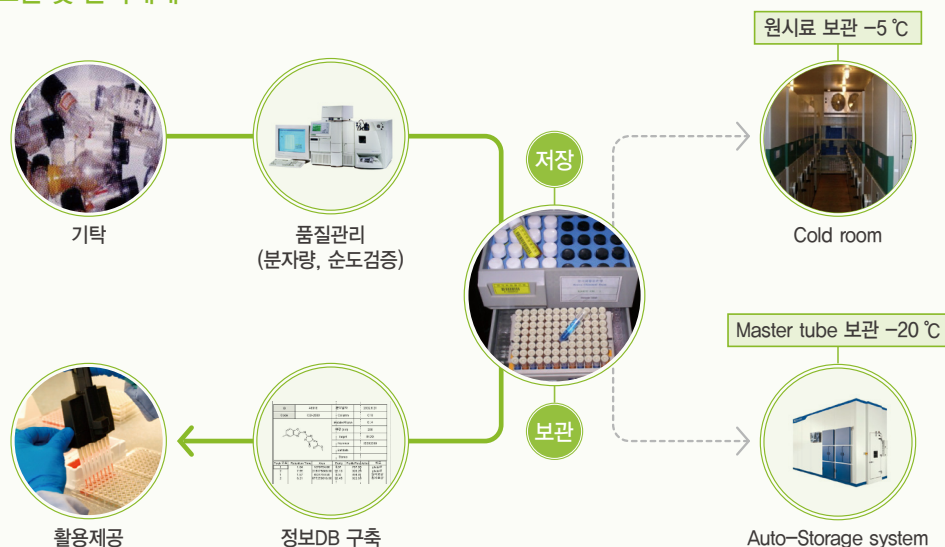
(기탁화합물의 활성여부와 상관 없음. 최종 화합물 뿐만 아니라 합성 중간체도 기탁 가능)

- 품질기준 화합물 양 >10mg(권장), 순도 >80%(한국화학물은행에서 UPLC-MS로 검증), 구조를 확인할 수 있는 NMR 자료 제출

기탁자 혜택

- 기탁화합물에 대한 활용 결과 도출시 그 결과를 통보 받을 수 있다.
- 기탁화합물의 활용결과 Hit이 도출된 경우
 - 활용자와 협의하여 후속연구에 공동연구자로 참여할 수 있다.
 - 활용결과를 논문 또는 특허로 출판할 경우: 기탁자의 기여도에 따라 공동저자 또는 공동출원인으로 참여 할 수 있다.
 - 기탁화합물의 활용결과 수익이 발생할 경우: 기탁자의 기여도에 따라 수익의 일부를 분배 받을 수 있다.

화합물 보관 및 관리체계



한국화학물은행 보유 화합물 품질관리

- 제공화합물의 품질관리가 활용자들의 실험결과에 큰 영향을 미칠 수 있으므로 한국화학물은행에서는 보유 화합물의 품질을 최상으로 유지하기 위하여 노력하고 있다.
- 장기 보관에 따른 화합물의 품질 저하를 최소화하기 위하여 원시료(순수) 화합물은 -5 °C 저온창고에 보관하며, 반출 요청에 신속하게 대응하기 위하여 소량의 화합물을 DMSO에 녹인 마스터 튜브(5 mM)를 제작하여 -20 °C에 보관하고 있다.

반입, 보관, 반출 단계별 화합물 품질 관리

1) 반입 화합물 품질 관리

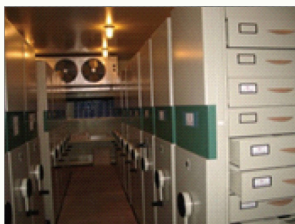
반입되는 모든 화합물은 UPLC-MS 분석을 통하여 분자량 및 순도를 확인하고 있으며, UPLC-MS 분석으로 분자량 확인이 되지 않거나 순도가 80% 이하인 화합물은 화합물과 함께 제출된 ¹H-NMR 자료를 검토하여 화합물 구조 및 순도(80% 이상)가 확인되는 화합물만 기탁받고 있다.

2) 보관 화합물 품질관리

화합물을 장기 보관함에 따라 필연적으로 생길 수 있는 화합물의 품질 저하를 체크하기 위하여 매년 보유 화합물의 일부 (약 20,000종)를 선별하여 UPLC-MS로 품질 검증을 하고 있으며 순도가 80% 이하인 화합물은 반출을 중지하고 별도로 관리하고 있다.

3) 반출 화합물 품질관리

화합물이 반출된 후 약효시험을 거쳐 Hit 화합물로 확인된 경우, 다음 단계의 실험을 진행하기 전에 Hit 화합물의 품질을 UPLC-MS로 다시 분석하여 그 결과를 활용자에게 제공하고 있다.



Cold room (-5 °C)



Auto-Storage (-20 °C)
(Brooks SS- II)



Barcoded
Tubes



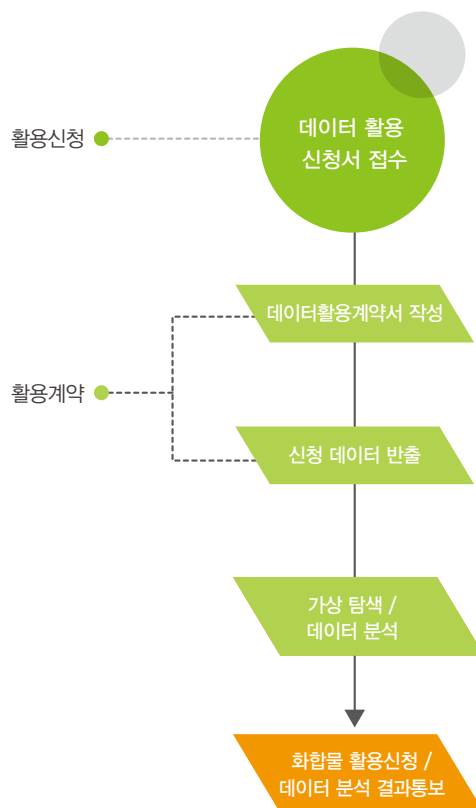
UPLC-MS
Waters Acquity UPLS-SQD2

한국화학물은행은 보유 화합물의 품질을 최상으로 유지하기 위하여 노력하고 있습니다.

데이터 활용 절차

문의

E-mail. chembank@kricr.re.kr Tel. 042-860-7190 Web. http://www.chembank.org



정보제공 데이터 종류

분류	소분류
화합물 구조 정보	구매화합물 라이브러리
	임상 화합물
	Focused 라이브러리
	GPCR 화합물
	PPI 화합물
	반출 가능한 라이브러리
약효시험결과 정보	공개 가능한 정보

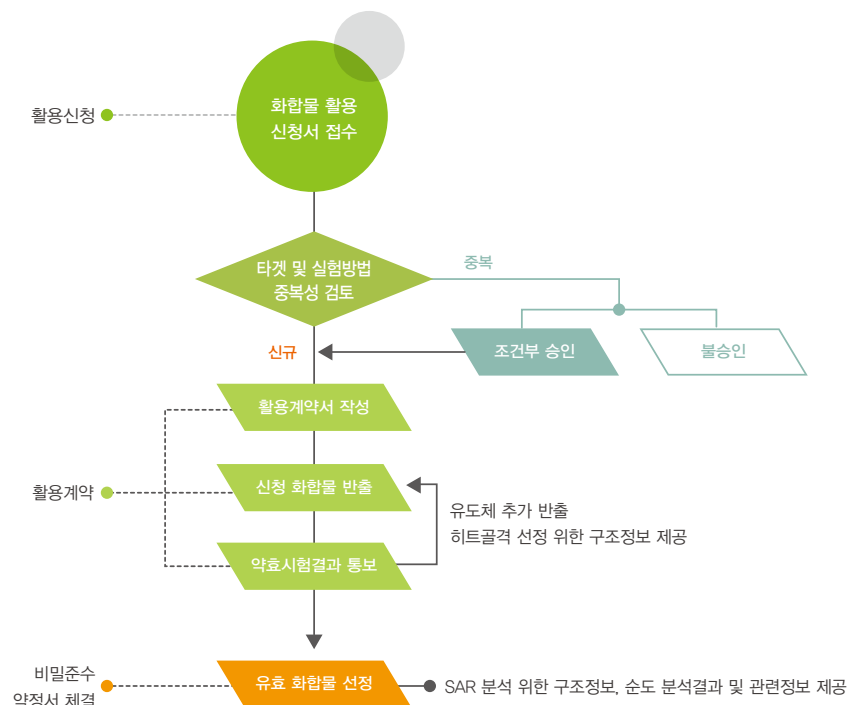
* 데이터 정보는 USB로 제공됩니다.

데이터 활용 계약의 종료 시, 반드시 데이터의 폐기 및 해당 USB를 반납하여야 합니다.

화합물 활용 절차

문의

E-mail. chembank@krcit.re.kr Tel. 042-860-7190 Web. http://www.chembank.org



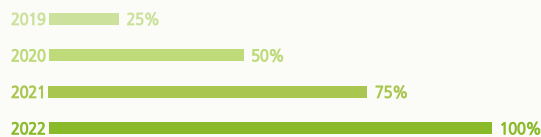
화합물 사용료 규정

- 화합물은 무상으로 제공합니다.
- 화합물 제공에 수반되는 실비용(plate, 분주기 tip, 튜브 cap, 택배 비용 등)은 사용자 부담합니다.(2019년 1월부터 부과)
- 중소 / 벤처 / 대학교 / 연구원은 실비의 50%만 활용자가 부담합니다.

화합물 반출실비 부과

한국화학물은행 단계적 화합물 반출 실비 부과 계획(2019년 1월 시행)

- 2019년 1월1일 부터 실비의 25% 활용자 부담
- 2020년 1월1일 부터 실비의 50% 활용자 부담
- 2021년 1월1일 부터 실비의 75% 활용자 부담
- 2022년 1월1일 부터 실비의 100% 활용자 부담



2019년 기준

* 부가세 별도

Plate 종류	Plate 개수	단가	대기업 / 중견기업	중소 / 벤처 / 대학교 / 연구원
96-well (80종 / 1Plate)	1	23,300원	5,800원	2,900원
384-well (320종 / 1Plate)	1	89,300원	22,300원	11,150원

* 2019년 대표화합물(6,948종) 중소 / 벤처 / 대학교 / 연구원 기준

- 96-well : 252,300원 | • 384-well : 245,300원

한국화학물은행 제공 화합물의 활용결과 (논문,특허 등)에 대한 권리관계 규정

I 규정목적

- 한국화학물은행은 기탁자들이 기탁한 화합물을 바탕으로 운영되고 있습니다.
- 기탁자들의 화합물 기탁을 장려하기 위하여 기탁자들에게 최소한의 혜택을 드리며, 동시에 화합물 사용자들의 불편함을 최소화하기 위하여, 한국화학물은행이 제공한 화합물을 활용하여 도출된 연구결과(논문, 특허 등)의 권리관계에 대하여 아래와 같은 규정을 적용합니다.

I 규정내용

• 활용결과(논문, 특허 등) 권리관계 규정 내용

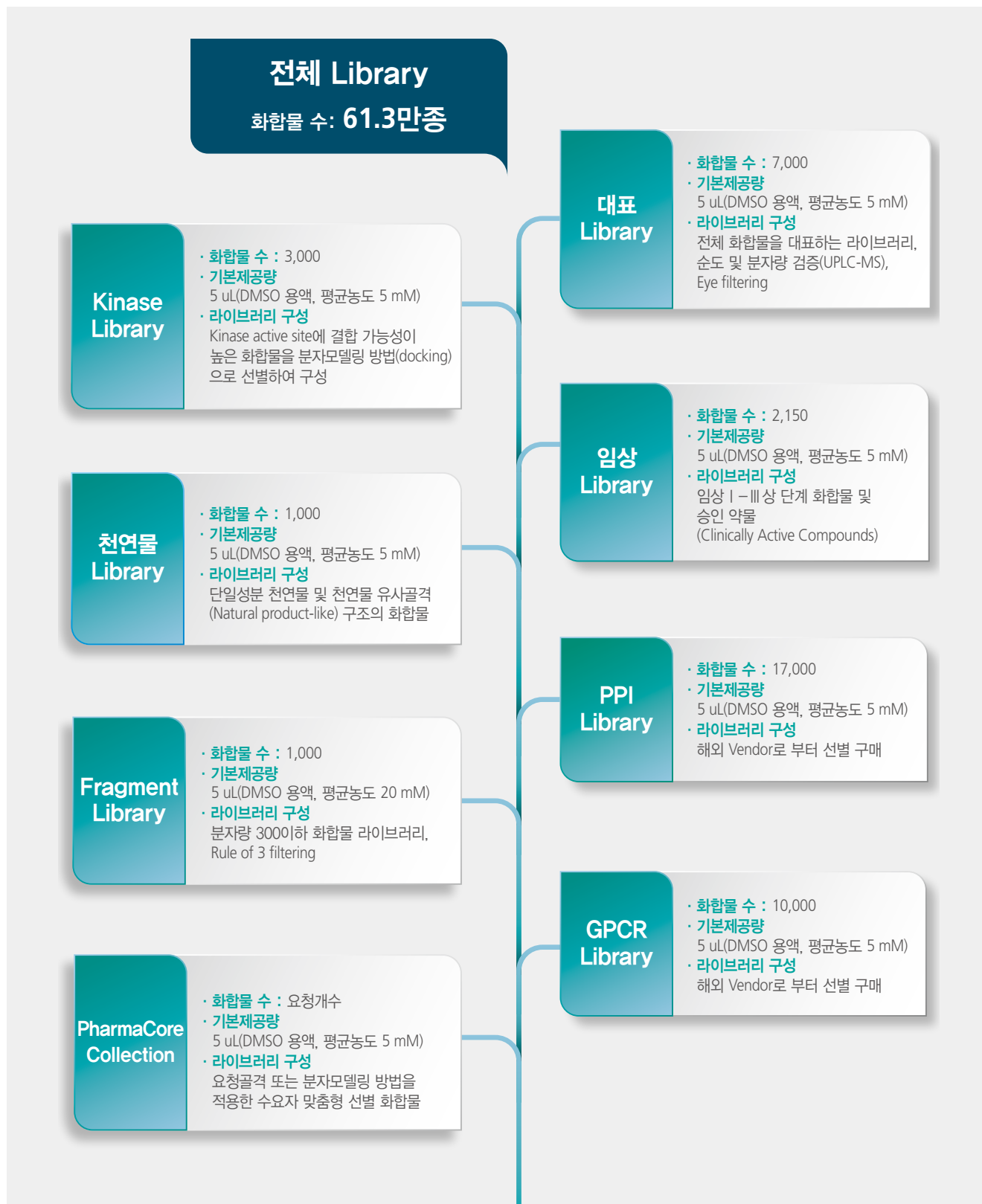
1. 기탁자가 단순기탁 이외의 추가적인 기여가 없는 경우에는 화합물 기탁자로서 논문의 사사(acknowledgement)에 포함되는 것이 "**원칙**"입니다.
 2. 기탁자가 추가적인 기여(유도체 합성 제공, 관련 정보제공 등)를 하였을 경우에는 기여정도에 따라 기탁자를 논문 공저자 및 특허 공동 발명인(공동출원인)으로 "**포함**"하여야 합니다.
 3. 기탁자의 추가적인 기여가 없는 경우, 활용결과에 대한 "**용도특허**"는 사용자(발견자)에게 귀속되며 "**물질특허**"에 관해서는 기탁자와 협의하여야 합니다.
- * 모든 논문 및 학회 발표에는 한국화학물은행 제공 화합물을 사용하여 연구가 진행되었다는 사사(acknowledgement)를 포함하여야 합니다.

한국화학물은행 활용결과의 권리 규정은 과학계에서 등용되는 연구결과 기여도에 대한 "연구윤리 기본원칙"을 따르고 있습니다.

후속연구를 위한 화합물 추가합성 및 구매 진행 안내

1. 한국화학물은행의 역할은 활용자의 요청에 따라 primary screening에 필요한 화합물을 제공하고 활성을 보이는 경우, 해당 화합물의 구조와 관련정보를 보내드리는 것 입니다.
한국화학물은행은 소량의 화합물만 기탁 받아서 보유하고 있기 때문에 확인 실험에 필요한 정도의 화합물 양까지는 제공할 수 있지만 진단단계(*in vivo* 실험 및 약물성/독성 시험 등)의 실험에 지원할 정도의 화합물량은 보유하고 있지 않습니다.
2. 진단단계의 실험을 위하여 화합물이 추가적으로 더 필요한 경우에는 아래의 방법을 이용하시기 바랍니다.
 - 1) 구매 가능한 화합물은 구매하여 사용(구매처 정보제공)
 - 2) 화합물 원기탁자와 협의하여 공동연구 진행
 - 3) 합성전문가를 물색하여 공동연구 진행
3. 한국화학물은행(KCB) 활용 화합물 중에서 상용 화합물의 구매를 원하시는 경우, KCB는 구매 가능한 vendor 정보만 제공해 드립니다.
개정된 화관법, 화평법 규정이 강화되고 있으므로, **화학물질의 해외 구매 시 신고 및 등록 의무를 사전에 반드시 확인**바랍니다.
(화학물질정보처리시스템, <http://kreachportal.me.go.kr>)

한국화학물은행 제공 라이브러리 종류



공동 활용을 통한 화합물 및 활용데이터 가치 재창출

We take care of your Compounds and Create new value!



화합물 기탁 · 활용 · 데이터 문의

- 홈페이지 : <http://www.chembank.org>
- 주소 : 대전광역시 유성구 가정로 141 한국화학연구원 한국화합물은행 E2연구동 (우.34114)
- 전화(대표) 042-860-7190 / 팩스 042-860-7096 / E-mail chembank@krcr.ac.kr

이름	담당업무	전화	E-mail
이선경	사업책임자	042-860-7148	leesk@krcr.ac.kr
이현규	의약화학	042-860-7016	leehk@krcr.ac.kr
이윤호	분자모델링 / 화학정보학	042-860-7453	yunolee1@krcr.ac.kr
강영목	화합물은행DB / 화학정보학	042-860-7015	ymkang@krcr.ac.kr
황순희	대외협력 / 정보관리	042-860-7190	chembank@krcr.ac.kr
김선우	화합물관리 / 장비관리	042-860-7171	swkim@krcr.ac.kr
김선호	화합물관리 / 저온창고관리	042-860-7090	shkim@krcr.ac.kr
이유리	화합물정보 / DB관리	042-860-7092	yurilee@krcr.ac.kr
이수연	LC-MS, 물성분석	042-860-7747	suyoun@krcr.ac.kr
박희정	LC-MS, 물성분석	042-860-7070	huijeong@krcr.ac.kr



과학기술정보통신부 지정

연구성과 관리·유통 전담기관
- 화합물 전담기관 -

한국화학물은행은 국가연구개발사업 수행을 통해 창출된 연구성과 중 화합물 및 관련 정보의 관리·유통 전담기관으로 지정되어 있습니다.

[과학기술정보통신부고시 제2017-7호]

No.02

KCB newsletter

한국화학물은행
뉴스레터

2019년 제2호



한국화학물은행

Korea Chemical Bank

대전광역시 유성구 가정로 141 한국화학연구원 한국화학물은행 (우.34114)

Tel : 042-860-7190 Fax : 042-860-7096 E-mail : chembank@krikt.re.kr Web : <http://www.chembank.org>