

## 약물 유사 화합물의 물리화학적 성질 기반 다양성 평가

채한화\*, 박원철, 임다정  
국립축산과학원 동물유전체과

### Evaluation of Diversity based on Physicochemical properties of Drug-like compounds

Han-Ha Chai\*, Woncheoul Park, Dajeong Lim

Animal Genomics and Bioinformatics Division, National Institute of Animal Science

**요약** 약물 유사 화합물은 동일한 목표단백질에 작용한다는 가정 하에서 특정 약물과 유사한 구조와 물리화학적 특성을 가진 화합물을 말한다. 이때 약물 유사성은 약물의 흡수, 분포, 대사, 배설, 독성과 관련된 몇 가지 분자 매개변수(물용해도, 옥탄올-물 분배계수, 극성표면적, 회전결합수 등)에 의해 정의된다. 본 연구의 목적은 Pubchem BioAssays에 등록된 11,144개 화합물에 대해 약물유사성, 물리화학적 성질과 구조에 대한 유사성을 평가하기 위해 약물 유사 화합물의 물리화학적 성질 기반 다양성을 알아보고자 하는 데 있다. Pubchem 데이터베이스 내 11,144개 화합물의 2D-구조로부터 물리화학적 성질과 ADMET 분자표현자를 계산했다. 이를 통해 화합물의 약물 가능성(QED score, Lipinski and Veber et al)을 측정하였고, 물리화학적 특성과 분자구조 다양성을 기반으로 약물 유사도를 평가하여 우선순위를 결정하였다. 사례 연구로, 표적단백질에 대한 정보 없이 항녹내장제 유효화합물(AL-3750A, 세로토닌 5HT2 수용체 조절)과 유사한 물리화학적 성질 및 구조를 갖는 화합물을 발굴하였다. 상위 5%내 유사한 화합물들 중에서 구조유사성을 가진 화합물들은 세로토닌을 포함한 멜라토닌의 생성과 분해를 조절하는 역할을 하는 반면에, 물리화학적 성질이 유사한 화합물들은 다른 구조와 기능을 가졌다. 결과적으로, 약물 유사 화합물의 물리화학적 성질 기반 다양성 평가는 화합물을 군집으로 분류하는 데 적용할 수 있으며, 다양한 구조의 선도물질을 발굴 및 최적화 하는 데 활용할 수 있다.

**Abstract** Drug-like compounds can be defined as compounds having a structure and physicochemical properties similar to a known drug, assuming that both act on the same target protein. Currently, drug-likeness is defined by several molecular parameters (viz., aqueous solubility, octanol-water partition coefficient, polar surface area, number of rotation bonds, etc.) related to drug absorption, distribution, metabolism, excretion, and toxicity. This study investigates the diversity evaluation based on the physicochemical properties of drug-like compounds obtained through PubChem BioAssays, and the screening priority of compounds for various target proteins. Physicochemical properties and ADMET molecular descriptors were calculated from the 2D structures of 11,144 compounds within the PubChem database. This allowed for measuring the drug potential (QED score, Lipinski, and Veber et al.) of compounds and prioritizing them by evaluating their drug-likeness based on the physicochemical characteristics and molecular structural diversity. As a case study, compounds with similar physicochemical properties and structures to the active compound of antiglaucoma drugs (AL-3750A, serotonin 5HT2 receptor regulation) were discovered without prior information regarding the target protein. Among the top 5% of similar compounds, those with structural similarities were found to play a role in regulating the production and degradation of melatonin (including serotonin), while compounds with similar physicochemical properties were determined to have different structures and functions. Thus, our results indicate that diversity evaluation based on the physicochemical properties of drug-like compounds can be applied to classify compounds into groups and can be used to discover and optimize lead compounds with diverse structures.

**Keywords** : Drug-like Compounds, Druggability, Chemical similarity, Diversity, Virtual lead screening

본 논문은 2022년도 국립축산과학원 1차 예비시험과제(세부과제번호: PJ01681903, "in silico 기반 화학정보 DB 활용 메타저각 후보 물질(화합물) 발굴" 연구과제로 수행되었음.

\*Corresponding Author : Han-Ha Chai(National Institute of Animal Science)

email: hanha@korea.kr

Received April 25, 2023

Revised May 22, 2023

Accepted June 2, 2023

Published June 30, 2023

## 1. 서론

분자 조합화학 등의 기술을 통해 다양한 화합물을 합성할 수 있게 되었다. 2007년 기준, 하루에 1개 화합물 라이브러리에서  $10^5$ 개 화합물 이상의 데이터를 얻을 수 있었다. 현재  $10^{14}$ 개 화합물 데이터가 필요한 경우,  $10^9$ 개 화합물로 구성된 라이브러리의 1mM 농도로 6,000번 초고속효율 스크리닝(HTS, high throughput screening)을 통해 가능하다[1]. 더 나아가서 8억개 합성화합물을 초고속효율 스크리닝 하는 데, DNA로 인코딩된 조합 라이브러리의 합성 비용을 포함하여 150,000\$로 할 수 있다[1].

이렇게 생산된 화합물의 빅데이터를 공용 데이터베이스(PubChem[2], ChEMBL[3], MMsINC[4]등)로 구축할 뿐만 아니라, 정부 주도형 소재 게놈 이시셔브(MGI, Material Genome Initiatives, www.mgi.gov)를 진행하고 있다. 1960년도에 구축한 화합물 및 합성 가능한 화합물의 데이터를 활용하여 구조-활성간의 관계를 의약 화학 분야[5]에 적용하기 시작하였다.

이후, 2022년 Binding DB(<https://www.bindingdb.org>)에는 PubChem DB(<https://pubchem.ncbi.nlm.nih.gov>)내 BioAssays로 확인된  $301 \times 10^6$ 개 생리활성 화합물 및 ChEMBL엔트리(<https://www.ebi.ac.uk/chembl>)에서 선별된 8,942개 목표 단백질과 1,123,939개 약물 유사 화합물간의 2,613,813개 결합 데이터가 구축되어 있다. 게다가, 축적된 화합물 데이터를 가상탐색 하는 CHARMM (Chemistry at HARvard Macromolecular Mechanics) 프로그램을 개발한 과학자들에게 2013년 노벨화학상이 수여된 바 있다[6]. 이로부터 가상탐색을 통해서 직접 실험을 하지 않고도 목표 단백질의 움직임과 화학반응을 예측할 수 있는 계기가 되었다.

데이터 기반 가상탐색 방법의 대표적인 적용은 화합물의 구조와 물리화학적 성질 또는 활성간의 정량적인 관계(QSPR, QSAR)를 모델화 하는 것이다. 방대한 화합물의 생리활성 데이터를 가지고 1) 화합물의 구조적 특징을 수치로 표현한 분자표현자(molecular descriptor)[7]와 물리화학적 성질간의 정량적인 관계를 통계적인 방법으로 설명할 수 있다. 이때 통계적인 방법에는 다중선형회귀분석, 기계학습 알고리즘[8], 조건부 확률 알고리즘, k-nearest neighbors 분류 등이 적용된다. 분자표현자는 화합물의 구조를 화학그래프(결합인접 또는 결합길이 행렬)로 나타낸다. 예를 들어 원자연결성을 나타내는 것부터 화합물의 2D 또는 3D 좌표를 이용한 'scaffold hops'로 구조적 변화[9]를 수학적으로 해석한다. 해석

결과로 3D-분자구조변화에 따른 에너지변화 등을 수치적 값으로 나타낼 수 있다. 이런 분자표현자의 값은 화합물 데이터 마이닝, 화합물 다양성 분석 및 활성 등을 예측하는 데 적용된다[10]. 특히, 활성에 영향을 주는 화합물의 물리화학적 성질을 중심으로 다양한 골격구조를 갖는 후보화합물 탐색에 사용된다[11]. 다른 측면에서 2) 분자표현자를 사용하지 않고 화합물의 substructure 구조 패턴과 물리화학적 성질간의 상호연관성을 패턴인식 알고리즘 사용하여 수학적 관계를 식별한다. 이 방법은 사전에 substructure 라이브러리를 정의하고 substructure의 존재 여부를 0과 1로 나타내어 화합물의 구조를 일정한 길이의 바이너리 fingerprint(0과1), ASCII 코드를 사용하여 복잡한 구조를 저장한다. 이는 물리화학적 특성에 영향을 주는 substructure를 갖는 분자를 특성화 하거나, 구조 유사성 검색 및 구조-활성관계(SAR)를 쉽게 설명할 수 있다. 일반적으로 공용 화합물 데이터베이스 PubChem등에서 활성에 영향을 주는 substructure를 갖는 후보화합물을 검색할 때 많이 쓰인다. 약물개발 단계과정에서 후보화합물 실패, 특히 임상 중 탈락의 비율(10%미만 성공)이 높아지면서[12] 전임상 단계에서 안전성을 개선하는 것 이외에도, 후보물질의 발굴 단계부터 ADMET(흡수, 분포, 대사, 배설, 독성)를 고려한다. 이로부터 다음단계로 진행 여부(on/off)에 대한 의사결정(Fig. 1)을 하는 전략을 취하게 되었다. 약물의 후보물질을 발굴하는 초기단계부터 약물로서 승인 및 시판될 때까지 단계별로 많은 의사결정들을 하게 되고, 이러한 과정들마다 가상탐색이 관련되어 적용된다[13]. 주로 이용되는 가상탐색방법은 데이터 기반 분자모델링(substructure 패턴 인식, QSPR, QSAR 등[14,15]), ADMET 관련된 표적 단백질[16]를 대상으로 분자도킹(molecular docking), 분자 동역학(molecular dynamics)등이 있다.

이번 연구에서는 생리활성 정보(표적 단백질 등)가 존재하는 화합물 데이터베이스 PubChem내 축적된 데이터들 중에서 PubChem BioAssays (<https://pubchemdocs.acs.ncbi.nlm.nih.gov/bioassays>)에 포함된 작은 분자크기를 갖는 11,144개 화합물을 대상으로 하였다. 본 연구의 목적은 Pubchem BioAssays에 등록된 11,144개 화합물에 대해 약물유사성, 물리화학적 성질과 구조에 대한 유사성을 평가하기 위해 약물 유사 화합물의 물리화학적 성질 기반 다양성을 알아보려고 하는 데 있다.

Fig. 1에서 보는 바와 같이, 화합물 데이터로부터 2D와 3D 구조정보 기반 분자표현자를 계산하여 ADMET 관련된 물리화학적 성질들의 정보(용해도[17], 혈관-뇌

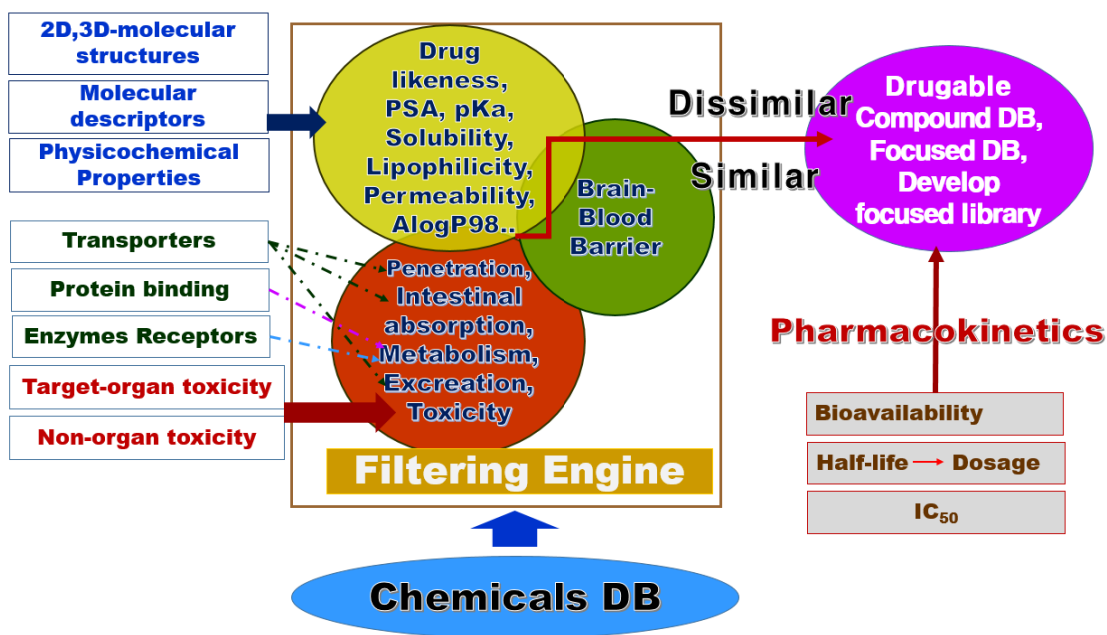


Fig. 1. ADMET properties depended on many factors such as molecular structures and physicochemical properties of drug-like molecules, their interactions with some proteins (e.g. transporters, enzymes, receptors), contribute to the pharmacokinetics.

장벽 투과도, 장에서 흡수도[18], 간독성[19], CYP2D6 저해도, 혈장 단백질 결합력)를 생성하였다. 생성한 정보를 이용하여 화합물간의 유사성을[20] 평가하여 유사성 filtering 기준으로 화합물을 분류하였다. 이는 분류된 화합물 그룹간의 다양성을 고려하여 Screening 대상 화합물 선별의 우선순위 결정을 목적으로 한다. 다양성 선택(diversity selection) 방법으로 특정 분자와 유사한 ADMET 범위와 약물유사성 관련 물리화학적 성질 내에 존재하지만, 화학적 구조가 다른 후보화합물을 가상탐색하는 데에 적용하고자 하였다.

## 2. 실험재료 및 연구방법

### 2.1 분자의 구조로부터 물리화학적 특성 계산

PubChem에서 활성이 측정된 11,144개 화합물의 2D-structure 와 3D-conformer 분자구조파일(SDF, Structural Data Files)을 수집하였다. 이로부터 분자내 원자의 개수, 원자정보, 그리고 분자 내 결합의 수, 결합된 원자들 및 결합종류, 각 원자들의 상대적인 좌표, 원자와 분자단위의 전하량 등을 molecular viewer 프로 그래프로 확인하였다. 특히, PubChem DB내 화합물의

3D-conformer 분자구조 파일에는 3D-원자 좌표 값을 MMF94 분자력 장 에너지[21] 계산한 값으로 결정이 되어 있다. 이 정보를 이용하여 원자단위에서 양전하/음전하, 분자의 총 전하량, 이온-이온 간의 상호작용에 의한 결합길이가 짧아 졌는지 또는 반발력 때문에 길어 졌는지도 확인 하였다. 또한, PubChem DB내에서 활성과 연관된 substructure로부터 2D-structure 와 3D-conformer 분자구조 유사성이나 substructure의 fingerprint 패턴인식으로부터 특정 substructure를 갖는 다른 분자를 검색하는 데 적용하였다(Fig. 2).

이번 연구에서는 각 화합물의 2D-structure와 3D-conformer 분자구조를 확인을 한 다음, 화합물들이 신체 내에서 작용하는 조건(pH = 7.0, 목표 단백질에 작용 가능성)을 주기 위하여, MMF94 분자력 장 대신에 CVFF[22] 원자 및 분자력 장(molecular force field)에서 분자표현자를 적용하였다. CVFF 분자력 장에서 11,144개 화합물들의 흡수 분포-대사-배설-독성(ADMET)과 관련된 물리화학적 성질[23-26] 및 약물 가능성 측정과 관련된 대표적인 8개 분자표현자를 계산 하였다. ADMET 관련 성질들은 AlogP98과 2D-PSA, 물에서 용해도 등 분자표현자를 적용하여 특정 값으로 수치화 하였고 ADMET 성질의 수준에 따라 그룹으로 분류하였다.

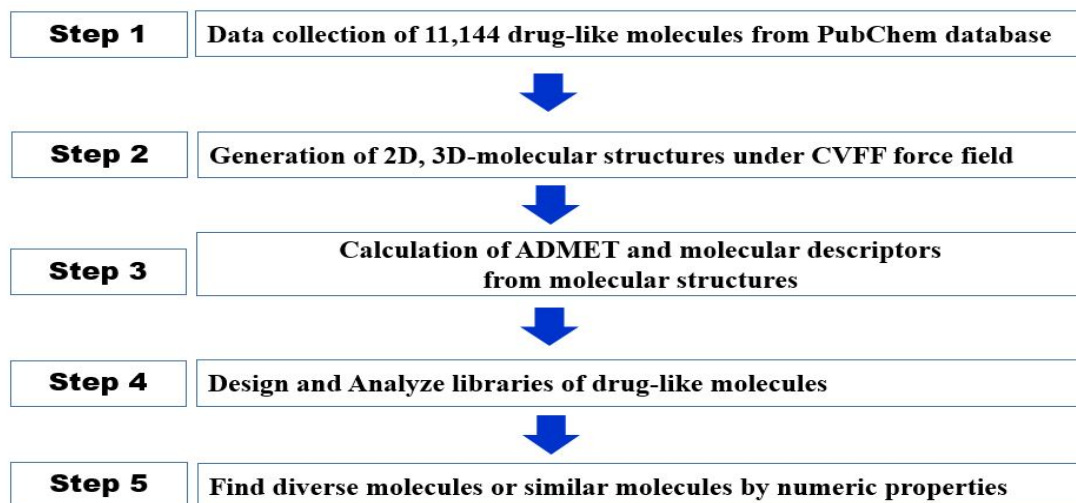


Fig. 2. Schematic diagram of clustering a set of drug like molecules into subsets with similar properties based on the root mean square (RMS) differences of both ADMET and molecular descriptor properties.

AlogP98은 두 용매(옥탄올과 물)로 구성된 한 개의 시스템에서 분배계수( $\log P_{\text{octanol-water}}$ ) [27]이며 대부분의 중성 유기 화합물을 구성하는 각 원자(C, H, O, N, S, Se, P, Si, 그리고 할로젠 등)의 기여도를 더 함으로써 계산이 되는 분자표현자이다. Eq. (1)과 같이

$$\log P_{\text{octanol-water}} = \sum n_i a_i \quad (1)$$

여기서  $n_i$ 는 I 형태의 원자개수이고,  $a_i$ 는 I형태의 원자가 기여하는 소수성 값이다. AlogP98(atomic level hydrophobicity as log of the octanol-water partition coefficient)은 화합물의 소수성(hydrophobicity) 또는 친지용성(lipophilicity) 성질을 표현한 것이다. AlogP98 분자표현자 값이 클수록 옥탄올 용매에 분포하며 비극성의 특징을 나타내어 세포표면의 지질막을 통과하기 용이하다. 하지만, 너무 높은 AlogP98 값은 큰 비극성을 가져서 물에 녹지 않아 세포막의 지질층까지 도달하지 못한다. 반면에, 너무 낮은 AlogP98 값은 물에 잘 녹지만, 세포막의 지질층을 통과하지 못한다. AlogP98 분자표현자는 약물유사 화합물의 세포막 투과도를 예측할 때 많이 적용된다.

또 다른 분자표현자 2D-PSA는 2차원 분자구조에서 극성을 가지는 작용기(functional group)가 물에 노출된 표면적의 합이다(Eq. (2)).

$$2D\text{-PSA} = \sum_i^{n\text{-types}} n_i \cdot C(\text{fragment}_i) \quad (2)$$

여기서 n-types는 극성을 가지고 있는 작용기 종류의 수이고,  $C(\text{fragment}_i)$ 는 분자 내 I형태의 작용기가 기여하

는 물에 노출된 표면적에 상수 C의 곱이며,  $n_i$ 는 분자 내 I 형태의 작용기 빈도수이다.

Eq. (2)과 같이 질소와 산소원자에 결합되는 수소원자들을 포함한 작용기가 물에 노출이 된 표면적을 계산[24]한다. 이 때문에 다른 화합물의 분자크기와 친지용성을 갖는 화합물에서 동일한 2D-PSA 분자표현자 값을 가지면, 화합물의 극성에 대해 해석을 하지 못한다. 결과적으로 동일한 2D-PSA를 가진다면, 비극성 화합물이나 극성 화합물 유래 극성작용기 인지를 명확하게 구별하지 못하는 단점이 있다.

$\log Sw$  분자표현자는 25°C, pH=7.0 조건에서 화합물의 물 용해도(aqueous solubility)를 의미한다[26]. 물 용해도는 약물유사 화합물의 흡수와 관련된 매우 중요한 물리화학적 성질이다. 장에서 흡수 될 수 있는 화합물의 양과 상관관계가 있고, 또한 물 용해도가 낮은 화합물은 일반적으로 친지용성의 성질을 갖는다(높은 AlogP98 값을 갖는 경향이 있음). 게다가, 화합물이 이온화 가능성이 있을 경우에는 화합물의 물 용해도가 pH에 따라 변화된다. 이는 각 화합물의 이온화상수(pKa)가 수용액에 수소이온( $H^+$ ) 주개와 관련이 있어서 강산/약염기에 대한 정보도 필요하다. 장과 같은 흡수기관의 pH 조건에 따른 화합물 구조의 이온화 경향을 반영하여, AlogP98,  $\log Sw$  분자표현자를 계산해서 화합물의 용해도, 친지용성, 생체막, 기관의 투과도를 해석한다.

Table 1. Aqueous solubility levels from computed logSw molecular descriptor[26]

Level	Value	Description
0	$\log Sw < -8.0$	Extremely low
1	$-8.0 < \log Sw < -6.0$	No, very low, but possible
2	$-6.0 < \log Sw < -4.1$	Yes, low
3	$-4.1 < \log Sw < -2.0$	Yes, good
4	$-2.0 < \log Sw = 0.0$	Yes, optimal
5	$0.0 < \log Sw$	No, too soluble
6	-1000	Molecules with one or more unknown AlogP98 types

다만, 물 용해도 분자표현자로 수치화를 하는 데 있어서, AlogP98를 계산할 수 없는 작용기나 치환기 그룹을 가진 경우에는 물 용해도에 각 원자의 기여도는 더해지지 않는다(Eq. (1)). 물 용해도는 알칸, 알켄, 할로젠, 아민계열, 알코올, 질소를 갖고 있는 화합물, 그리고 케톤, 알데히드, 유기산등의 화합물에 대해 계산할 수 있다. Table 1에서 보는 바와 같이, 다양한 화합물들의 물 용해도는 logSw 분자표현자로 계산한 값과 이에 따른 분류단계(0~6)로 용해도 수준이 해석되었다.

## 2.2 분자표현자로부터 ADMET 성질 계산

사람에서 장 흡수도(human intestinal absorption)는 약물 유사 화합물을 경구 투여 후, 흡수되는 정도(%)를 나타낸다. 장 흡수도가 높은 화합물은 혈류에서 90% 이상 흡수된다. 장 흡수도는 AlogP98과 2D-PSA 분자표현자를 적용하여 다중회귀 모델로부터 예측될 수 있다[18]. 이때 AlogP98과 2D-PSA로 구성된 물리화학적 성질 좌표에서 마할라노비스 거리(Mahalanobis distance)인 2D-T2 개념을 사용하여 평균과의 거리가 표준편차의 몇 배인지를 나타내는 값을 추가하여 흡수도를 4단계 수준(0~3)으로 분류하였다(Table 2).

약물유사 화합물에 대한 생체막의 투과도는 장 외막 투과도와 뇌-혈관장벽(BBB, Blood Brain Barrier) 투과도들이 포함된다. 약물유사 화합물이 경구 투여 후, 장 흡수도(% HIA)와 다르게 뇌-혈관장벽 투과도는 중추신경계(CNS)로 유입되는 특성을 혈류-뇌 투과농도의 비(a ratio of concentration within blood-brain penetration)로 설명하였다[28]. 뇌-혈관장벽 투과도는 장 흡수도처럼 2개 AlogP98과 2D-PSA 분자표현자로 설명이 될 수 있고, 이들의 좌표에서 95%와 99% 신뢰 타원(confidence ellipses)으로 경향성을 나타낼 수 있다.

약물유사 화합물의 장 흡수도(Table 2)와 뇌-혈관장

벽 투과도(Table 3)는 동일한 AlogP98과 2D-PSA 분자표현자의 좌표로 나타낼 수 있지만, 유사한 투과도 수준을 갖지 않는다. 예를 들어, 수동수송의 기작에 따라 투과되는 화합물은 장 흡수의 경우 2D-PSA가  $120 \text{ \AA}^2$  이상, 뇌-혈관장벽 투과도의 경우  $60-70 \text{ \AA}^2$  이면 낮은 투과도를 갖는 다고 알려져 있다[29].

Table 2. Four prediction level of human intestinal absorption from computed properties[18]

Level	Value	Description
0	Absorption 2D-T2 < 6.1261 (inside 95%)	Good absorption
1	$6.1261 \leq \text{Absorption 2D-T2} < 9.6026$ (inside 99%)	Moderate absorption
2	$9.6026 < \text{Absorption 2D-T2}$ (outside 99%)	Low absorption
3	$2D-PSA \geq 150.0$ or $AlogP98 \leq -2.0$ or $AlogP98 \geq 7.0$	Very low absorption

Table 3. Four prediction level with logBBB values within the 95% and 99% confidence ellipsoids[28]

Level	Value	Description
0	$\log BBB \geq 0.7$	Very high penetrant
1	$0 \leq \log BBB < 0.7$	High penetrant
2	$-0.52 < \log BBB < 0$	Medium penetrant
3	$\log BBB \leq -0.52$	Low penetrant
4	Undefined level	Undefined

약물유사 화합물의 약효를 결정하는 데에서 대사(metabolism)작용은 분포부피(volume of distribution)에 부합되고 약물의 반감기와 청소율(clearance)을 예측하는 것도 연계된다. 또한, 대사 작용에 관여하는 사이토크롬 P450(CYP450)과 같은 효소들의 작용, 이로부터 분해되는 대사물질의 효과 등이 모두 복합적으로 연관되어 있다. 더 나아가서, 서로 다른 경로를 통해 생성되는 수많은 대사물질은 간 독성과도 직접적으로 관련되어 있다. 이번 연구에서 대사 작용을 Acceryls사의 Metabolism 화합물 데이터베이스를 통해 지식 기반 rules에 의해 Bayesian learning 방법으로 2D-분자구조로부터 사이토크롬 P450(CYP450) 이소효소 2D6 (CYP2D6) 저해능력을 예측하였다[27]. 이 대사모델을 통해서 각 화합물의 CYP2D6 효소 저해능력을 확률변수(모수)로 해서 그

Table 4. Computed properties and their descriptions of ADMET molecular descriptors

Properties	Description
2D-PSA	2D-Fast polar surface area[24]
AlogP98	Atom-based logP of the octal-water partition coefficient[27]
Solubility	The base 10 logarithm of molar solubility expressed as a logS, where S is the solubility in mol/L predicted by the regression model[26]
pKa	The pka of all ionizable sites [15]
Blood Brain Barrier (BBB)	Base to logarithm of (brain concentration)/ (blood concentration) as predicted by a robust (least-median-of-squares) regression derived from literature in vivo brain penetration data[28]
CYP2D6	Bayesian score from the cytochrome P450 2D6 (CYP2D6) model[25]
Hepatotoxicity	Bayesian score from the hepatotoxic model[30]
Plasma Protein binding (PPB)	Bayesian score from the plasma protein binding model[25,31]

변수의 확률분포를 추정한 베이지안 점수(Bayesian score)로 계산하였다. 이때, 베이지안 점수의 값이 0.161기준으로 각 화합물이 CYP2D6 효소의 저해제로 작용 가능성 유/무(Bayesian score > 0.161 then true, or Bayesian score < 0.161 then false)를 결정하였다(Table 4).

화합물의 2D-분자구조를 이용하여 물리화학적 성질과 관련된 분자표현자와 25개 구조 표현자(간 독성이 존재하는 것으로 알려져 있는 fragment와 구조 유사도가 반영된 표현자 등)를 이용하여 간 독성의 발생 유무(독성/무독성)[30]를 예측하였다. 간독성 유/무를 예측하는 데에 흡수에 영향을 주는 AlogP98, 수소결합 받개 수, 생체이용률과 혈장 단백질의 결합력에 영향을 주는 화합물 내 회전 가능한 결합수 등이 포함되었다. 또한, Bayesian learning 방법으로 간독성 값을 베이지안 점수로 예측하였다. 예측과정에서 화합물의 간독성의 발생유무에 대한 기준 값은 -4.154 베이지안 점수(각각 귀무가설을 잘못 기각하는 1종 오류와 귀무가설을 잘못 채택하는 2종 오류를 최소화함으로써 얻어진 기준 점수)로 간독성/무독성(Bayesian score > -4.154 then true, or Bayesian score < -4.154 then false)을 분류하였다(Table 4).

투여된 화합물의 혈장단백질에 결합여부는 약물의 효율에 영향을 준다. 혈장 단백질에 결합이 되면 대사과정으로부터 일시적으로 차단된다. 다른 측면으로 혈장 단백질에 결합을 하지 않은 일부 화합물만이 약리작용을 나타낸다[31]. 각 화합물들의 상대적 혈장 단백질 결합력을 베이지안 점수로 나타내었다. 예를 들어 탄소고리 화합물, 카르복실산염, 아민, 하이드록실기, 할로젠 작용기를 가진 화합물들 중에서 -2.209 베이지안 점수를 기준으로 혈장 단백질에 90%이상 결합유무(Bayesian score > -2.209 then true, or Bayesian score < -2.209 then false)를 해석하였다(Table 4).

### 2.3 분자표현자로부터 약물가능성 측정

약물유사성은 화합물의 우선순위를 결정하는 데 매우 중요한 고려사항이다. 그러나 약물유사성의 평가는 화합물의 전체적인 품질(quality)을 평가하는 것은 아니다. 다만, 11,144개 화합물들의 약물가능성(druggability)을 측정하기 위하여 2012년 Hopkins에 의해 제안된 분자표현자를 이용하여 정량적으로 경구용 약물유사성(Quantitative Estimate of Drug-likeness, QED)을 계산[32]하였다. 약물유사성과 관련된 대표적인 8개 분자표현자는 분자량(MW), 옥탄올-물 분배계수(AlogP98), 수소결합 주개 수 (HBD)와 수소결합 받개 수(HBA), 분자의 극성표면적(2D-PSA), 회전결합수(ROTB), 방향족 고리수(AROM), 화합물 내 독성유발 substructure 구조 개수(ALERT)를 포함한다(Table 5).

약물가능성을 평가하는 데 있어 각 분자표현자의 요구되어 지는 값의 편향(bias)으로 기인한 변동 등을 최소화하는 데 만족도 함수(desirability function)를 이용하였다.

$$\text{unweighted QED} = \exp(1/n \sum_{i=1}^n \ln d_i) \quad (3)$$

여기서 11,144개 화합물 내에서  $n$ 은 분자표현자의 수이고,  $d_i$ 는 각 분자표현자의 만족도 함수이다.

약물가능성을 평가하는 함수(QED)는 각 분자표현자의 값에 대하여 만족도 함수의 기하평균을 구하여 그것을 최대로 하는 변수의 조합을 찾아 가중치(weighted)를 뒤서 계산한다(Eq. (4)).

$$\text{weighted QED}_w = \exp([\sum_{i=1}^n w_i \ln d_i] / (\sum_{i=1}^n w_i)) \quad (4)$$

여기서 11,144개 화합물 내에서  $n$ 은 분자표현자의 수이고,  $d_i$ 는 각 분자표현자의 만족도 함수이고  $w_i$ 는 각 만족도함수에 적용된 가중치이다(Table 5). 결과적으로 경구용 약물가능성을 평가하는 QED 값은 0(모든 물리화

Table 5. Computed molecular properties and their descriptions of QED (Quantitative Estimate of Drug-likeness)

Properties	Description
QED <sub>w</sub>	Quantitative Estimate of Drug-Likeness (QED) is calculated [32], when the default weight factors for QED_MW, QED_AlogP, QED_HBA, QED_HBD, QED_PSA, QED_ROTb, QED_AROM, and QED_ALERTS are 0.60, 0.46, 0.05, 0.61, 0.06, 0.65, 0.48, and 0.95, respectively.
QED_unweighted	Unweighted estimate of drug likeness (desirability) score.
QED_MW	Molecular weight contribution to the QED score.
QED_AlogP	AlogP contribution to the QED score.
QED_HBA	Hydrogen bond acceptor contribution to the QED score.
QED_HBD	Hydrogen bond donor contribution to the QED score.
QED_PSA	Polar surface area contribution to the QED score.
QED_ROTb	Contribution to the QED score from the number of rotatable bonds.
QED_AROM	Contribution to the QED score from the number of aromatic rings.
QED_ALERTS	Contribution to the QED score from the number of undesirable substructures found in the molecule, which is calculated based on a list of 113 SMARTS queries.

학적 성질이 비적합)과 1(모든 물리화학적 성질 적합) 범위 내에 존재한다.

약물가능성 평가에서 가중치를 두었을 경우(weighted)와 가중치를 두지 않고(unweighted) 계산한 QED 점수간의 관계의 강도를 상관관계 계수로 살펴보았다. 두 QED 점수간의 어떤 선형적인 상관관계(Correlation, Correlation coefficient)를 갖고 있는지를 이변량 상관분석(피어슨 상관 계수 r<sup>2</sup>로 가중치를 두지 않은 QED 점수로부터 가중치를 둔 QED 점수를 예측할 수 있는 정도)하였다.

또한, 약물유사성 관련 8개 분자 표현자를 이용하여 11,144개 화합물에 대해 주성분 분석(Principle component analysis, PCA)을 통해 서로 연관 가능성이 있는 고차원 데이터들을 선형연관성이 없는 저차원 공간(주성분, 공분산 행렬의 고유벡터)으로 직교변환을 하였다. 이때 데이터들의 분산이 가장 큰 방향벡터를 이용하여 3차원 데이터들을 3개 서로 수직인 주성분 벡터로 변환하였다. 그 결과로, 높은 약물유사성을 갖는 화합물들을 8개 분자표현자로부터 직교 변환된 3개 축(PC1, PC2, PC3)에서 분포를 확인하였다.

## 2.4 화합물간의 유사성 기반 다양성 평가

화합물의 유사성(molecular similarity)은 두 분자간의 유사성이 얼마나 존재하는지에 대하여 정량적으로 설명한다. 이는 구조적으로 유사한 분자들은 비유사한 분자들보다 더 물리화학적 성질들이 유사하다는 가정이 존재한다. 그래서 기능적으로 유사한 분자들을 찾거나 한다면, 구조적으로 유사한 분자를 검색한다. 화합물의 물리화학적 성질(분자량, AlogP98 등)을 이용하여 유사한 분자를 검색할 때에는 구조적 또는 분자표현자 공간

(molecular descriptor space)에서 측정된 상대적 거리 매트릭스로 나타낸다. 예를 들어, 물리화학적 성질의 분자표현자 값으로 분자간의 유사성을 측정하는 상대적 거리를 맨해튼 거리(Manhattan distance), 유클리드 거리(Euclidean distance), 두 분자표현자간의 R 상관관계를 이용하여 (1-correlation) 또는 (1-correlation)<sup>2</sup> 거리로 측정할 수 있다.

$$\text{Manhattan distance } D_{ij} = \sigma[|X_{ik} - X_{jk}|] \quad (5)$$

맨해튼 거리는 두 분자의 각 좌표 차를 모두 더한 것 (Eq. (5))으로 나타내는 반면에, 유클리드 거리는 각 분자에게 주어진 물리화학적 성질의 분자표현자에 대한 벡터 값의 차이를 나타낸다(Eq. (6)).

$$\text{Euclidean distance } D_{ij} = \text{sqrt}(\sigma[(X_{ik} - X_{jk})^2]) \quad (6)$$

맨해튼 거리는 유클리드 거리보다 크거나 같다는 특성이 있다. 이외에도 두 분자표현자(예. i와 j) 간의 R 상관관계를 이용하여 OneMinusCorrelation(1-correlation)은

$$\text{OneMinusCorrelation } D_{ij} = 1 - R \quad (7)$$

OneMinusCorrelationSquared (1-correlation)<sup>2</sup> 거리로서

$$\text{OneMinusCorrelationSquared } D_{ij} = 1 - R^2 \quad (8)$$

거리 매트릭스가 짝을수록 두 분자간의 유사성은 높다. 두 분자의 물리화학적 성질을 기반으로 유사성 평가와 달리, 특정 화합물 내 fragment의 fingerprint로 구조적으로 유사한 분자를 찾거나 할 때는 타니모토 계수(Tanimoto coefficient)를 적용한다[33]. 분자내 원자간의 결합의 정보를 이용하여 유사성을 계산하며, 0~1 범위의 값을 가진다. 타니모토 계수의 값이 1에 가까울수록 가장 유사한 분자이다(dissimilar distance와 반대

경향을 가짐).

다른 측면으로, 분자의 다양성은 11,144개 화합물들의 구조에 대해 fingerprint 특징들의 수, 작용기들로 구성된 어셈블리들의 수, fingerprint 거리, 물리화학적 성질의 거리 등을 기반으로 다양성 매트릭스로 비교한다. 이중에서 화합물들의 다양성 평가는 두 화합물간의 fingerprint 거리(최소, 최대, 평균)와 유클리드 물리화학적 성질(Eq. (6))의 평균거리를 적용하였다. 그 결과로 최대 상이성(dissimilarity)중심으로 각 그룹으로 클러스터링 하였다. 본 연구과정에서 수행된 계산 및 분석은 상용화 discovery studio 분자모델링 프로그램(BIOVIA 2017 R2 version[34])에서 수행하였다.

### 3. 실험결과 및 고찰

#### 3.1 분자의 2D-구조로부터 ADMET 성질 계산

11,144개 화합물의 2D-구조 기반 분자표현자 (AlogP98, 2D-PSA, logSw 등)를 이용하여 ADMET와 연관된 6개 성질인 ① 물에서 용해도(solubility), ② 장 흡수 수준, ③ 뇌-혈관장벽(BBB) 투과도, ④ 간 대사 관련 CYP2D6 효소의 저해능력, ⑤ 혈장단백질 결합력 (Plasma Protein Binding, PPB), ⑥ 간독성 유/무 등을 예측하였다(Fig. 3). 이들 중에서 분자표현자 AlogP98과 2D-PSA의 좌표 공간에서 ② 장 흡수 수준과 ③ 뇌-혈관장벽(BBB) 투과도에 따른 분포를 각각 2개 신뢰타원형(95%, 99%)을 적용하여 나타냈다(Fig. 4).

흡수(Absorption)는 화합물 투여 후 세포막이나 점막을 통해서 위벽이나 위장관의 내장흡수가 되어야 한다. 낮은 화합물의 용해도, 위내 체제시간이 길어진다면 흡수시간을 줄일 수 있는 방향으로 화합물의 구조 및 물리화학적 성질(분자의 크기, 용해도, 양극성 등)을 최적화하여야 한다. 이에 따라, 신체 내 흡수와 관련된 ① 물의 용해도(logSw)와 ② 장 흡수 경우에는 분자표현자를 계산할 수 있는 8,655개 화합물(11,144개 화합물의 77.67%)중에서 물에 잘 용해되는 767개 화합물을 포함(용해도 최적인 4그룹)하여 5,436개 화합물이 위장관의 흡수도가 높았다(Fig. 3(a)). 장에서 흡수는 물에서 용해도와 위벽의 투과도에 연관된 화합물의 지용성(옥탄올-물 분배계수 AlogP98)에도 영향을 받기 때문이다. 다른 측면으로 ③ 뇌-혈관장벽(BBB) 투과도는 분자의 크기가 지용성(AlogP98)보다 더 중요하게 작용한다. Fig. 4에서 보는 바와 같이, 분자표현자 좌표공간에서 2D-PSA:50.525,

AlogP98:2.442의 위치에 해당하는 화합물154번(95% 신뢰 타원의 중심에 존재)은 장 흡수가 매우 잘되는 0그룹에 해당하지만, 뇌-혈관장벽 투과도가 중간정도인 2그룹(logBBB:-0.199)에 속한다. 따라서 장에서 흡수와 뇌-혈관장벽 투과도간의 차이가 존재한다.

장에서 흡수되는 화합물은 뇌-혈관장벽 투과도보다 더 적은 값의 지용성과 더 극성 표면적을 갖는 경향이 있다. 장 흡수의 99% 신뢰 타원형이 뇌-혈관장벽 투과도의 99% 신뢰 타원형보다 더 넓게 존재한다(Fig. 4). 이와 연관되어 분포(distribution)는 화합물이 혈류를 통해 이동해서 세포, 근육, 장기에 혈장의 농도를 낮추는 경향으로 분포하게 된다. 가역적인 생리학적 반응이다. 분자량이 작을수록, 지용성이 높을수록 조직의 세포막을 통과하기 쉽고 확산이 되어 조직에 많이 분포하게 된다. 하지만, 세포막을 경계로 세포에 필요한 이온이나 당, 아미노산 등의 농도를 유지하기 위한 세포에서 농도차이를 극복(저 농도에서 고 농도로 이동)하는 능동수송(active transport)과 외부 유출 펌프(efflux pump)로 흡수되는 화합물의 흡수도는 위와 같은 물리화학적 성질로만 설명되지 않는다. 다만, 친지용성인 세포막으로 화합물이 투과 할 때 수동적 이동(passive transport)이 주로 속도 결정단계이기 때문에, 화합물의 소수성, 친지용성을 나타내는 AlogP98 분자표현자 등으로 세포막투과도 뿐만 아니라 흡수 수준을 예측 할 수 있는 것이다.

신체 내 화합물이 투여되면, 간의 산화환원효소 사이토크롬 P450(CYPs)과 CYPs 이소효소(isoform)인 CYP1A, 2C9, 2C19, 2D6, 3A4 등에 의해 대사(metabolism)작용을 통해 분해되어 대사산물이 생성된다. 승인된 약물의 90%가 CYPs 이소효소에 의해 대사되기 때문에 이소효소별로 기질의 특이성, 대사체로 분해되는 위치, 기질(substance)과 저해 효과를 갖는 화합물(inhibitor) 간의 비교를 통해 구별할 수 있는 substructure를 규명한다. 이 접근의 어려운 점은 대사체로 분해되지 않는 substructure를 정의하는 것이다. 이를 극복하기 위해, 화합물이 작용부위에서 충분히 노출이 되어서 효과를 이끌어 내기위한 전제조건인 substructure fragment들의 공개된 정보, 그리고 CYPs P450 저해 화합물을 <http://www.lmmd.org/database/cheminformatics/p450i> 검색하여 연구에 적용하였다.

신체 내에서 화합물의 효과를 결정(반감기 등)하는 데 중요한 역할을 하는 것은 간 대사과정이다. ④ CYP2D6 효소의 작용에 의해 분해되는 대사물질을 예측하는 것 이외에도, 서로 다른 대사경로에 따라 생성되는 대사물



질의 구조 다양성 및 화합물의 효과를 나타내는 데 방해 요인인 ⑤ 혈장단백질 흡착, ⑥ 대사물질의 간독성 유/무 예측 등은 밀접한 인과관계로 연관성이 있다. 이에 따라, 동시에 Bayesian learning 방법을 적용하여 각 Bayesian score 값으로 주어진 논리적 기준을 판단하여 참(true)과 거짓(false)으로 나타내었다(Fig. 3(b)). 예를 들어, 화합물 154번(관용명: AL-37350A)은 선택적으로 세로토닌 5-HT<sub>2</sub> 수용체에 작용(EC<sub>50</sub>=28.6nM, Emax=103%)하여 안구의 압력을 낮추는 효과를 갖는다[35]. 그래서 녹내장에 의한 안압이 급속도로 높아지면서 시력감소, 두통, 구토, 출혈 등 증상에 대한 완화효과를 가진 항녹내장제의 유효물질로 고려대상(FDA Global Substance Registration System, GSRS ID: KT54N4CC67)이다.

화합물 154번의 분자량(molecular weight)은 230.31g/mol이고 수소결합 주개 수(HBD) 2개, 수소결합 받게 수(HBA) 2개를 가지고 있으며, 이 화합물의 2D-구조에 따른 극성표면적(2D-PSA) 50.53Å<sup>2</sup>를 가지고 있다(Fig. 4). 이러한 물리화학적 성질들을 적용하여 Bayesian learning 방법으로 간 대사·분해·독성과 관련된 논리적 기준 값을 설정하였다. 설정된 논리적 기준 값에 따라, 화합물 154번은 간 산화환원 이소효소 CYP2D6의 저해제로 작용하지 않는다(Bayesian score < 0.161 then false). 혈장단백질에 90%이상 결합하지 않고(Bayesian score < -2.09 then false) 효과를 나타

내다가, 간독성을 나타내지 않는다(Bayesian score < -4.154 then false in Fig. 3(b)). 만일 화합물이 매우 높은 소수성(logP)을 갖는다면, 응집이 되어 간 독성을 유발하기도 한다. 경구용 약물은 개발과정에서 중단된 후보 화합물보다 분자량(MW)과 소수성(logP)이 더 적은 경향이 있다[36]. 친유성 후보 화합물이 경구용 약물 개발과정에서 중단되고 있는 빈도수가 더 높다.

작은 분자크기의 화합물인 경우에는 간 대사, 신장 및 담즙 배설을 통해 신체 내에서 제거되지만, 단백질 약물들은 리소솜에서 이화작용을 통해 제거될 수 있다. 다만, 60kDa미만의 분자크기를 갖는 단백질 약물은 신장 배설도 중요한 제거(clearance) 경로이다. 투여된 화합물의 경우와 달리, 단백질 약물은 작은 화합물의 수용체와 관련이 없으며, 대사 효소 또는 혈장 단백질에 결합, Fc 수용체 결합에 의해 체내 흡수 속도에 영향을 준다. 투여된 화합물의 배설(Excretion)은 소변을 통해서 배출된다. 생성물들이 배설되는 신장이 가장 중요한 장기이다. 그리고 간에서 대사산물들은 담즙 배설 외에 소화관을 거쳐 대변과 함께 배설이 된다.

위와 같이 투여된 화합물이 신체 내 흡수되고, 심장박동으로 일어나는 혈류에 의한 말초조직으로 분포되어 있다. 이후, 간이나 신장을 통한 신체 밖으로 배설될 때까지 혈액과 조직에서 약물 농도가 시간에 따라 변화하는 과정이 약동학(PK, Pharmacokinetics)이다[37]. 이때



Fig. 3. Physicochemical profiles of the 11,144 compounds from Pubchem DB with histogram of some commonly used molecular descriptors(i.e. AlogP98, 2D-PSA, logSw) with a red or blue star to highlight the accepted values of properties (a) on the corresponding descriptors and (b) on the cutoff Bayesian scores after oral administration.

약동학 파라미터들로 살펴보면, 흡수에는 일차-흡수속도 상수(Ka)와 생체이용률(F), 분포에는 분포용적(Vd), 대사나 배설을 통한 혈장에 있는 화합물을 제거하는 능력은 청소율(CL)이다. 시간당 제거되는 화합물의 양(청소율)인 경우, 약물의 혈장 농도가 높을 때에는 제거 능력이 상대적으로 크다가 농도가 떨어지면 줄어든다.

결과적으로, 작은 화합물의 신체 내 흡수·분포·대사 배설 그리고 독성과 관련된 물리화학적 성질(수용액 용해도, 세포 투과도 등)과 약동학(PK) 파라미터(흡수속도 상수, 생체이용률 등)간의 결합은 약물유사(drug-likeness) 화합물이 나타내는 성질(약물유사성을 정량적으로 평가하는 8개 분자표현자)들의 좌표 공간(molecular property space)을 정의하게 되었다. 약물유사 화합물의 좌표 공간은 판매되고 있는 약물의 공간과 완벽하게 일치하지 않는다. 약물유사 화합물은 목표단백질에 작용기전이 동일하다는 조건하에 특정한 약물에 화합물이 얼마나 유사한지를 평가하기 때문이다. 따라서 약물유사성과 관련된 대표적인 8개 분자표현자로 구성된 분자성질 좌표공간은 약동학적 파라미터를 최적화 하는 과정에서 간격이 좁아지게 된다. 그 결과, 판매되고 있는 약물분자의 특성 공간에 채워질 수 있다.

### 3.2 8개 분자표현자 적용 약물가능성 평가(QED)

11,144개 화합물의 경구용 약물가능성(druggability)은 지식기반 물리화학적 성질들의 분포를 직관적으로 8개 분자표현자를 사용하여 하나의 척도(QED score)로 변환하고 이를 최적화하여 상대적 우선순위를 결정한다 (Eq. (4)). 이때, 8개 분자표현자로 고려하는 것은 분자량, 옥탄올-물 분배계수, 수소결합 주개 수와 받게 수, 분자의 극성표면적, 회전결합수, 방향족 고리 수, 독성유발 부분구조의 수이다(Table 5). 승인된 약물의 8개 분자표현자로 표현된 물리화학적 성질의 범위는 좌표공간에서 좁은 영역에 분포되어 있다. 화합물의 특성이 해당 영역에 존재를 하면 약물유사(drug-like) 화합물이라고 한다.

약물유사성은 8개 분자표현자들이 신체 내에서 용해도, 투과도, 대사 안정성, 능동·수동 수송의 효과, 생체이용률 등에 어떻게 반응이 될지를 고려하였다. 그 결과, 약동학적 측면에서 부적합 화합물은 약물가능성도 낮다.

승인된 약물의 정량적 약물유사성(QED) 점수는 0.67 (Standard Deviation S.D.=0.16)이고, 약물가능성이 낮은 화합물의 점수는 0.49(S.D.=0.16)이다. 특히, 화합물의 구조가 복잡해서 약물로서 부적합한 경우에는 QED 점수가 0.34(S.D.=0.24)이다. 더 나아가서, DrugBank (<http://www.drugbank.ca/>)에 등록 되어 있는 1,124개 승인된 약물의 평균 QED 점수는 0.57이고 ChEMBL

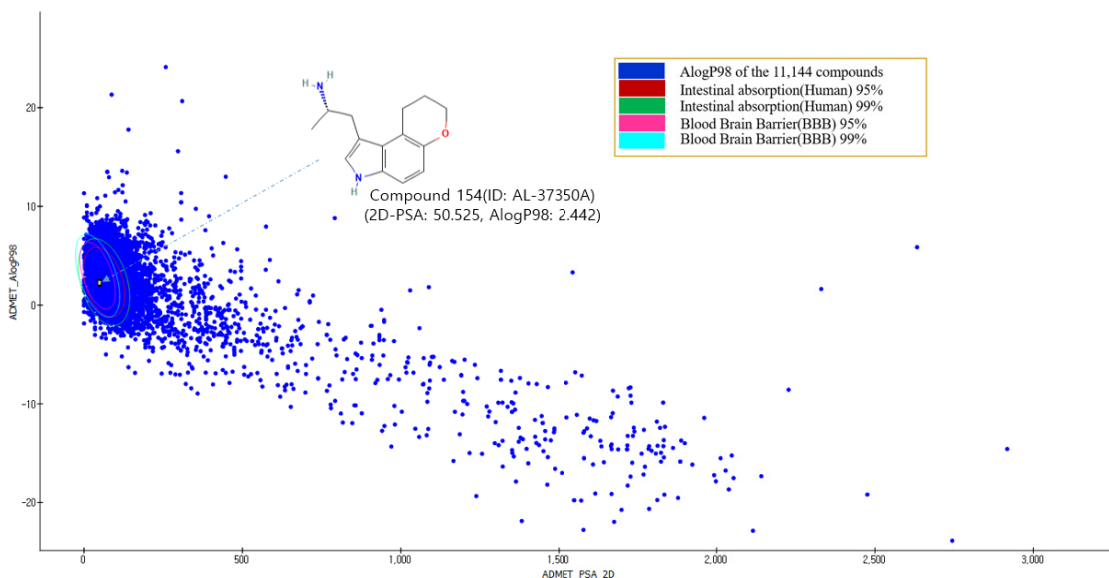


Fig. 4. The predicted profiles of human intestinal absorption and blood-brain penetration in the 11,144 compounds from PubChem DB, as well as 95% and 99% confidence ellipses in the ADMET\_2D-PSA, ADMET\_AlogP98 plane.

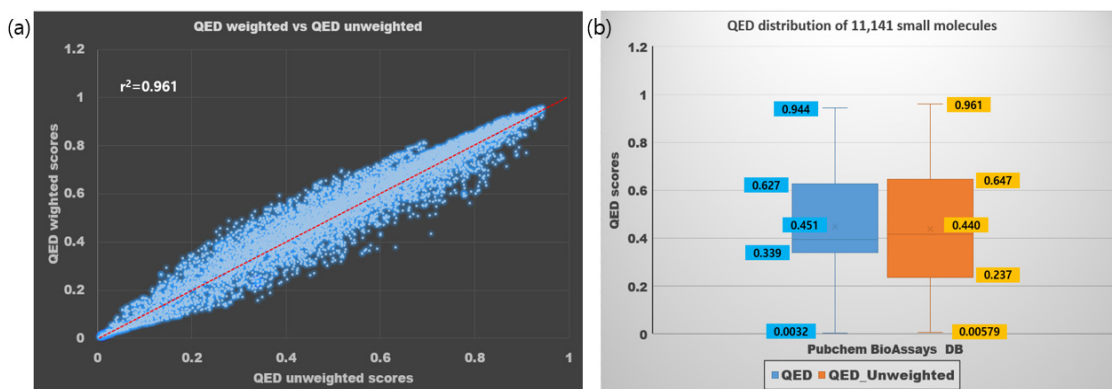


Fig. 5. Weighting dependencies of QED. (a) Scatter plot of between weighted and unweighted QED scores for the set of 11,141 small molecules ranked in order. (b) Variation in between QED scores, due to different weighting eight QED descriptors.

Table 6. Comparison between weighted and unweighted QED values corresponding number and percentage of 11,141 molecules with respected to the PubChem BioAssays database.

QED <sub>w</sub> values	Number(Percentage %) of molecules in the PubChem BioAssays DB	QED <sub>unweighted</sub> values	Number(Percentage %) of molecules in the PubChem BioAssays DB
> 0.90	118(1.06)	> 0.90	202(1.81)
> 0.80	762(6.84)	> 0.80	844(7.58)
> 0.70	1,101(9.88)	> 0.70	1,129(10.13)
> 0.60	1,103(9.90)	> 0.60	1,147(10.30)
> 0.50	1,175(10.55)	> 0.50	1,201(10.78)
> 0.40	1,253(11.25)	> 0.40	1,204(10.81)
> 0.30	3,543(31.82)	> 0.30	986(8.85)
> 0.20	738(6.62)	> 0.20	3,100(27.82)
> 0.10	530(4.76)	> 0.10	482(4.33)
> 0.00	816(7.32)	> 0.00	846(7.59)
Total	11,141(100.00)	Total	11,141(100.00)

(<http://www.ebi.ac.uk/chembl/db>)에 포함이 된  $5 \times 10^5$ 개 생리활성 화합물들의 평균 QED 점수는 0.404이다. 공용화합물 DB내 화합물의 다양성에 따라 약물유사성의 평균 QED 점수는 달라진다. 또한 약물가능성이 있는 화합물이 개발단계에서 중단된 화합물보다 상대적 QED 점수는 더 높다. 약물 유사성의 QED 점수는 서로 다른 생물학적 반응경로에서 목표단백질의 작용위치 및 기능을 고려하지 않기 때문에, 각 기준 점수는 달라진다. 유의적으로, 약물 유사성 화합물의 가능성 여부에 대해서 QED 점수를 절대평가 하지 않는다. 예를 들어, ChEMBL DB에서 1,729개 목표단백질에 대해 binding affinity  $< 10 \mu\text{M}$ 을 갖는 167,045개 유효화합물의 평균 QED 점수는 0.478이다[32]. 이를 고려하여, 본 연구에서 약물유사성의 QED 점수는 화합물을 가상탐색(e.g. ligand based screening) 할 때 11,144개 화합물의 상

대적 우선순위를 결정하기 위해 적용하였다. PubChem BioAssays에 포함된 11,144개 화합물 중에서 독성을 유발하는 부분구조를 정의할 수 없는 3개 화합물을 제외하고, 11,141개 화합물의 약물 유사성 QED 점수를 계산하였으며, 평균점수는 0.480(S.D.=0.24)이었다(Fig. 5). 이때 약물가능성과 관련된 8개 분자표현자에 각각 가중치를 둔 QED<sub>w</sub> 점수(Eq. (4))의 범위는 0.0032(Min) ~ 0.944(Max)에서 중간 값은 0.451이었다(Fig. 5(b)). QED<sub>w</sub> 점수의 25%~75%에 해당하는 값은 0.339~0.627범위에 있다. 반면에 가중치를 두지 않은 QED<sub>unweighted</sub> 점수(Eq. (3))는 0.0058(Min) ~ 0.961(Max)에서 중간 값은 0.440이었고 가중치를 둔 QED<sub>w</sub> 점수와 선형적으로 높은 0.961 상관도로 연관성을 가졌다(Fig. 5(a), Table 6). 이때 약물가능성의 QED 점수와 8개 분자표현자간의 일대일 상관관계 행렬(pair-wise correlation

matrix) 분석결과에 따르면, 가중치를 둔 QED<sub>w</sub> 점수와 화합물 내 회전결합수(QED\_ROT<sub>B</sub>)간의 상관관계 0.75 계수로 가장 높았지만, 방향족 고리수(QED\_AROM)가 0.41로 낮았다(Fig. 6).

이와 다르게, 가중치를 두지 않은 QED\_unweighted 점수와 분자의 극성표면적(QED\_PSA)간의 상관도가 0.83으로 가장 높았지만 독성유발 부분구조의 수(QED\_ALERTS)와 0.25로 가장 낮은 상관관계를 가졌다. 특히, 독성유발 부분구조의 수는 분자 내 방향족 고리수(QED\_AROM)와 수소결합 받개 수(QED\_HBA), 분자량(QED\_MW)과 음의 상관관계이며 매우 낮은 상관도 (< -0.001)를 가졌다(Fig. 6).

추가적으로, 11,141개 화합물의 8개 물리화학적 성질 값에 대해 주성분 분석(Principle Component Analysis, PCA)을 통해 3개축으로 구성된 좌표계(PC1, PC2, PC3)로 직교변환을 하여 약물가능성이 높은 분자들의 (QED 점수가 상대적으로 높은 순서를 가짐) 분포를 살펴해보았다(Fig. 7). Fig. 7에서 보는 바와 같이, 약물가능성이 높은 화합물은 PC2와 PC3축으로 구성된 평면에 밀도 있게 존재를 한다. 반면에 낮은 가능성을 갖는 화합물(QED 점수 < 0.25, 빨강색 표기)의 경우에는 PC1과 PC2 축 평면에 넓게 분포되어 있다. 특히, 약물유사 화합물로 우선순위가 높은 화합물인 경우(QED<sub>w</sub> 점수 > 0.90, 파랑색 표기)에는 PC2-PC3 축 평면의 중심에

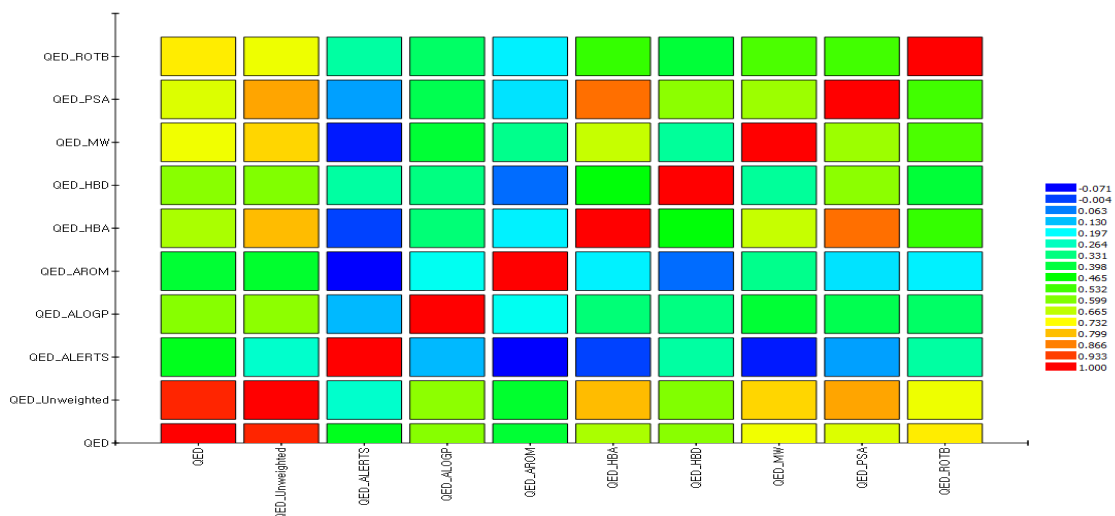


Fig. 6. Pair-wise correlation matrix for 8 molecular descriptors of QED scores with respect to 11,141 compounds from the PubChem BioAssays database.

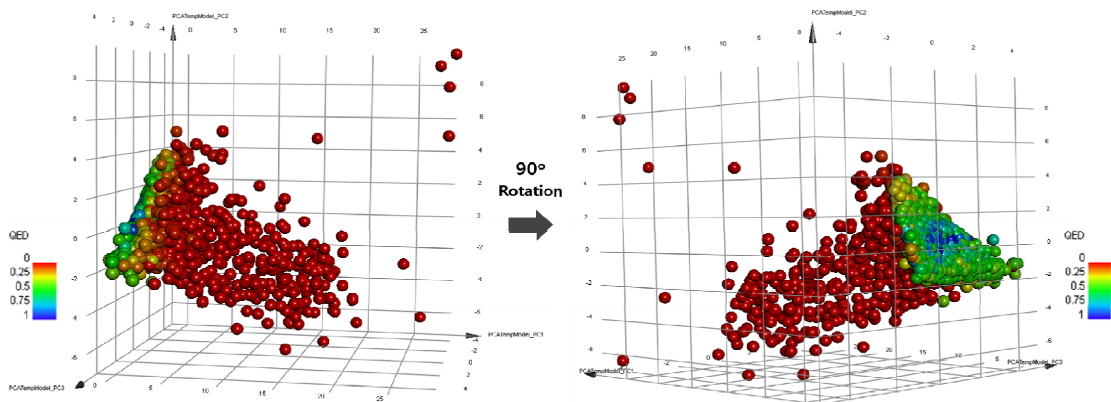


Fig. 7. Ranked ordered weighted QED<sub>w</sub> scores containing the equation coefficients for each principle component on the selected 8 molecular descriptors across 11,144 compounds.

존재한다. 놀랍게도, PC2-PC3 축 평면을 구성하는 데, 주요한 물리화학적 성질은 분자의 극성표면적, 방향족 고리 수, 옥탄올-물 분배계수가 11,141개 화합물의 약물 가능성 QED 점수를 구분하는 변수(feature)로 작용하였다. 이러한 물리화학적 성질이 임상 2상을 통과한 경구용 약물 후보에 대한 생체이용률(bioavailability)의 중요한 예측인자인 것으로 밝혀져 있다[38,39].

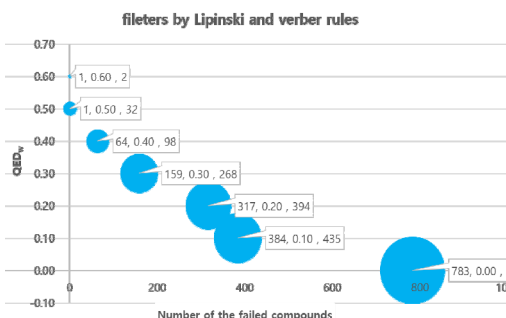


Fig. 8. The ordered QED<sub>w</sub> scores based on the failed compounds by filters of (left) Lipinski and (right) Veber rules.

투과율은 생체이용률의 전제조건이다. 감소된 극성표면적은 친유성(옥탄올-물 분배계수, logP)보다 증가된 투과율과 더 잘 연관되며, 증가된 회전결합수는 투과율을 감소시키는 경향이 있다. 방향족 고리 수 또는 회전결합수로 측정되는 감소된 분자의 유연성과 함께 극성표면적, 수소 결합 수가 분자량에 따라 증가될 수 있다. 이 때문에 분자량 500이하(MW≤500)라는 범위(Table 7)는 결정적인 변수로 작용을 하지 않았다. Table 7에서 보는 바와 같이, 9,433개(85%) 화합물은 Lipinski 기준에 적합하였고, 4 종류 물리화학적 성질의 기준범위에서 1가지 이상 적합하지 않은 1,711개(15%) 화합물이 filtering 되었다.

Table 7. Filter by Lipinski and Veber rules for drug-likeness

Rule	Parameter	Pass	Fail
Lipinski [94]	- Molecular weight no more than 500(MW≤500)	9,433 (84.65%)	1,711 (15.35%)
	- LogP no more than 5(logP≤5)		
	- No more than 5 hydrogen bond donors(HBD≤5)		
	- No more than 10 hydrogen bond acceptors(HBA≤10)		
Veber [95]	- No more than 10 rotatable bonds(RB ≤10)	9,177 (82.35%)	1,967 (17.65%)
	- Polar surface area no more than 140Å <sup>2</sup> (PSA≤140Å <sup>2</sup> )		
	- No more than 12 hydrogen bond donors and acceptors (HBD + HBA ≤12)		

경구용 약물의 생체이용률과 연관된 물리화학적 성질의 범위를 설정하기 위해 2가지 기준(Lipinski and Veber rules)을 적용하였다. 각 기준마다 설정된 물리화학적 성질의 범위로 11,141개 화합물을 filtering 하여 약물유사성이 높은 화합물을 선별하였다.

이때, Lipinski 기준에 따라 분자량과 수소결합능력으로 분자유동성(molecular flexibility)을 고려하되, 분자유동성이 생체이용률에 미치는 영향은 고려되지 않았다. Veber에 의해 제시된 경구용 약물유사성의 평가 기준 [39]은 Lipinski 기준의 친유성(logP)을 제한하는 것[38] 대신에 극성표면적(PSA) 감소를 통한 투과율을 높여 생체이용률을 증가시키는 것에 맞춰져 있다. 동시에 회전결합수를 적용하여 분자의 유동성과 크기를 제한하여 경구용 약물의 흡수와 투과율을 추가적으로 높였다. Veber 기준을 적용하여 11,144개 화합물에 대해 filtering한 결과, 9,177개(82%)만 통과 하였다(Table 7). 이 중에서 Lipinski 기준에 부적합한 화합물들의 약물 가능성 QED<sub>w</sub> 점수가 0.3미만이 86%(1,484개)이었고, Veber기준에 대해 83%(1,634개)를 차지하였다. Lipinski and Veber 기준에 따라 약물유사성이 낮은 화합물들은 QED<sub>w</sub> 점수도 낮았다. 이를 통해 약물가능성 QED<sub>w</sub> 점수를 적용하여 약물유사성 평가에도 활용할 수 있음을 재확인 하였다.

### 3.3 약물유사성 기반 화합물의 다양성 평가

PubChem BioAssays에 등록된 11,144개 화합물을 대상으로 화합물 154번(AL-37350A)과 같은 물리화학적 성질 또는 구조를 가진 화합물을 각각 탐색하였다. 이때 물리화학적 성질은 Lipinski 기준에 포함된 약물 유사성의 파라미터(Table 7)를 적용하여 유사성이 높을수록 상대적 dissimilar 거리가(Eq. (5)~ Eq. (8))가 짧은 화합물을 우선순위로 정렬하였다[33]. 최단거리가 0의 값을 가진 경우에는 기준 화합물(화합물 154번)이며, 0 값에 가까운 값을 가질수록 물리화학적 성질이 유사한 화합물이다. 반면에 dissimilar 거리가 1인 경우에는 가장 다른 성질을 가진 화합물이다. dissimilar distance < 0.1 조건에서 기준 화합물 154번과 유사한 물리화학적 성질을 갖는 19개 화합물을 발굴하였다. 이들 중에서 우선순위 5위 이내에 해당하는 화합물을 선별하였다(Fig. 9). 그 결과, 유사한 화합물들은 2D-구조뿐만 아니라 다른 기능을 하는 화합물이다. 우선순위 1위인 화합물 7023번(OG-L002)은 히스톤 탈메틸화 효소의 억제제로서 허피스 바이러스의 복제와 잠복기에서 재활성화를 차단한다[40]. 이와 다르게 4순위 화합물인 경우에는 도파

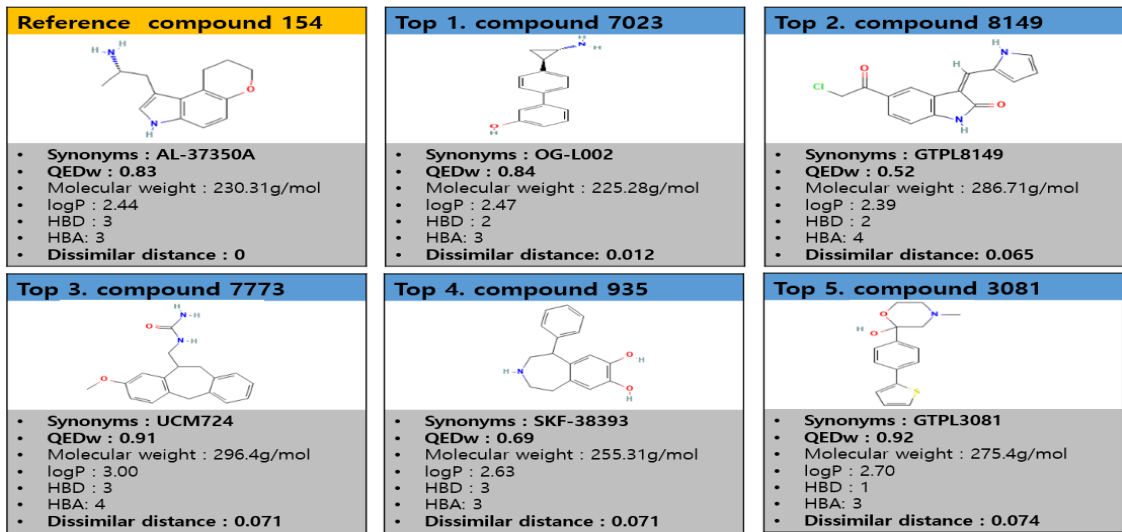


Fig. 9. The high ranked similar molecules on the basis of distance matrices from physicochemical properties for all pairs of 11,144 compounds to the reference compound 154, without any kind of a priori knowledge about the molecules.

인 D1 수용체의 선택적인 작용제(IC<sub>50</sub> =110nM) 이다 [41]. 3순위 화합물(화합물 7773)은 멜라토닌 수용체 리간드로서 길항제 역할을 한다[42]. 2순위와 5순위에 해당하는 GTPL 화합물은 세로토신을 갖는 비가역적 NeK2 인산화효소의 기능 억제제(화합물 8149번)[43]와 간에서 주로 합성하는 콜레스테롤의 전구체(화합물 3081번)[44]이다. 이와 같이, 물리화학적 성질이 유사한 화합물 탐색은 목표단백질(target)을 고려하지 않는다.

만일 동일한 목표 단백질에 대해 작용하는 유효화합물 세트로부터 추출된 물리화학적 성질의 범위로 유사성이 높은 화합물을 검색한다면, 동일한 작용점에 작용할 수 있는 선도물질(hit)의 도출 확률은 더 높아진다. 실제로 이 접근은 공용 화합물 데이터베이스에서 유효화합물과 같은 물리화학적 성질들의 범위에 적합한 화합물을 찾고자 할 때 filtering 기준으로 사용된다. 또한 동일한 목표 단백질에 작용하는 유효화합물 대상으로 약물 유사성의 파라미터를 최적화할 때도 활용한다. 다른 관점에서 화합물의 구조가 유사하면 기능도 유사한지를 확인하기 위해, 기준 화합물 154번과 유사한 구조를 갖는 화합물을 Tanimoto index[33]를 적용하여 탐색하였다. Tanimoto index는 dissimilar distance와 반대로 1의 값이면, 동일한 구조를 가진 화합물(기준 화합물 154번)이며, 0의 값을 가지면, 구조의 유사성을 갖지 않는다. 이는 구조 유사성을 측정하는 Tanimoto index의 개념을 적용하면 similarity = [1/(1+distance)]로 측정되기 때문이다.

기준 화합물 154번에 대해 Tanimoto index가 0.6 이상인 기준에서 유사한 구조를 갖는 110개 화합물을 발굴하였다. 이 중에서 상위 5위에 포함된 화합물들의 구조는 Fig. 10에서 보는 바와 같다. 놀랍게도, 기준 화합물 154번(항녹내장제 유효화합물)과 구조 유사성을 갖는 5위 이내 화합물은 멜라토닌의 기능을 조절하는 역할을 하며, 세로토닌(화합물 145번, 화합물 152번)과 멜라토닌 생합성 과정(화합물 224번) 및 분해과정(화합물 107번)에 포함되었다[45-47]. 구조 유사성 1순위인 화합물 1345번은 멜라토닌 수용체 1A에는 작용제이다[48]. 구조 유사성 화합물을 탐색할 때, 물리화학적 성질이 유사한 화합물을 발굴하는 과정처럼 목표단백질을 고려하지 않는다. 그럼에도 불구하고, 기준 화합물 154번과 물리화학적 성질 유사성을 갖는 화합물 7773번(Fig. 9)과 구조 유사성을 갖는 화합물 1345번(Fig. 10)은 멜라토닌 수용체 1A의 기능을 조절하는 길항제(antagonist)와 작용제(agonist)이다.

이번연구에 적용한 11,144개 화합물을 구조와 물리화학적 성질의 다양성을 고려하여 10개 군집으로 화합물을 분리하였다. 이때 물리화학적 성질의 다양성은 분자간의 성질들의 값에 대한 유클리드 평균 거리(Eq. (6))를 계산하였다. 동시에 구조적 특성을 화합물 내 원자와 작용기를 path-based fingerprint로 고려하여 fingerprint 간의 Tanimoto index로 다양성(diversity)을 계산하였다. 그 결과, 10개 군집으로 분류될 때 11,144개 화합물

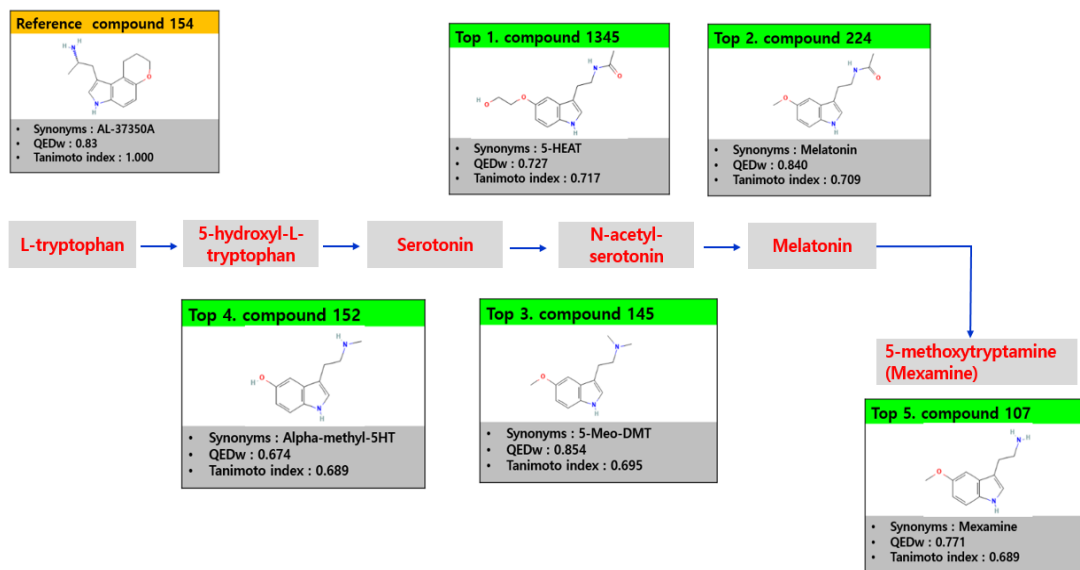


Fig. 10. The structural similarity of compounds within the 5th priority calculated by the Tanimoto index, compared to reference compound 154.

의 물리화학적 성질에 대한 유클리드 평균거리는 1.04 (최대거리 12.97)이고 fingerprint 간의 평균거리는 0.94(최대거리 1.00)의 군집 다양성을 갖는다. 각 군집에 속한 화합물의 수는 Fig. 11에서 보는 바와 같다. 가장 많은 화합물이 속한 1그룹에는 2,488개 화합물들이 포함되어 있지만 8그룹이 제일 적은 수의 188개 화합물이 포함되었다.

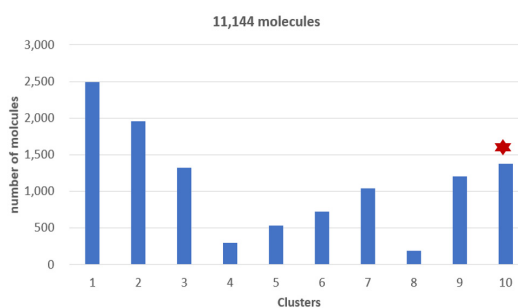


Fig. 11. A set of 11,144 molecules into 10 clusters of molecules based on maximal dissimilarity partitioning.

항녹내장제 유효성분인 화합물 154번은 10그룹에 포함이 되어 있으며, 이 유효성분과 구조유사성을 가진 순위 5이내 화합물 모두가 동일한 10그룹에 존재를 하였다. 그러나 물리화학적 성질이 유사한 화합물인 경우에는 10그룹이 아닌 각각 다른 그룹(2, 5, 9그룹)에 포함이

되었다. 유효화합물 154번이 갖고 있는 공통의 패턴(구조와 물리화학적 성질)을 갖는 화합물 그룹을 찾기 위해 수행을 하였다. 10그룹에는 구조유사성을 갖는 우선순위 5위에 포함된 화합물 이외에도 1,350개 화합물을 추가로 발굴 하였다. 다양성기반 클러스터링은 물리화학적 성질보다는 구조의 패턴에 가중치를 줘서 그룹을 분류하는 경향이 있다. 또한 분자의 크기에 의존하여 dissimilar distance가 계산되기 때문에 기준 화합물 154번보다 분자의 크기가 큰 화합물일수록 다른 그룹에 존재를 하게 된다. 이러한 단점을 극복하기 위해서는 우선순위를 결정할 때 물리화학적 성질이 유사한 화합물들의 subset과 구조 유사성을 갖는 화합물의 subset으로 각각 분리하여야 한다. 다음단계로 동일한 목표단백질에 작용할 수 있는 지를 각각 subset 화합물의 우선순위에 따라 structure-based 분자모델링 과정(docking 결합력, 결합모드 등) 또는 실험적 검증과정을 추진해야 한다.

#### 4. 결론

분자 조합화학과 더불어, 많은 화합물 중에서 연구목적에 따라 어떤 성질을 어떤 기준으로 고려할 것인가에 대한 의사결정에 도움이 될 수 있도록 가상탐색기술도 발전 하였다. 많은 수의 화합물을 무작위로 효능평가 실

힘을 하기 보다는 먼저 가상탐색을 통해 가능성이 높은 hit 화합물만을 선별한 후, 효과 확인 실험을 함으로써 시간과 비용을 줄일 수 있다. 이를 통해 개발된 대표적 약물은 백혈병 치료제 글리벡, 독감 치료제 타미플루, 에이즈 치료제 등이 있다.

이번연구에서 Pubchem BioAssays에 등재된 11,144개 화합물들을 대상으로 ADMET 관련 물리화학적 성질(용해도, 옥탄올-물 분배계수, 극성표면적 등)을 축으로 하여 새로운 좌표 공간을 정의하였다. 이는 약동학적 파라미터(대사안정성, 생체이용률 등)의 특성공간과 일치하지 않는다. 이를 극복하고자 화합물의 흡수, 투과도와 연관된 성질 이외에 분자량, 수소결합 주개·받개 수, 회전결합 수, 방향족 고리수, 그리고 독성유발 부분구조의 수 등과 같은 약물가능성(QED)에 관련된 분자표현자(molecular descriptor)를 추가하였다. 특히, 경구용 약물 유사 화합물의 QED 점수와 관련된 8개 분자표현자를 통해 특성공간이 더 밀도 있게 채워졌다. 이런 정보를 적용하여 항 녹내장제 유효물질인 화합물 154번(AL-37350A)의 물리화학적 성질과 구조패턴 정보를 이용하여 유사성이 있는 화합물들을 발굴하였다. 이때, 화합물 154번이 작용하는 세로토닌 5-HT<sub>2</sub> 수용체를 고려하지 않았다. 놀랍게도, 구조 유사성을 갖는 5순위 이내 화합물들은 세로토닌을 포함한 멜라토닌 생성과 분해 기능을 조절 하는데 관여를 한다. 반면에, 물리화학적 성질유사성을 갖는 5순위 이내 화합물들은 다양한 구조와 기능을 가졌다. 게다가 이들 화합물은 다양성을 기반으로 군집을 분류하였을 때도 화합물154번에 포함된 10그룹에 포함이 되지 않았다(구조유사성이 있는 화합물은 모두 10그룹). 다양성 분석은 두 가지 요인(구조와 특성)을 고려하였음에도 불구하고 구조패턴의 dissimilar distance에 가중치를 더 두는 경향이 있다. 이로 인해 화합물이 154번보다 분자의 크기가 큰 화합물은 dissimilar distance가 더 커졌다. 이런 다양성 분석의 경향성을 고려하여 동일한 목표단백질에 작용하는 유효화합물들의 물리화학적 성질의 범위를 정의하는 데 적용할 수 있다. 또한, 정의된 물리화학적 성질의 profile 좌표공간에 위치하는 다른 구조의 hit를 발굴할 수 있다. 추가적으로 구조-활성간의 관계를 적용하여 후보물질을 최적화 하는 데에도 활용할 수 있을 것이다.

우리는 공용 화합물 DB를 활용하여 다양한 목표 단백질에 각각 기능을 할 수 있는 화합물을 탐색할 때 약물 유사 화합물의 우선순위에 따라 가상탐색 하려고 한다. 또한, 판매되고 있는 약물 또는 기능성 성분에 대한 알려

진 유효성이외에 새로운 효과를 규명하여 이미 안전성이 확보된 제품의 활용 영역을 확대하는 방향으로 연구를 진행할 것이다.

## References

- [1] F. Árpád, "Forty years of combinatorial technology", *Drug Discovery Today*, Vol.27, No.10, pp.103308, October 2022.  
DOI: <https://doi.org/10.1016/j.drudis.2022.06.008>
- [2] Y. Wang, J. Xiao, T.O. Suzek, J. Zhang, J. Wang, S. H. Bryant, "Pubchem: a public information system for analyzing bioactivities of small molecules", *Nucleic Acids Research*, Vol.37(Web Server issue), pp. W623-33, July 2009.  
DOI: <https://doi.org/10.1093/nar/gkp456>
- [3] M. M. Nowotka, A. Galutón, D. Mendez, A. P. Bento, A. Hersey, A. Leach, "Using ChEMBL web services for building applications and data processing workflows relevant to drug discovery", *Expert opinion on drug discovery*, Vol.12, No.8, pp.757-767, August 2017.
- [4] J. Masciocchi, G. Frau, M. Fanton, M. Sturless, M. Floris, L. Pireddu, P. Palla, F. Cedratì, P. Rodriguez-Tome, S. Moro, "MMsINC: a large-scale chemoinformatics database", *Nucleic Acids Research*, Vol.37(Database issue), pp. D284-90, January 2009.  
DOI: <https://doi.org/10.1093/nar/gkn727>
- [5] C. Hansch, "The physicochemical approach to drug design and discovery(QSAR)", *Drug Development Research*, Vol.1, No.4, pp.267-309, 1981.
- [6] J. L. Miler, "Chemistry nobel honors computer simulation of biomolecules", *Physics Today*, Vol.66, No. 12, pp.13 2013.  
DOI: <https://doi.org/10.1063/PT.3.2197>
- [7] T. Nakamura, S. Sakaue, K. Fujii, Y. Harabuchi, S. Maeda, S. Iwata, "Selection molecules with diverse structures and properties by maximizing submolecular functions of descriptors learned with graph neural networks", *Scientific Reports*, Vol. 12, No. 1, pp.1124, January 2022.  
DOI: <https://doi.org/10.1038/s41598-022-04967-9>
- [8] P. Carracedo-Reboredo, J. Liñares-Blanco, N. Rodríguez-Fernández, F. Cedrón, F. J. Novoa, A. Carballedo, V. Maojo, A. Pazos, C. Fernandez-Lozano, "A review on machine learning approaches and trends in drug discovery", *Computational and Structural Biotechnology*, Vol.19, pp.4538-4558, August 2021.  
DOI: <https://doi.org/10.1016/j.csbi.2021.08.011>
- [9] H. Yang, L. Sun, Z. Wang, W. Li, G. Liu, Y. Tang, "ADMETopt: A web server for ADMET optimization in drug design via scaffold hopping", *Journal of Chemical Information and Modeling*, Vol. 58, No.10, pp.2051-2056, October 2018.



- DOI: <https://doi.org/10.1021/acs.icim.8b00532>
- [10] K. Szilágyi, B. Flachner, I. Hajdú, M. Szaszko, K. Dobi, Z. Lórinz, S. Cseh, G. Dormán, "Rapid identification of potential drug candidates from multi-million compound's repositories. Combination of 2D similarity search with 3D ligand/structure based methods and in vitro screening", *Molecules*, Vol.26, No.18, pp.5593, September 2021.  
DOI: <https://doi.org/10.3390/molecules26185593>
- [11] D. Rogers, M. Hahn, "Extended-connectivity fingerprints", *Journal of chemical information and modeling*, Vol.50, No.5, pp.742-754, May 2010.  
DOI: <https://doi.org/10.1021/ci100050t>
- [12] J. Wang, "Comprehensive assessment of ADMET Risks in Drug discovery", *Current Pharmaceutical Design*, Vol.15, No.19, pp.2195-2219, 2009.  
DOI: <https://doi.org/10.2174/138161209788682514>
- [13] H. van de Waterbeemd, E. Gifford, "ADMET in silico modelling: towards prediction paradise?", *Nature reviews Drug Discovery*, Vol.2, No.3, pp.192-204, March 2003.  
DOI: <https://doi.org/10.1038/nrd1032>
- [14] A. B. Yongye, J. L. Medina-Franco, "Systematic characterization of structure-activity relationships and ADMET compliance: a case study", *Drug Discovery Today*, Vol.18, No.15/16, pp.732-739, August 2013.  
DOI: <https://doi.org/10.1016/j.drudis.2013.04.002>
- [15] R. Fraczekiewicz, M. Lobell, A. H. Göller, U. Krenz, R. Schoenneis, R. D. Clark, A. Hillisch, "Best of both worlds: combining pharma data and state of the art modeling technology to improve in silico pKa prediction", *Journal of Chemical Information and Modeling*, Vol.55, No.2, pp.389-397, February 2015.  
DOI: <https://doi.org/10.1021/ci500585w>
- [16] L. Huang, X. Wu, X. Fu, H. Wang, B. Tang, Y. Xiao, C. Zhou, Z. Zhao, Y. Wan, H. Chen, Z. Tang, H. Yao, Z. shan, "Ligand based 3D-QSAR model, pharmacophore, molecular docking and ADME to identify potential fibroblast growth factor receptor 1 inhibitors", *Journal of Biomolecular Structure & Dynamics*, Vol.40, No.16, pp.7584-7597, October 2022.  
DOI: <https://doi.org/10.1080/07391102.2021.1899049>
- [17] C. A. S. Bergström, "In silico predictions of drug solubility and permeability: two rate-limiting barriers to oral drug absorption", *Basic & Clinical Pharmacology and Toxicology*, Vol.96, No.3, pp.156-161, March 2005.  
DOI: <https://doi.org/10.1111/j.1742-7843.2005.pto960303.x>
- [18] W. J. Egan, G. Lauri, "Prediction of intestinal permeability", *Advanced Drug Delivery Review*, Vol.54, No.3, pp.273-289, March 2002.  
DOI: [https://doi.org/10.1016/s0169-409x\(02\)00004-2](https://doi.org/10.1016/s0169-409x(02)00004-2)
- [19] A. Cheng, S. L. Dixon, "In silico models for the prediction of dose-dependent human hepatotoxicity", *Journal of Computer-Aided Molecular Design*, Vol.17, No.12, pp.813-823, December 2003.  
DOI: <https://doi.org/10.1023/b:jcam.0000021834.50768.c6>
- [20] A. K. M. Azad, M. Dinarvand, A. Nematollahi, J. Swift, L. Lutze-Mann, F. Vafaee, "A comprehensive integrated drug similarity resource for in silico drug repositioning and beyond", *Briefings in Bioinformatics*, Vol.22, No.3, pp.bb126, May 2021.  
DOI: <https://doi.org/10.1093/bib/bbaa126>
- [21] E. E. Bolton, K. Sunghwan, S. H. Bryant, "PubChem3D: conformer generation", *Journal of Cheminformatics*, Vol.3, No.1, pp.4, January 2011.  
DOI: <https://doi.org/10.1186/1758-2946-3-4>
- [22] R. L. M. Robinson, D. Geatches, C. Morris, R. Mackenzie, A. G. P. Maloney, K. J. Roberts, A. Moldovan, E. Chow, K. Pencheva, D. R. M. Vatvani, "Evaluation of force field calculations of lattice energies on a large public dataset, assessment of pharmaceutical relevance, and comparison to density functional theory", *Journal of chemical information and modeling*, Vol.59, pp.4778-4792, November 2019.  
DOI: <https://doi.org/10.1021/acs.icim.9b00601>
- [23] A. de la Nuez, R. Rodríguez, "Current methodology for the assessment of ADME-Tox properties on drug candidate molecules", *Biotechnología Aplicada*, Vol.25, No.2, pp.97-110, April 2008.
- [24] P. Ertl, B. Rohde, P. Selzer, "Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to prediction of drug transport properties", *Journal of medicinal chemistry*, Vol. 43, No.21, pp. 3714-3717, October 2000.  
DOI: <https://doi.org/10.1021/jm000942e>
- [25] X. Xia, E. G. Maliski, P. Galluant, D. Rogers, "Classification of kinase inhibitors using a bayesian model", *Journal of medicinal chemistry*, Vol. 47, No.18, pp. 4463-4470, August 2004.  
DOI: <https://doi.org/10.1021/jm0303195>
- [26] A. Cheng, K. M. Merz Jr, "Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships", *Journal of medicinal chemistry*, Vol. 46, No.17, pp.3572-3580, August 2003.
- [27] A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, "Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of AlogP and ClogP methods", *Journal of Physical Chemistry A*, Vol.102, No.21, pp.3762-3772, May, 1998.  
DOI: <https://doi.org/10.1021/jp980230o>
- [28] A. Cheng, D. J. Diller, S. L. Dixon, W. J. Egan, G. Lauri, K. M. Merz jr, "Computation of the physico-chemical properties and data mining of large molecular collections", *Journal of Computational Chemistry*, Vol.23, No.1, pp.172-183, January 2002.  
DOI: <https://doi.org/10.1002/jcc.1164>
- [29] J. Kelder, P. D. Grootenhuys, D. M. Bayada, L. P. Delbressine, J. P. Pleomen, "Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs", *Pharmaceutical Research*,

- Vol.16, No.10, pp.1514-1519, October 1999.  
DOI: <https://doi.org/10.1023/a:1015040217741>
- [30] A. Cheng, S. L. Dixon, "In silico models for the prediction of dose-dependent human hepatotoxicity", *Journal of Computer-aided molecular design*, Vol.17, No.12, pp.811-823, December 2003.  
DOI: <https://doi.org/10.1023/b:icam.0000021834.50768.c6>
- [31] J. R. Votano, M. Parham, L. M. Hall, L. H. Hall, L. B. Kier, S. Oloff, A. Tropsha, "QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation", *Journal of Medicinal Chemistry*, Vol.39, No. 24, pp.7169-7181, November 2006.  
DOI: <https://doi.org/10.1021/jm051245v>
- [32] G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, "Quantifying the chemical beauty of drugs", *Nature Chemistry*, Vol.4, No.2, pp.90-98, January 2012.  
DOI: <https://doi.org/10.1038/nchem.1243>
- [33] D. Bajusz, A. Rácz, K. Héberger, "Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations?", *Journal of Cheminformatics*, Vol.7, Article number 20, May 2015.  
DOI: <https://doi.org/10.1186/s13321-015-0069-3>
- [34] Dassault Systéms BIOVIA Discovery Studio modeling environment release R2 dassault system, San Diago 2017.  
<https://www.3ds.com/products-services/biovia/products>
- [35] J. A. May, H. H. Chen, A. Rusinko, V. M. Lynch, N. A. Sharif, M. A. McLaughlin, "A novel and selective 5-H2 receptor agonist with ocular hypotensive active: (S)-(+)-1-(2-aminopropyl)-8,9-dihydropyranol[3,2-e]indole", *Journal of Medicinal Chemistry*, Vol.46, No.19, pp.4188-4195, August 2003.  
DOI: <https://doi.org/10.1021/jm030205t>
- [36] C. M. Wenlock, R. P. Austin, P. Barton, A. W. Davis, P. D. Leeson, "A comparative of physicochemical property profiles of development and marketed oral drugs", *Journal of Medicinal Chemistry*, Vol.46, No.7, pp.1250-1256, March 2003.  
DOI: <https://doi.org/10.1021/jm021053p>
- [37] L. Gaohua, X. Miao, L. Dou, "Crosstalk of physiological pH and chemical pKa under the umbrella of physiologically based pharmacokinetic modeling of drug absorption, distribution, metabolism, excretion, and toxicity", *Expert opinion on drug metabolism & toxicology*, vol.17, No. 9, pp.1103-1124, September 2021.  
DOI: <https://doi.org/10.1080/17425255.2021.1951223>
- [38] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings", *Advanced Drug Delivery Reviews*, Vol.23, No.1-3, pp.3-25, January 1997.  
DOI: [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
- [39] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, "Molecular properties that influence the oral bioavailability of drug candidates", *Journal of Medicinal Chemistry*, Vol.45, No.12, pp.2615-2623, June 2002.  
DOI: <https://doi.org/10.1021/jm020017n>
- [40] Y. Liang, D. Quenelle, J. L. Vongel, C. Mascaro, A. Ortega, T. M. Kristie, "A novel selective LSD1/KDM1A inhibitor epigenetically blocks hepes simplex virus lytic replication and reactivation from latency", *mBio*, Vol.4, No.1, pp.e00558, February 2013.  
DOI: <https://doi.org/10.1128/mbio.00558-12>
- [41] N. Reichenbach, U. Hermann, T. Kähne, H. Schicknick, R. Pielot, M. Naumann, D.C. Dieterich, E. D. Gundelfinger, K. H. Smalla, W. Tischmeyer, "Differential effects of dopamine signalling on long-term memory formation and consolidation in rodent brain", *Proteome Science*, Vol.13, pp.13, March 2015.  
DOI: <https://doi.org/10.1186/s12953-015-0069-2>
- [42] V. Lucini, M. Pannacci, F. Scaglione, F. Franchini, S. Rivara, M. Mor, F. Bordi, P. Vincenzo, P. Gilberto, S. A. Bedini, G. Piersanti, G. Diamantini, G. Tarzia, "Tricyclic alkylamides as melatonin receptor ligands with antagonist or inverse agonist activity", *Journal of Medicinal Chemistry*, Vol.47, No. 17, pp.4202-4212, August 2004.  
DOI: <https://doi.org/10.1021/jm040768k>
- [43] J. C. Henise, J. Taunton, "Irreversible Nek2 kinase inhibitors with cellular activity", *Journal of Medicinal Chemistry*, Vol.54, No. 12, pp.4133-4166, June 2011.  
DOI: <https://doi.org/10.1021/jm200222m>
- [44] S. P. Alexander, D. Fabbro, E. Kelly, A. Mathine, J. A. Peters, E. L. Veale, "Lanosterol biosynthesis pathway (version 2019.4) in the IUPHAR/BPS guide to pharmacology database", September 2019.  
DOI: <https://doi.org/10.1093/nar/gkab1010>
- [45] A. M. Arújo, F. Carvalho, M. D. L. Bastos, P. G. D. Pinho, M. Carvalho, "The hallucinogenic world of tryptamines: an updated review", *Archive of Toxicology*, Vol.89, No.8, pp.1151-1173, August 2015.  
DOI: <https://doi.org/10.1007/s00204-015-1513-x>
- [46] J. M. Corkery, E. Durkin, S. Elliott, F. Schifano, A. H. Ghodse, "The recreational tryptamine 5-meo-dalt (n,n-diallyl-5-methoxytryptamine): a brief review", *Progress in neuro-psychopharmacology & biological psychiatry*, Vol.39, No.2, pp.259-262, December 2012.  
DOI: <https://doi.org/10.1016/j.pnpbp.2012.05.022>
- [47] S. Pal, C. Haldar, R. Verma, "Melatonin attenuates LPS-induced ovarian toxicity via modulation of SIRT-1, PI3K/pAkt, pErk1/2 and NFkB/CoX-2 expressions", *Toxicology and applied pharmacology*, Vol. 451, pp.15, September 2022.  
DOI: <https://doi.org/10.1016/j.taap.2022.116173>
- [48] R. Nonno, V. Lucin, G. Spadoni, M. Pannacci, A. Croce, D. Esposti, C. Balsamini, G. Tarzia, F. Fraschini, B. M. Stankov, "A new melatonin receptor ligand with mt1-agonist and mt2-antagonist properties", *Journal of Pineal Research*, Vol.29, No.4, pp.234-240, November 2000.  
DOI: <https://doi.org/10.1034/j.1600-0633.2002.290406.x>

채 한 화(Han-Ha Chai)

[정회원]



• 2010년 11월 ~ 현재 : 농촌진흥청  
국립축산과학원 농업연구사

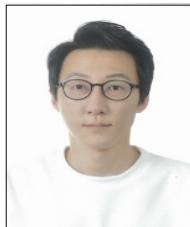
<관심분야>

동물 기능유전체, 화학·생명정보 응용연구(분자모델링)

---

박 원 철(Woncheoul Park)

[정회원]



• 2019년 2월 ~ 현재 : 농촌진흥청  
국립축산과학원 농업연구사

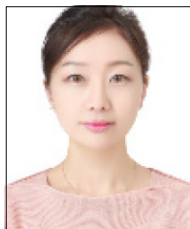
<관심분야>

집단 유전체, 전사체, 다중-오믹스

---

임 다 정(Dajeong Lim)

[정회원]



• 2007년 11월 ~ 현재 : 농촌진흥청  
국립축산과학원 농업연구관

<관심분야>

가축 집단 유전체, 생명정보 적용 유전자 선발