

- AI 신약개발 경진대회 2nd

## Train 내 목차 의미

2024년 08월 09일

# Agenda

---

- 1. Molecule ChEMBL ID
- 2. Standard Type/ Standard Relation/ Standard Value/ Standard Units
- 3. pChEMBL Value
- 4. Assay ChEMBL ID
- 5. Target ChEMBL ID/ Target Name/ Target Organism/ Target Type
- 6. Document ChEMBL ID
- 7. IC50\_nM
- 8. pIC50
- 9. Smiles

# • 1. Molecule ChEMBL ID

---

• 큐레이트 = 자료를 수집, 선별 -> 의미 부여하는 과정

## ▶ ChEMBL이란?

: 약물과 유사한 특성을 가진 생리활성분자(bioactive molecular)의 수동 큐레이트 데이터베이스

ChEMBL에서 특정 target (단백질)에 대한 정보를 검색할 수 있다.

- 유럽 생물정보학 연구소에서 관리하는 화학데이터베이스

- 효과적인 신약으로서 게놈 정보의 번역을 돕기 위해 화학적, 생물학적 활성 및 게놈 데이터를 결합하여 제공한다.

ChEMBL => 하나의 분자구조마다 ChEMBL 번호를 매긴 백과사전 같은 것  
Molecule ChEMBL ID = 분자 ChEMBL 명

참고 ChEMBL 홈페이지 사용법:  
<https://bdsi.jbnu.ac.kr/blog/chembl-database/>

## ► Example

-> Name: Heparan sulfate = ChEMBL 명: CHEMBL2364677

<input type="checkbox"/>	ChEMBL ID	Search Hit	Name	UniProt Accessions	Type	Organism	Compounds	Activities
<input type="checkbox"/>	CHEMBL2364677		Heparan sulfate		OLIGOSACCHARIDE	Homo sapiens	No Data	No Data
<input type="checkbox"/>	CHEMBL4630725		Nitric oxide synthase	P29475, P35228, P29474	PROTEIN FAMILY	Homo sapiens	No Data	No Data
<input type="checkbox"/>	CHEMBL2364158		Tumor necrosis factor ligand superfamily member 13B	Q9Y275	SINGLE PROTEIN	Homo sapiens	No Data	No Data
<input type="checkbox"/>	CHEMBL3140342		Chondroitin-6-sulfate		OLIGOSACCHARIDE	Homo sapiens	No Data	No Data

Retrieve [P29475](#)

<input type="checkbox"/>	Entry	Entry name		Protein names	Gene names	Organism	Length	
<input type="checkbox"/>	P35228	NOS2_HUMAN		Nitric oxide synthase, inducible	NOS2 NOS2A	Homo sapiens (Human)	1,153	
<input type="checkbox"/>	P29474	NOS3_HUMAN		Nitric oxide synthase, endothelial	NOS3	Homo sapiens (Human)	1,203	
<input type="checkbox"/>	P29475	NOS1_HUMAN		Nitric oxide synthase, brain	NOS1	Homo sapiens (Human)	1,434	

## • 2. Standard Type/ Standard Relation/ Standard Value/ Standard Units

---

Standard Type	Standard Relation	Standard Value	Standard Units
IC50	'='	0.022nM	
IC50	'='	0.026nM	

- Standard Type: 표준 타입  
= 모두 IC50(반수저해량)으로 통일
- Standard Relation: 표준 관계  
= 모두 동일함
- Standard Value: 표준 값  
= 실제 IC50값
- Standard Units: 표준 단위  
= 모두 nM(나노 몰)으로 통일

## • 3. pChEMBL Value

---

: ChEMBL을 역상용로그화 한 것 (= -log<sub>10</sub> 붙임)  
여러 유형의 생체 활성 측정 값을 용이하게 보기 위해 값을 변환함.

$$pChEMBL = -\log_{10}(Effective\ Value)$$

*Effective Value* : molar concentrations of IC<sub>50</sub>, XC<sub>50</sub>, EC<sub>50</sub>, AC<sub>50</sub>, K<sub>i</sub>, and K<sub>d</sub>

## • 4. Assay ChEMBL ID

Molecule ChEMBL ID	Assay ChEMBL ID	Document ChEMBL ID
CHEMBL4443947	CHEMBL4361896	CHEMBL4359855

: 관련 시험에 대한 문서번호

ChEMBL 내 홈페이지에서 CHEMBL4443947 검색하면 관련 시험에 CHEMBL4361896 가 나타남

-> 클릭 시 시험 정보 확인 가능

**ID:** CHEMBL4361896

**Type:** Binding

**Description:** Binding affinity to human IRK4 using myelin basic protein as substrate by [gamma-33P]-ATP assay

**Format:** BAO\_0000357

**Reference:** ---

**Organism:** Homo sapiens

**Strain:** ---

**Tissue:** ---

**Cell type:** ---

**Subcellular Fraction:** ---

**Target:** [CHEMBL3778](#)

**Document:** [CHEMBL4359855](#)

## • 5. Target ChEMBL ID/ Target Name/ Target Organism/ Target Type

---

Target ChEMBL ID	Target Name	Target Organism	Target Type
CHEMBL3778	Interleukin-1 receptor-associated kinase 4	Homo sapiens	SINGLE PROTEIN
CHEMBL3778	Interleukin-1 receptor-associated kinase 4	Homo sapiens	SINGLE PROTEIN

- Target ChEMBL ID: 목표 타겟 단백질의 ChEMBL ID  
= CHEMBL3778
- Target Name: 타겟 실제 이름  
= Interleukin-1 receptor-associated kinase 4
- Target Organism: 타겟 개체  
= Homo sapiens = 인간
- Target Type: 타겟 타입  
= Interleukin-1 receptor-associated kinase 4는 SINGLE PROTEIN 범주에 속함



## • 6. Document ChEMBL ID

Molecule ChEMBL ID	Assay ChEMBL ID	Document ChEMBL ID
CHEMBL4443947	CHEMBL4361896	CHEMBL4359855

: 관련 논문 문서번호

ChEMBL 내 홈페이지에서 ChEMBL4443947 검색하면 관련 논문에 ChEMBL5137104 가 나타남

-> 클릭 시 논문 정보 확인 가능

**ID:** ChEMBL5137104

**Journal:** ACS Med Chem Lett

**Title:** Novel IRAK4 Inhibitors for Treating Asthma.

**Authors:** Sabnis RW.

**Abstract:** The innate immune receptor nucleotide-binding oligomerization-domain-containing protein 2 (NOD2) represents an important target for the development of structurally defined small molecule immunomodulatory compounds that have great potential to be used either as vaccine adjuvants or as general immunostimulatory agents. We report here the investigation of the structure-activity relationship of a series of novel desmuramylpeptide NOD2 agonists. Extensive exploration of chemical space culminated in the discovery of a lipophilic adamantane-moiety-featuring compound 40, the first single-digit nanomolar and the most potent NOD2 agonist in its structural class to date. Moreover, 40 acted synergistically with lipopolysaccharide and interferon- $\gamma$  to induce the production of cytokines in human peripheral blood mononuclear cells and enhance their nonspecific cytotoxic activity against K562 cancer cells. These findings provide initial insight into its immunostimulatory potential, especially when used in combination with other immunopotentiators.

**CiteXplore:** [35978689](#)

**DOI:** [10.1021/acsmedchemlett.2c00324](https://doi.org/10.1021/acsmedchemlett.2c00324)

**Patent ID:** ---

## • 7. IC50\_nM

: 약물 1nM에서의 표적단백질 절반을 저해하는 농도(단위: 나노몰)

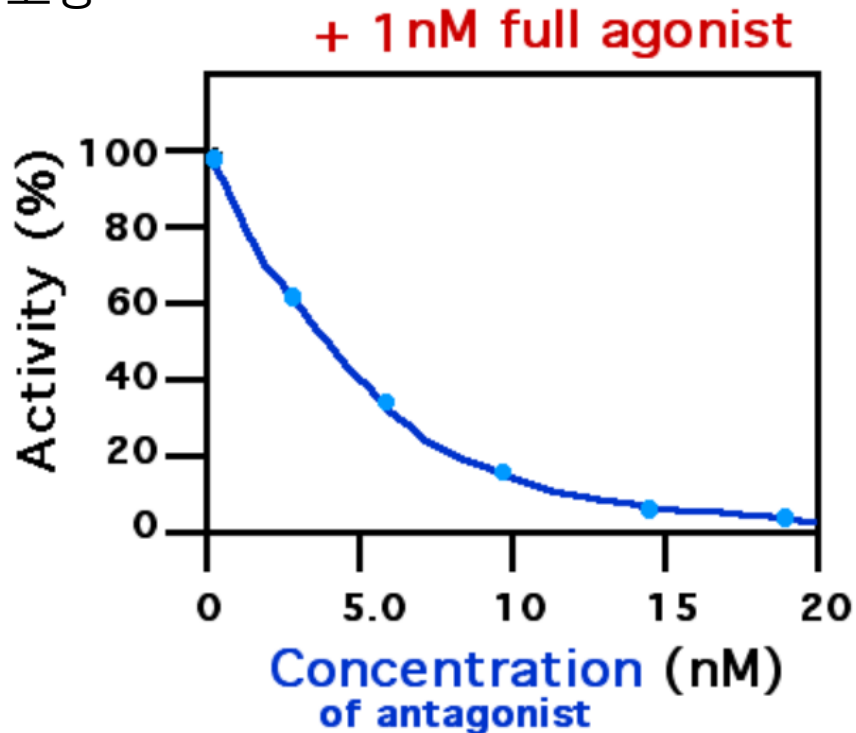
IC50 : 반수저해량(Half maximal inhibitory concentration)

= 표적단백질 절반을 저해하는 농도

= 최대 억제 농도의 절반

(약물의 효과와 관련된 지표, 길항제의 약물 효능 척도!)

\*참고용



- nM : 나노 몰(mole)

- IC50: Activity 가 50일 때의 Concentration 값

=1nM의 agonist가 있을때 antagonist가 타겟을 저해하는 농도  
(-> 10nM이면 투여해야 하는 antagonis의 농도도 올라감)

# IC50을 구하는 공식

Competitive inhibitor

$$IC_{50} = K_i \left(1 + \frac{[S]}{k_m}\right)$$

Noncompetitive inhibitor

$$IC_{50} = K_i$$

Uncompetitive inhibitor

$$IC_{50} = K_i \left(1 + \frac{k_m}{[S]}\right)$$

- $K_i$ : inhibition constant
- $K_m$ : 미카엘 상수(효소 활성이 최대의 50%일 때 기질의 농도)
- $[S]$ : 기질 농도(보통 대괄호 안에 들어가면 농도다)

\*참고용

**train 엑셀 내 IC50\_Nm**란에 IC 50이 계산되었음  
(해당 분자식 약물 1nM 내 몇 농도 사용 시  
**인터루킨-1 수용체 관련 키나아제 4의**  
**절반을 저해** 가능한 지)

# 대회가 원하는 것

---

= 어떤 분자식을 가진 약물 1nM 사용 시  
IC50이 자동으로 계산되어지는 머신러닝 프로그램

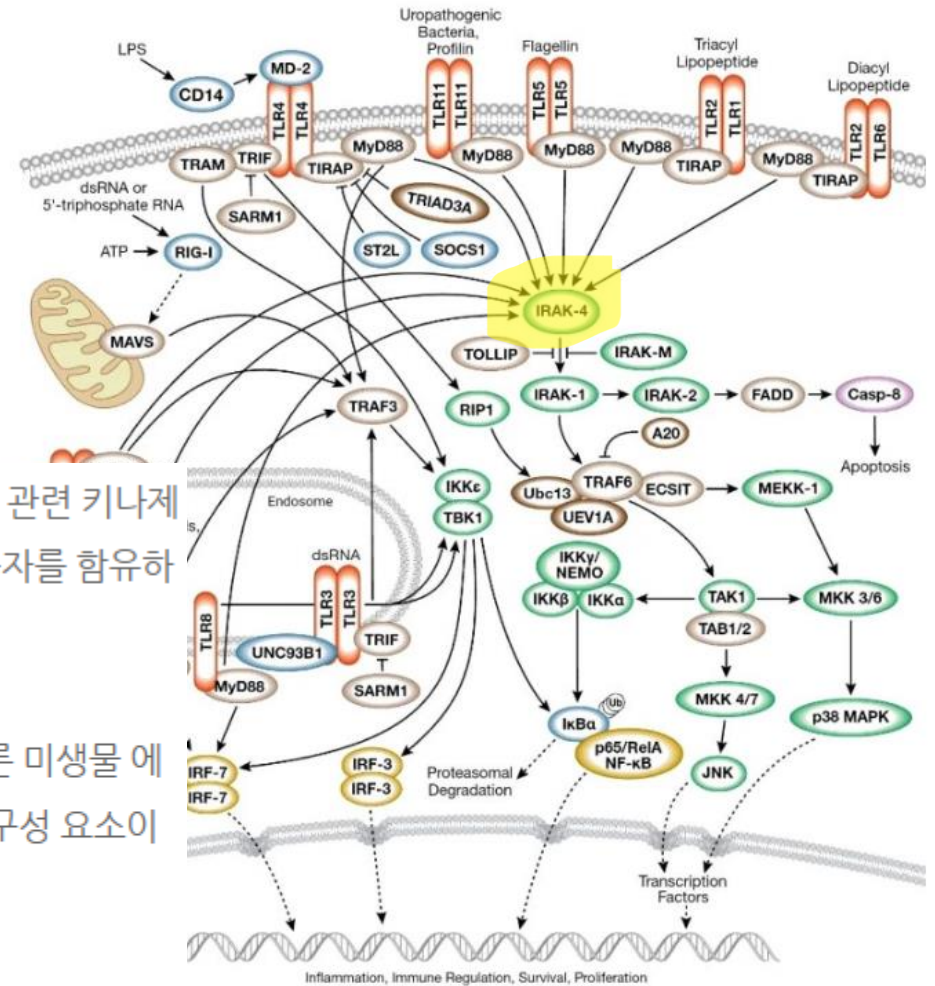
(인터루킨-1 수용체 관련 키나아제 4의 절반을 저해 가능한 농도)

## 개발된 프로그램 통해 발견하고자 하는 약물

= 인터루킨-1 수용체 관련 키나아제 4의 절반을 저해하여 **자가면역 활동을 조절(억제)하는 약물**  
(중에서 본인들이 원하는 IC50 가진 것)



그림 5-1 : 자가면역 질환에서의 면역



인터루킨-1 수용체( IL-1R )와 Toll 유사 수용체 수퍼패밀리의 구성원은 TIR을 통해 인터루킨-1 수용체 관련 키나제 (IRAK) 복합체의 모집을 중재하는 세포질 내 Toll-IL-1 수용체( TIR ) 도메인을 공유합니다. -어댑터 분자를 함유하고 있습니다.

TIR-IRAK 신호 전달 경로는 특정 박테리아에 대한 보호 면역에 중요한 것으로 보이지만 대부분의 다른 미생물 에 대해서는 중복됩니다 . IRAK4 는 IL-1/TLR 신호 전달 경로에서 그 기능에 절대적으로 중요한 유일한 구성 요소이 기 때문에 포유류 IRAK 계열의 "마스터 IRAK"로 간주됩니다.

이러한 경로 중 하나가 자극되면 세포는 염증 유발 신호를 방출하고 선천적 면역 활동을 유발하도록 촉발됩니다. IRAK4 또는 그 고유의 키나제 활성이 상실되면 이러한 경로를 통한 신호 전달이 완전히 중단될 수 있습니다. [15]

= IRAK-4를 비활성화 시키면 오른쪽 신호 체계 전체 중단 = 자가면역 활동 중단  
 -> IC50(절반 저해 농도)이 적절한 약을 사용하여, 비정상적 자가면역 활동을 약물로 조절 예정

## • 8. pIC50

---

: IC50를 역상용로그화 한 것 ( $= -\log_{10}$  붙임)

때때로 IC50 값은 pIC50 척도로 변환할 수 있으며, 식은  $pIC_{50} = -\log_{10}(IC_{50})$  이다.

빼기 기호로 인해 pIC50 값이 높을수록 기하급수적으로 더 강력한 억제제를 나타낸다. pIC50은 일반적으로 몰 농도(mol/L 또는 M)로 표시되므로 M 단위의 IC50이 필요하다.<sup>[2]</sup>

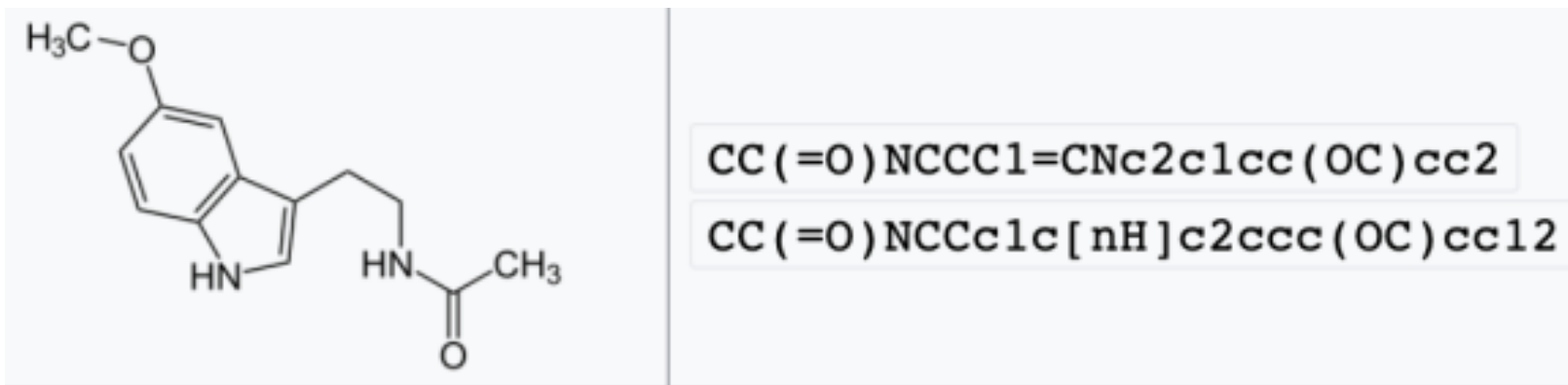
# • 9. Smiles

: simplified molecular-input line entry system

-> [?] [?] [?] [?] [?] [?] [?] [?] [?]

## ► chemical "SMILES"

: molecule [?] [?] [?] [?] (structure) [?] [?] [?] [?] [?] [?] [?] [?] [?] [?] [?] [?] [?] [?] [?]





# 끝

\* 읽어볼 만한 글

전북대\_딥러닝-신약개발\_약물예측모델 만들기

<https://bds1.jbnu.ac.kr/blog/mid-term-project-deep-learning-and-drug-discovery/>

약물이 HIV 표적에 대해 얼마나 활성화될지를 예측하는 머신러닝 개발 방법

<https://blog.naver.com/issuewatch/222600618700>