KCB

newsletter

한국화합물은행 뉴스레터

2024년 제1호

KRICT 한국화학연구원
Korea Research Institute of Chemical Technology

한국화합물은행
Korea Chemical Bank

과학기술정보통신부지정
연구성과 관리·유통 전담기관
- 화합물 전담기관 -

한국화합물은행 뉴스레터
2024 Vol.1

발행일 : 2024년 1월
발행인 : 이영국
편집인 : 이계형
발행처 : 한국화합물은행

한국화합물은행은 국가연구개발사업의 화합물 연구성과
관리·유통 전담기관으로 지정되어 있습니다.
[과학기술정보통신부 고시 제2022-41호]

CONTENTS

한국화합물은행은 국내에서 합성되는 화합물(유기화합물 및 천연물) 및 관련정보를 범국가적으로 수집·관리하고 공동으로 약효시험에 활용함으로써 국내 신약개발 및 바이오 연구를 지원하기 위하여 2000년에 설립되었으며 75만종 이상(2023년)의 화합물을 확보하고 있습니다.

한국화합물은행 소식

제1회 신약개발 AI 경진대회 성황리 종료

한국화학연구원(한국화합물은행)은 국내 신약소재 화합물·데이터의 활용 촉진과 우수한 AI 인재 발굴을 위하여 제1회 신약개발 AI 경진대회를 한국생명공학연구원(국가생명연구자원정보센터), 한국제약바이오협회(인공지능신약개발연구지원센터)와 공동으로 주최하였다. 2023년 8월 7일부터 9월 25일까지 총 1,447명(1,254팀)이 참가한 가운데 열띤 분위기 속에서 진행되어 최종적으로 5개 팀이 수상자로 선정되었다.

참가자들은 한국화합물은행(KCB)이 학습용으로 제공한 3,498종 화합물의 구조 및 대사안정성(Human/Mouse Liver Microsomal Stability) 실측데이터를 사용하여 대사안정성 값을 예측할 수 있는 AI 모델을 개발하고, 예측 모델을 사용하여 화합물 구조만 제공된 경진용 483종에 대한 대사안정성 예측값을 결과물로 경진사이트에 제출하였다. 예측값과 실측값을 비교한 AI 모델의 예측정확도와 순위를 리더보드를 통해 실시간으로 확인할 수 있어 참가자들은 제출 마감 직전까지 높은 성능의 모델 개발을 위해 노력하였다.

1차 모델 예측 성능 심사를 통해 선정된 상위 10개 팀의 발표평가로 모델 구축 전략과 데이터 전처리, 다양한 접근 전략 및 적용 가능성 등을 평가하여 최종 5개 팀을 수상자로 선정하였다. 발표평가 대상자들은 대부분 모델의 성능을 높이기 위하여 ChEMBL, PubChem 등의 외부 공개 데이터를 병합하여 사용하였으며 다양한 오픈소스 프레임워크, 화합물 구조표현, 알고리즘, 모델, 코드 등을 조합하여 활용함으로써 모델 성능을 향상시켰다.

수상자들에 대한 시상식은 11월 9일 국립과천과학관에서 개최된 '2023 대한민국 과학기술대전-빅데이터·인공지능(AI) 융합 첨단바이오 심포지엄'에서 진행되었다.

경진대회 시상식과 함께 개최된 '빅데이터·인공지능(AI) 융합 첨단바이오 심포지엄-화합물신약분야'에서는 경진대회 수상자들의 발표 세미나와 전문가 초청 강연, 간담회 등을 통하여 데이터 분석 방법 및 노하우를 공유하고 데이터 기반 신약 연구의 미래, 향후 목표 등을 논의하는 시간을 가졌다.

<수상자 명단>

대상 2팀 (상금 각 1,000만원/팀)

- 과학기술정보통신부 장관상: 'datu팀' (중앙대학교, 아주대학교)
- 보건복지부 장관상: 'suleezard팀' (성균관대학교)

우수상 3팀 (상금 각 300만원/팀)

- 한국화학연구원장상: '실력으로말해팀' (GIST)
- 한국생명공학연구원장상: '약과도넛팀' ((주)HITS)
- 한국제약바이오협회장상: '다이머팀' ((주)스탠다임)

경진대회 시상식 기념촬영

첨단바이오 심포지엄 간담회

주요 방문 인사

방문일자	방문기관
2023. 07. 04	(주)아이엔테라퓨틱스 이승범센터장
2023. 07. 12	일동제약(주) 오창목 선임연구원, 정지은 선임연구원
2023. 07. 13	질병관리청 백신연구개발총괄과 이유경 과장 외 6인
2023. 07. 24	특허청 서을수 국장, 백영란 주무과장, 방성철 공업사무관 외 3인
2023. 07. 27	KOBIC 소재클러스터 강태호 박사 외 3인
2023. 08. 11	국가과학기술인력개발원 주관 이공계 장애학생 연구기관 탐방캠프 20인
2023. 08. 17	중외제약 박월성 선임연구원, 유태규 매니저, 홍우상 팀장
2023. 08. 31	UST캠퍼스 투어(전북대학교)
2023. 09. 19	NIST 이해정 박사 외 2인
2023. 09. 22	한양대학교 김주연 교수
2023. 10. 11	주한 EU 및 유럽대사관 20인

방문일자	방문기관
2023. 10. 16	Symeres: Norimasa Morita
2023. 10. 17	TNO: Martijn de Koning
2023. 10. 19	한남대학교 이광섭 총장 및 아시아 대학 관계자 약 20인
2023. 10. 23	한남대학교 이진아 교수
2023. 10. 26	㈜미주씨앤디 송명근 책임연구원 외 2인
2023. 11. 02	유한양행 중앙연구소 김태균 팀장 외 4인
2023. 11. 15	UST캠퍼스 투어(한양대학교)
2023. 11. 15	질병관리청 국립보건연구원 이정민 과장, 김성한 과장
2023. 11. 16	가천대학교 심원식 교수
2023. 11. 28	이화여자대학교 차선신 교수, IPK 장수진 박사
2023. 12. 01	(주)대현케미칼 김대훈 대표


㈜아이엔테라퓨틱스


질병관리청

특허청


중외제약


UST 캠퍼스 투어

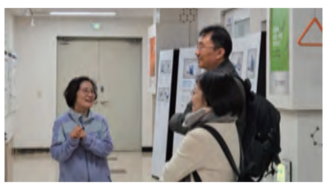

한양대학교


주한 EU 및 유럽대사관


TNO 네덜란드 응용과학연구기구


㈜미주씨앤디


가천대학교


이화여자대학교 및 IPK


㈜대현케미칼

화합물 확보 현황

한국화합물은행은 2023년 12월 31일 기준 75만종 이상의 화합물을 보유하고 있으며 다양성, 약물성, 독창성이 우수한 고수준 신약 소재 화합물 라이브러리를 제공하기 위하여 다양한 방법으로 우수한 화합물을 확보하고 있다.

1 연구성과 기탁[의무 기탁]

「국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률」 제26조 및 「국가연구개발혁신법」 시행령 제33조 3항에 따라 국가연구개발 사업 수행을 통해 창출된 화합물은 연구성과 관리·유통 전담기관인 한국화학연구원 한국화합물은행에 의무적으로 기탁하도록 되어 있다. 연구성과 기탁 효율제고를 위하여 2017년부터 국가연구 과제 및 사업평가에 연구성과 기탁실적을 반영하고, 전담 기관에 기탁된 성과만 인정하도록 "국가연구개발 과제평가 표준지침"이 개정되었다.

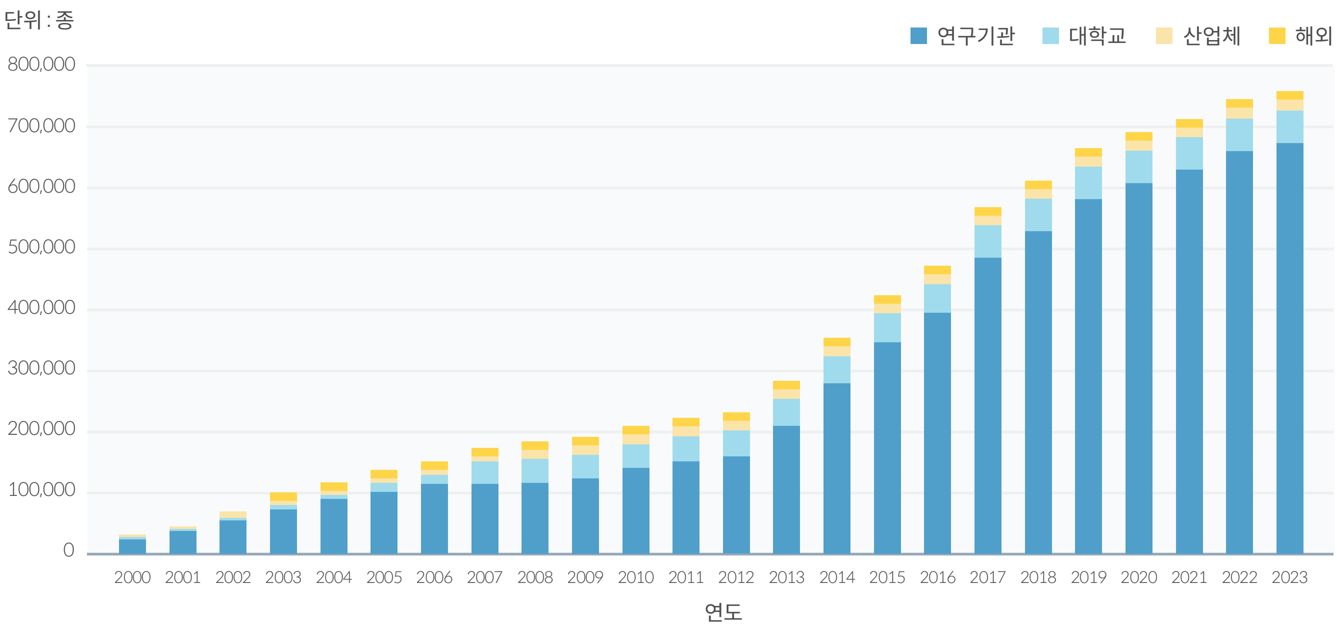
2 국내 전문가 공모 위탁합성

보유화합물의 구조 다양성 확대를 위하여 국내 합성 전문가를 대상으로 위탁합성 과제를 공모, 선정을 통하여 화합물을 확보하고 있다. 2014년부터 매년 15개 내외의 전문가 위탁합성 과제를 선정하고, 이를 통해 2023년까지 누적 20,500여종을 확보하였다. 2024년에도 전문가 위탁합성 공모를 통하여 지속적으로 화합물을 확보할 예정이다.

3 특수 골격 화합물 해외 구매

단순 기탁만으로 충족하기 어려운 화합물 다양성을 신속하게 확보하기 위하여 화학정보학 및 분자모델링 기술을 활용하여 약물성 및 다양성 극대화를 고려한 특수 골격 화합물을 해외 vendor로부터 선별하여 구매, 확보하고 있다. 2024년에도 가치가 높은 라이브러리를 지속적으로 확보하여 연구자들에게 제공할 예정이다.

화합물 확보 누계 현황(2000 ~ 2023): 75만종 보유



데이터 활용 현황

한국화합물은행은 2018년부터 KCB가 보유하고 있는 화합물의 구조 및 공개 가능한 약효시험 데이터 정보를 제공하고 있다. 정보 제공 서비스 건수도 지속적으로 증가하고 있으며, 임상화합물, 천연물 정보 외에도 세포독성 및 대사안정성 데이터(KCB 생산데이터), Drug-likeness정보(계산 데이터)등도 연구자들에게 제공하고 있다.

데이터 지원 현황

분류	소분류		2018-2022년	2023년
화합물 구조 정보	전체(분양 가능) 라이브러리		84건	34건
	구매화합물 라이브러리		62건	33건
	Focused 라이브러리	대표	33건	31건
		임상	50건	31건
		천연물	34건	29건
		Fragment	32건	32건
		Kinase	14건	29건
		PPI	45건	31건
		GPCR	41건	30건
	PharmaCore Collection		5건	1건
약효시험결과 정보	공개 가능한 정보		53건	23건

정보 제공 현황

정보 종류	2023년							
	산업체		대학교		연구기관		계	
	횟수	개수	횟수	개수	횟수	개수	횟수	개수
Hit 구조정보 및 분석 보고서	5	38	9	24	2	11	16	73
SAR 분석 구조정보	135	22,130	99	9,008	79	3,210	313	34,348
임상화합물 정보	14	116	35	1,267	15	652	64	2,035
천연물 정보	2	15	15	178	6	164	23	357
세포독성(KCB 생산 데이터)	3	27	7	20	3	7	13	54
대사안정성(KCB 생산 데이터)	3	26	7	19	3	7	13	52

화합물 활용 현황

매년 100건 내외의 신규 작용점에 대하여 20만점 이상의 화합물이 국내 산·학·연에 제공되어 활용되고 있다. 신규과제 활용지원은 2021년 116건, 2022년 126건, 2023년 95건이다. 분양 화합물 수도 2022년 대량분양을 제외하면 매년 꾸준히 증가하고 있다. 2023년 신규과제 95건 중 산업체가 33건, 대학교가 40건, 연구기관이 22건이었다. 2023년 KCB 화합물 활용 누적 과제수는 1,379건이다.

최근 4년간 화합물 활용 현황

연도	활용과제 건수		화합물 분양 횟수		분양 화합물 수	
	전체과제	신규과제	전체과제	신규과제	전체과제	신규과제
2020	223건	105건	414회	254회	479,332점	423,120점
2021	213건	116건	387회	238회	491,985점	340,996점
2022	246건	126건	442회	244회	1,000,752점	245,653점
2023	197건	95건	403회	181회	493,064점	240,540점

2023년 : 403회에 걸쳐 총 493,064점 화합물 분양(평균 2,020점/일)

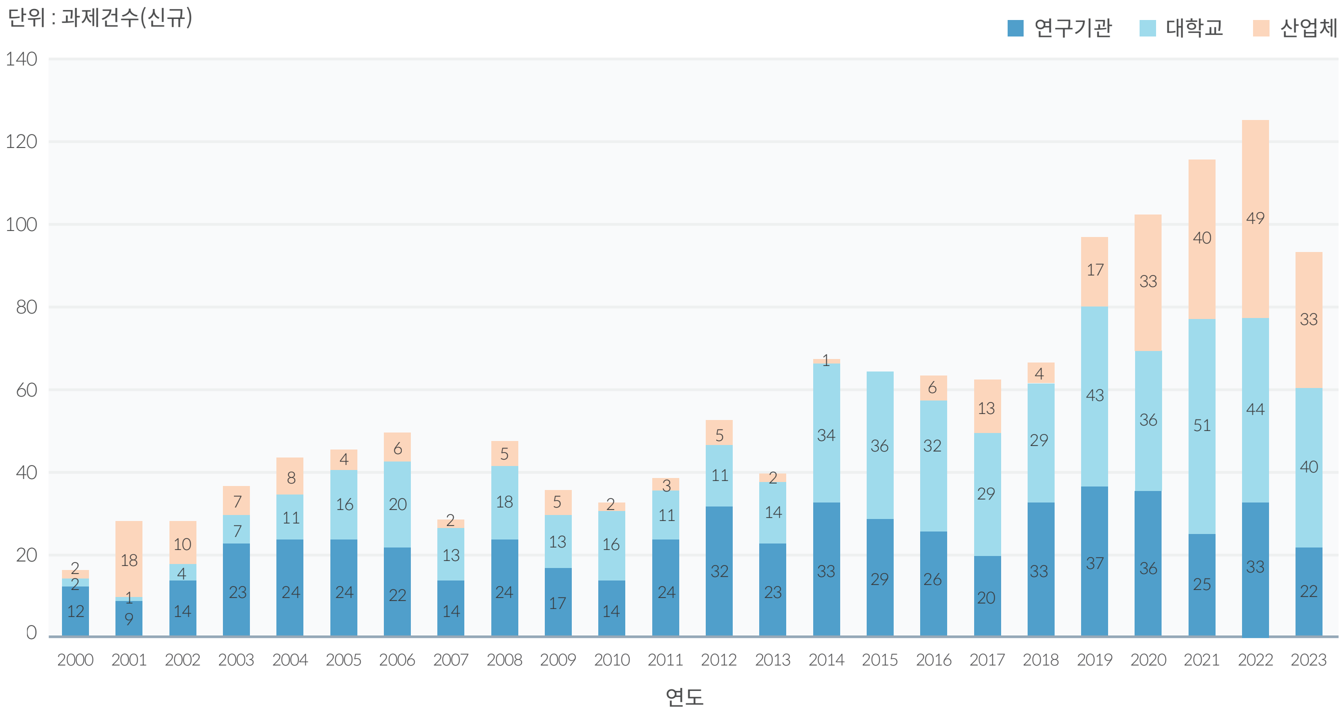
기관별 화합물 활용 신규과제 현황

(건수)

연도	2020	2021	2022	2023	전체누적
산업체	33	40	49	33	**275**
대학교	36	51	44	40	**534**
연구기관	36	25	33	22	**570**
계	105	116	126	95	**1,379**

2018년 하반기부터 전체 분양 가능한 화합물 라이브러리의 구조정보를 포함하여 활용자의 동의를 얻은 약효정보 등 KCB가 보유하고 있는 데이터를 연구자들에게 제공하고 있다. 산업체의 경우에는 제공받은 데이터 기반으로 최적의 화합물 선별과 지식재산권 확보 전략을 수립할 수 있으므로 데이터 활용 지원 사업 확대 이후에 산업체의 활용이 증가된 것으로 사료된다.

화합물 신규 활용 현황(2000 ~ 2023)



기관별 화합물 분양 현황(2023년)

구분	분양횟수	분양 화합물 수(점)						
		Pre-made	일반분양	Active compound (AC)	Active derivative (AD)	Hit compound (HC)	Hit derivative (HD)	계
산업체	170회	63,299	21,988	1,093	2,356	3	0	88,739
대학교	115회	157,275	38,067	3,025	812	0	0	199,179
연구기관	118회	70,079	129,686	4,386	914	15	66	205,146
총계	403회	290,653	189,741	8,504	4,082	18	66	493,064

2023년 분양된 화합물은 총 493,064점이었고, 자동선별(cherry-picking) 화합물이 202,411점이었으며, 그중 active compound (AC) 화합물 8,504점, active derivative (AD) 화합물 4,082점, hit compound (HC) 18점, hit derivative (HD) 66점이 분양되었다.

사전제작(Pre-made) 라이브러리 분양 현황(2023년)

Pre-made 라이브러리의 제작 목적은 사용자의 활용 편의를 위하여 사용목적별로 focused library를 구축하여 활용자의 요구를 충족하는 화합물을 신속하게 제공하는 것이다. 다수의 plate를 한번에 제작함으로써, freeze-thaw cycle의 반복에 따른 용해도 감소나 화합물의 변질 등을 최소화할 수 있는 장점이 있다. 대표화합물 19건, 임상화합물 13건, 천연물 7건 Kinase 5건, Fragment 4건, PPI 4건, GPCR 2건의 pre-made set이 분양되었다.
KCB는 화학정보학 역량을 집중하고 수요자의 의견과 전문가의 자문을 적극 수용하여 pre-made set의 질적개선을 위해 지속적으로 노력하고 있다.

라이브러리 종류	96 well	384 well	분양 수(Set)
대표	14	5	19
임상화합물	12	1	13
천연물	5	2	7
Kinase	3	2	5
Fragment	3	1	4
PPI	4	0	4
GPCR	2	0	2
총계	43	11	54

화합물 활용 지원의 경제적 가치평가(2023년)

· 화합물 : 연 50만점 분양(1만원 X 50만점 = **50억원 수입대체효과**)
· 신약개발 히트 도출 등을 고려한 화합물 당 경제적 가치*는 2만원 : 연 50만점 분양 시 100억 분양가치
* KOBIC 합성화합물 자원의 경제적 가치에 관한 조사연구('21)

화합물 약효데이터 수집 현황

KCB에서 분양 받은 화합물은 약효시험 후 결과정보를 KCB에 제출해야 한다. 매년 분양 화합물이 증가함에 따라 축적된 약효시험 데이터도 증가하고 있다. 축적된 데이터는 활용자의 동의 과정을 거쳐 연구자들에게 제공하고 있다.

연도	2019	2020	2021	2022	2023	누적
데이터 건수	198,528	457,565	657,670	680,363	467,830	8,884,906

스크리닝 type 및 표적질환별 활용 현황(2023년)

구분	현황
스크리닝 type	Target 기반 스크리닝: 49건 Function 기반 스크리닝: 46건
표적질환	암(35건), 감염병(14건), 면역/염증(12건), 신경계(11건), 대사질환(3건), 심혈관계(1건), 기타(19건)

2023년 신규 약효시험 활용과제는 Target 기반 스크리닝과 Function 기반 스크리닝이 각각 49건과 46건이었다. 항암제 연구가 35건으로 가장 많았고, 감염병 치료제 연구가 14건이었다. 면역/염증 치료제 연구가 12건, 신경계 연구 11건, 대사질환 3건, 심혈관계 연구가 1건이었다.

화학정보학 전문지원 현황(2023년)

KCB는 수요자의 요구에 맞는 화학정보학 전문 지원을 통해 화합물 선별 서비스를 제공하고 있다.

정보 종류		산업체	대학교	연구기관	계
지원 과제 수		**7**	**8**	**4**	**19**
지원 횟수	Pharmacophore	3	4	0	7
	분자도킹 결합모드 분석/예측	7	6	7	20
	유도체 선별	11	6	0	17
	Homology model 구축	2	1	0	3
	분자동역학 모의실험	0	1	0	1
	화합물 선정	1	3	4	8
	계	**24**	**21**	**11**	**56**

화합물 활용 신규과제 현황(2023년 하반기 49건)

약효시험명	구분	적용질환
*** 저해제 스크리닝	대학교	항암제
***화합물을 이용한 ***의 제어	대학교	항생제
화합물을 활용하여 *** 반응의 in vitro 및 in vivo *** 효능	대학교	염증
화합물 라이브러리를 이용한 *** 활성 평가	산업체	살균제
***의 *** 저해제 발굴	대학교	감염증
***의 결합 ligand 탐색	대학교	면역
*** 키나아제 저해활성 평가 시험	산업체	염증
*** 물질 발굴	연구기관	신경계
신규 *** 억제효능의 *** 개발	대학교	항암제
***의 치료제 개발	대학교	항암제
*** 효소 저해제 발굴	산업체	항암제
*** binder 스크리닝	연구기관	항암제
*** 활성 평가	산업체	면역
*** 상호작용 저해 약물 탐색	대학교	기타
***효소와 ***의 상호작용 저해 물질 스크리닝	대학교	항균제
*** 저해제 약효평가	산업체	항암제
신규 골격의 *** 길항제 개발 연구	대학교	기타
*** 세포에서 *** 화합물 검색	연구기관	항암제
*** 억제제 스크리닝 분석	산업체	면역
*** 반응성 Cell painting	산업체	항암제
***을 이용한 활성 검증	산업체	항암제
*** 치료를 위한 선택적 *** 억제제 개발시험	산업체	신경계
*** 결합 화합물의 알레르기 염증 반응	대학교	염증
in vitro *** 활성 검정	산업체	살균제
*** 상호 작용을 촉진하는 ***의 가상 스크리닝	산업체	항암제

약효시험명	구분	적용질환
*** 세포에서의 약물 스크리닝	대학교	항암제
*** 활성 평가	연구기관	항암제
*** 물질 탐색	산업체	살균제
*** 저해제 발굴	산업체	항암제
*** 저해제 발굴	산업체	신경계
***의 작용제 화합물 선별 연구	대학교	대사성질환
*** agonist	연구기관	신경계
*** 억제제 스크리닝	대학교	항암제
***의 HTS assay system 사용	산업체	신경계
*** 조절제 발굴	대학교	항암제
*** 타겟 저해제 스크리닝	산업체	항암제
*** 세포 분석	산업체	항암제
***의 *** 저해제 개발을 위한 유효물질 발굴	연구기관	항생제
***의 과발현을 유도하는 화합물 선별	연구기관	기타
*** 저해제	산업체	항암제
*** 억제제 선별	대학교	항암제
***에서의 전산약물분석	대학교	대사성질환
*** 기능 저해 물질 스크리닝	대학교	감염증
*** 개발을 위한 후보물질 발굴	대학교	면역
*** 치료 효능 스크리닝	대학교	간염
*** 저해제를 통한 *** 조절	연구기관	항암제
*** inhibitor 발굴	연구기관	신경계
***세포 활성화 테스트	대학교	신경계
*** 저해제를 통한 *** 생성 저감	대학교	기타

활용 논문(2023년 하반기 17건)

한국화합물은행 제공 화합물 라이브러리를 활용하여 발표된 논문 목록

No	논문명	저자	저널
1	Stress induces behavioral abnormalities by increasing expression of phagocytic receptor, MERTK, inastrocytes to promote synapse phagocytosis	Youkyeong Gloria Byun, Nam-Shik Kim, Gyuri Kim, Yi-Seon Jeon, Jong Bin Choi, Chan-Woo Park, Kyungdeok Kim, Hyunsoo Jang, Jinkyeong Kim, Eunjoon Kim, Yong-Mahn Han, Ki-Jun Yoon, Seung-Hee Lee, Won-Suk Chung*	*Immunity* **2023**, *56(9)*, 2105-2120.e13
2	Identification and preliminary structure –activity relationship of brain-penetrant quinoxaline-based compounds with in vitro anti-glioblastoma activity	Seohyeon Ahn, Eun Hye Kim, Chaemi Lee, Yoon Chae Nam, Joo-Youn Lee*, Jin Sook Song*, Seong Hwan Kim*, Moon-Kook Jeon*	*Bulletin of the Korean Chemical Society* **2023**, *44(11)*, 932-938
3	Novel Quinazoline Derivative Induces Differentiation of Keratinocytes and Enhances Skin Barrier Functions against Th2 Cytokine-Mediated Signaling	Yukyung Park, Huddar Srigouri and Dongwon Kim*	*Molecules* **2023**, *28(16)*, 6119
4	Synthesis and evaluation of 2-NMPA derivatives as potential agents for prevention of osteoporosis *in vitro* and *in vivo*	Zhihao Chen, Devaneyan Joseph, Mina Ding, Swapnil Pandurang Bhujbal, Robin Prakash Sirvin Rajan, Eunae Kim, Sang-wook Park, Sunwoo Lee*, Tae-Hoon Lee*	*European Journal of Medicinal Chemistry* **2023**, *260*, 115767
5	Farnesol prevents aging-related muscle weakness in mice through enhanced farnesylation of Parkin-interacting substrate	Ju-Hyeon Bae, Areum Jo, Sung Chun Cho, Yun-Il Lee, Tae-In Kam, Chang-Lim You, Hyeon-Ju Jeong, Hyebeen Kim, Myong-Ho Jeong, Yideul Jeong, Young Wan Ha, Yu Seon Kim, Jiwoon Kim, Seung-Hwa Woo, Minseok S. Kim, Eui Seok Shin, Sang Ok Song, Hojin Kang, Rin Khang, Soojeong Park, Joobae Park, Valina L. Dawson, Ted M. Dawson*, Sang Chul Park*, Joo-Ho Shin*, Jong-Sun Kang*	*Science Translational Medicine* **2023**, *15(711)*
6	Anti-proliferative Profiling of 6,000 Representative Compounds from the Korean Chemical Bank Diversified Compound Library in De-differentiated Schwann Cells	Mohammad Ashrafuddin Khan, Se Yeon Lee, Youngbuhm Huh, Ja-Eun Kim, Kyo Seon Hwang, Ji Hyun Lee, Hwajin Lee, Yoo Lim Chun, Na Young Jeong, Junyang Jung	**대한체질인류학회** *(Anatomy & Biological Anthropology)* **2023**, *36(3)*, 103-110
7	Discovery of *(E)-2-(hydroxyimino)-N-(2((4methylpentyl)amino)ethyl) acetamide* (KR-27425) as a non-pyridinium oxime reactivator of paraoxon-inhibited acetylcholinesterase	Avinash G. Vishakantegowda, Berehe Solomon Girmay, Jin Soo Shin, Joo-Youn Lee, Sunjoo Ahn, Young-Sik Jung*	*Bioorganic & Medicinal Chemistry Letters* **2023**, *96*, 129504
8	Benzbromarone Induces Targeted Degradation of HSP47 Protein and Improves Hypertrophic Scar Formation	Jung Gyu Park, Dong Chul Lim, Jeonghwan Park, Seoah Park, Jongsoo Mok, Keon Wook Kang*, Joonghoon Park*	*Journal of Investigative Dermatology* **2023**, *online*

No	논문명	저자	저널
9	Prostaglandin F2α analogue, bimatoprost ameliorates colistin-induced nephrotoxicity	Lina Jooa, Hye Yun Jeong, Dong Hyuck Bae, Joo Hyun Jee, Woo Hee Choi, Hye-Youn Kim, Sejoong Kim, Dong-Ho Yang, Heon Yung Gee, SeongGyeong Jeon, Yun-Gil Roh, Jongman Yoo*	*Biomedicine & Pharmacotherapy* **2023**, *168,* 115446
10	Microtubule-destabilizing agents enhance STING-mediated innate immune response via biased mechanism in human monocyte cells	Songhyun Lim, Hee Ra Jung, Hyelim Lee, Yeonjeong Chu, Hyejin Kim, Eunha Kim, Sanghee Lee*	*Biomedicine & Pharmacotherapy* **2023**, *169,* 115883
11	How Deep Learning in Antiviral Molecular Profiling Identified Anti-SARS-CoV-2 Inhibitors	Mohammed Ali, In Ho Park, Junebeom Kim, Gwang-hee Kim, Jooyeon Oh, Jin Sun You, Jieun Kim, Jeon-Soo Shin,* and Sang Sun Yoon*	*Biomedicines* **2023**, *11(12),* 3134
12	What drives researcher preferences for chemical compounds? Evidence from conjoint analysis	Bongsuk Sung, Kang-Min Park, Chun Gun Park, Yong-Hee Kim, Jaeyong Lee, Tae-Eun Jin*	*PLoS ONE* **2023**, *18(11):* e0294576
13	DW71177: A novel [1,2,4]triazolo[4,3-a]quinoxaline-based potent and BD1-Selective BET inhibitor for the treatment of acute myeloid leukemia	Imran Ali, Hyung Jin Cha, Byungho Lim, Chong Hak Chae, Jihyun Youm, Whui Jung Park, Sang Ho Lee, Jung Hwan Kim, Docgyun Jeong, Jae Kyung Lim, Yun-Ha Hwang, Jae-Seok Roe, Jae-Sung Woo*, Kwangho Lee*, Gildon Choi*	*European Journal of Medicinal Chemistry* **2024**, *265,* 11605
14	Highly potent and selective phosphatidylinositol 4-kinase IIIβ inhibitors as broad-spectrum anti-rhinoviral agents	Avinash G Vishakantegowda, Dasom Hwang, Prashant Chakrasali, Eunhye Jung, Joo-Youn Lee, Jin Soo Shin*, Young-Sik Jung*	*The Royal Society of Chemistry* **2023**, *online*
15	Discovery of GABA Aminotransferase Inhibitors via Molecular Docking, Molecular Dynamic Simulation, and Biological Evaluation	Muhammad Yasir, Jinyoung Park, Yuno Lee, Eun-Taek Han, Won Sun Park, Jin-Hee Han, Yong-Soo Kwon, Hee-Jae Lee and Wanjoo Chun*	*International Journal of Molecular Sciences* **2023**, *24(23),* 16990
16	Homoharringtonine, a novel senotherapeutic, mitigates diet- and age-associated obesity and insulin resistance	Eok-Cheon Kim, Han-Byul Jung, Hye-Na Cha, Yash Patel, Ju Hee Lee, Minah Choi, Yu-kyoung Park, Soyoung Park, Il-Kug Kim, Lauren Pickel, Min-Kyung Shin, Hoon-Ki Sung*, Jae-Ryong Kim*, So-Young Park*	*Research Square Posted 19 Jul,* **2023**
17	Fluoroquinolone derivatives have a synergistic effect on the antifungal activity of echinocandins in Aspergillus fumigatus	Suzie Kang, Jin-Ju Choi, Yoonseo Lee, Dong-Hyun Lee, Cheol-Won Yun	*Research Square Posted 23 Oct,* **2023**

통합데이터플랫폼 기능 개선 소개 (https://korea.chembank.org)

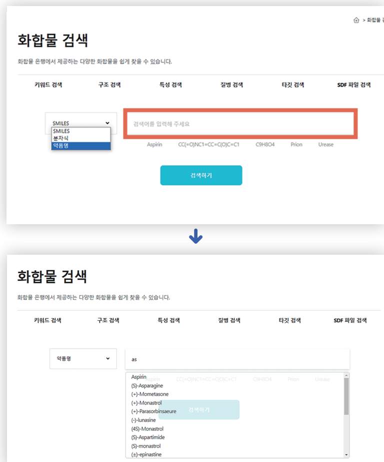
KCB 통합데이터플랫폼은 온라인에서 활용자가 직접 KCB의 서비스(화합물 검색, 선별, 기탁, 화합물 및 데이터 활용신청, 약효시험 결과 업로드 등)를 신청하거나 수행할 수 있도록 개발한 웹-플랫폼으로 지난 2019년 서비스를 시작한 이후 사용자들의 편의를 위하여 지속적으로 기능 보완 및 개선 작업을 수행하고 있다.

2023년에 개선된 기능을 아래에 소개한다.

1 검색 기능

약품명 검색 기능 보완

동일한 약품이라 하더라도 상품명, 성분명, 일반명 등의 다양한 명칭들이 존재한다. 기존 약품명 검색에서는 일반명만 검색이 되었는데, DrugBank에서 제공하는 동의어(Synonyms) 명칭이 모두 검색되도록 개선하였다. 또한 아래 그림과 같이 aspirin의 앞 두 글자(as)만 입력하면 활용자들이 많이 검색했던 순서대로 as로 시작하는 명칭들의 자동완성 리스트가 보여줌으로써 쉽게 원하는 약품명을 선택할 수 있다.



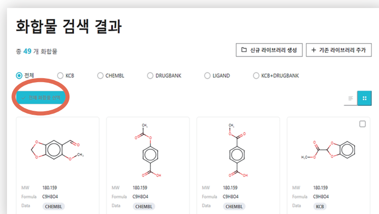
약품명 검색기능 보완

타깃/질병 검색 속도 향상

타깃 또는 질병명으로 검색 시 내부적으로 많은 테이블과 내장함수를 사용하게 되어 검색시간이 오래 걸리는 단점이 있었다. 구조 매핑 테이블을 백엔드에 미리 제작해 둠으로써 10배 가량 검색 속도를 향상시켰다.

화합물 검색 결과의 선택 및 저장 기능 개선

화합물 검색 결과화면에는 검색 조건에 해당하는 화합물이 한 페이지당 16개, 최대 1,000개까지 나타난다. 기존에는 동일 페이지에 있는 화합물에 대한 선택 및 저장 기능만 있어서 다른 페이지에 있는 화합물을 선택하여 저장할 때에는 별도의 선택 및 저장 작업이 필요하였다. 기능 개선을 통하여 검색된 화합물에 대해서는 페이지에 상관 없이 원하는 화합물을 한꺼번에 선택 및 저장할 수 있게 되었다. 또한 현재 페이지 전체 선택 기능만 아니라 검색된 전체 화합물 선택 기능을 추가하여 활용자의 편의성을 증대시켰다.



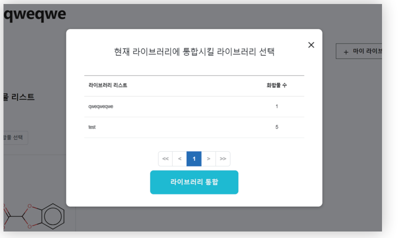
화합물 검색결과의 전체선택 기능

2 활용 신청 및 관리 기능

마이 라이브러리 통합 기능

KCB의 마이 라이브러리 기능은 활용 신청 시 각각의 마이 라이브러리가 각각의 배치로 생성되도록 설계되어 있다. 따라서 저장해 둔 마이 라이브러리를 동시에 2개 이상 활용 신청할 경우, 신청 배치도 2개 이상 생성되는 불편함이 있었다. 마이 라이브러리 통합 기능을 추가하여, 신청 시 하나의 배치만 생성되도록 기능을 개선하였다. 마이 라이브러리 메뉴에서 기존에 만들어 놓은 라이브러

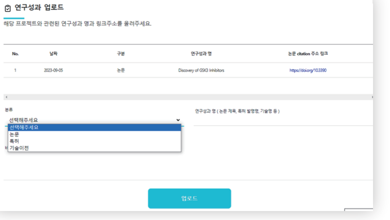
리 목록 중 한 개의 라이브러리 명칭을 클릭하면 해당 라이브러리의 리스트창으로 넘어가고, 오른쪽 위 상단에 [마이 라이브러리 통합] 버튼이 있다. 버튼을 누르면 저장된 마이 라이브러리 목록이 나타나고 통합을 원하는 라이브러리 목록을 선택하면 2개가 통합된 라이브러리가 생성된다. 추가적으로 다른 라이브러리의 통합을 원할 경우에는, 반복 작업을 진행하면 된다.



마이 라이브러리 통합 기능

KCB 소재 및 데이터 활용으로 창출된 연구성과 등록 기능

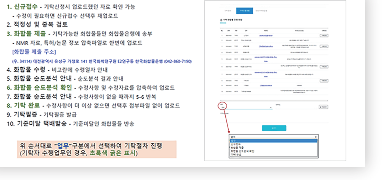
KCB 보유 소재와 데이터의 활용 성과의 관리를 위해 [연구성과 업로드] 기능을 추가하였다. 분양을 통해 발생한 논문, 특허, 기술이전 등의 연구성과를 연구자가 직접 업로드하여 관리할 수 있도록 하였다.



연구성과 등록 기능

기탁자/KCB 양방향 편집 기능

화합물 기탁 신청 및 진행의 편이성과 정확성을 위하여 [기탁진행현황] 탭에서 기탁자와 KCB의 양방향 편집 기능을 개발하였다. 이를 통해 화합물 기탁 및 검증 단계에서 요청하는 분석자료 및 정보 등의 등록과 수정, 삭제 등의 편집을 양방향에서 수행하고 확인, 검증할 수 있게 되어 기탁자와 KCB 실무자간 소통이 원활하게 되었다.


기탁자/KCB 양방향 편집 기능

3 사용자 기능 개선 및 UI 환경 개선

마이페이지 기능 개선

기존에는, 분양 신청할 때마다 소속기관 정보와 배송지 주소를 작성해야 하는 불편함이 있었다. 마이페이지에 사용자가 사용 중인 소속 정보를 등록하면, 분양 신청 시마다 소속기관과 배송지 정보를 자동으로 불러오는 기능을 추가하였다. 또한 활용자의 소속변경 시 마이페이지에서 새로운 소속정보를 등록할 수 있는 기능과, 학교와 창업회사 등 소속이 2곳 이상일 경우 주배송을 원하는 기관과 주소를 사용 중으로 등록할 수 있는 기능을 추가하였다. 로그인 화면의 우측 상단에 있는 본인의 이름을 클릭하면 마이페이지 화면으로 이동하고, 마이페이지에서 소속 정보 추가 버튼을 클릭하면 등록 팝업이 나타난다. 정보 입력 후 등록을 누르면 KCB 관리자가 기관 이메일을 포함한 기관정보 등을 확인하여 승인하면 등록이 완료된다. 등록 후 마이페이지에 나타나는 기관을 선택하고 사용 버튼을 누르면 분양 신청 시 해당 기관 및 배송지 정보가 자동 생성된다,

우측상단 본인의 **"성함"**을 누르면 마이페이지 화면으로 이동

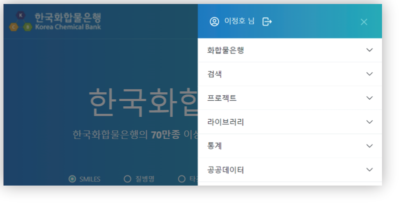
"소속추가" 선택→ "기관등록" 화면으로 이동

"기관등록" 후 목록에 추가

"기관선택" 후 사용버튼

반응형 UI 적용으로 태블릿에서도 사용 가능한 환경 구축

Bootstrap 라이브러리를 사용하고 반응형 UI를 적용하여 데스크탑이 아닌 태블릿/모바일 환경에서도 통합데이터플랫폼을 사용할 수 있도록 개선하였다.

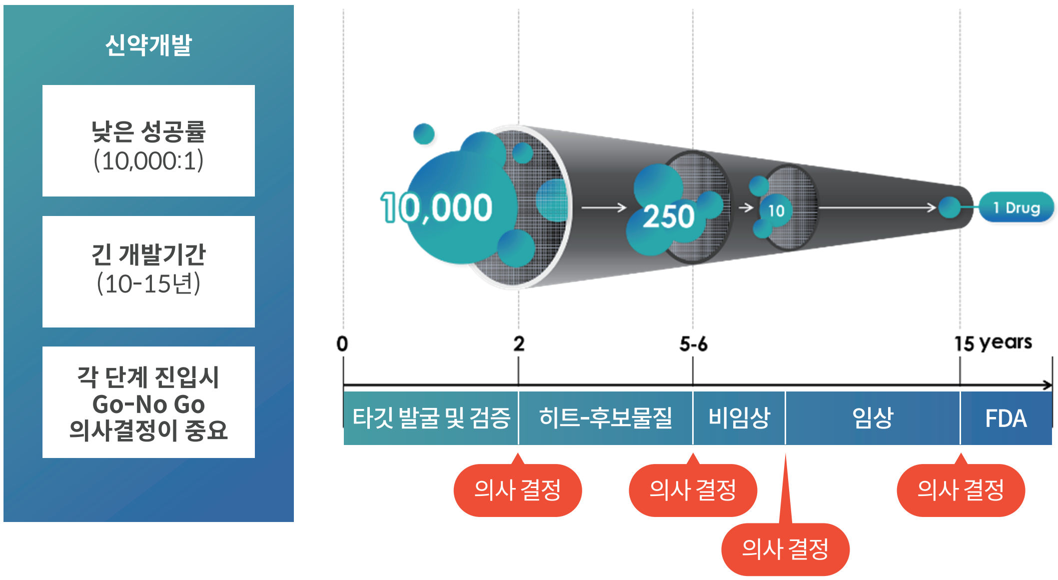


한국화합물은행(KCB)을 활용한 데이터 기반 신약연구

한국화학연구원
이선경 박사

<목차>

'만에 하나'는 실현 가능성이 아주 희박한 경우를 가정하여 하는 말인데, 합성신약개발이 바로 평균적으로 10,000개 화합물에서 1개의 화합물이 신약으로 허가를 받고 있어서, 성공확률이 매우 낮다. 허가를 받아 개발에 성공한 화합물이 시장에서 모두 성공하는 것도 물론 아니며, 개발기간도 10년 이상으로 매우 길다. 그럼에도 인류는 신약개발 연구를 통해 감염에 의한 영아사망을 거의 완벽하게 방어하였고, 고혈압과 당뇨 등 대사증후군의 관리로 평균 수명을 연장하였으며, 불치의 대상이던 암과 에이즈 등을 통제 가능한 질환으로 상당부분 전환하였다. 아직 치료제가 없는 질환도 상당수 있고, 신종 바이러스에 대한 대응태세도 갖추어야 하며, 나에게 꼭 맞는, 약효는 우수하고 부작용이 적은 개인 맞춤형 치료제도 요구되므로 신약개발은 중단될 수 없다. 신약개발의 성공률이 낮은 이유와, 개발기간 내내 수많은 의사결정이 필요한 신약연구에 데이터 분석연구를 통해 효율성을 높일 수 있는 전략들을 살펴보고, 그동안 고민해 왔던, 한국화합물은행이 기여할 수 있는 방법들을 공유하며 함께 길을 찾을 수 있기를 기대한다.



1 합성 신약개발의 성공률이 낮은 이유

최근 국내 제약기업과 바이오 벤처의 사례에서 보듯이 연구원들의 지속적이고 도전적인 노력들이 없으면 성공자체가 불가능하지만, 성공 확률이 워낙 낮다 보니 10년 이상의 축적된 기술과 노하우를 갖춘 연구팀이나 1년 밖에 안 된 연구팀이나 겉으로 나타나는 연구성과는 그다지 다르지 않게 보인다. 합성 신약개발은 실패하면 실패한 시점의 결과에서 연구를 계속하여 성공에 다가가는 것이 아니라 처음부터 똑같이 다시 시작해야 하는 "비가역성(Irreversibility)"의 특징이 있다. 연구 초기단계이든 임상 1상이든 임상 3상에 있든 과제에 문제가 생기면 다시 시작해야 하고 그때까지의 경험과 학습이 더 좋은 화합물을 고안하는데 획기적인 도움이 되지 않는다. 새로 시작해야 하니 무형 자산으로 가치 측정이 불가능한 상태로 머물러 있던 업적은 실패가 확인되는 순간 사라지고 이 속성을 "휘발성(Intangible, perishable nature of assets)"이라고 한다.[1]

신약개발은 왜 이렇게 어려운 것일까? 근원적으로 합성신약 연구가 어려운 가장 큰 이유는 서로 독립변수인, 아니 일반적으로는 서로 상충하는 다음의 조건들을 만족시켜야 하기 때문이다. 후보 화합물 하나를 열심히 만들었을 때 그 화합물이 약효, 독성, 약동력학적 특성, 물성의 목표를 충족할 확률을 각각 P1, P2, P3, P4 이라고 하자. 약은 이 4가지가 동시에 충족되어야 승인을 받을 수 있기

때문에 최종성공확률 (P) = (P1. 약효가 좋을 확률) * (P2. 독성이 없을 확률) * (P3. 흡수, 대사, 분포, 배설이 좋을 확률) *(P4. 물성이 좋을 확률)이 될 것이다.

P1 & P2: 약은 당연히 약효가 좋고, 독성이 없어야 한다는 것은 누구나 알고 동의한다.

P3: 그런데 약을 하루에 6번 먹을 수는 없다. 최악이라도 아침, 점심, 저녁, 취침 전 4번이고, 보통이면 아침, 저녁 그리고 이상적으로는 하루 한 번이 가장 좋은데 이걸 결정하는 것이 3번 확률이다. 또한 같이 복용할 때 다른 약이 흡수가 안되거나 또는 너무 흡수되어 부작용이 나타난다면 위험한 약으로 간주되어 개발을 포기하게 만드는 요인, 즉 약물 상호작용에 영향을 주는 것도 3번이다. 말기 신부전환자나 간 경화증 환자가 복용할 수 있는 약인지 없는 약인지를 결정하는 것도 3번이다. 즉 P3는 약의 상품성을 총괄 대변한다.

P4: 연구단계에서 간과하기 쉽지만 4번 용해도와 세포막 투과도 등의 물성이 나쁘면 개발이 정말 어렵다. 약제학적 분류에서 "물에 대한 용해도가 높고, 위장관 투과도도 높은 클래스 1"에 해당하는 약은 꿈의 물성이다. 용해도도 낮고 투과도도 낮은 클래스 4에 해당하는 약은 임상에서 통계적인 유의성을 확보하기 매우 힘들다.

약제학적 약물 분류

용해도가 높으면 흡수(투과도)가 나쁘고, 흡수가 좋으면 물에 잘 안 녹는 경향이 있다. 이 저울추의 중심에 원하는 약효를 구현하는 물질을 개발하는 것이 의약화학자의 숙명이다. 약이 사람 몸 속의 위장관을 통과하고 목표 장기의 세포막을 뚫고 들어가려면 특수 채널이 아닌 이상 일단은 지용성이어야 한다. 물질은 유유상종이라 지용성인 세포막을 뚫고 들어가려면 약도 지용성이어야 하고, 이렇게 "기름에 잘 녹는 지용성"이면 당연히 물에 잘 안 녹는다. 약효 만을 쫓아서 화합물을 만들다 보면 용해도가 낮고 흡수도도 천차만별인 물질만 골라서 만드는 경우가 대부분이다. 후보물질로 선정되면 본격적인 약리, 흡수-대사-분포, 독성 실험 등이 실험동물을 대상으로 시행된다. 독성을 알려면 약효 용량의 수십배까지 용량을 늘려가며 일단 "녹여서", 쥐나 개에게 먹이든 정맥주사를 하든 투여하여야 치사용량과 약효용량의 비율인 안전역을 구할 수 있다. 약효도 약효-용량 곡선을 얻기 위해 고용량 투여가 반드시 수반된다.

그런데 약이 물에 잘 안 녹는다면? 실험하는 연구자들이 괴롭다. 특히 입원환자를 위한 주사제도 동시에 개발해야 하는 약이라면 용해도는 경제성 평가의 아주 중요한 고려 요소가 된다. 1 mL 주사 한번 맞으면 되는 것을 1 L 링거 병으로 장시간 투약하여야 하는 약은 시장성이 떨어지고 출시되더라도 편의성이 떨어져 그만큼 매출이 줄어들기 때문이다. 1,000여개의 FDA승인 신약을 물성에 따라 분류한 논문에 의하면 클래스 1(37%), 클래스 2(31%), 클래스 3(26%), 클래스 4(6%) 순으로 대략 배분된다고 하는데,[2] 최근 들어와 신약의 분자량이 점점 커지고 지용성이 높아져서 물에 대한 용해도가 떨어지는 클래스 2가 많아지고 있다. 가장 큰 문제는 이 P1, P2, P3, P4가 독립변수라는 것이다!

즉 약효가 좋다고 독성이 없다는 법이 없다. 독성이 없다고 흡수 대사 분포 배설이 좋다는 법이 없다. 약효가 좋다고 물성이 좋다는 법이 없다. 오히려 약효와 물성의 관계는 태생적으로 반비례 관계가 더 보편적이다. 그리고 각각의 확률이 매우 낮다. 예를 들어 각각 10% 확률이라도 최종 성공확률은 0.01%가 된다. 즉 신약개발은 만에 하나의 확률을 안고 시작하는 것이다.

신약개발이 종합학문인 것도 성공률이 낮은 원인 중의 하나일 것이다. 신약연구개발은 크게 연구부문과 개발부문의 두 파트로 나누어진다. 연구부문은 비임상 개발 후보를 도출하는 파트이고, 개발부문은 후보물질을 가지고 비임상 임상시험을 거쳐 허가를 받는 파트이다. 연구부문도 개별전공으로 들어가면 분야별로 과학적 토대가 전혀 다르다. 대부분의 생물학, 약리학, 독성 연구원들에게 의약합성 자료를 들여다보라고 하면 최종 화합물의 결과만 볼 뿐이지 내용을 이해하지 못한다. 그런데 이걸 모르면 합성신약 프로젝트의 진척을 파악하기 어렵고 과제 전체의 조망과 평가가 어렵다. 반대로 합성 연구원은 생화학, 약리, 세포와 동물 실험, 독성, 약물동력 자료를 그냥 결과가 좋다 나쁘다 수준으로만 이해하는 경우가 많다. 앞서 설명한 용해도와 투과성이 나쁜 화합물들의 경우에는 더군다나 세포나 동물의 개체차이가 있는데 작은 차이에 의해 실험결과의 재현성이 나오지 않을 가능성이 크다.

재현성이 없을 경우, "화합물의 색이나 입자도 등의 외형이 다르다", "생물실험에 무언가의 실수가 있었을 것이다" 하는 서로 다른 해석은 아마 많은 연구자들이 경험했을 것이다. 임상을 알아야 신약개발 전략이 제대로 정립된다고 주장하는 후단의 개발부문 연구원들도 전단의 연구내용을 파악하는 경우가 드물고, 연구부문 연구원들이 각 임상 단계 별 연구목표, 임상 디자인이나 환자 모집 난이도, 임상통계 등등을 아는 경우도 드물다. 허가를 받고 최종적으로 보험 약가를 받아야 시판을 할 수 있는데 이 과정을 이해하는 연구원은 극히 드물다. 같은 신약 연구분야인데도 치매 연구자와 바이러스 연구자, 심혈관질환 연구자가 상호 연구문제에 대해 논의하고 교류하는 것은 쉽지 않다.

이런 전문분야 장벽으로 인해 종합학문의 통찰이 필수인 프로젝트의 Go-No go 의사결정이 쉽지 않고 효율성 저하의 원인이 된다. 특히 학문간 진입장벽이 높고, 상호 협력하는 문화가 없는 경우에는 신약개발이 거의 불가능하다고 생각되고, 실질적으로도 선진화된 문화기반이 없는 국가의 신약개발 성과는 거의 전무한 것으로 나타난다. 미국 FDA 승인신약 보유국가는 미국, 유럽, 일본등의 신약 선진국 이외에, 2023년 10월 27일 중국의 '준시 바이오사이언스(Junshi Biosciences)'가 미국 '코헤러스 바이오사이언스(Coherus Biosciences)'와 공동개발한 PD1 면역관문억제 항암제로 처음으로 미국 FDA 승인을 얻기 전까지는 한국이 유일한 국가였다. 유럽의 각개국은 모두 단일국가와 마찬가지인 시스템으로 운영되므로 인구가 1억명이 되지 않아 자체 내수시장의 한계가 있는 나라 중에서 글로벌 신약을 만든 유일한 나라가 대한민국이다.

신약 연구의 불확실성, 휘발성, 장기 진행과 같은 근본적 속성은 수십년 전에도 존재했던 특성이었다. 따라서 이런 배경을 제외하고 생명과학 및 주변 기술의 눈부신 발전과는 달리 어떤 사유로 인해 2000년대 초반부터 신약개발이 특별히 더 어려워지고 투자 대비 생산성 저하라는 단어가 생겨났을까? 조금 무책임한 답변 같지만 "만들기 쉬운 약은 이미 다 만들었다."는 것에 공감하는 연구자들이 많이 있을 것이다. 연구하기 어려운 질환, 예를 들면 치매처럼 노화인지 질병인지를 구분하기 어렵고 그 원인도 개인에 따라 다를 뿐만 아니라 여러가지 유전자가 복합적으로 관여하는 만성, 난치성, 노인성 질환만 남게 된 것이다. 블록버스터 약들의 특허만료까지 도래한 거대 제약사들이 어려움을 돌파하기 위한 마땅한 수단을 찾지 못하고 연구의 혁신을 이루지 못하여, 인수합병을 통해 임기응변 식으로 대응한 것도 현실이었다.[3] 단순한 미국 FDA 승인 신약의 숫자 외에 연구비 투자 대비 가성비로 본다면 이론적 연구가 아닌 관찰연구 또는 현상연구에 기반하여 대부분의 신약을 도출한 1900년대 중반이 신약연구의 황금기였다.

즉 항생제나 면역억제제 등과 같이 정확한 인체 내 타깃을 모르는 상태에서 세포나 장기, 또는 동물 실험의 효과만 보고 승인을 받은 약이 많았다. 생화학적 실험없이 동물실험과 장기실험으로 소수의 화합물을 테스트하고 혁신적인 다수의 약을 만들었다. 혹은 이미 화합물의 특징으로 약효를 유추할 수 있는 선행지식을 기반으로, 예를 들면 핵산 유도체는 바이러스 약이나 항암제로 개발가능하다는 식의 연구가 진행되었다.

1970년대에 이르러 이런 연구방식은 서서히 생산성이 저하되었고 생화학적 연구가 뒷받침이 되자 타깃기반 연구로 선회하였다. 우연을 바라면서 혹은 호기심 기반의 창의적 개인적연구에서 벗어나 이론적 신약개발인 타깃기반 연구가 시작되었다.

이런 타깃기반 연구는 다시 생산성을 높였으며, 특히 ACE 저해제인 Captopril 과 같은 고혈압 치료제는 이론적 신약연구의 대표적 성과물이라 할 수 있겠다. 이상지질혈증, 당뇨병 등 유병률이 높은 질환의 블록버스터 신약들로 다국적 제약사는 호황기를 누렸다.

높은 영업이익률의 신약을 보유한 다국적 제약사는 막대한 연구개발비 투자를 통해, 사람의 유전자 연구로 도출된 수많은 타깃들에 대하여 조합화학 및 화합물 라이브러리와 고효율 스크리닝(HTS, high throughput screening)의 기술을 이용하여 공격적인 연구를 진행하였다. 바코드 튜브의 등장과 자동합성 및 보관 장비와 Liquid Handler 등의 자동화 장비가 대규모 약효 스크리닝의 속도를 엄청나게 향상시켰다.

그러나 유전체 연구에서 도출되어 아직 임상에서는 검증되지 못한 많은 타깃들은 실패확률이 임상에서 검증된 타깃보다 40% 정도 높았다.[4,5] 특히 1999~2008년 사이 승인을 받은 신약을 분류해 보면 동급최초 혁신신약은 현상기반연구(function based research) 37%, 생물의약 33%로 타깃기반 합성신약 23%를 압도하였다. 반대로 추종신약은 타깃기반 연구에서 도출된 신약이 51%로 압도적 우위를 점하였다. 화합물이 앞서 언급한 여러 조건들을 만족시키지 못하여 개발에 실패하면 화합물 구조의 디자인과 합성부터 다시 시작해야 하듯이, 타깃이 조건을 충족하지 못하여 실패하면 새로운 타깃에 대한 연구로 방향을 전환하여야 한다.

화합물이 실패한 경우, 타깃의 약효 및 약리연구에 대한 기술은 지속적으로 활용이 가능하지만, 타깃이 실패한 경우에는 타깃 연구를 위해 개발된 스크리닝 기술도 무용지물이 될 수 있어서 타격이 더 클 수도 있다. 이후, 임상에서 유용한 타깃 검증을 위한 중개연구가(translational research) 활성화되면서 사람의 유전체를 조절하는 모든 타깃에 대한 연구에서 약물로 질병과 관련된 증상의 조절이 가능한 타깃을 선정하여 진행하는 방향으로 연구전략이 전환되었다.

그 이외에도, 심의기관의 약물 독성 검증 강화, 만성 질환 치료제의 임상 실험 시 임상 환자수 증가와 기간 연장에 의한 연구개발비 증가, 약가 인하 압력 등 생산성을 저하시키는 환경들이 있다.

2 신약개발 효율성 제고 전략-데이터 기반 연구 현황

약으로 조절되지 않으면서 고통을 주는 질병들이 다수 존재하고, 신종 바이러스가 언제 출몰할지 모르며, 결핵과 항생제 내성은 아직도 숙제이므로 건강한 인류사회를 위해 신약개발은 필수적이다. 효율성을 높일 수 있는 다양한 방법들이 있겠지만 데이터 기반 연구 전략에 대해 살펴보겠다.

화합물과 타깃이 조건을 충족하지 못하여 실패하면 처음부터 다시 시작해야 하는 비가역성과 해당 연구 과정에서 축적된 성과의 휘발성이 신약개발의 속성이라고 하였다. 하지만 실패한 화합물과 타깃에 대하여 평가된 데이터들이 후속 신약개발 연구에 활용된다면 분명하게 후속연구의 성공률 향상에 기여할 수 있을 것이다. 또한 개인이나 개별 연구팀의 데이터로 한정되는 것이 아니라 데이터 공유를 통해 다른 연구원과 연구팀의 경험도 활용할 수 있으면 효율성은 배가될 것이며, 인공지능 알고리즘의 발전과 함께 증폭될 수 있을 것이다. 아래에 세계적으로 유명한, 공개되어 있는 의약정보관련 데이터베이스를 소개한다.

(1) 데이터

· 공개 화합물 데이터 저장고(2015년)[6]

DB	화합물 (종수)	실험 데이터		주요 데이터
		건수	실험 종류	
ChEMBL v.21	1,592,191	13,968,617	1,212,831	PubChem HTS, 문헌 발췌 데이터
BinidngDB	529,618	1,207,821	6,265	단백질-저분자화합물 결합 실험데이터
PubChem	> 60M	> 157M	> 1M	HTS 수행한 Bioactivity 데이터
Reaxys	> 74M	> 500M	-	물성, 활성 및 반응 논문 데이터
SciFinder (CAS)	> 111M	> 80M	-	실험 반응, NMR 데이터
GOSTAR	> 3M	> 24M	> 5K	특허 및 문헌의 타깃 관련 데이터
AZ IBIS	-	> 150M	-	AstraZeneca 내부 SAR 데이터
OCHEM	> 600K	> 1.2M	> 400	문헌 수집 ADMET 데이터

· 공개 생물 및 임상 데이터 저장고[7]

정보 종류	데이터베이스
생물 현상	CTD(Comparative Toxicogenomics DB), Entrez Gene, Gene Ontology
단백질	Bgee, CTD, DisGeNET, DrugBank, Entrez Gene, UMLS, Human Phenotype Ontology, Human PPI, Network, Reactome
질병	CTD, DisGeNET, Disease Ontology, Drug Central, Mayo Clinic, Human Phenotype Ontology, MONDO Disease Ontology, Orphanet
Phenotype (현상)	DisGeNET, Human Phenotype Ontology, SIDER
Anatomy	Bgee, UBERON
분자 기능	CTD, Entrez Gene, Gene Ontology
약물	DrugBank, Drug Central, SIDER
세포 조성물	CTD, Entrez Gene, Gene Ontology
신호전달 경로	Reactome
노출	CTD

· 공개 유전자-단백질-약물 상호관계 분석 데이터베이스[8]

정보 종류	데이터베이스
DTI	ChEMBL, STITCH (search tool for interactions of chemicals)
PPI	STRING (Search Tool for the Retrieval of Interacting Genes), HPRD (Human Protein Reference Database)
DDI	Online Mendelian Inheritance in Man (OMIM)
DTI, PPI	Comparative Toxicogenomics Database
DTI, PPI, 신호전달	KEGG (Kyoto Encyclopedia of Genes and Genomes)
DTI, DDI	PubChem (NCBI), DrugBank
신호전달 경로	Reactome
생물학적 경로	GeneGo MetaBase
암 유전체	Cancer Genome Atlas (TCGA), METABRIC (Molecular Taxonomy of Breast Cancer International Consortium)
유전자 조절, 상호작용	GEO (Gene Expression Omnibus), GO (Gene Ontology), CCLE (the Cancer Cell Line Encyclopedia)
유전자 발현 및 전사	UniGene (NCBI)

DTI: drug target interaction **DDI**: Drug-Drug Interaction **PPI**: Protein-protein Interaction

· **데이터 정보 연계**

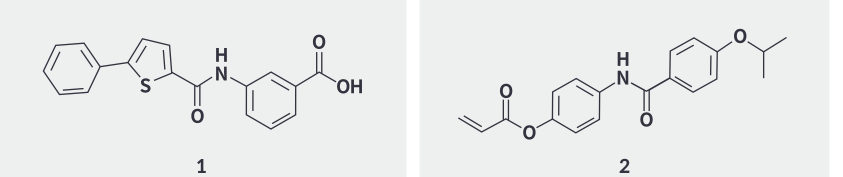
각기 다른 종류의 DB로부터 정보를 추출하여 상호관계를 연계함으로써 통합 정보를 활용한다. 대표적인 예로 Havard-MIT연구진이 세계적으로 저명한 20개 DB에서 추출된 17,080건의 질병과 4,050,249건의 상호관계를 분석하여 precision medicin을 연구할 수 있는 정보관계망인 ‘PrimeKG’를 개발하여 공개한 바 있다.[7]

· **데이터 표준화**

질병 분류의 경우에도 목적에 따라 비체계적으로 중첩되고, 종종 서로 충돌도 있으며, 표준화된 질병 표현방법이 없다. 세계 보건기구(WHO)의 국제질병분류(International Classification of Diseases, ICD)는 사람의 질병 및 사망원인에 관한 표준 관리규정으로 의료비 청구에 최적화되어 있다.[7] MedGen, PhenoDB, Orphanet 등은 희귀질환 또는 유전 질환에 특화되어 있다. 일례로 자폐 스펙트럼 장애(autism spectrum disorder)는 진단을 고려한 용어인 반면, 임상 증상과 연계된 많은 아형들이 있는데 Unified Medical Language System (UMLS) 에서는 193 아형이, Monarch Disease Ontology (MONDO)에서는 37개의 아형, Orphanet에는 6개 질환이 존재한다. 따라서, 연구자들이 본인들의 목적에 맞도록 최적화하여 활용해야 하는 것이 현재 상황이다.

(2) 화합물 구조 디자인

화합물의 물리화학적 특성 및 약효 정보를 분석하여 구조 디자인 모델 연구 즉 생성형 인공지능(Generative AI) 모델 연구가 활발히 진행되고 있다. 아래 그림에 나와 있는 화합물 1,2는 생성형 인공지능 모델(deep LSTM network)에 의해 디자인된 화합물로 실제 합성하여 약효시험을 수행한 결과 우수한 활성을 나타내었다.
(화합물 1: EC_{50} = 0.06 μM as RXRγ agonist, 화합물 2: EC_{50} = 19.1 μM)[9]



1 2

생성형 인공지능에 의한 화합물 디자인[9]

(3) 약물-질병-단백질-유전체 연구

단일 분자 단위의 오믹스(omics)와 약물로부터 출발하여 질병과의 관련성을 규명하는 연구들을 역시 생성형 인공지능 모델을 활용하여 활발하게 연구 중에 있다. 대표적인 예로 질병과 약물 영향 분석을 omics data layer를 통하여 분석하고 computational systems biology를 이용하여 질병 모델링과 조절에 관한 연구 및 약물과 타깃 상호작용 분석에 의한 병용투여와 약물 재창출 기회에 관한 연구 등을 소개한 총설 논문이 2022년에 발표된 바 있다.[8]

3 데이터 기반 신약연구를 위한 한국화합물은행의 역할

한국화합물은행이 국내 데이터 기반 신약개발 연구에 현재까지도 일정부분 기여하였으며 향후 그 역할이 커질 것도 확실하지만 데이터의 수집과 관리, 표준화, 분석을 위한 분양 시스템의 선진화와 전략에 대한 고민이 필요하다. 2023년, 인력 자원을 포함하여 국내 인공지능 신약연구의 인프라와 관심도를 알아보기 위하여 KCB 보유 화합물을 활용하여 생산한 데이터를 제공하고, 과학기술정통부와 보건복지부의 후원으로 AI 신약개발 경진대회를 한국화학연구원, 한국생명공학연구원과 한국제약바이오협회가 공동 주최하였다. 예상을 훨씬 뛰어 넘는 총 1,447명(1,254팀)이 참가하여 크게 흥행함으로써 국내 인공지능 신약개발 연구에 대한 관심과 가능성을 확인할 수 있었다.

(1) 2023 AI 신약개발 경진대회

· 경진대회 개요

KCB가 2018년부터 2022년까지 대표 라이브러리 중 3,981종 화합물을 활용하여 생성한 사람(HLM, human liver microsome) 및 마우스의(MLM, mouse liver microsome) 간 마이크로좀(대사효소)에 대한 안전성 데이터를 제공하여 AI 신약개발 경진대회 개최하였다. 3,498종의 학습용 데이터(training set)는 모델구축을 위해 화합물 구조와 대사 안정성 데이터를 모두 제공하고, 483종의 경진용 데이터(test set)는 화합물 구조만 제공하여 구축된 모델의 대사안정성 예측 값을 결과물로 제출하도록 하여 실측 값과의 근제곱 평균오차(RMSE, root mean square error)로 모델의 성능을 평가하였다.

· 대사안정성(metabolic stability) 데이터

대사안정성 데이터는 약의 체내 노출 농도에 영향을 주는 핵심 정보이다. 약물 대사반응은 크게 Phase I 과 Phase II 로 나뉘는데 Phase I 은 약물이 위장관에 흡수 후 혈액으로 이행하기 전에 간의 hepatocyte에서 주로 Cyp 효소에 의해 산화, 환원, 가수분해 등을 일으켜 극성물질로 대사되어 혈액으로 흡수되지 않고 몸밖으로 배설되게 하는 대사반응이다.[10] 경진대회에 제공된 데이터는 사람과 쥐의 간 microsome과 화합물을 37 °C에서 30분 동안 반응시킨 후, 간 대사효소에 의해 대사되지 않고 남아있는 화합물의 양을 LC-MS-MS로 측정한 값(%)이다. In vitro에서의 약동력학적인 여러 요인에 의해 달라질 수 있지만, 일반적으로 30분 반응 후, 남아있는 화합물의 양이 50% 이상이면 비교적 안정한 화합물이다.

< 대사안정성 판단기준 >

30분 반응 후 남아있는 양(%)	판단
>90	반감기가 3시간 이상으로 예상되는 매우 안정한 화합물
70~90	반감기가 1-3시간 이상으로 예상되는 안정한 화합물
50~70	반감기가 30-60분으로 예상되는 비교적 안정한 화합물
30~50	반감기가 15-30분으로 예상되는 비교적 불안정한 화합물
<30	반감기가 15분 이하로 예상되는 불안정한 화합물

대사안정성 실험에 활용된 KCB 화합물 약 4,000종은 대표 라이브러리에서 추출하여 구조적 다양성이 확보되어 있고, 한국화학연구원의 송진숙 박사 연구팀에서 동일 조건으로 실험하였으며, 실험 결과, 50%이상 되는 안정한 화합물이 마우스에서는 37.2%, 사람에서는 55.8%로 종간 차이가 있기는 하지만, 두 종 모두 불안정과 안정한 화합물의 비율이 크게 편향되게 나타나지 않아서 데이터 분석에 적합할 것으로 판단되었다.

1) 대회 준비

· **대회 기획('23년 1월~3월): 주관기관의 협의로 경진대회 기본 기획안 도출**

- 주최: 한국화학연구원, 한국생명공학연구원, 한국제약바이오협회
- 주관: 한국화합물은행, 국가생명연구자원정보센터, 인공지능신약개발지원센터
- 후원: 과학기술정통부, 보건복지부
- 운영 시스템: 리더보드(실시간 순위 차트)가 운영되고, 회원수가 많은 ㈜데이콘(https://dacon.io)에 위탁
- 운영위원회 구성: 각 주관기관 책임자, 데이터 도메인과 분석 전문가 총 9인
- 일정, 예산, 홍보 등에 대한 기본 기획안 도출

· **운영위원회: 4회 개최 (04.04, 05.03, 06.13, 07.03)하여 대회 일정 및 규정, 방법 결정**

- 대회일정 및 기간: 하기 방학 휴가 영향을 고려하여, 2학기 개강 전에 실시하는 것이 참여 유도에 바람직함.
- 평가방법: 평가모델(회귀 또는 binary 분류), 정성 및 정량 평가 방법 및 비율
- 데이터셋 구성의 적절성 검토
- 참여대상, 홍보방법, 수상인원 및 상금

· **데이터셋 구성: 운영사인 ㈜데이콘이 수행하고, 최종 운영위에서 결정. 데이터 셋 구성을 바꾸어 가며 베이스 모델로 점수 실험을 하여, MLM과 HLM 값 회귀 예측 시 어느정도 예측성능이 나올 수 있도록 데이터셋을 구성하였음.**

- 화합물 유사도를 기준으로 군집화를 진행하여 Training data set과 Test data set을 0.87:0.13의 비율로 구성
- GNN (Graph Neural network) 계통의 4개 베이스 모델을(GAT, GCN, MPNN, Attentive FP) 바탕으로 test data set에서 오차가 큰 outlier(MLM과 HLM 각각에 대해 오차 절대값이 전체 오차에서 삼분위수 이상인 경우)의 70%는 training data set으로 옮겨서 구성
- 최종 데이터셋 구성(training data set: 3,498종, test data set: 483종) 및 베이스 모델로 성능 확인

2) 대회 규정 및 진행

· **대회일정**

- 대회기간: 08.07 ~ 09.25(하루 3회까지 결과 제출 가능)
- 코드 및 발표자료 제출: 10.06 (발표평가 대상팀, 수상팀의 2배수)
- 온라인 발표평가: 10.13, 코드평가: 10.06 ~ 10.23, 최종 수상자 발표: 10.23

· **대회규칙**

- 외부데이터 활용: 출처가 있는 외부 데이터 활용이 가능하며, training data로만 활용해야 함.
- 회귀모델(Regression), 평균 제곱근 오차(RMSE, Root Mean Square Error)
- 리더보드 평가산식: 0.5*RMSE (MLM) + 0.5*RMSE (HLM)

Public Score: 전체 테스트 데이터의 약 35%, Private Score: 전체 테스트 데이터
- 평가방법: 정량평가 점수와 정성평가 점수(발표평가)를 6:4로 반영

평가항목	심사기준	점수
리더보드 private 점수	모델 예측 성능	60
모델링 전략	도메인을 고려한 모델 활용	10
데이터 전처리 및 validation 데이터 구축 전략	데이터 특성 고려한 전략	10
실험관리	다양한 실험 시도와 논리성	10
적용성(모델의 범용성)	모델의 다른 분야 적용 가능성	10

· **참여대상: 한국에 거주하는 누구나(외국인 포함) 및 재외 외국인**
· **운영사(데이콘) 제공 베이스라인 모델: Morgan Fingerprint + MLP (RMSE = 33.8433)**

3) 대회 결과 및 시상행사

· **1,447명(1,254팀) 참여, 764팀 결과 제출**

- RMSE 30 이하: 265팀
- 대회마감 전 10일 동안의 리더보드 순위변동: 상위그룹은 큰 변동이 없었지만, 일부 변동은 있었으며, 특히 도메인 분야 데이터 분석 경험자들이 후반으로 갈수록 약진하는 경향이 있었음.

· **수상자 명단**

순위	표창	인원	상금
대상	과기정통부 장관상: datu (중앙대, 아주대) 복지부 장관상: suleezard (성균관대)	2인(팀)	1,000만원/각팀
우수상	한국화학연구원 원장상 : 실력으로 말해(GIST) 한국생명공학연구원 원장상: 약과 도넛((주)HITS) 한국제약바이오협회장상: 다이머((주)스탠다임)	3인(팀)	300만원/각팀
	합계	5인(팀)	2,900만원

· **시상식 및 연계 행사**

- 2023 대한민국 과학기술대전: “빅데이터·인공지능(AI) 융합 첨단바이오 심포지엄 - 화합물신약분야”
  시상: 과기정통부장관상, 화학(연)원장상 및 생공(연)원장상
  연계행사: 수상자 구두발표 및 포스터 발표, 초청강연, 간담회 등 (11.09, 국립과천과학관/창조홀, 15:00~17:20)

- AI Pharma Korea Conference 2023: (11.02, 몬드리안 서울 이태원, 9:30~17:50)
  시상: 복지부장관상, 한국제약협회장상
  연계행사: 포스터 발표, 기조강연 및 특별강연, 세미나

4) 대회 후기

전공자가 아닌 경우에는 화합물 구조와 대사안정성 데이터가 생소하고, 특히 SMILES로 처리된 구조에 대한 이해가 쉽지 않기 때문에 대회 개최 전에는 참가팀이 적지 않을까 우려 했었는데, 첫날에만 100팀이상이 참가 신청을 하였고 최종 1,447명의 참가 하였다. 실제로, 경영학 전공자가 추가 데이터를 PubChem에서 추출하여 모델구축에 활용하고, 수소결합 공여체와 수여체 및 Log-D, 극성표면적 등 화합물의 특성(feature)를 추출하고, 오픈소스인 Uni-Mol로 3D 구조를 예측하고, ChemBERTa와 xgboost 등을 활용하여 높은 성능의 모델을 개발하기도 하였다. 대사안정성 데이터에 대한 지식과 화학구조 특성과의 상관성을 ChatGPT를 활용한 학습도 시도하였다.

전공이 다른, 전혀 일면식이 없는 참가자들에게 서로 연락하여 함께 팀을 만들고 협동함으로써 좋은 성과를 내는 것을 보며 신세계를 경험하였다. 대회기간 내내 설레는 마음으로 리더보드를 확인하였고, 코드와 경험을 공유하는 현장을 바라보며 선진화한 문화에 뿌듯하였다. 발표 평가 대상자가 아니었음에도 대회 참가하여 학습하고 경험한 것을 정리하여 ppt 파일을 만들어 공유한 참가자도 있었다.

데이터에 대한 질문들을 보며, 데이터도 소재와 마찬가지로 많이 활용하여야 다듬어지고 빛이 날 수 있다는 평범한 진리도 다시 느꼈다. 생각만큼 후반부로 갈수록 모델 성능이 개선되지 않은 것은 절대적으로 데이터의 양이 부족해서인 것 같다. 4,000종 화합물의 대사안정성 데이터를 얻기 위해서도 상당한 예산과 노력이 투자되었으므로 과학 데이터를 빅데이터 수준으로 확보하는 것은 불가능하고, 데이터 증강(augmentation)과 분석기술의 획기적 개선만을 기다려야 하는지 아쉬운 부분이다. 대회 준비와 연계행사를 위해 많은 노고를 하신 주관기관의 연구원들, 운영위원들 및 최종 평가위원들, 과기정통부 첨단바이오기술과 과장님과 사무관님께 감사인사를 드리고, 대회가 매끄럽게 진행하도록 운영을 맡아준 ㈜데이콘 관계자분들께도 감사드린다.

- **대회 준비: 참가율을 제고한 대회 일정 조율, 양성 및 음성 데이터의 비율 확인, 구조 다양성 확인, 베이스라인 모델을 이용하여 경진대회에 적합하도록 training data set과 test data set의 구성, 난이도를 고려하여 회귀모델(regression) 및 binary 분류(classification) 모델의 결정 및 평가방법(R^2 또는 RMSE, MLM과 HLM의 반영 비율, 정량 및 정성 비율 및 항목) 확정, 리더보드의 public score 비율 확정(대회기간 중 리더보드의 순위를 나타내는 test data set 중 35%의 데이터에 대한 성능인 public score 점수에만 맞추어 모델을 최적화할 경우 35%에만 과적합(ovefitting)되어 대회 종료 후 전체 test data set에 대한 성능인 private score는 좋지 않을 수 있음), 정성평가 기준 등 대회 준비를 하면서 문제를 내는 것도 푸는 것 만큼 쉽지 않고, 많은 사항들을 고려해야 한다는 것을 알았다. ㈜데이콘의 경진대회 운영 경험이 큰 도움이 되었다.**

- **데이터 이슈: 대회기간 참가자들의 데이터 관련 질문 및 이슈를 보며 데이터는 많이 활용되어야 정제될 수 있음을 느꼈고, 참가자들이 현명하게 데이터 전처리를 수행함을 알았다.**
  - 편차가 있는 중복 데이터: 동일 화합물에 대하여 다른 배치로(동일 배치는 3회 반복 실험이며, 평균값만 제공하였음) 2회 실험한 데이터가 학습용 데이터에 소수 포함되어 있었고, 그 중 일부는 배치간 편차가 컸다. 대회를 준비하며 편차가 있는 중복 화합물의 데이터 처리에 고민이 있었지만, 생물 데이터에 다양한 오류를 반영한 이러한 편차는 늘 발생할 수 있는 것이고, 실험 과정 중에 발생할 수 있는 오류와 편차를 포함하여 데이터를 전처리하는 방법도 데이터 분석에 중요하기 때문에, 문제에 포함하

기로 결정하였다. 발표평가 대상자들은 편차가 있는 데이터를 매우 논리적으로 처리함을 확인할 수 있었다. 일부는 학습 데이터에서 모두 제거하고 모델을 구축하였고, 일부는 데이터를 제거하고 모델을 구축한 후에 모델에 적용하여 더 잘 맞는 데이터를 포함시킨 후 다시 모델을 만들었다.

- 100%가 넘는 대사안정성(30분 배양 후 대사되지 않고 남아있는 화합물의 양%) 데이터: 논리적으로 100%가 넘는 값은 나올 수가 없으나 실험에서는 발생할 수 있고 역시 실험값을 보정없이 그대로 제공하였으며 관련된 질문이 있었다. 다음과 같이 답변하였고, 앞에서 언급된 용해도가 좋지 않은 화합물의 경우 재현성이 떨어지는 이유 중의 하나이다. (답변: 간혹 용해도가 극히 좋지 않으면서 대사안정성이 높은 물질의 경우, 37 °C에서 30분간 반응하며 주어진 시간동안 실험 전보다 용해도가 더 증가하여 반응 후 % remaining값이 반응 전 대비 더 큰 값으로 계산되는 경우도 발생될 수 있습니다. 모든 실험값들은 본 상황과 같은 문제를 포함하여 편차가 있을 수 있는 다양한 가능성이 있으므로 실측값을 그대로 게시하였습니다. 실험값 데이터의 분석에 항상 수반되는 문제이므로 이런 데이터의 처리에 대하여서도 분석하시는 분들이 고려하셔야 할 것입니다.) 100%로 보정하여 활용한 참가자들의 비율이 높았다.

· **외부 데이터 및 공개 베이스 모델 활용: 발표평가 자료분석 결과**

- 발표평가 대상자들은 거의 대부분 ChEMBL, PubChem 등 외부 데이터들을 병합하여 사용하였으며, 외부 데이터가 binary 모델 등 병합이 어려운 경우에는 pretraining 모델에 활용하였다. 데이터 추출과 병합 처리 훈련이 잘 되어있고 경험이 많은 것을 확인할 수 있었다.
- NCATS등의 공개 모델을 베이스 모델로 활용하거나 본인팀의 다른 신약분야 데이터 모델을 pretraining 모델로 활용하는 등 다양한 베이스 모델을 활용하였다.

· **오픈 소스 활용: 참가자들의 다수가 DeepChem 라이브러리, ChemBERT 활용 코드, pytorch-geometric 프레임워크, Uni-mol 등 화합물 구조, 알고리즘, 모델, 코드 등 다양한 오픈소스들을 조합하여 활용함으로써 모델 성능을 향상시켰다.**

· **모델 구축 경향: 참가자 다수가 그래프 모델을 활용하였으며, pretraining 모델을 만들어 전이학습을 수행하였다. Cyp 효소에 의한 대사반응은 효소의 활성부위가 매우 평평하여 선택성이 적고 화합물의 크기와 모양의 영향은 적은 반면, 주로 탈알킬화 반응에 의한 알코올의 생성과 산화반응 등 치환기(functional group)의 소수성과(hydrophobicity) 전기음성도(electronegativity) 및 치환기 주변의 bulkiness 등의 영향을 많이 받는다. 수상팀들은 대사안정성 반응에 대한 이해를 기반으로 화합물의 특성을 추출하고 반영하여 모델을 구축함으로써 좋은 성과를 도출한 것으로 생각된다.**

(2) KCB 보유 데이터 현황

· **구조정보**

KCB는 전체 75만여의 화합물 구조정보를 보유하고 있으며, 이 중 분양 가능한 60여만종의 구조 정보 및 계산 물성 정보를 공개하고 있다. 구조정보 이외에 분석정보(NMR, LCMS 등)를 보유하고 있다.

· **약효 정보**

KCB는 2023년 11월 현재 1,374 과제에 화합물을 분양하여 850만건 이상의 약효 데이터를 보유하고 있다. 이 중 활용자가 공개에 동의한 189 과제에 대한 150만건 이상의 약효 데이터를 공개하고 있다.

· 약효 데이터 관리 현황

- 화합물 활용을 위한 약효시험 신청 접수 시 작용점, assay 시스템, 스크리닝 방법 등에 대한 분류체계 및 실험방법 등을 확인하여 관리하고 있음
  · 작용점 분류: Enzyme_kinase, Enzyme_protease, Receptor_GPCR, Receptor_Nuclear, PPI, Transporter, Transcription Factor
  · 스크리닝 분류: Function based assay, Target based assay
  · Assay System: protein, cell, microorganism

· 데이터 이슈

- 대부분이 단일 농도의 %inhibition 값으로 정량화 어려움
- 1~2%를 제외한 데이터는 약효가 없는 음성 데이터임: 논문 등의 데이터는 양성 데이터가 많으므로 통합 시 활용도가 증가할 것임

· 데이터 강점

- 동일 실험 조건으로 수행된 화합물 수가 비교적 많음(평균 3,000~7,000종)
- 대표 라이브러리 등 분양 횟수가 많은 화합물들은 다수의 약효시험 정보가 있음

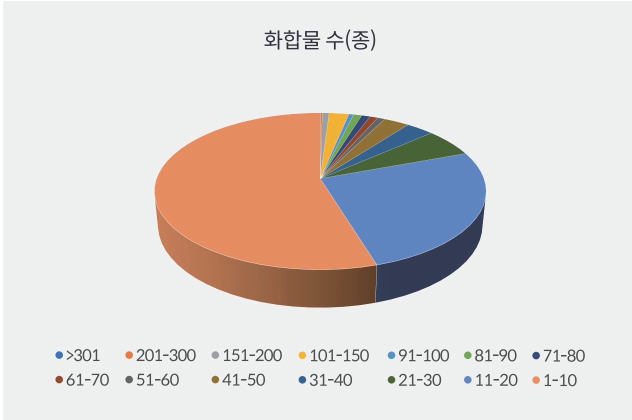
(3) KCB의 향후 데이터 활용 전략

컴퓨터에 저장된 데이터는 전기만 낭비할 뿐이고, 활용되어야 그 가치가 발휘될 수 있다. 데이터가 잘 활용되기 위해서는 추가적인 데이터의 관리 시스템이 필요할 것으로 생각되며 다음과 같은 작업들을 우선적으로 수행하고자 한다. 데이터 분석 전문가를 포함한 다양한 분야의 전문가들과 협의를 통해 데이터 활용을 활성화할 수 있는 방법을 찾고 시도해야 할 것이다.

1) KCB 데이터 관리 시스템 고도화

- 양성 및 음성 데이터 분류: 약효 데이터의 90% 이상이 약효가 없는 음성 데이터이므로 실험 농도를 기재한 후 양성 및 음성으로 데이터를 분류하고, IC_{50} 등의 정량값이 있는 경우는 별도 관리한다.
- 비공개 데이터 활용 전략: 현재 전체 데이터 중 13%만 공개 가능하다. 대부분의 활용자들이 타깃을 공개하기를 원하지 않으므로 타깃 class로 모두 분류하여 데이터의 통합 관리가 필요하다. Protease 실험을 통합하여 약효의 양성 및 음성으로 데이터를 분류하면 활용자의 정보는 보호하면서 데이터 분석에 활용이 가능할 것이다.
- 다수 실험된 화합물 정보의 관리: KCB에는 대표 라이브러리 등 51건 이상의 약효시험에 활용된 화합물들이 상당수(48,936종) 존재한다. 이들 화합물들에 대해서는 타깃 class별로 실험된 데이터를 정리하여 별도로 공개하면 활용도가 증대될 것이다.

화합물별 분양횟수	화합물 수(종)	누적(종)
> 301	61	61
201-300	1,476	1,537
151-200	5,092	6,629
101-150	14,818	21,447
91-100	3,557	25,004
81-90	6,114	31,118
71-80	6,424	37,542
61-70	5,934	43,476
51-60	5,460	48,936
41-50	19,109	68,045
31-40	20,830	88,875
21-30	38,348	127,223
11-20	170,200	297,423
1-10	359,438	656,861
계	**656,861**	



2) 외부 데이터와의 통합

- 타깃 기반 정보 통합: 공개 가능한 타깃 데이터에 대해서는 KCB 미보유 화합물 정보를 포함하여 외부 데이터와 통합하면 활용도가 증대될 것이다.
- 화합물 기반 정보 통합: 보유 화합물에 대한 외부 본문이나 특허 정보가 있을 경우에 화합물 기반으로 정보를 통합한다.
- 화합물 및 약효 정보 이외 정보와의 통합: 타깃과 관련된 유전체 정보 등 KCB가 관리하는 데이터 이외의 정보 등과 통합을 추진하면 유용한 지식이 창출될 수 있을 것이다.

4 참고문헌

Research and Development Project Valuation and Licensing Negotiations. *Interfaces*, **2007**, *Vol. 37*, 472-398.

BCS class IV drugs: Highly notorious candidates for formulation development. *J. of Controlled Release*, **2017**, *Vol. 248*, 71-95.

https://en.wikipedia.org/wiki/List_of_largest_pharmaceutical_mergers_and_acquisitions

How were new medicines discovered? *Nature Reviews Drug Discovery*, **2011**, *Vol.10*, 507−519.

Why and how have drug discovery strategies in pharma changed? What are the new mindsets. *Drug Discov Today*, **2016**, 239-49.

BIGCHEM: Challenges and Opportunities for Big Data Analysis in Chemistry. *Mol. Inf.* **2016**, *35*, 615−621.

Building a knowledge graph to enable precision medicine. *Scientific data*, **2023**, *Vol. 10*, 1-16.

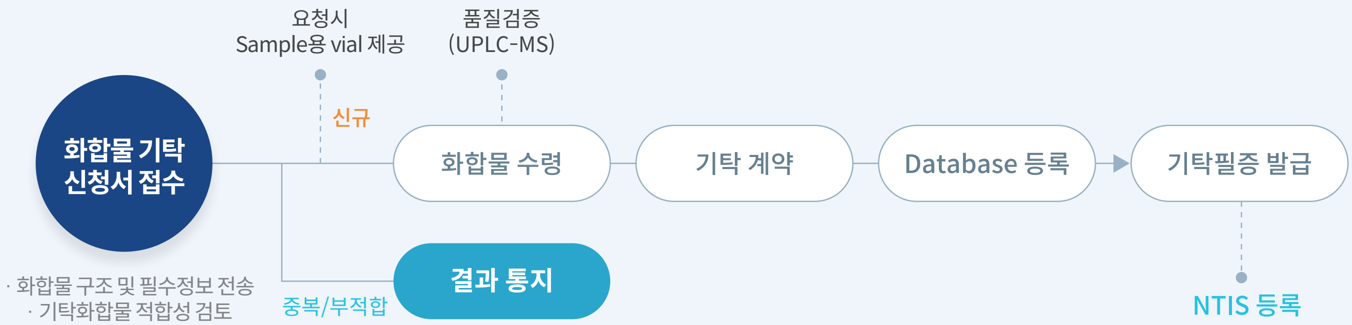
Computational systems biology in disease modeling and control, review and perspectives. *npj systems biology and applications*, **2022**, *Vol. 8*, 37.

Rethinking drug design in the artificial intelligence era. *Nature Review/ Drug Discovery*, **2020**, *Vol. 19*, 353.

An Automated High-Throughput Metabolic Stability Assay Using an Integrated High-Resolution Accurate Mass Method and Automated Data Analysis Software. *Drug Metab Disos.* **2016**, *1653-61*.

화합물 기탁 절차

문의 **E-mail.** chembank@krict.re.kr **Tel.** 042-860-7190 **Web.** www.chembank.org **통합데이터플랫폼.** korea.chembank.org



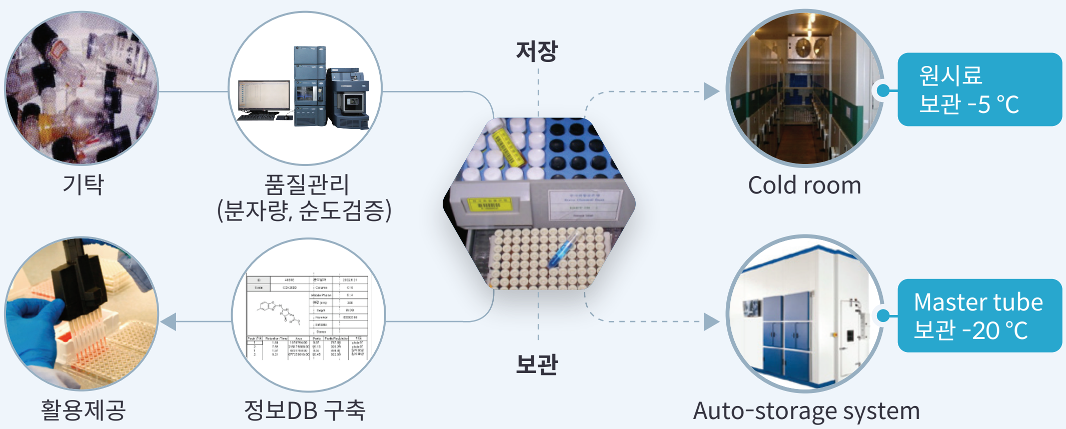
기탁화합물 범위 및 기준

- **범위** 유기합성 화합물 및 단일성분 천연물
- **품질기준** 화합물 양 > 10 mg (권장), 순도 > 80% (한국화합물은행에서 UPLC-MS로 검증)
  ^{1}H-NMR 등 분석자료 제출

기탁자 혜택

- **기탁화합물에 대한 활용 결과 도출시 그 결과를 통보 받을 수 있다.**
- **기탁화합물의 활용결과 Hit가 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



데이터 활용 절차

문의 **E-mail.** chembank@krict.re.kr **Tel.** 042-860-7190 **Web.** www.chembank.org **통합데이터플랫폼.** korea.chembank.org

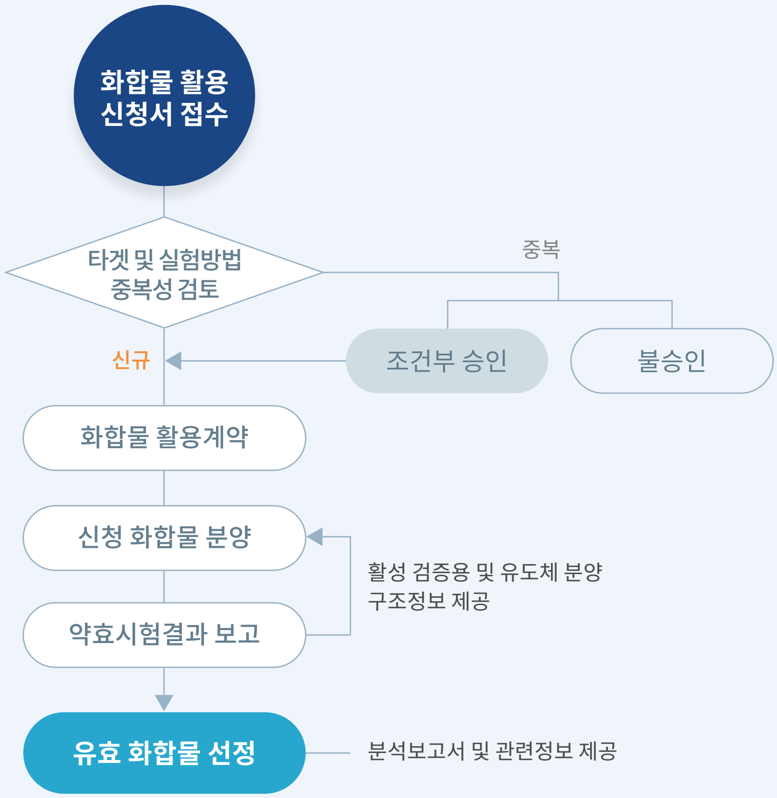
정보제공 데이터 종류

분류	소분류	
화합물 구조 정보	전체(분양 가능) 라이브러리	
	구매화합물 라이브러리	
	Focused 라이브러리	대표 라이브러리
		임상화합물 라이브러리
		천연물 라이브러리
		Fragment 라이브러리
		Kinase 라이브러리
		PPI 라이브러리
		GPCR 라이브러리
	PharmaCore Collection	
약효시험결과 정보	공개 가능한 정보	

· 데이터 정보는 USB로 제공됩니다.
· 데이터 활용 계약의 종료 시, 반드시 데이터의 폐기 및 해당 USB를 반납하셔야 합니다.

화합물 활용 절차

문의 **E-mail.** chembank@krict.re.kr **Tel.** 042-860-7190 **Web.** www.chembank.org **통합데이터플랫폼.** korea.chembank.org



화합물 사용료 규정

· 화합물은 무상으로 제공합니다.
· 화합물 제공에 수반되는 실비용(plate, 분주기 tip, 튜브 cap 등)은 활용자가 부담합니다.
· 중소기업 / 벤처 / 대학교 / 공공연구기관은 실비의 50%만 활용자가 부담합니다.

화합물 분양실비 부과

* 부가세 별도

Plate 종류	단가	대기업 / 중견기업	중소기업 / 벤처 / 대학교 / 공공연구기관
96-well (80종 / 1Plate)	30,400원	30,400원	15,200원
384-well (320종 / 1Plate)	117,100원	117,100원	58,550원

· 화합물 분양 배송비용은 활용자가 부담합니다.
· 2024년 3월 1일부터 화합물 분양 실비가 상기가격으로 인상됩니다.

제공화합물의 활용 결과(논문, 특허 등) 권리관계 규정

규정목적

· 한국화합물은행은 기탁자들이 기탁한 화합물을 바탕으로 운영되고 있습니다.
· 기탁자들의 화합물 기탁을 장려하고, 동시에 화합물 사용자들의 불편함을 최소화하기 위하여 한국화합물은행이 제공한 화합물을 활용하여 도출된 연구결과(논문, 특허 등)의 권리관계는 아래와 같은 규정을 적용합니다.

규정내용

활용결과(논문, 특허 등) 권리관계 규정 내용

1. 기탁자가 단순기탁 이외의 추가적인 기여가 없는 경우에는 화합물 기탁자를 논문의 사사(acknowledgement) 에 기재하는 것이 **원칙**입니다.
2. 기탁자가 추가적인 기여(유도체 합성 제공, 관련 정보제공 등)가 있을 경우에는 기여정도에 따라 기탁자를 논문 공저자 및 특허 공동발명인(공동출원인)으로 **포함**을 권장합니다.
3. 기탁자가 **물질특허**가 있는 경우, 활용결과에 대한 **용도특허**는 사용자(발견자)가 취득할 수 있습니다.

* 활용기관은 한국화합물은행의 라이브러리를 활용한 연구결과를 논문, 특허 등에 발표 또는 공개할 경우 한국화합물은행 제공 화합물을 사용하여 연구가 진행되었다는 사사 또는 문구를 기재하여야 합니다.

* 한국화합물은행은 화합물 활용에 관한 지적재산권(intellectual property)에 관여하지 않습니다.

> 한국화합물은행 활용결과의 권리 규정은
> 과학계에서 통용되는 연구결과 기여도에 대한 **연구윤리 기본원칙**을 따르고 있습니다.

후속연구를 위한 화합물 추가합성 및 구매 진행 안내

1. 한국화합물은행은 소량의 화합물(10 mg)만 기탁 받아서 보유하고 있으며, 최대한 많은 연구자들의 활용을 위해 1차 스크리닝 및 Hit화합물의 검증에 필요한 최소량만 제공하고 있습니다. 후속단계(*in vivo* 실험 및 약물성/독성 시험 등) 실험에 필요한 충분한 양의 화합물은 서비스하고 있지 않습니다.

2. 후속단계의 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) 자체 연구진 또는 전문가를 활용한 합성 진행

3. 한국화합물은행(KCB) 활용 화합물 중에서 상용 화합물의 구매를 원하시는 경우, KCB에서 구매한 vendor 정보를 제공해 드립니다. 화학물질의 해외 구매 시 화평법 및 화관법에 대한 신고 및 등록 의무를 사전에 반드시 확인 바랍니다.
   (화학물질정보처리시스템, https://kreach.me.go.kr/)

한국화합물은행 제공 라이브러리 종류

전체 Library

화합물 수 : 75만종

기본 제공량 : 5 μL
(DMSO 용액, 평균농도 5 mM)

KCB Library

1 대표 Library

- **화합물 수** 7,000
- **라이브러리 구성**
  전체 화합물을 대표하는 라이브러리,
  순도 및 분자량 검증 (UPLC-MS),
  Eye filtering

2 임상화합물 Library

- **화합물 수** 3,200
- **라이브러리 구성**
  임상 I - III 상 단계 화합물 및 승인
  약물(Clinically applied compounds)

3 천연물 Library

- **화합물 수** 1,600
- **라이브러리 구성**
  단일성분 천연물 및 천연물 유사골격
  (Natural product-like) 구조의 화합물

4 Fragment Library

- **화합물 수** 1,600
- **라이브러리 구성**
  분자량 300이하 화합물 라이브러리,
  Rule of 3 filtering (20 mM in DMSO)

5 Kinase Library

- **화합물 수** 2,700
- **라이브러리 구성**
  Kinase active site에 결합 가능성이
  높은 화합물을 분자모델링 방법
  (Docking)으로 선별하여 구성

6 PPI Library

- **화합물 수** 17,000
- **라이브러리 구성**
  해외 vendor로 부터 선별 구매화합물

7 GPCR Library

- **화합물 수** 8,600
- **라이브러리 구성**
  해외 vendor로 부터 선별 구매화합물

8 PharmaCore Collection

- **화합물 수** 요청개수
- **라이브러리 구성**
  요청골격 또는 분자모델링 방법을
  적용한 수요자 맞춤형 선별 화합물

공동 활용을 통한 화합물 및 활용데이터 가치 재창출

We take care of your compounds and Create new value!

화합물 기탁 · 활용 · 데이터 문의

· 홈페이지 : www.chembank.org / 통합데이터플랫폼 : korea.chembank.org
· 주소 : (34114) 대전광역시 유성구 가정로 141 한국화학연구원 한국화합물은행
· 전화 : 042-860-7190 / 팩스 : 042-860-7096 / E-mail : chembank@krict.re.kr

담당업무	이름	전화	E-mail
사업책임자	이계형	042-860-7173	ghlee@krict.re.kr
의약화학	이선경	042-860-7148	leesk@krict.re.kr
의약화학	이현규	042-860-7016	leehk@krict.re.kr
화학정보학 / 분자모델링	이윤호	042-860-7453	yunolee1@krict.re.kr
화학정보학 / 분자모델링	조남철	042-860-7093	nccho@krict.re.kr
연구전략 / 대외협력	황슬기	042-860-7097	sghwang@krict.re.kr
대외협력 / 정보관리	황순희	042-860-7190	chembank@krict.re.kr
화합물관리	안은숙	042-860-7648	esan@krict.re.kr
화합물관리	권호진	042-860-7149	hjkwon89@krict.re.kr
화합물관리	김선우	042-860-7171	swkim@krict.re.kr
화합물관리	김선호	042-860-7090	shkim@krict.re.kr
화합물관리	유순종	042-860-7011	sunjong@krict.re.kr
데이터 / 화합물 정보관리	이유리	042-860-7092	yurilee@krict.re.kr
데이터 / 화합물 정보관리	강수민	042-860-7014	smkang@krict.re.kr
LC/MS 분석	이수연	042-860-7747	suyoun@krict.re.kr
LC/MS 분석	김수희	042-860-7070	rlatngml@krict.re.kr

과학기술정보통신부지정
연구성과 관리·유통 전담기관
- 화합물 전담기관 -

Vol.1
KCB newsletter
한국화합물은행
뉴스레터
2024년 제1호

한국화합물은행은 국가연구개발사업의 화합물
연구성과 관리·유통 전담 기관으로 지정되어 있습니다.
[과학기술정보통신부 고시 제2022-41호]

한국화합물은행
Korea Chemical Bank

(34114) 대전광역시 유성구 가정로 141 한국화학연구원 한국화합물은행
Tel. 042-860-7190 **Fax.** 042-860-7096 **E-mail.** chembank@krict.re.kr **Web.** www.chembank.org **통합데이터플랫폼.** korea.chembank.org