



한국화합물은행 소식

이와
같이
한
과
정
이
다

한양대학교
한양대학교
한양대학교

빅데이터 소개

화합물 기탁 절차

화합물 활용 절차

활용결과 권리 규정

제공 라이브러리 종류

한국화합물은행

Korea Chemical Bank newsletter

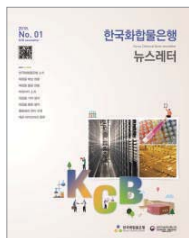
뉴스레터



2018. No. 01

KCB newsletter

한국화학물은행 뉴스레터
2018년 제1호



[한국화학물은행 뉴스레터]

발행일 2018년 1월

발행인 김성수

편집인 이현규

발행처 한국화학물은행

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

[과학기술정보통신부 고시 제2017-7호]

Contents

한국화학물은행 소식

03 대외활동

03 주요방문인사

04 화합물 확보 현황

05 화합물 활용 현황

06 약효시험 활용 현황

07 화합물 활용성과

11 [빅데이터 소개]

신약개발관련 화학정보학 데이터베이스 소개 (I)

22 화합물 기탁 절차

23 한국화학물은행 보유 화합물 품질관리

24 화합물 활용 절차

25 활용결과 권리 규정

26 제공 라이브러리 종류

Korea Chemical Bank newsletter_한국화학물은행 소식

KCB newsletter

한국화학물은행 소식

대외활동

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

번호	일정	내용	장소
1	2017-04-25	연구성과물 전달기관 실적 및 성과 발표	서울 양재메타워
2	2017-05-17 ~ 19	생화학분자생물학회 홍보부스 운영	부산 벡스코
3	2017-06-16	R&D 성과관리 및 활용관련 교육 강의	KIRD 대전교육센터
4	2017-07-06 ~ 07	대한화학회 의약화학분과회 하계 워크샵 홍보 광고	동국대학교 만해마을
5	2017-08-16 ~ 19	태국화학물은행 설립을 위한 전문가미팅에 참석하여 한국화학물은행 설립 및 운영에 관한 세미나 발표	태국 방콕
6	2017-08-21 ~ 22	유기화학분과회 하계 워크샵 홍보 광고	문경 STX 리조트
7	2017-10-18 ~ 20	대한화학회 추계총회 및 학술발표회 홍보부스 운영	광주 김대중컨벤션센터
8	2017-12-14	연구소재은행 교육 프로그램 세미나 발표	KTX 대전역 회의실

한국화학물은행 소식

주요방문인사

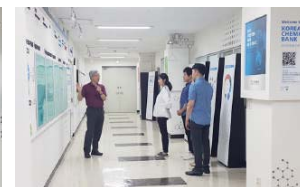
번호	방문일자	기관	방문자
1	2017-02-17	LG화학기술연구원	유진녕사장 외 5명
2	2017-03-24	전국연구소장협의회	손진형회장 외 30명
3	2017-06-08	미국 조지아공과대학교	Prof. Anna Stenport, Prof. Jong Hyun Lee 외 17명
4	2017-06-12	벨기에 VIB	Dr. Els Beirnaert (Senior Manager of New Ventures)
5	2017-06-26	한국과학기술원	강경화교수 외 1명
6	2017-06-29	삼성서울병원	신용재팀장 외 1명
7	2017-08-30	과학기술정보통신부 지역연구진흥과	지운하사무관
8	2017-09-05	스탠다임	송상욱이사
9	2017-07-26	(주)씨앤드씨 신약연구소	이순욱책임연구원 외 2명
10	2017-09-26	질병관리본부	이정아박사 외 3명
11	2017-11-01	사우디아라비아	식약청 장관 외 9명
12	2017-11-03	일본 Niigata University of Pharmacy and Applied Life Sciences	Hiroshi TERADA, Ph.D.(총장), Kouki KITAGAWA Ph.D., Kaori ONITA, MBA
13	2017-11-14	과학기술정보통신부, 한국과학기술기획평가원	윤희수사무관, 이종선연구원 외 1명



〈 사우디아라비아 식약청장 일행 〉



〈 일본 Niigata University 총장 일행 〉



〈 질병관리본부 일행 〉

한국화학물은행 소식

화합물 확보 현황

한국화학물은행은 2017년 12월 말 현재 약 55만종의 화합물을 보유하고 있으며 다양성, 약물성, 특이성이 우수한 고수준 신약소재 화합물 라이브러리를 제공하기 위하여 다양한 방법으로 우수한 화합물을 확보 하고자 노력하고 있습니다.

1. 연구성과물 기탁 [법정의무 기탁]

「국가연구개발사업의 관리 등에 관한 규정」제 25조 13항에 따라 국가연구개발사업 수행을 통해 창출된 화합물은 연구성과물 관리·유통 전담기관인 한국화학연구원 한국화학물은행에 의무적으로 기탁하도록 되어 있습니다. 연구성과물 기탁 효율제고를 위하여 2017년부터 국가연구과제 및 사업평가에 연구성과물 기탁실적을 반영하고, 전담기관에 기탁된 성과만 인정 하도록 “국가연구개발 평가지침”이 개정 되었습니다.

2. 국내 전문가 공모 위탁합성

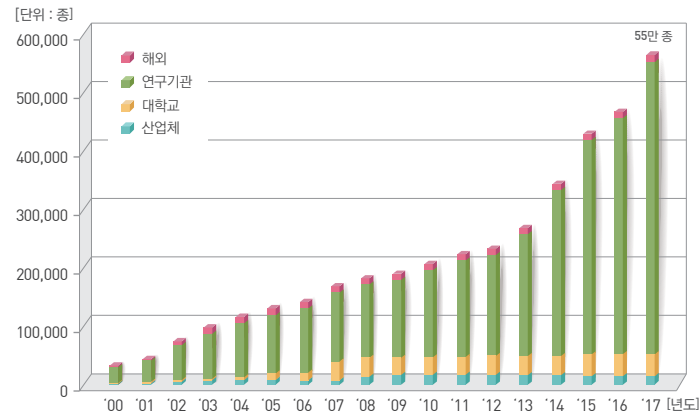
보유화합물의 구조 다양성 확대를 위하여 국내 합성 전문가를 대상으로 위탁 합성 과제를 공모, 선정을 통하여 화합물을 확보하고 있습니다. 전문가 위탁합성은 2014년 15개 과제에서 1,600여종, 2015년 18개 과제에서 2,400여종, 2016년 13개 과제에서 1,600여종, 2017년 14개 과제에서 1,800여종의 화합물을 확보하였습니다. 2018년에도 전문가 위탁합성공모를 통하여 우수한 화합물을 계속 확보할 예정입니다.

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 2017년의 경우 약 89,000여종의 화합물을 선별 구매하여 연구자들에게 제공하고 있습니다.

화합물 확보 현황 (누계)

2017년 12월 말 현재 총 55만종 보유



한국화학물은행 소식

화합물 활용 현황

연 평균 60건 이상의 신규 작용점에 대하여 40만종 이상의 화합물이 국내 산·학·연에 제공되어 활용되고 있으며 현재까지 770건 이상의 특정 작용점에 대한 약효검색에 활용되었습니다. Hit 화합물에 대한 구조 정보 제공시 LC-MS 분석에 의한 순도 및 분자량 검증 자료를 기본적으로 제공하고 있으며 그 외 요청시 유도체 검색, 약물성 예측, off-target 정보 등도 함께 제공하고 있습니다.

2014년 도입한 화합물 선별 자동화 관리 장비[Brooks Sample Store II]를 활용하여 요청 화합물을 신속하게 제공하고 있으며 확보 화합물의 증가에 따른 반출시간 지연에 대응하기 위하여 2017년에 자동화 관리장비[96만종 보관, 관리]를 추가로 도입하여 화합물 활용기관에 신속하게 화합물을 제공하고 있습니다.

최근 4년간 화합물 활용 현황

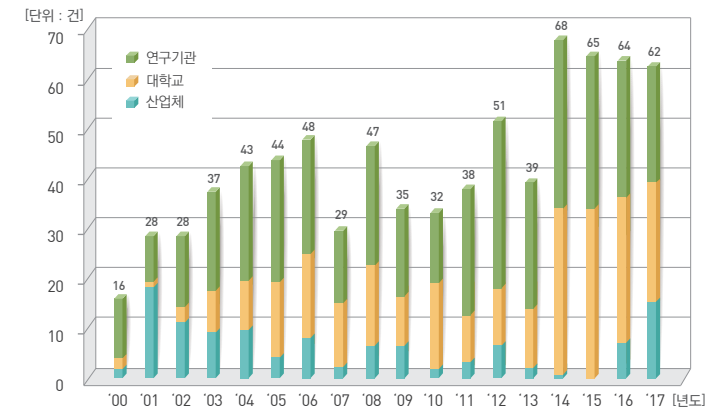
년도	활용과제 건수		반출 횟수		반출화합물 개수	
	전체과제	신규과제	전체과제	신규과제	전체과제	신규과제
2014	122건	68건	291회	161회	414,185종	293,566종
2015	153건	65건	320회	129회	480,677종	225,566종
2016	239건	64건	411회	184회	1,236,527종	1,148,136종
2017	195건	62건	371회	137회	361,561종	181,032종

- 2017년 : 371회에 걸쳐 총 361,561종 화합물 반출 (평균 1,400개/일)
- 2017년 유도체검색(의약화학) 지원 : 122회 (평균 2.3건/week 지원)

화합물 활용 지원의 경제적 가치평가(2017년)

- 36만종 x 4만원 = 144억원 수입대체효과 (Chemdiv 2016년 최저 구매가격 \$36/개)
- 국내 고유 화합물은 구입 불가능 → 가치 산정 어려움.

연도별 신규 약효시험 활용 현황 (2000~2017)



한국화학물은행 소식

약효시험 활용 현황

[2017년 62건]

약효시험명	약효시험기관	적용질환분류
*** 저해 효과	산업체	항암제
*** 활성 조절물질 탐색	연구기관	항암제
*** 매개 기판 *** 약물 스크리닝	연구기관	염증
*** 제어 물질 개발	연구기관	항암제
*** 억제 화합물 스크리닝	연구기관	감염증
*** 유도물질 스크리닝	대학교	항암제
*** 억제제 탐색	연구기관	감염증
*** 저해제 개발	연구기관	항암제
*** 사용한 *** 유효물질 발굴	대학교	감염증
***의 *** 유도제 발견	대학교	신경계
*** 조절제 개발을 위한 약물 스크리닝	대학교	대사성질환
*** 에세이	산업체	신경계
*** 효과 발굴	산업체	대사성질환
*** 억제제를 통한 *** 치료제 개발	대학교	대사성질환
*** 변화를 유도하는 화합물 동정	대학교	대사성질환
*** 억제제 스크리닝	대학교	항암제
*** 저해제 검색	대학교	항암제
*** 약물 발굴	대학교	항암제
*** 에세이	산업체	항암제
새로운 *** 발굴	대학교	대사성질환
***에 의한 *** 저해제 탐색	대학교	대사성질환
*** 억제제 개발	연구기관	항암제
*** 스크리닝	연구기관	살충제
***를 타겟으로 하는 선도물질 스크리닝	연구기관	살균제
*** 억제제 개발	연구기관	항암제
*** 에세이	산업체	신경계
*** 억제제 탐색	연구기관	항암제
*** 인자 동정	대학교	기타
*** 억제물질 스크리닝	대학교	항암제
*** 활성 측정	산업체	파킨슨병
*** 저해제 검색	산업체	항암제
*** 에세이	산업체	항암제
*** 저해제 탐색	대학교	항암제
*** 조절 약물탐색	대학교	항암제
*** 억제 물질 스크린	대학교	대사성질환
*** 저해제	산업체	항암제

약효시험명	약효시험기관	적용질환분류
*** 유도제 약효시험	대학교	항암제
*** 약효시험	대학교	결핵
*** 화합물 발굴	대학교	염증
***에 작용하는 *** 후보물질의 발굴	대학교	대사성질환
*** 화합물들의 *** 조사	대학교	살충제
*** 개발	연구기관	항암제
*** 저해제 검색	연구기관	항암제
*** 억제제 탐색	대학교	항암제
*** 신규활성 *** 후보 물질 발굴	연구기관	항생제
*** 저해제	대학교	항암제
***에 대한 유효물질 발굴 *** 약효시험	대학교	감염증
*** 에세이	산업체	신경계
*** 억제하는 *** 억제제 발굴	대학교	항암제
*** 유도제의 활성성 검색	연구기관	감염증
*** 저해제 탐색	연구기관	신경계
*** 특이적 *** 물질 개발	대학교	감염증
*** 억제물질 탐색	연구기관	기타
*** 후보물질로 *** 억제제 개발	연구기관	감염증
*** 유도물질 동정	대학교	염증
*** 개발을 위한 유효 화합물 탐색	산업체	제초제
*** 활성을 저해하는 화합물 스크리닝	대학교	감염증
*** 저해제 검색	산업체	항암제
*** 라이브러리에 의한 ***의 변경된 *** 확인	대학교	항암제
*** 치료제 발굴	연구기관	신경계
*** 저해제 검색	산업체	간염
*** 저해제 탐색	대학교	감염증



한국화학물은행 소식

화합물 활용 논문 및 특허 성과

한국화학물은행 제공 화합물 library를 활용하여 발표된 논문 및 특허 목록

발표논문 [2017년 28건]

- "한국화학물은행 제공 화합물로 연구진행" 문구가 포함된 논문

번호	논문명	저자	저널
1	Antimicrobial activity of novel 4H-4-oxoquinolizine compounds against extensively drug-resistant <i>Acinetobacter baumannii</i> strains	Seok Hyeon Na, Hyejin Jeon, Yoo Jeong Kim, Hyeon Il Kwon, Gati Noble Selsa, Asimwe Nicholas, Chang-Soo Yun, Sang Ho Lee, Je Chul Lee*	<i>International Journal of Antimicrobial Agents</i> 2017, 49, 107-111
2	On-Chip Peptide Mass Spectrometry Imaging for Protein Kinase Inhibitor Screening	Young-Lai Cho, Young-Pil Kim, Jin Gyeong Son, Miyoung Son, and Tae Geol Lee*	<i>Analytical Chemistry</i> 2017, 89(1), 799-806
3	Replacing the terminal piperidine in ceritinib with aliphatic amines confers activities against crizotinib-resistant mutants including G1202R	Gangadhar Rao Mathi, Chung Hyeon Kang, Heung Kyoung Lee, Raghavendra Achary, Ha-Yeon Lee, Joo-Youn Lee, Jae Du Ha, Sunjoo Ahn, Chi Hoon Park, Chong Ock Lee, Jong Yeon Hwang, Chang-Soo Yun, Hee Jung Jung, Sung Yun Cho, Hyoung Rae Kim*, Pilho Kim*	<i>European Journal of Medicinal Chemistry</i> 2017, 126, 536-549
4	Anti-cancer effect of novel PAK1 inhibitor via induction of PUMA-mediated cell death and p21-mediated cell cycle arrest	Tae-Gyun Woo, Min-Ho Yoon, Shin-Deok Hong, Jiyoung Choi, Nam-Chul Ha, Hokeun Sun, Bum-Joon Park*	<i>Oncotarget</i> 2017, 8(14), 23690-23701
5	Design and Synthesis of Novel Pyrazolo [3,4-d]pyrimidin-1-yl piperidine Derivatives as Bruton's Tyrosine Kinase Inhibitors	Hyebin Park, Chi Hoon Park, Seung-Tae Kang, Jeong Hee Jeon, Raghavendra Archary, Joo-Youn Lee, Pilho Kim, Hee Jung Jung, Chang-Soo Yun, Jong Yeon Hwang, Do Hyun Ryu, and Sung Yun Cho*	<i>Bulletin of the Korean Chemical Society</i> 2017, 38, 278-281
6	Development of an HTS system to identify natural chemicals that specifically inhibit <i>Escherichia coli</i> O157:H7 adhesion to host cells	Seung Min Kim, Hyun Jung Kim*	<i>Food Control</i> 2017, 73, 185-192
7	The Discovery of Novel Protein Tyrosine Phosphatase ε Inhibitors Using a High-throughput Screening Approach	Hye-Yeoung Yun, Bonsu Ku, Hye Seon Lee, Ho-Chul Shin, Jun-Beom Park, Chang Hyen Kim, * and Seung Jun Kim*	<i>Bulletin of the Korean Chemical Society</i> 2017, 38, 44-53
8	Suppression of NFAT5-mediated Inflammation and Chronic Arthritis by Novel κB-binding Inhibitors	Eun-Jin Han, Hyun Young Kim, Naeun Lee, Nam-Hoon Kim, Seung-Ah Yoo, H. Moo Kwon, Dae-Myung Jue, Yune-Jung Park, Chul-Soo Cho, Tran Quang De, Dae Young Jeong, Hee-Jong Lim, Woo Kyu Park, Ge Hyeong Lee, Heeyeon Cho*, Wan-Uk Kim*	<i>EBioMedicine</i> 2017, 18, 261-273
9	Synthesis of Fluorescent Indazoles by Palladium-Catalyzed Benzannulation of Pyrazoles with Alkynes	Og Soon Kim, Jin Hyeok Jang, Hyun Tae Kim, Su Jin Han, Gavin Chit Tsui*, and Jung Min Joo*	<i>Organic Letters</i> 2017, 19(6), 1450-1453

번호	논문명	저자	저널
10	Activity of novel inhibitors of <i>Staphylococcus aureus</i> biofilms	Seung-Gyun Woo, So-Yeon Lee, So-Min Lee, Kyoung-Hee Lim, Eun-Ju Ha, Yong-Bin Eom*	<i>Folia Microbiologica</i> 2017, 62(2), 157-167
11	A clinical drug library screen identifies clobetasol propionate as an NRF2 inhibitor with potential therapeutic efficacy in KEAP1 mutant lung cancer	E-J Choi, B-J Jung, S-H Lee, H-S Yoo, E-A Shin, H-J Ko, S Chang, S-Y Kim and S-M Jeon*	<i>Oncogene</i> 2017, 36, 5285-5295
12	3-Amino-1 <i>H</i> -pyrazolopyridine Derivatives as a Maternal Embryonic Leucine Zipper Kinase Inhibitor	Chul Min Park, Vithal B. Jadhav, Jong-Hwan Song, Sunkyoung Lee,* Hee Young Won, Sang Un Choi, and You Hwa Son	<i>Bulletin of the Korean Chemical Society</i> 2017, 38(6), 595-602
13	Stereoselective Synthesis of Functionalized 1, 3-Disubstituted Isoindolines via Rh(III)-Catalyzed Tandem Oxidative Olefination-Cyclization of 4-Aryl-cyclic Sulfamidate-5-Carboxylates	Raghavendra Achary, In-A Jung, Se-Mi Son, and Hyeon-Kyu Lee*	<i>The Journal of Organic Chemistry</i> 2017, 82, 7223-7233
14	A Novel Series of Highly Potent Small Molecule Inhibitors of Rhinovirus Replication	Jinwoo Kim, Yu Kyoung Jung, Chonsaeng Kim, Jin Soo Shin, Els Scheers, Joo-Youn Lee, Soo Bong Han, Chong-Kyo Lee, Johan Neyts, Jae-Du Ha, and Young-Sik Jung*	<i>Journal of Medicinal Chemistry</i> 2017, 60, 5472-5492
15	Identification of Antipneumococcal Molecules Effective Against Different <i>Streptococcus pneumoniae</i> Serotypes Using a Resazurin-Based High-Throughput Screen	Hyung Jun Kim, Namyoul Kim, David Shum, Srigouri Huddar, Chul Min Park, and Soojin Jang*	<i>ASSAY and Drug Development Technologies</i> 2017, 15(5), 198-209
16	Acceleration of osteoblast differentiation by a novel osteogenic compound, DMP-PYT, through activation of both the BMP and Wnt pathways	Su Jung Bae, Hye Joo Kim, Hee Yeon Won, Yong Ki Min* and Eun Sook Hwang*	<i>Scientific Reports</i> 2017, 7, 8455-8464
17	Enterovirus inhibitory activity of C-8- <i>tert</i> -butyl substituted 4-aryl-6,7,8,9-tetrahydrobenzo[4,5]thieno [3,2- <i>e</i>][1,2,4]triazolo[4,3- <i>a</i>]pyrimidin-5(4 <i>H</i>)-ones	Bishyajit Kumar Biswas, Yashwardhan R. Malpani, Neul Ha, Do-Hyun Kwon, Jin Soo Shin, Hae-Soo Kim, Chonsaeng Kim, Soo Bong Han, Chong-Kyo Lee, Young-Sik Jung*	<i>Bioorganic & Medicinal Chemistry Letters</i> 2017, 27, 3582-3585
18	Synthesis and biological evaluation of <i>N</i> -(<i>cis</i> -cyclobutyl)purine derivatives for use as cyclin-dependent kinase (CDK) inhibitors	Sun Jun Park, Eunjin Kim, Miyou Yoo, Joo-Youn Lee, Chi Hoon Park, Jong Yeon Hwang*, Jae Du Ha*	<i>Bioorganic & Medicinal Chemistry Letters</i> 2017, 27(18), 4399-4404
19	Antihelminthic drug niclosamide inhibits CIP2A and reactivates tumor suppressor protein phosphatase 2A in non-small cell lung cancer cells	Myeong-ok Kim, Min Ho Choe, Yi Na Yoon, Jiyeon Ahn, Minjin Yoo, Kwan-Young Jung, Sungkwan An, Sang-Gu Hwang, Jeong Su Oh*, Jae-Sung Kim*	<i>Biochemical Pharmacology</i> 2017, 144(15), 78-89

번호	논문명	저자	저널
20	Development of a High-Throughput Assay for Inhibitors of the Polo-Box Domain of Polo-Like Kinase 1 Based on Time-Resolved Fluorescence Energy Transfer	Tae Gi Kim, Ju Hee Lee, Mi Young Lee, Ka-UL Kim, Jeong Hyun Lee, Chi Hoon Park, Byung Ho Lee, and Kwang-Seok Oh*	<i>Biological and Pharmaceutical Bulletin</i> 2017, 40(9), 1454-1462
21	Discovery of novel [1,2,4]triazolo[4,3- <i>a</i>]quinoxaline aminophenyl derivatives as BET inhibitors for cancer treatment	Imran Ali, Jooyun Lee, Areum Go, Gildon Choi*, Kwangho Lee*	<i>Bioorganic & Medicinal Chemistry Letters</i> 2017, 27, 4606-4613
22	An Efficient Copper-catalyzed Nucleophilic Addition to <i>N</i> -Acyliminium Ions Derived from <i>N</i> -Benzyloxycarbonylamino Sulfones: A Novel Approach to C-3 Functionalization of 2-Phenylimidazo[1,2- <i>a</i>]pyridine	Hyeong Jin Park and Jong-Gab Jun*	<i>Bulletin of the Korean Chemical Society</i> 2017, 38, 1123-1128
23	Identification of novel ALK2 inhibitors and their effect on cancer cells	Minsung Kim, Okran Choi, Suhkneung Pyo, Sang Un Choi*, Chi Hoon Park*	<i>Biochemical and Biophysical Research Communications</i> 2017, 492, 121-127
24	KRICT-9 inhibits neuroinflammation, amyloidogenesis and memory loss in Alzheimer's disease models	Do Yeon Lee, Chul Ju Hwang, Ji Yeon Choi, Mi Hee Park, Min Ji Song, Ki Wan Oh, Sang Bae Han, Woo Kyu Park, Hee Yeong Cho, Sung Yun Cho, Hye Byn Park, Min Jong Song and Jin Tae Hong*	<i>Oncotarget</i> 2017, 8(40), 68654-68667
25	BET Inhibitors as Anticancer Agents : A Patent Review	Imran Ali, Gildon Choi* and Kwangho Lee*	<i>Recent Patents on Anti-Cancer Drug Discovery</i> 2017, 12, 340-364
26	Naphthofuroquinone derivatives show strong antimycobacterial activities against drug-resistant <i>Mycobacteria</i>	Woong Sik Jang, Young-Sang Choi, Sukyoung Kim, Md. Anirban Jyoti, Hoonhee Seo, Juhye Han, Yong-Sik Kim, Jiwon Lyu, Kung-Woo Nam, Byung-Eui Lee, Kee-In Lee & Ho-Yeon Song*	<i>Journal of Chemotherapy</i> 2017, 39(6), 383-343
27	Synthesis and Molecular Modeling Studies of <i>N</i> -Hydroxyindazolecarboximidamides as Novel Indoleamine 2,3-Dioxygenase 1 (IDO1) Inhibitors	Dong-Ho Lee, Joo-Youn Lee, Jieun Jeong, Miok Kim, Kyung Won Lee, EunSeo Jang, Sunjoo Ahn, Chang Hoon Lee* and Jong Yeon Hwang*	<i>Molecules</i> 2017, 22, 1936-1945
28	A pivotal role of AMPK signaling in medicarbin-mediated formation of Brown and Beige adipocytes from C3H10T1/2 mesenchymal stem cells	Khan Mohammad Imran, Dahyeon Yoon, Yong-Sik Kim*	<i>BioFactors</i> 2017, 온라인 게재

제공 화합물 활용 특허 등록 현황 (2017년 7건)

번호	특허명	발명자	등록
1	신규한 하이드라진 유도제 및 이의 용도	김용철, 이원길	한국등록 10-1728898 (2017-04-14) 광주과학기술원
2	다제 내성 결핵균 억제용 조성물	정인필, 윤문영, 김민서, 고영관, 류재욱, 구동완	한국등록 10-1696352 (2017-01-09) 한양대학교 산학협력단, 한국화학연구원
3	비브리오 파루베이트 키나아제를 표적으로 하는 비브리오 항생제 및 이의 스크리닝 방법	이주연, 석영재, 김연란, 윤지희, 김혜민, 이창로	한국등록 10-1696605 (2017-01-10) 한국화학연구원, 서울대학교
4	암예방 및 치료용 조성물	박종선, 박지수, 김선환, 트란광돈	한국등록 10-1702468 (2017-01-26) 충남대학교 산학협력단
5	암예방 및 치료용 조성물	박종선, 박지수, 김선환, 트란광돈	한국등록 10-1702464 (2017-01-26) 충남대학교 산학협력단
6	5-니트로-1,10-페난트로린 유도제 및 이를 포함하는 결핵 질환의 예방 또는 치료용 약제학적 조성물	이월영, 경현혜, 라만디프싱, 박찬용, 사키브키드와이, 정영근	한국등록 10-1757629 (2017-07-07) 한국화학연구원
7	3-(4-클로로페닐)벤조[4, 5]이미다조[2, 1-b]싸이아졸-6-카복실산을 포함하는 항비만 또는 항우울증 치료용 조성물	김철희, 정승현, 황규석, 배영애, 김낙정, 이규양	한국등록 10-1719321 (2017-03-17) 한국화학연구원

제공 화합물 활용 특허 출원 현황 (2017년 2건)

번호	특허명	발명자	출원
1	p22phox 유래 N 말단의 8개 아미노산인 N8 펩타이드의 모방 화합물을 유효성분으로 포함하는 염증성 질환의 예방 또는 치료용 조성물	이주연, 양철수	한국출원 10-2017-0012350 (2017-01-26) 한국화학연구원, 한양대학교 에리카산학협력단
2	옥셀라딘시트레이트를 유효성분으로 포함하는 뇌암 예방 또는 치료용 약제학적 조성물	박재화	한국출원 10-2017-0134624 (2017-10-17) 성균관대학교 산학협력단

[빅데이터 소개]

“신약개발관련 화학정보학 데이터베이스 소개 (I)”

ChEMBL - 유럽분자생물학연구소 공개 데이터베이스

한국화학연구원 **채종학**

최근 들어 신약개발에 소요되는 시간과 경비를 줄이고 효율을 향상시키기 위하여 AI(인공지능) 및 빅데이터(Big Data)를 이용한 신약개발연구에 많은 관심이 집중되고 있습니다.

이러한 노력이 성공하기 위해서는 빅데이터를 효과적으로 분석하고 활용할 수 있는 인공지능의 개발도 중요하지만 우선적으로 고품질의 빅데이터 수집·관리가 매우 중요합니다.

이러한 필요에 따라 2000년대 후반부터 국공립 연구소 또는 대학 등의 공공 연구기관을 중심으로 수백만에서 수 천만 건의 신약개발 관련 화합물 및 활성데이터 등의 정보를 생산 또는 문헌에서 추출하고 큐레이션을 거쳐 데이터베이스화 하여 ChEMBL, PubChem과 같은 화학정보학 데이터베이스를 구축하여 일반에 공개하여 누구나 자유롭게 인터넷을 통하여 자료검색과 다운로드를 하여 사용할 수 있도록 하고 있습니다.

한국화학물은행은 신약개발 연구자들이 공개 빅데이터를 잘 활용할 수 있도록 세계적으로 잘 알려진 신약개발관련 화학정보학 데이터베이스에 대한 소개를 연재합니다. 이번호에서는 유럽분자생물학연구소에서 제공하는 신약개발관련 공개 데이터베이스인 ChEMBL에 대하여 소개합니다. [편집자 주]

차례

1. 신약개발과 화학정보학 데이터베이스

2. ChEMBL 개요

3. ChEMBL의 데이터

- 수집데이터 범위
- 데이터베이스 구조
- 데이터 수집방법
- 데이터 모델
- 데이터 기탁 및 교환

4. ChEMBL 제공정보

- Compound Structural Alert
- 비임상, 임상단계 화합물 정보
- ADMET 및 Pharmacokinetic 데이터
- 특허 데이터 (SureChEMBL)
- 농약분야 데이터
- Assay 및 타겟정보

5. 사용자 인터페이스

- ChEMBL Web Interface
- myChEMBL S/W 컨테이너
- ChEMBL Beaker
- RESTful 서비스
- KINASE SARfari
- GPCR SARfari
- ADMET SARfari
- SureChEMBL

6. 참고문헌



1. 신약개발과 화학정보학 데이터베이스

- 수십 년간의 연구개발에도 불구하고 신약개발은 여전히 고비용, 고위험도의 연구분야에 머물러 있음. 전체 신약개발 프로세스의 비전을 제시하고, 연구자들의 성공적인 신약개발의 요인에 대한 이해를 돕기 위하여 컴퓨터를 이용한 신약개발, 화학정보학, DB 개발연구가 진행중임.
- 특히 대학이나 국가연구소 등 공공기관의 신약개발 연구에서 화합물 구조, 활성 정보를 제공하는 공개DB는 구조-활성 관계 분석, 화학정보학 방법과 신약설계 방법 검증에 필수적임.
- 저분자 유기물이나 바이오 의약품의 활성정보는 문헌에 많이 공개되어 있으므로 이를 활용하면 특정 pathway나 타겟을 탐색하기 위한 tool compound의 선정, 관심 화합물의 위해성이나 부작용 예측, 기존 화합물의 부작용이나 새로운 작용점 예측, 특정 시리즈 화합물의 SAR 분석, in vivo ADMET 예측, 유사한 화합물의 유전자 표현형 탐색, 신규 타겟에 최적화된 focused 라이브러리의 구축 등의 연구가 가능함.
- 그러나 저널에 발표된 화합물 활성정보는 표준화되어 있지 않으므로 수집, 정제, 활용이 어렵고, 이미지로 표현된 화합물 구조, DB 태그가 없이 축약어, 유사어로 표현된 단백질 타겟, 탈락된 실험정보 등의 자료로부터 DB를 구축하는 데는 많은 인력이 소요됨.
- 유전체 염기서열, 단백질 3차원 구조, gene expression data 등 공공 DB가 잘 축적된 생물정보학 분야와는 달리 화합물 분야 연구자들은 화합물 및 활성정보의 DB 구축에는 관심이 적었으므로, 기존의 화학정보학 DB들은 상업적인 유료DB가 대부분임.
- 2010년대 이후 대학교, 국립연구소 등의 공공분야에서 이러한 화학정보학 DB에 대한 수요와 필요성에 대한 인식이 급증함에 따라 다양한 분야의 화학정보학 공공DB를 구축하려는 움직임이 시작됨. PubChem Bioassay, ChemBank에서는 수백만 개의 HTS 데이터를 공개함. BindingDB에서는 신약타겟 가능성이 높은 단백질에 대한 활성데이터를 수작업으로 수집함. PDBBind, Binding MOAD, AffinDB는 PDB에서 추출한 리간드활성정보, PDSP Ki DB는 NIH의 정신질환 스크리닝 데이터, BRENDA는 enzyme에 대한 결합상수 데이터를, IUPHAR는 수용체, 이온채널에 대한 데이터, GLIDA/GPCRDB에서는 GPCR 관련 데이터를 제공함.
- 그 외, PubChem (<https://pubchem.ncbi.nlm.nih.gov/>, 6천만개 화합물, 2.3억개 활성데이터), PDB(Protein Data Bank, <http://www.rcsb.org>; 14만개 단백질 3차원구조), ChemSpider(<http://www.chemspider.org>; 6천만개 화합물 구조, 속성 및 관련정보) 등 많은 공개 화합물, 타겟 단백질 DB가 구축되어 있음.
- EU에서도 2016년부터 Horizon2020의 BIGCHEM 프로젝트를 통하여 화학분야 빅데이터와 정보학을 결합한 화학정보학 시스템을 구축하고 있음. 다국적 제약기업, 중소벤처, 일반 연구자를 대상으로 대용량 화학데이터 생성, 데이터 분석기법 교육, 인력양성이 목표임.

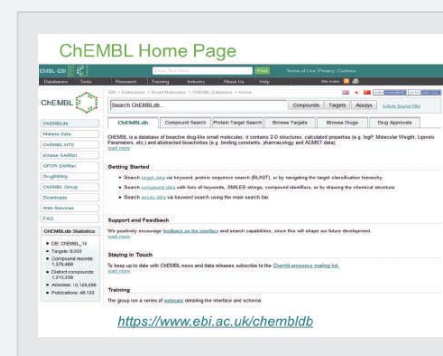
〈신약개발 관련 화합물 데이터베이스〉

DB 이름	내용	URL
ChEMBL	논문 등에서 수작업으로 2백만종 화합물, 15백만개 활성데이터를 수집정리한 신약개발용 DB	www.ebi.ac.uk/chembl
PubChem	NIH/MLSMR 프로젝트에서 9천만종 화합물, 1백만건 스크리닝, 2.3억개 활성데이터를 수집한 DB	pubchemdocs.ncbi.nlm.nih.gov
ChemSpider	영국 왕립화학회에서 6천만개 화합물 구조, 속성, 관련정보를 수집한 DB	www.chemspider.com
BindingDB	64만개 화합물, 7천개 타겟에 대한 활성 DB	www.bindingdb.org
BRENDA	논문에서 수작업으로 추출한 6,500개 효소를 EC 번호로 분류하고, 효소의 기능, 대사 네트워크와 연계한 효소 DB	www.brenda-enzymes.org
PDB	X-ray, NMR, cryo-EM 등으로 분석한 14만개 단백질, nucleic acid의 3차원구조 DB	www.rcsb.org
BigChem	EU Horizon2020의 화학데이터, 화학정보학 프로젝트; EU/IMI 산하에 eTox, EHR4CR, OpenPhACT, EMIF, eTriKS 등 화학정보학 프로젝트가 있음	bigchem.eu

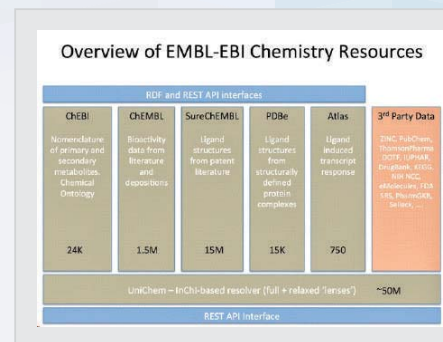
2. ChEMBL 개요

- ChEMBL(<https://www.ebi.ac.uk/chembl/>)은 의약화학 문헌으로부터 화합물 구조와 신약타겟에 대한 활성데이터를 수집하여 수작업으로 정리한 세계최대의 신약개발용 공공 데이터베이스로서, 누구나 자유롭게 웹을 통한 자료 검색과 ftp를 통한 다운로드 하도록 공개되어 있음. 2010년도 이후 발표된 리간드 기반 신약탐색 모델링, drug repositioning, ADMET/독성 예측, polypharmacology 연구, 인공지능 신약개발 등 화학정보학 분야 연구논문은 대부분 ChEMBL을 이용한 것임. (1천여건 이상)
- 2017년 현재 67,722개의 의약분야 논문에서 추출한 11,538개 타겟, 2,101,843종 화합물 구조, 1,320,147개 assay, 14,675,320개의 활성 데이터를 보유하고 있음.
- 영국 Hixton의 Wellcome TruGen Campus에 소재하고 있으며, 유럽생물정보학협회(European Bioinformatics Institute; EBI)에 의해 운영되고 있음. 원래 영국 제약기업인 Inpharmatica/Galapagos 사에서 구축한 StARlite DB 시스템을 2008년에 바이오의료자선기관인 WellcomeTrust에서 인수하여 EBI의 John Overington의 ChEMBL 화학유전체 그룹에서 개발하여, 2010년부터 일반에 공개함.
- 현재 EMBL-EBI에는 ChEBI, SureChEMBL, PDBe, Atlas 등 화합물 DB 이외에도 ZINC, PubChem, ThomsonPharma, IUPHAR, DrugBank, KEGG, NIH NCC, eMolecules, FDA SRS, PharmaGKB, Sellek 등 많은 외부 DB도 연계되어 있음.

〈ChEMBL 홈페이지〉



〈EMBL-EBI의 화합물 관련 DB〉



3. ChEMBL의 데이터

● 수집데이터의 범위

- 화합물 구조와 신약타겟에 대한 활성데이터를 보유하고 있음. 단일 타겟에 대한 활성과 함께 functional assay 데이터, ADMET 데이터, 물성데이터를 제공함. 활성데이터는 Ki, Kd, IC50, EC50 등의 종류로 분류되어 있으며, 데이터 정리를 거쳐 표준화된 활성 값도 제공됨.
- 초기의 ChEMBL_02 버전은 12개 저널의 34,000개 논문에서 수작업으로 추출된 8,603개 타겟, 1,195,368개 화합물, 1,060,258개 활성데이터로부터 시작하여, 현재 ChEMBL_23 버전은 67,722개의 논문에서 추출된 11,538개 타겟, 2,101,843종 화합물에 대한 1,320,147개 assay, 14,675,320개의 활성 정보를 보유함.
- 화합물의 2D 구조, 계산물성값 (logP, MW, Lipinski 등)을 포함하며, 활성 데이터는 화합물 구조, 단백질 타겟과 연계되며, 단백질 타겟은 분자구조와 UniProt의 단백질 서열과 연계되어 있음.
- J. Med. Chem. 등과 같은 의약화학 저널에 발표된 화합물의 활성 데이터, 말라리아 등과 같은 ND 스크리닝 데이터, PubChem으로부터 추출된 데이터 등 30년전 축적된 화합물 합성 및 생활성 데이터를 관리하고 있음.
- 2013년 PubChem의 활성 데이터 중에서 ChEMBL의 데이터 수준에 적합한 confirmatory assay 데이터를 선별하여 추가 하였으며, 특히 DB인 SureChem도 합병함.

● 데이터베이스의 구조

• 데이터베이스는 화합물 구조, 정보, 실험정보, 타겟정보, 결합 부위 정보, 대사정보, 질환정보, 승인약물정보 등을 포함하는 63개의 복잡한 테이블로 구성됨.

• 데이터의 생성, 열람, 수정, 삭제에 위한 ORM (Object Relational Mapping) 기반 툴이 제공됨. ORM 툴에서는 다양한 DB S/W에서 재사용가능한 툴로서 찾기 어려운 오류를 유발하는 원시 SQL 문의 사용을 피할 수 있음.

• ORM-모델 상위에 RESTful 웹서비스를 구축하여 http 프로토콜을 통한 다양한 종류의 데이터에 대한 자동화된 접근방법을 연구자들에게 제공함. 화합물, 스크리닝, 활성, 타겟 등 다양한 레벨에서 데이터가 구조화, 모듈화가 되어 있으므로 사용자는 한 번의 질의로 화합물의 구조와 관련 물질, ADMET 정보를 열람할 수 있음.

● 데이터 수집 방법

• 화합물 구조, 활성데이터는 Journal of Medicinal Chemistry, Bioorganic Medicinal Chemistry Letters, Journal of Natural Products 등 의학화학 저널에서 수작업으로 추출함. 저널의 선택은 완벽하지는 않지만 시간과 경비를 고려하여 가장 많은 양의 고품질 데이터를 수집하는데 효율적인 범위의 저널을 선택함. 각 저널에서 화합물, assay, 타겟에 대한 상세정보를 수집함.

• 화합물 구조는 논문에 제공되는 scaffold, R-group, 이름, 유사어 등 정보로부터 컴퓨터가 인식할 수 있는 전체구조를 추출하고, 구조검사, salt 제거, 약자분석 등의 표준화를 거쳐 DB로 입력함.

• assay 정보는 binding, functional, ADMET 등의 상세정보를 문헌에서 추출하고, 서로 다른 assay간의 데이터 비교를 위하여 IC50은 nM로, half-life는 시간으로 바꾸는 등의 표준화를 거침.

• 일관성있는 타겟정보를 제공하기 위하여 분자 타겟, organism 정보를 문헌에서 추출하고 전문가의 심사를 거쳐 DB에 수록됨. 단백질/바이오펙타 복합체에 대한 실험에서는 특정 타겟을 지정하기가 어려우므로 해당 단백질들과 함께 complex 태그로 표시하며, 세포/조직 실험에서는 특정 단백질군을 지정할 수 없으므로 multi 태그로 표시함. 세포독성실험에서는 대상 세포를, 박테리아 항생제 실험에서는 박테리아를 타겟으로 표시함. 단백질 타겟은 GPCR의 리간드기반 분류, protease/kinase/phosphatase 효소분류 등 의약연구에서 많이 쓰이는 군 계층으로도 세분되어 있음.

● 데이터 모델

• 가장 중요한 기본 데이터는 document (논문)와 이에 연계된 화합물, assay, 타겟, 화합물-타겟간의 활성정보임.

• 화합물은 InChI 규칙에 따라 중복성 검사를 거쳐 중복성없는 화합물사전에 저장되며, stereoisomer는 별도 화합물로, tautomer는 동일 화합물로 처리됨. 소수의 구조가 명확하지 않은 화합물이나 단백질 의약품도 포함됨. 화합물 정보는 Accelrys사의 Direct cartridge로 저장, 관리, 검색되고 있음.

• 타겟정보도 단백질, nucleic acids, subcellular fractions, cell-lines, tissues, organisms 등으로 구분되어 저장됨. assay 정보와 타겟군 분류 정보도 함께 저장됨.

• 각 화합물, 타겟, assay, document에는 특별한 ChEMBL_ID가 부여됨. 필요한 경우 화합물은 ChEBI, InChI 키, PubChem의 CID, AID, SID, PDB의 HET 코드, PubMed의 DOI 등 외부 ID도 함께 저장되며, 단백질 타겟은 UniProt ID, organism 타겟은 NCBI Taxonomy ID를 저장함.

● 데이터 기탁 및 교환

• ChEMBL은 문헌데이터 이외에도 데이터 교환을 원하는 비영리기관, 상업적 회사로부터 데이터를 수집하고 있음. Dundee 대학교로부터 천연물 유사화합물 라이브러리를 수집하여 소외질환 타겟에 대한 스크리닝에 활용할 계획이며, 현재 화합물 구조와 세포독성 데이터가 공개되어 있음.

• 또 말라리아 활성을 가진 400종 화합물로 구성된 Malarial Box Compound Set 라이브러리를 Medicine for Malaria Venture사로부터 기탁받아 20여개의 말라리아 스크리닝 센터에서 실험한 데이터를 공개함. (<http://www.mmv.org/research-development/open-access-malariabox>). 최근 MMV사로부터 더 확장된 적응타겟을 가지는 Pathogen Box 화합물 라이브러리를 기탁받아 스크리닝에 활용중임. (<http://www.pathogenbox.org>)

• 소외질환사업[Drugs for Neglected Diseases Initiative, DNDi, <http://www.dndi.org>]으로부터 Leishmaniasis and Chagas disease 등의 소외질환에 대한 스크리닝 데이터를 기탁받아 공개함.

• Vienna 대학교 Roche, GlaxoSmithKlinet 등 제약사에서도 데이터를 기탁함. AstraZeneca는 자사의 실험데이터중에서 ChEMBL에 이미 존재하는 화합물에 대한 혈장단백질결합, microsome and hepatocyte clearance, 용해도, 이온화도, 친지질도 등의 in vitro ADME, 물성 스크리닝 데이터를 기탁함.

〈 데이터 소스별 ChEMBL 데이터 수집 현황 〉

Short Names	Source	No. Cpds	No. Assays	No. Activities
LITERATURE	Scientific Literature	967,242	963,186	5,635,084
PUBCHEM BIOASSAY	PubChem BioAssays	489,575	2,937	7,559,601
GATES LIBRARY	Gates Library compound collection	68,490	2	69,444
BINDINGDB	BindingDB Database	68,149	1,317	99,061
GSK TCMDC	GSK Malaria Screening	13,467	6	81,198
ST JUDE LEISH	St Jude Leishmania Screening	13,422	6	42,105
USP/USAN	USP Dict USAN and International Drug	11,356	0	0
DNDI	Drugs for Neglected Diseases Initiative	7,053	233	14,452
ASTRAZENECA	AstraZeneca Deposited Data	5,799	15	11,687
NOVARTIS	Novartis Malaria Screening	5,614	6	27,888
ORANGE BOOK	Orange Book	2,016	0	0
SUPPLEMENTARY	Deposited Supplementary Bioactivity Data	1,786	13	4,817
CANDIDATES	Clinical Candidates	1,633	0	0
ST JUDE	St Jude Malaria Screening	1,524	16	5,456
TP TRANSPORTER	TP-search Transporter Database	1,434	3,592	6,765
DRUGMATRIX	DrugMatrix	930	113,678	350,929
METABOLISM	Curated Drug Metabolism Pathways	828	0	0
GSK TB	GSK Tuberculosis Screening	826	15	1,814
WHO TDR	WHO-TDR Malaria Screening	740	16	5,853
GSK TCAKS	GSK Kinetoplastid Screening	592	13	7,235
MMV MBOX	MMV Malaria Box	400	138	45,158
MMV PBOX	MMV Pathogen Box	400	0	0
ATLAS	Gene Expression Atlas Compounds	398	0	0
DRUGS	Manually Added Drugs	378	0	0
GSK PKIS	GSK Published Kinase Inhibitor Set	366	456	169,451
OSM	Open Source Malaria Screening	211	22	344
WITHDRAWN	Withdrawn Drugs	192	0	0
TG GATES	Open TG-GATES	160	158,199	158,199
SANGER	Sanger Inst Genomics of Drug Sensitivity	137	714	73,169
FDA APPROVAL	FDA Approval Packages	43	1,386	1,387
HARVARD	Harvard Malaria Screening	37	4	111

• ChEMBL은 기탁된 모든 데이터에 DOI를 부여하여 논문에서 이를 활용할 때 언급하도록 하였으며, 기탁된 데이터의 원본은 'Activity Source Filter' 항목에서 찾아볼 수 있음.

• PubChem의 BioAssay DB는 각 연구실, 스크리닝 센터의 HTS 장비에서 생산되는 수많은 화합물의 단일농도에 대한 실험데이터를 자동으로 수집하므로 데이터의 숫자는 많지만

IC50, Ki 등 dose-response 데이터는 제한적임. 이에 반하여 ChEMBL 데이터는 논문에서 수작업으로 추출된 데이터이므로 대부분 dose-response 데이터임. 2011년 자료에서 dose-response 데이터가 있는 단백질 타겟의 개수는 PubChem이 700개, ChEMBL은 4000개임.

- PubChem과 ChEMBL간의 데이터 상호보완을 위하여 데이터 교환시스템을 구축함. ChEMBL의 데이터는 PubChem의 BioAssay에 전송되며, PubChem에서 dose-response 데이터가 있는 confirmatory assay 데이터는 ChEMBL에 병합됨.
- 동일한 방식으로 BindingDB와 데이터가 상호교환됨.
- 현재 11,538개 타겟, 2,101,843종 화합물, 1,320,147개 assay, 14,675,320 개 활성 데이터가 수집되어 있음.

4. ChEMBL 제공정보

● Compound Structural Alerts

- 화합물의 약물성 특성을 이용한 선별방법을 제공하기 위하여 SMARTS, Pfizer LINT, Glaxo Wellcome Hard, BMS HTS Deck, NIH MLSMR Excluded Functionality, Dundee 대학 NTD 스크리닝 라이브러리, PAINS, Inpharmatica, SureChEMBL 필터 등 다양한 공개 화합물 필터를 구현함.

- 필터는 대부분 활성작용기, 독성, assay 방해, 모든 임의의 타겟에 결합하는 성질을 가지는 조각구조기반이며, 승인된 약물의 50% 정도가 한 개 이상의 필터에 의하여 걸러지기 때문에, 이러한 필터를 적용할 때는 충분한 주의가 필요함.

● 비임상, 임상 단계 화합물 정보

- ChEMBL에는 FDA Orange Book 21에 수록된 승인의약품의 상품명, 투약경로, 용량, 화합물 구조, 단백질의약품의 서열 등 정보도 있음.
- 의약품류 [합성, 천연물, 항체, 단백질, oligosaccharide, oligonucleotide, 무기물 등], 안전성, 이성질체, 적용증(치료제, 이미징, 진단 등)의 정보도 포함되어 있음.

- FDA 승인 약물에 대한 약물성, 타겟 활성 정보와 함께 적용질환 정보를 추가하여 타겟 검증연구를 지원하고 있음. 적용질환 정보는 DailyMed package inserts(<https://dailymed.nlm.nih.gov/>), Anatomical Therapeutic Chemical (ATC) classification (<http://www.whocc.no/atc-ddd-index/>), ClinicalTrials.gov

(<https://clinicaltrials.gov>) 등 다양한 소스로부터 텍스트마이닝 자동매핑, 수작업 검토, 검증을 거쳐 수집함.

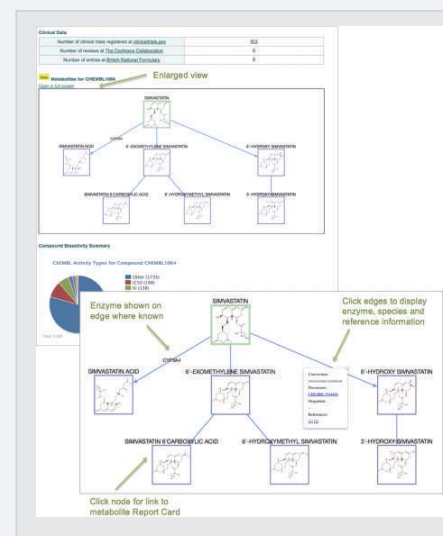
- NIH의 Illuminating the Druggable Genome 프로젝트 (<https://commonfund.nih.gov/idg>)에서 집중하고 있는 GPCR, kinase, nuclear hormone receptor, 이온채널 등의 타겟에 대한 임상 1-3상 단계의 1,023개 임상 후보물질에 대한 작용기전, 활성 타겟 정보를 보완중임.
- 현재까지 1,451개 임상후보물질의 작용기전, 적용질환 정보가 구축되어, withdrawn/discontinued/prescription-only/over the count 등으로 분류됨.
- 시판후 독성이나 활성부족 문제로 퇴출된 약물의 정보를 FDA(<http://www.fda.gov>), EMA(<http://www.ema.europa.eu/ema/>), WITHDRAWN database, the US Electronic Code of Federal Regulations, FederalRegister(<https://www.gpo.gov/fdsys/pkg/FR-2014-07-02/pdf/2014-15371.pdf>) 등의 DB로부터 퇴출일, 퇴출이유, 국가/지역 등의 정보와 함께 수집함.

〈Triglitazone의 작용기전과 퇴출이유〉

● ADMET 및 Pharmacokinetic 데이터

- 용량, 시간, 투여방법, 다른 화합물과의 상호작용 등 다양한 변수에 따라 복잡해지는 in vivo 데이터의 특성을 처리하기 위하여 ASSAY_PARAMETERS 테이블을 도입함.
- 2011년 이후의 260개 논문, 2200개 화합물, 1만5천개 assay에서 비임상 및 임상 화합물의 약물대사, PK, 독성 데이터를 수작업으로 추출하고, FDA 승인 신청시에 제출된 문헌에서 ADMET 정보를 수집하여 DB를 구축함. 약물 대사체의 활성, 독성문제를 해결하기 위하여 승인된 약물의 대사체와 대사 효소를 탐색하는 연구가 진행되어 METABOLISM 테이블에 관련정보가 저장됨.

〈Simvastatin 대사체 정보〉



● 특허데이터 수집

- Digital Science사에서 개발한 SureChem 특허 화합물 DB를 병합하여 SureChEMBL로 발전 시킴으로써 문헌에 아직 발표되지 않은 화합물의 활성정보를 연구자들에게 미리 제공함. 그러나 아직 SureChEMBL에서는 특허에서는 화합물 구조만 가져오고, 활성데이터는 BindingDB로부터 가져옴.

- BindingDB는 저널에서 수집한, 저분자 화합물의 단백질 타겟에 대한 활성DB임. 최근에는 2013년부터 공개된 1천개의 미국 특허자료에서 600개 타겟, 7만개 화합물, 10만개의 활성 데이터를 추출하여 포함시킴.

- ChEMBL은 BindingDB에 비하여 화합물 구조와 타겟 단백질에 대한 주석이 집중되어 있으며, 타겟단백질은 트리계측구조로 구성되어 타겟 연관성 이해나 타겟군별 분류에 유리함. BindingDB는 타겟의 3차원 구조에서 시작했기 때문에 enzyme 타겟에 더 특화되어 있으며, GPCR 리간드는 ChEMBL에 비하여 상대적으로 적음.

● 농약분야 데이터 수집

- 현재까지는 제약산업계가 ChEMBL DB의 가장 큰 사용자가지만 농업계에서도 점차 agrochemicals DB의 구축이 진행중임. 작물보호제 분야의 저널에서 제초제, 살충제, 항진균제 등 4만개 화합물, 25만개 활성 데이터가 추출되어 DB에 포함됨.

- 기존 ChEMBL에 포함되어있던 화합물에 대해서도 Fungicide Resistance Action Committee (FRAC: <http://www.frac.info/publications>), Herbicide Resistance Action Committee (HRAC: <http://hrac.tsstaging.com/tools/classification-lookup>) or Insecticide Resistance Action Committee (IRAC: <http://www.irac-online.org/documents/moa-brochure/?ext=pdf>) 시스템의 규칙에 따라 작용기전이 표기됨.

- ChEMBL의 타겟 분류 계층도 시스템에서 agrochemical 분야의 타겟과 화합물 정보를 열람할 수 있음.

● Assay 및 타겟 정보

- ChEMBL의 activity 데이터는 화합물이나 타겟 기반으로 탐색하는 것이 체계적임. 그러나 ChEMBL의 절반이상의 assay 데이터는 functional 혹은 phenotype 스크리닝 데이터이므로 특정한 분자타겟을 지정할 수 없기 때문에 이들은 cell-line, tissue, assay format 등의 필터를 이용하여 탐색가능하도록 ontology를 구성함.

- BioAssay Ontology (BAO)는 ChEMBL의 많은 저분자 화합물 스크리닝 데이터를 처리하기 위하여 개발되었으며, 데이터 상호교환을 촉진하기 위하여 외부 신약개발협회, 연구기관, 기업계 등에서도 채택됨. ChEMBL의 표준활성 type은 BAO의 BAO_ENDPOINT로, 표준활성 unit은 UO_UNITS로 매핑되었으며, 이러한 방식으로 90% 이상의 활성데이터가 매핑됨.

- ChEMBL의 assay는 BAO의 BAO_FORMAT으로 매핑되어 biochemical, cell-based, tissue-based, organism-based 방식으로 구분됨. 축적된 assay는 세부정보와 타겟분류에 따라 자동화된 규칙기반 방식으로 구분함.
- cell, tissue 기반 assay는 Cell Line Ontology, Experimental Factor Ontology, Cellosaurus, LINCS cell dictionary, Uberon, Brenda Tissue Ontology, CALOHA 등 기존의 ontology DB를 참고하여 CELL_DICTIONARY, TISSUE_DICTIONARY를 도입하여 분류함.
- 단백질 타겟데이터의 열람 검색을 위하여 이온채널, 막수송체에 대한 단백질 군 분류를 기존의 IUPHAR/BPS, TCDB 등과 일관성을 유지하도록 개발하였으며, Gene Ontology Term으로 ChEMBL 타겟을 검색할 수 있도록 GO Slim 시스템을 구축함.

5. 사용자 인터페이스

● ChEMBL Web Interface (<https://www.ebi.ac.uk/chembl>)

- ChEMBL은 웹브라우저 접속을 통하여 화합물 구조, 타겟, 스크리닝에 대한 검색을 제공함.
- 특정 타겟에 대한 tool 화합물 정보는 타겟 이름, 유사명칭, UniProt ID, ChEMBL ID를 이용한 키워드 검색으로 구할 수 있음. 타겟은 단백질군 hierarchy, organism 등의 키워드로 검색하거나, 단백질 서열정보를 이용한 BLAST 검색으로 유사타겟 검색을 할 수 있음. 검색된 타겟에 대하여 수행된 assay 상세정보, 사용된 화합물의 구조, 활성도 종류 (IC50, Ki 등), 측정단위, 타겟 상세정보, 원본 문헌정보 등을 얻을 수 있으며, text나 excel 등의 형태로 저장할 수 있음.

〈 ChEMBL의 단백질 타겟 종류 〉

Target type	Description	Example	# of targets
Single protein	Single protein chain	Phosphodiesterase 5A (CHEMBL1827)	5,518
Protein family	Group of closely related proteins	Muscarinic receptors (CHEMBL2094109)	188
Protein complex	Defined protein complex, consisting of multiple subunits	GABA-A receptor (CHEMBL2094120)	159
Protein complex group	Poorly defined protein complex, subunit composition is unclear	GABA-A receptor (CHEMBL2093872)	43
Protein-protein interaction	Disruption of a PPI	p53/Mdm2 (CHEMBL1907611)	12
Chimeric protein	Fusion of two different proteins, synthetic or natural	Bcr/Abl fusion protein (CHEMBL2096618)	2
Selectivity group	Pair of proteins for which selectivity has been assessed	Muscarinic receptors M2 and M3 (CHEMBL2095187)	96
Protein-nucleic acid complex	Complex consisting of both protein and nucleic acid components	70S ribosome (CHEMBL2363965)	6
Nucleic acid	DNA, RNA or PNA	Apo-B 100 mRNA (CHEMBL2364185)	28
Oligosaccharide	Oligosaccharide	Heparin (CHEMBL2364712)	4
Small molecule	Small molecule, such as amino acid, sugar or metabolite	Glutamine (CHEMBL2366039)	20
Macromolecule	Large biological molecule other than protein complex	Hemozoin (CHEMBL613898)	4
Metal	Metal or ion	Iron (CHEMBL2363058)	8

- 특정 화합물에 대하여 substructure/유사도 검색을 통하여 활성도, 선택성, ADMET 정보를 얻을 수 있음. 검색된 화합물에 대하여 logP, pKa, PSA 등의 물성에 대한 필터링을 적용하여 원하는 물질범위의 화합물 정보만 얻을 수도 있음.
- ChEMBL 웹 인터페이스에서 일반 연구자들이 필요로 하는 대부분의 기능을 제공하지만, 독자적인 DB 시스템을 구축하거나, 대규모의 데이터마이닝을 원하는 연구자를 위하여 모든 데이터 파일을 다운로드하는 기능도 제공함. 화합물 DB는 Oracle, MySQL, SD 파일 형태로 제공되며, 단백질 서열은 FASTA 형태로 제공됨.

● myChEMBL S/W 컨테이너

- myChEMBL은 PostgreSQL DB, web 서비스, RDKit, OSAR, OpenBabel, Jupyter Notebook 등 오픈소스 화학정보학 도구를 포함하는 가상머신으로서 VMDK, QCOW2, IMG 등의 가상머신 디스크 이미지나 Docker 컨테이너 형식으로 배포됨.
- 연구자는 가상머신 이미지를 다운로드 받아 Ubuntu, CentOS 등의 리눅스/가상머신 시스템에서 로딩하여 ChEMBL의 기능을 연구실 내에서 쉽게 구현, 활용할 수 있음.

● ChEMBL Beaker

- 중앙서버에 설치된 오픈소스기반의 신약개발용 화합물 API(Application Programming Interface)를 말하며, 이를 통한 서비스를 제공함으로써 새로운 화학정보학 플랫폼 개발을 지원하는 도구임. Python으로 개발되었으며 scikit-learn,

pandas, matplotlib and RDKit 등의 오픈소스 라이브러리를 포함하고 있음.

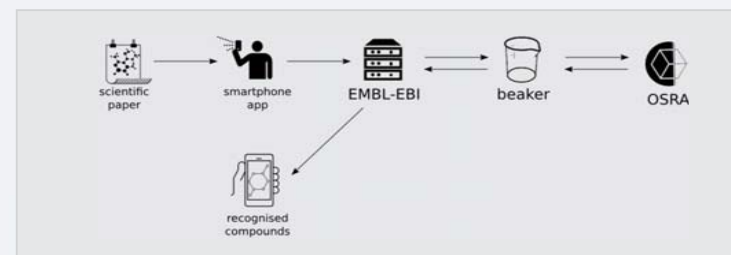
- 분자 파일 포맷 상호변환, 평거프린트/디스크립터 표현자 계산, 2D/3D 화합물 그리기, InChI/InChIKey 생성, 이미지로부터 분자구조 생성하기 등의 기능이 있으며, 사용자가 제출한 데이터를 처리하는 것이 주 목적임.

- Beaker 소스는 다음에서 다운로드 할 수 있음.
https://github.com/mnowotka/chembl_beaker

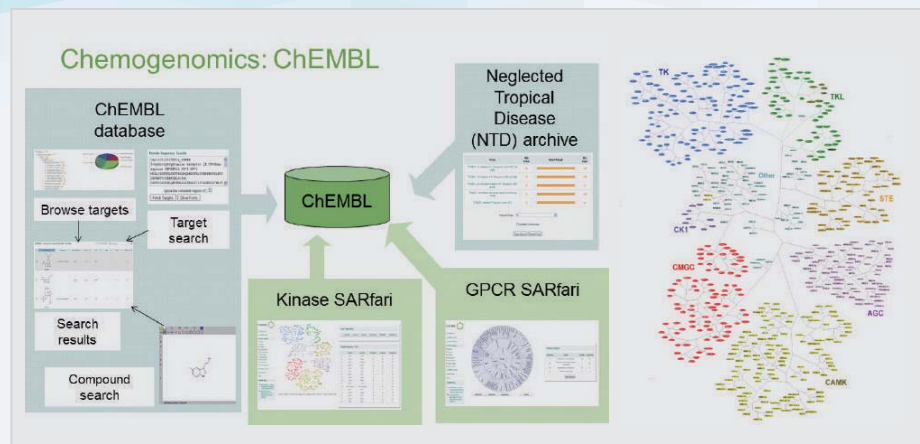
● RESTful 서비스 (<https://www.ebi.ac.uk/chembl/api/utis/docs>)

- 화학정보학 시스템 개발자를 위하여 Python, Java, Perl 등 프로그래밍 도구를 이용한 자동화된 Http 프로토콜 접속으로 JSON, XML, YAML 형태의 결과파일을 받을 수 있는 RESTful 서비스도 제공됨. (<https://www.ebi.ac.uk/chembl/ws>)
- 단순한 텍스트 질의를 통하여 화합물, 타겟, assay 정보를 탐색할 수 있는 Solr-기반의 검색 기능도 제공됨. 예를 들어, 다음의 질의를 통하여 angiogenesis와 관련된 모든 assay 정보를 얻을 수 있음.
(<https://www.ebi.ac.uk/chembl/api/data/assay/search?q=angiogenesis>)
- ChEMBL DB는 RESTful 서비스를 통하여 PubChem BioAssay, BindingDB, CanSAR, Open PHACTS, Open Targets, Target Central Resource Database / PHAROS 등의 외부 DB에서 활용되고 있음.

〈 Beaker기반의 모바일 앱 〉



〈 ChEMBL의 화학유전체학 시스템 〉



● KINASE SARfari (<https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari>)

- Kinase 서열, 3차원 구조, 화합물, 결합부위, 스크리닝 데이터를 망라하는, kinase에 특화된 화학유전체학 데이터베이스, 마이닝 툴, 분석 도구를 포함하는 SW 시스템임.
- 989개 kinase, 54,189개 화합물, 532,155 개 활성데이터를 관리하고 있음.
- 홈페이지에서 kinase 패밀리별 리간드와 활성도 검색, 단백질 3차원구조, 결합부위 분석 등을 할 수 있으며, 화합물, 활성도, 단백질, 서열 등을 다운로드 할 수 있음.

● GPCR SARfari (<https://www.ebi.ac.uk/chembl/sarfari/gpcrsarfari>)

- GPCR의 서열, 3차원구조, 화합물, 스크리닝 데이터를 관리하는 화학유전체학 시스템으로서 912개 GPCR, 147,292 개 화합물, 1,037,273개 스크리닝 데이터를 보유하고 있음.
- 홈페이지에서 GPCR 패밀리별 리간드와 활성도 검색, 단백질 3차원구조, 결합부위 분석 등을 할 수 있으며, 화합물, 활성도, 단백질, 서열 등을 다운로드 할 수 있음.

● ADMET SARfari (<https://www.ebi.ac.uk/chembl/admesarfari>)

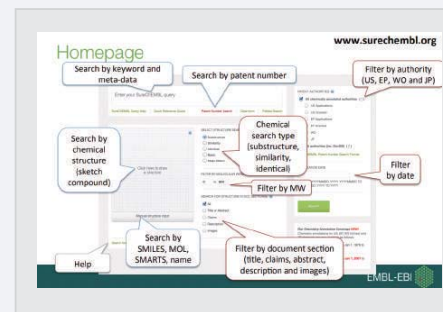
- 특정 화합물의 가능한 ADME 타겟 검색, 입력 서열과 유사한 ADME 타겟 검색, PK 데이터 검색, 입력과 유사한 화합물의 ADME 데이터 검색, 특정 타겟의 인체 expression level 탐색등을 할 수 있음.
- 입력 화합물에 대한 PK 예측은 ChEMBL의 PK타겟에 대한 activity와 Multiclass Naive Bayesian Classifier로 학습한 예측모델을 이용함.

● SureChEMBL

- Digital Science / Macmillan사에서 상업적인 특허화합물 마이닝 프로젝트로 개발한 SureChem 프로젝트를 2013년에 EMBL-EBI에 기부함. EBI는 후속개발, 일반공개 및 OpenPHACT와 연계하여 활성, 생물 데이터를 연계하는 업무를 수행함.
- US Applications, US Granted, EP Applications, EP Granted, WO, JP 특허에서 텍스트, 이미지를 처리하여 화합물 구조와 관련정보를 수집함.

- 1997년부터 공개된 1억2천만건의 특허에서 매월 8만개 화합물 정보를 수집하여 16백만개 화합물 정보를 구축함. 공개된 특허는 1-4일 후에 SureChEMBL에서 검색가능함.

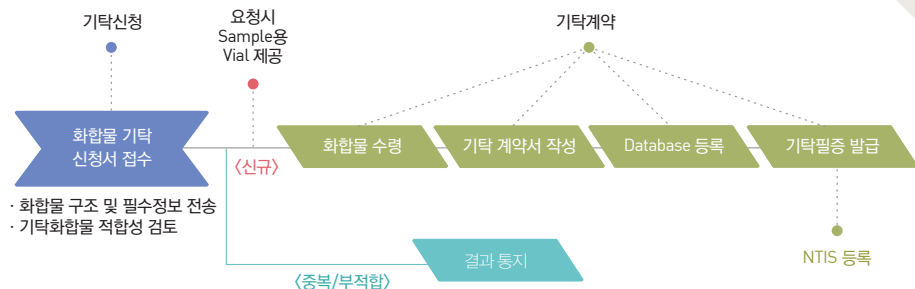
〈 SureChEMBL 홈페이지 〉



6. 참고문헌

- A. Gaulton, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* (2012) 40, D1100-D1107
- A. P. Bento, et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* (2014) Mol Inform. 2016 35, 615-621, D1083-D1090
- A. Gaulton, et al. The ChEMBL Database in 2017, *Nucleic Acids Research Nucleic Acids Res.* (2017) 45, D945-D954
- M. Nowotka, et al, ChEMBL Beaker: A Lightweight Web Framework Providing Robust and Extensible Cheminformatics Services, *Challenges*, (2014) 5, 444-449
- M. Michal, et al, Using ChEMBL web services for building applications and data processing workflows relevant to drug discovery, *Expert Opinion on Drug Discovery* (2017) 12, 757-767
- M. Davies, et al, ChEMBL web services: streamlining access to drug discovery data and utilities, *Nucleic Acids Res.* (2015) 43, 612-620
- Wikipedia, <https://en.wikipedia.org/wiki/ChEMBL>
- I. V. Tetko, et al, BIGCHEM: Challenges and Opportunities for Big Data Analysis in Chemistry, *Mol. Inform.* (2016) 35, 615-621

화학물 기탁 절차



• 기탁화학물 범위 및 수준

- (범위) 저분자 유기합성 화학물(분자량: 1700이상 1,000이하) 및 단일성분 천연물
(기탁화학물의 활성여부와 상관 없음, 최종 화합물 뿐만 아니라 합성 중간체도 기탁 가능)
- (수준) 화합물 양 >10mg(권장), 순도 >80%(한국화학물은행에서 LC-MS로 검증), 구조를 확인할 수 있는 NMR 자료 제출

• 기탁자 혜택

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

• 화학물 관리 체계



한국화학물은행 보유 화학물 품질관리

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

• 반입, 보관, 반출별 화학물 관리

1) 반입 화학물 품질 관리

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

2) 보관 화학물

- 화학물을 장기 보관함에 따라 필연적으로 생길 수 있는 화학물의 품질 저하를 체크하기 위하여 매년 보유화학물의 일부(약 20,000종)를 선별하여 LC-MS로 품질 검증을 하고 있으며 순도가 80% 이하인 화학물은 반출을 중지하고 별도로 관리합니다.

3) 반출 화학물

- 화합물이 반출된 후 악효실험을 거쳐 Hit 화합물로 확인된 경우, 다음 단계의 실험을 진행하기 전에 Hit 화합물의 품질을 LC-MS로 다시 분석하여 그 결과를 활용자에게 제공합니다.

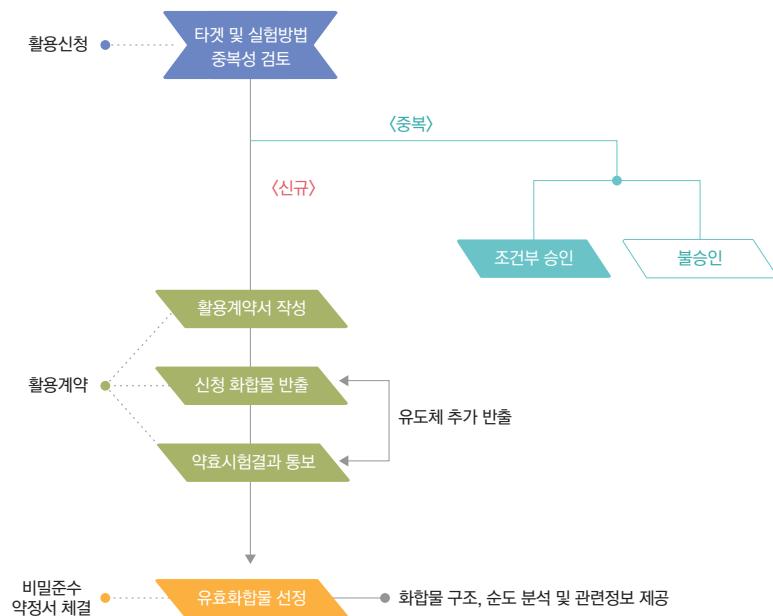


한국화학물은행은 보유 화학물의 품질을 최상으로 유지하기 위하여 노력하고 있습니다.

화학물 활용 절차

문의

E-mail : chembank@krcit.re.kr · 대표번호 : 042-860-7190 · 홈페이지 : http://www.chembank.org



화학물 사용료 규정

- 화학물은 무상으로 제공합니다.
- 그러나 화학물 제공에 수반되는 실비용(plate, 분주기 tip, 튜브 cap 비용 등)은 사용자 부담이 원칙입니다

활용결과 권리 규정

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

■ 규정목적

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

■ 규정내용

· 활용결과(논문, 특허 등) 권리관계 규정 내용

- ① 기탁자가 단순기탁 이외의 추가적인 기여가 없는 경우에는 화학물 기탁자로서 논문의 사사(acknowledgement)에만 포함되는 것이 "원칙"입니다.
- ② 그러나 해당 화학물이 논문/특허의 핵심 화학물이고 구조까지 공개 되는 경우에는 기탁자를 논문공저자 및 특허 공동발명인(공동 출원인)으로 "적극 고려"하여야 합니다.
- ③ 기탁자가 추가적인 기여(유도체 합성 제공, 관련 정보제공 등)를 하였을 경우에는 기여정도에 따라 기탁자를 논문공저자 및 특허 공동 발명인(공동출원인)으로 "포함"하여야 합니다.
- ④ 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 화학물을 사용하여 연구가 진행되었다는 사사(acknowledgement)를 포함하여야 합니다.

※ 한국화학물은행 활용결과와 권리 규정은 기본적으로 과학계에서 통용되는 연구결과 기여도에 대한 "연구윤리 기본원칙"을 따르고 있습니다.



한국화학은행 제공 라이브러리 종류

임상화합물 Library

- 화합물수 : 2,150
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 5 mM)
- 라이브러리 구성 : 임상 I - III 상 단계 화합물 및 승인 약물 (Clinically Active Compounds)

Fragment Library

- 화합물수 : 1,000
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 20 mM)
- 라이브러리 구성 : 분자량 300이하 라이브러리, 순도 및 분자량 검증 (LC-MS)

GPCR 및 PPI Library

- 화합물수 : 각 10,000
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 5 mM)
- 라이브러리 구성 : 해외 Vendor로 부터 선별 구매

Virtual Screening

- 화합물수 : 협의
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 5 mM)
- 화학정보학 및 분자모델링 방법을 적용한 가상탐색
- 타겟 연구관련 별도협의 필요

대표 Library

- 화합물수 : 6,560
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 5 mM)
- 라이브러리 구성 : 전체 화합물을 대표하는 라이브러리, 순도 및 분자량 검증 (LC-MS) 기탁/구매 분리 구성 (2015.6 신규구성)

Kinase Library

- 화합물수 : 3,000
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 5 mM)
- 라이브러리 구성 : 분자모델링방법 (Docking)을 적용하여 kinase 타겟의 active site에 결합할 가능성이 높은 화합물로 구성

천연물 Library

- 화합물수 : 1,020
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 5 mM)
- 라이브러리 구성 : 단일성분 천연물 및 천연물 유사골격 (Natural product-like) 화합물

PharmaCore Collection

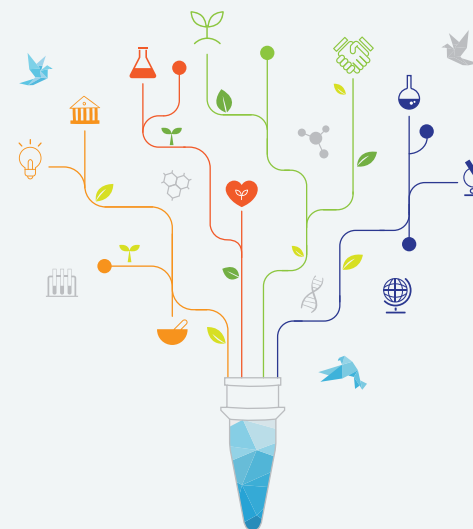
- 화합물수 : 요청개수
- 기본 제공량 : 5 uL (DMSO 용액, 평균농도 5 mM)
- 라이브러리 구성 : 요청골격으로 선별 구성한 화합물

전체 Library

- 화합물수 : 55만종
- 농도 : 5mM, 5uL

공동 활용을 통한 화합물의 가치 재창출

We take care of your compounds and
Create new value for you!!



화합물 기탁 및 활용 문의

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

이름	담당업무	전화	E-mail
이현규	센터장	[042]860-7016	leehk@krcit.re.kr
이주연	화학정보학/분자모델링	[042]860-7181	leejy@krcit.re.kr
채종학	화학정보학/분자모델링	[042]860-7451	chchae@krcit.re.kr
황순희	대외협력/정보관리	[042]860-7190	chembank@krcit.re.kr
김선우	화합물관리/장비관리	[042]860-7171	swkim@krcit.re.kr
김선호	화합물관리/저온창고관리	[042]860-7090	shkim@krcit.re.kr
이유리	화합물정보/DB관리	[042]860-7092	yurilee@krcit.re.kr
이수연	LC-MS, 물성분석	[042]860-7747	suyoun@krcit.re.kr
박희정	LC-MS, 물성분석	[042]860-7747	huijeong@krcit.re.kr