

베이지안 모델 기반 약물유사 화합물의 간 독성 예측

채한화¹, 박원철^{1*}, 진영배², 임다정^{1,3}

¹국립축산과학원 동물유전체과, ²경상국립대학교 수의과대학, ³충남대학교 동물자원과학부 동물바이오시스템학과

Predicting Hepatotoxicity of Drug-like compounds based on Bayesian model

Han-Ha Chai¹, Wonchoul Park^{1*}, Yeoung Bae Jin², Dajeong Lim^{1,3}

¹Division of Animal Genomics and Bioinformatics, National Institute of Animal Science

²Department of Laboratory Animal Medicine, College of Veterinary Medicine, Gyeongsang National University

³Department of Animal Resources Science, College of Agriculture and Life Science Chungnam National University

요약 간 독성을 예측하는 것은 약물유사-화합물의 안전성 평가에 중요하다. 간 독성은 약물유사-화합물의 물리화학적 성질, 특히 구조적 특성과 관련이 있다. 이번 연구에서 정량적 구조-독성 상관관계(QSAR)로 498개 약물유사-화합물의 간 독성을 예측하기 위하여 베이지안 모델을 개발했다. 베이지안 모델은 25개 구조 표현자(예. ECFP6 화학결합 지문)를 사용하여 498개 약물유사-화합물의 간 독성을 예측했으며, 그 결과 민감도, 특이도 및 정확도가 각각 97.2%, 86.9%, 90.6%였다. 이 베이지안 모델은 498개 약물유사-화합물의 구조로부터 간 독성/무독성 화합물 분류를 결정하는 데 유용하였다. 베이지안 모델 기반으로 간 독성 예측 후, TOPKAT[®] 독성모델을 사용하여 독성학적 종말점(예. LD₅₀, LOAEL) 및 Ames 변이원성, 수컷과 암컷 마우스 대상 발암성, 발생 독성 가능성과 연관된 간 독성 영향을 예측했다. 약물 관련 부작용을 최소화하기 위해서, 베이지안 모델 기반 간 독성 예측과 TOPKAT[®] 독성모델로부터 간 독성의 영향을 예측하는 것은 약물유사-화합물의 안전성 평가를 수행하는 데에 실험적 접근법을 대체할 수 있다. 결과적으로, 우리의 독성모델은 동물실험모델 기반의 독성 연구에서 소요 되는 시간과 비용을 줄이고 안전성과 유효성을 고려하여 선도화합물을 최적화하는 데 활용될 것이다.

Abstract Predicting hepatotoxicity is an important component of safety-related evaluations of drug-like compounds. Hepatotoxicity is related to the physicochemical properties of drug-like compounds, especially their structural alerts. In this study, we developed a Bayesian model to predict the hepatotoxicities of 498 drug-like compounds based on their quantitative structure-toxicity relationships (QSAR). The devised model predicted the hepatotoxicity of these compounds using 25 structural descriptors (such as the ECFP6 fingerprint) and provided a sensitivity, specificity, and concordance of 97.2%, 86.9%, and 90.6%, respectively. The model also successfully classified the 498 drug-like compounds by hepatotoxicity. In addition, TOPKAT[®] toxicity models were used to predict hepatotoxic effects related to toxicological endpoints (i.e., LD₅₀, LOAEL) and liver injury-related potentials, such as Ames mutagenicity, carcinogenicity in male and female mice, and developmental toxicity potentials. Our results indicate that combined use of the devised Bayesian hepatotoxicity prediction model and the TOPKAT[®] hepatotoxicity model could replace experimental safety assessments of drug-like compounds. Accordingly, adopting the devised Bayesian toxicity model would reduce the time and cost of animal-based toxicity research by considering the safety and effectiveness of candidate compounds.

Keywords : Hepatotoxicity, Drug-like Compounds, Bayesian Model, Quantitative Structure-toxicity Relationship, TOPKAT[®]

본 논문은 농촌진흥청 연구사업(PJ01720401, 화합물 소재의 반추가속 장내발효 메탄저감 효능 및 안전성 평가) 연구과제로 수행되었음.

*Corresponding Author : Wonchoul Park(National Institute of Animal Science)

email: wcpark1982@korea.kr

Received November 15, 2023

Accepted January 5, 2024

Revised December 20, 2023

Published January 31, 2024

1. 서론

의약품, 합성 화합물 소재(10^{30} 이상 신규화합물 합성 가능)를 개발하는 과정에서 실험동물을 적용한 유효성·안전성 평가가 많이 수행되었다[1]. 하지만, 실험동물과 사람간의 생물학적 차이로 인해 동물실험의 연구 결과를 인체에서 직접 재현하는 데 한계가 있다[2]. 이로 인해 동물실험 결과를 정확하게 해석하고 인체에 적용하는 데에 제한성을 갖는다(동물실험의 정확도가 43.5%~66.7%로 제시된 바 있음[3]). 그 결과, 임상시험 단계 또는 의약품 시장출시 후, 부작용으로 중단되기도 한다. 예를 들어, 독일 제약회사 그뤼넨탈에서 개발한 탈리도마이드(임산부 입덧 방지용 약)는 동물실험에서 독성을 보이지 않았지만, 실제 복용한 임산부들은 기형아와 사산아를 출산하는 등 심각한 부작용이 있었다.

동물실험에 대한 윤리 의식 확산 및 종(species)간의 차이로 유래된 한계효용으로 인해, 유럽은 2007년 'REACH'/3Rs[4] (Registration Evaluation and Authorization of Chemical) 신화학물질관리정책을 시행하였다. 동물 대체 시험법을 화학물질 평가에 활용할 뿐만 아니라, 대형 연구프로젝트 'Horizon 2020 (2014-2020)'을 통해 장기 칩 개발[5]과 2016년부터 진행된 'EU-Tox Risk project'[6] 통해 동물 대체 독성시험법과 함께 화합물의 흡수·분포·대사·배설(ADME) 데이터가 포함된 생리학 기반 동역학 모델링 적용까지 확대하였다. 특히, 'ONTOX'[7] 프로젝트는 위험성 평가를 위한 화합물의 온톨로지 및 인공지능 기반 반복투여 독성효과를 예측하는 시스템에서 작용하는 반응경로, 온톨로지 프레임워크(framework) 등이 통합되어 있다. 이외에도 화합물의 대사 능력과 독성물질에 대한 민감도를 세포와 조직(폐, 간, 신장 등) 절편(slices)을 적용함으로써, 실험동물 대체법으로 제안하는 'PrecisionTOX'[8], 'RISK-HUNT3R'[9]등을 통해 화학물질 유해성 평가·예측 기술 개발을 지원하고 있다.

또 다른 주요국인 미국도 2009년부터 'Tox21'[10] 프로젝트를 통해 1억 개가 넘는 데이터 포인트(Tox21 10K 라이브러리)를 생성하였다. 다양한 화학구조, 용도 범주 및 특성을 포괄할 뿐만 아니라, 재현성 평가를 위해 구현된 데이터흐름이 잘 설명되었다. 화학정보 프로파일링은 규제 목록, 예상 독성 종말점 및 물리화학적 특성의 적용 범위를 보완하여 라이브러리 범위를 확장하고 있다. 모든 데이터는 미국 환경보호청(Environmental Protection Agency, EPA), 국립 환경 보호센터의 웹사

이트를 통해 공개적으로 제공한다. 게다가, ToxCast/Tox21 데이터베이스[11]는 독성 메커니즘 식별, 화합물 우선순위 결정, 환경 모니터링에 대한 분석기술 개발 및 이를 이용하여 독성을 예측한다. 이를 통해 실험실 동물을 사용하지 않고도 화합물 우선순위 지정 및 규제 적용이 가능하다. 더 나아가서, 미국식품의약국 FDA는 동물 대체시험법 관련 전략보고서[12] 발간과 더불어 개발과 활용 촉진을 위해 로드맵을 수립하여 추진하고 있다. 더 나아가서, 2022년 12월 「식품·의약품·화장품법」 개정을 통해 신약개발 단계에서 동물실험 의무사항을 삭제하고, 비임상 시험의 예로 동물대체시험법 등을 제시하였다.

또한, 국내도 「화학물질 등록 및 평가 등에 관한 법률」 제14조 및 시행령 13조제3호에 의거, 연간 10톤 미만의 화학물질 등록 건에 대해 "산업체를 위한 QSAR 자료 제출안내서(2021. 국립환경과학원)"에 따라 컴퓨터를 활용한 구조-독성간의 정량적 관계(Quantitative Structure-Activity Relationship, QSAR)자료로 동물시험을 대체하고 있다. 게다가, '동물대체시험법 활성화 법률안'이 발의(2022.12.23.)되어 관련 연구 활성화와 지원을 도모하고 있다.

동물대체시험법은 실험동물을 사용하지 않고 원료물질의 유효성과 안전성을 평가하는 시험법으로 ①세포 기반 시험법과 ②비세포 기반 시험법으로 분류할 수 있다. ①세포기반 시험법에는 시험관 시험법(In vitro assay)[13], 장기 칩(Organ-on-a-chip)[14], 오가노이드(Organoid) 장기 유사체[15] 등이 포함되며, ②비세포 기반 시험법에는 컴퓨터 모델링(QSAR 모델, ADME/tox 예측 모델 등), *in silico* 플랫폼 기반 장기모델(Physiome)을 이용한 시뮬레이션[16,17] 등을 통해서 원료물질의 위해성을 예측하는 기술이 포함된다. 동물대체시험법으로 생성된 데이터까지 포함하여 구축된 대표적 데이터베이스는 화학물질의 독성효과 등록(Registry of Toxicity Effects of Chemical Substrates, RTECS) 데이터베이스, FDA의 간에서 부작용 효과에 대한 데이터베이스(Human Liver Adverse Effects Database, HLAED)이다[18,19]. 화학물질의 구조를 기반으로 위해성을 예측하는 기술로 OECD QSAR 프로젝트를 통해 개발된 QSAR toolbox[20-22]를 이용하여 화학물질의 성질 및 독성에 대한 종말점(예, LD₅₀과 같은 반수치사용량 등)값의 계산이 가능하다. 화학물질의 독성을 decision tree 형식 의사결정법으로 평가하여 화학물질을 Class 1, 2, 3 독성등급에 해당하는 화학구조를 분류하는 Toxtree[23], 미국 환경 보호청(EPA)에서 제공하는 QSAR 관련 T.E.S.T.(Toxicity

Estimation Software Tool)[24] 및 PreADME로 약물의 흡수·분포·대사·배설, 독성(발암성, 유전독성 등)[25]을 화학구조로부터 예측할 수 있다.

이번 연구에서는 직접적인 동물실험 없이 FDA에서 공개한 간 독성이 결정된 2,889개 약물유사-화합물 데이터를 활용하였다. 이 중에서 염(salt)과 무기물 형태의 화합물을 제외한 2,780개 2D-화학구조를 생성하였다. 화학구조를 생성한 2,780개에는 1,429개 간 독성이 있는 화합물과 1,351개 무독성 화합물로 분류되었다. 2,780개 화합물에서 베이지안 모델의 498개 원료물질로 구성된 training set 선별은 화합물 2D-구조와 물리화학적 특성의 최대 비유사성(Maximum Dissimilarity) 거리[26]를 기반으로 다양성을 갖도록 선택했다. 즉, 단순 무작위 표본에 속한 화합물은 다양성을 갖는 간 독성/무독성 화합물을 각각 우선순위 정렬 후, 상위 18%에 해당하는 화합물 그룹이다. 이를 이용하여 498개 약물유사-화합물의 2D-구조적 성질을 기반으로 컴퓨터 모델링 기술을 적용하여 간 독성/무독성을 예측하였다.

2. 실험재료 및 연구방법

2.1 약물의 화학구조로부터 물리화학적 특성 계산

FDA의 간 독성 지식 베이스(LTKB, Liver Toxicity Knowledge Base)[27,28]를 통해 개별 약물과 관련된 다양한 정보(예. 약인성간손상 DILI 메커니즘, 약물대사, 조직병리학적, 치료용도, 표적, 부작용 등)를 수집하였다. 특히, LTKB 간 독성 지식 베이스에는 미국 국립보건원(National Institute of Health, NIH) LiverTox 데이터베이스[29]의 약인성간독성 DILI의 간 손상(표현형)의 주요임상(간 손상 빈도, 주요 원인, 역학, 진단 및 관리 등) 및 조직학적 패턴에 대한 설명과 함께 분류체계와 심각도 점수 등이 포함되어 있다. 또한, LTKB 간 독성 지식 베이스에서 시스템 생물학 분석을 적용하여 공개하는 약인성간손상(DILI) 평가 및 예측 데이터[30]를 수집한 자료에 통합하였고, Pubchem ID, InChI 정보를 이용하여 겹치는 화합물을 제외하였다. 최종적으로, LTKB 프로젝트를 통해 축적된 데이터 중에서 약물유사-화합물의 간 독성 유·무(독성 1, 무독성 0)가 결정된 2,889개 데이터 목록이 수집되었다. 2,889개 데이터 목록에서 화합물의 2D-분자구조로부터 물리화학적 성질을 계산할 수 없는 염과 무기질 형태의 화합물 목록을 제외한 2,780개 화합물의 2D-화학구조를 생성하였다. 생성

된 2,780개 화합물의 2D-화학구조를 4개 구조 표현자(ECFP4, ECFP6, FCFP4, FCFP6 fingerprint)의 타니모토 계수(Tanimoto coefficient)[31] 기반 최대 비유사성 거리를 기반으로 다양성[26]을 갖는 6개 세트(1개 training set과 5개 test sets)로 분류하였다. 이 중에서 가장 다양성이 높은 그룹을 training set으로 하여 단순 무작위표본 498개 약물유사-화합물을 대상으로 베이지안 QSAR모델 기반 간 독성/무독성을 예측하였다. 이는 2,780개 약물유사-화합물의 구조 다양성의 범위를 포함하게 해서 베이지안 모델의 정확도와 검증 세트(5개 test sets)에서 예측 정확도를 유지하기 위함이다. 그래서, 가장 다양성을 갖는 약물유사-화합물의 그룹(최대 비유사성 거리의 순으로 상위 18%에 해당하는 training set)을 가지고 간 독성 예측 모델을 개발하였다.

498개 약물유사-화합물의 식별자로 화학명(IUPAC name), SMILES 형식의 화학구조와 분자정보를 인코딩하는 표준적인 방법 InChI를 이용하여 만들어진 문자열 InChI Key, 공용 화합물 데이터베이스 Pubchem ID, 간 독성 유·무의 바이너리 값(독성 1, 무독성 0) 정보로 구성하였다. 이로부터 각 약물의 2D, 3D-화학구조를 생성하여 SDF(Structure Data File) 파일형태로 총 원자수, 총 화학결합수, 원자의 x, y, z 공간적 좌표, 화학결합 정보, 화합물의 성질에 대한 정보를 저장하였다.

약물이 작용하는 신체와 같은 생체환경을 조성하기 위해, pH=7.0 조건, CVFF[32] 원자 및 분자력 장(molecular force field)에서 498개 약물유사-화합물의 화학구조로부터 물리화학적 성질을 분자매개변수(분자표현자, molecular descriptors)로 적용하여 수치화하였다.

2.2 약물구조로부터 물리화학적 성질 공간분포

상용화 프로그램인 discovery studio 분자모델링 프로그램[33]의 ADME/Tox 프로그램 모듈에서, 각 약물유사-화합물마다 2D-화학구조로부터 흡수·분포·대사·배설·독성 관련 40개 물리화학적 성질(예. 소수성 AlogP, 극성표면적 PSA 등)[34,35]과 25개 간 독성이 존재하는 것으로 알려진 fragment가 반영된 구조 표현자[36]등과 같은 데이터를 생성하였다. 이로부터 498개 약물유사-화합물의 2D-화학구조 연관성이 있는 고차원 데이터(498 molecules x 65 molecular descriptor)들을 주성분 분석(Principal Component Analysis, PCA)하였다. 주성분 분석을 통해 고차원의 데이터를 선형 연관성이 없는 저차원 공간(주성분, 교차 공분산을 가장 잘 설명하는 좌표시스템) 3개축(PC1, PC2, PC3)로 직교변

환하여 차원을 환원시켰다. 이때 3개축은 8개 분자매개 변수와 상수로 구성된 단위벡터로, PC1축은 41.3%, PC2축은 24.6%, PC3축은 14.4%로 498개 약물유사-화합물의 화학적 공간분포를 설명하였다(Fig. 1).

2.3 약물의 화학구조로부터 간 독성 유·무 예측

498개 약물유사-화합물의 2D-화학구조로부터 계산된 구조·물리화학적 성질과의 간 독성간의 통계적 상관관계(QSAR)를 베이지안 모델[37,38]을 통하여 분석하였다. 베이지안 QSAR 모델은 주어진 498개 약물유사-화합물의 구조·물리화학적 성질 데이터를 기반으로 간 독성 유·무인 모수를 확률변수로 해서 조건부 확률분포(확률밀도함수)를 계산한다.

$$p(h|d) = \frac{P(d|h)P(h)}{P(d)} \quad (1)$$

여기서 추론대상(모수) h에 대하여 h에 대한 사전확률 p(h)가 주어진다. 추론대상 h와 관계된 d의 확률분포 p(d|h)와 같이 주어진다고 할 때, 베이지안 모델은 d가 추가적으로 주어졌을 때, h의 확률분포 p(h|d)를 계산한다.

베이지안 QSAR 모델의 신뢰도를 검증하기 위해 교차 검증(leave-one-out cross validation)을 통해 데이터 추출하는 동안, 베이지안 QSAR 모델의 견고성을 평가하였다. 실제적으로 498개 약물유사-화합물의 간 독성 데이터를 5개 검증세트(5-fold cross validation)로 나눠 최소 한번씩은 test 세트로 쓰이도록 하였다. 총 5개 검증세트로부터 각 ROC Curve(Receiver Operating Characteristic Curve)로 y축은 민감도(Sensitivity)와 x축은 1-특이도(Specificity)로 그려진 곡선을 통하여 베이지안 QSAR모델의 유용성과 정확도를 평가하였다. 또한, 간 독성 유·무를 결정하는 기준값을 설정하는 데에 적용하였다. Table 2에서 보는 바와 같이, 베이지안 모델의 training set의 민감도는 실제 179개 간 독성이 있는 약물을 간 독성이 있다고 예측한 비율(%)을 의미하는데, 98.3%였다. 특이도는 319개 간 독성이 없는 약물을 간 독성이 없다고 예측한 것의 비율(%)이며, 85.3%였다. 따라서, 1-특이도는 간 독성이 있는 약물과 간 독성이 없는 약물을 잘못 예측한 것의 비율(%)로 위양성율(False Positive Rate, FPR)을 의미하는 데, 10.0%였다. 따라서 민감도와 특이도가 클수록 베이지안 QSAR 모델의 예측 정확도는 높아져 일치도(Concordance)는 90.0%였다.

결과적으로, 베이지안 QSAR 모델에서 간 독성 유·무를 결정하는 기준값(Bayesian score)>-1.698, 간 독성

에 따라 분류된 약물의 2개 그룹(간 독성/무독성)으로 간 독성 유·무를 결정하였다. 베이지안 QSAR 모델을 개발하는데 적용한 training set과 검증단계의 5-fold cross validation set에서 민감도, 특이도, 일치율(concordance rate), 위양성율 비교하였다(Table 2). 검증세트(5-fold cross validation set)에서 위양성율이 9.44%로 training set 10.0% 보다 더 적었다. 반면에 특이도는 검증세트 86.9%로 training set 85.3%보다 더 높아서, 90.6% 일치율도 검증세트에서 약간 높아졌다. 최종적으로 ROC score로 베이지안 QSAR 모델의 정확도와 예측 신뢰도를 평가하였다(Fig. 3). 게다가, 약물유사-화합물의 간 독성 존재 여부에 따라 분류된 2개 그룹에서 상위 20%에 기여하는 부분적인 화학구조(substructure)를 위상적 지문(topological fingerprint) 계열의 분자표현자인 ECFP (Extended-Connectivity Fingerprint) [39]를 적용하여 시각화하였다(Fig. 4). 이를 통해 약물유사-화합물의 간 독성에 기여하는 부분적인 화학구조를 확인 할 수 있었고, 이를 약물개발과정에서 선도물질(hit)을 최적화(약인성간독성을 유발하지 않는 작용기로 치환 등) 하는 데에 활용할 수 있다.

2.4 약물의 2D-화학구조로부터 유해영향 예측

498개 약물유사-화합물의 2D-화학구조로부터 TOPKAT® (Toxicity Prediction by Computer Associated Technology) 프로그램[40,41]을 적용하여 설치류 대상 급성경구독성(반수치사용량 LD₅₀, 최대허용량), 발생독성 가능성(Developmental toxicity potential), 발암독성 유·무 등과 같은 독성을 통계적 방법으로 인체 유해영향을 예측하였다. 발생독성은 정상적인 성장, 분화, 발달 또는 행동을 방해하는 등 가역적 또는 비가역적 구조적 또는 기능적 변형이다. 발생독성 가능성은 화학물질이 정상적인 발달을 방해하고 자손에게 부작용을 일으킬 수 있는 정도를 평가한다. FDA 연구자료를 이용한 발암독성 유·무의 예측결과 산출은 기존데이터로부터 도출된 발암잠재력이 있는 2D-화학구조로부터 통계적 확률을 계산하는 방식으로 예측 결과를 얻었다.

TOPKAT® 은 약물유사 화합물의 2D-구조 정보만을 이용하여 전자위상화적인 원자상태(electrotopological state atom, E-state)[42], 화학구조 형상(shape), 대칭적인 구조 특성과 분자량, VlogP(옥탄올-물의 분배계수, 소수성)와 화학물질의 수송과 관련 분자표현자(transport-related descriptors)를 부분구조(substructure) 매개함수로 구조-독성간의 상관관계를 분석한다. 따라서 두

화학물질이 유사성을 갖는다면, 위의 분자표현자의 값이 유사해서 화학구조의 분자표현자 공간에서 동일한 영역 또는 근접한 영역에 존재하는 것이다. 이때 성질 유사성을 의미하는 것은 아니다. 대상 약물유사-화합물의 화학구조 분자표현자 공간에서 구조유사성 거리의 값은 0에서 1값을 가지며, 유사성이 클수록 1에 가까운 값을 가진다.

TOPKAT[®] 프로그램의 경우 기존 FDA의 의약품평가 연구센터와 국립 독성 프로그램(NTP, National Toxicity Program)에서 제공받은 데이터로부터 개발되었다. 이들의 데이터로부터 화학물질의 독성결과를 회귀분석, 확률 기반 베이지안 모델방식으로 알려진 독성물질의 구조유사성 등을 기계학습 한다. 이를 통하여 학습한 정보에 기반하여 대상 약물유사-화합물의 독성을 예측한다[43]. TOPKAT[®] 은 정량적인 구조-독성의 관계식으로부터 약물유사-화합물의 독성 예측값을 계산한다.

약물유사-화합물의 화학구조 특성을 설명하는 분자표현자와 해당 분자표현자 계수의 곱은 독성에 대한 화학구조 특성의 기여도를 의미한다. 양적인 기여도(양의 상관관계)와 음의 기여도(음의 상관관계)가 존재할 수 있다. 양적인 상관관계인 경우, 선택된 화학구조의 특성 확률이 증가하면 기여도가 높아지고 음의 상관관계라면, 화학구조의 특성 확률이 증가하면 기여도는 감소된다.

약물유사-화합물의 예측된 독성 값은 화학구조 특성 관련 분자표현자들의 기여도를 합한 것이다. 이 기여도의 합한 값은 0.0과 1.0사이의 확률값으로 변환된다. 발암독성은 독성이 있는 화합물에서 1의 값을 가지며, 무독성인 경우에는 0의 값으로 선형 판별 분석(Linear Discriminant Analysis, LDA)으로 데이터 분포를 학습한다. 이를 통해 결정경계(Decision boundary)를 만들어서 데이터를 분류하여 발암독성 유·무를 결정하였다. 이때, 발암독성의 확률값(0.0~1.0)이 0.3미만이면 무독성이고, 확률값이 0.7이상이면 발암독성이 있는 약물유사-화합물이다. 다만, 확률값이 0.3이상~0.7미만일 경우에는 불확실한 영역(indeterminate zone)에 존재한다고 간주하였다.

만일, 기존 독성자료 데이터베이스에 없는 자료에 대해서는 최적 예측 공간(OPS, Optimal Predictive Space) 검증법을 통해 약물유사-화합물의 TOPKAT[®] 독성모델에 적용 가능성을 검토하였다. 또한, 최적 예측 공간 내에 있다면, 계산된 독성 예측값을 수용하였다. 반면에, 최적 예측 공간에 벗어나 있는 경우에는 최적 예측 공간 OPS로부터 대상 약물유사-화합물질의 2D-구조특성(원자형태, 결합형태, 결합수등 고려)간의 마할라노비

스 거리(mahalanobis distance p-value) 허용한도에 의존하여 계산된 독성 예측값의 수용 여부를 결정하였다. 마할라노비스 거리의 허용한도보다 큰 경우에는 독성 예측값을 수용하지 않는다. 마할라노비스 거리는 다변량 공간에서 두 변수 간의 상관관계까지 고려한 거리를 의미하는 것으로서,

$$d(\mu, \nu) = \sqrt{(\mu - \nu) \sum_{i=1}^{-1} (\mu - \nu)^T} \quad (2)$$

여기서 μ 는 각 데이터(raw), ν 는 각 데이터의 평균을 의미한다. 이는 공분산 매트릭스를 이용하여 거리를 측정하는 개념이다. 이를 통해서 약물유사-화합물의 화학구조가 기존 FDA 또는 NTP 독성모델의 데이터 공간의 정보가 풍부한 영역에 있는지와 유사한 화합물이 TOPKAT[®] 독성모델에 의해 잘 예측이 되는 경우인지를 확인(신뢰도 % 측면에서) 할 수 있었다. 하지만, 알려진 독성물질의 구조유사성은 일반적으로 대사 또는 발암성의 잠재적 메커니즘과 관련하여 약물유사-화합물과 명확한 관계를 갖고 있지 않다.

이번 연구에서 498개 약물유사-화합물 중에서 15개 약물유사 화합물(염, 무기물, 고분자 등)에 대해서는 독성예측에서 제외되었다. 이는 FDA 또는 NTP 데이터베이스 내 화학구조가 유사한 화학물질의 정보가 부족한 경우이거나 예측을 할 수 없기 때문이다. 게다가, 498개 약물유사-화합물에 대한 독성정보는 간 독성 유·무를 알고 있는 데이터만을 사용하였기에, TOPKAT[®] 모델은 베이지안 QSAR 모델로 간 독성 유·무 예측한 것과 직접적인 비교를 하지 못한다. 다만, 다른 측면의 유해성 유·무를 보충하는 자료로 483개 약물유사 화합물의 구조별 예측값으로, 선도물질(hit) 발굴 단계의 filtering 기준(독성 유발 잠재성 제거)에 적용하기 위해 수행하였다. 본 연구과정에서 수행된 계산 및 분석은 상용화 discovery studio 분자모델링 프로그램(BIOVIA 2022 version[33])에서 수행하였다.

3. 실험결과 및 고찰

3.1 간독성/무독성 결정인자 : 약물의 화학구조

약물의 대사과정에서 가장 중요한 기관은 간이다. 장에서 흡수된 화학물질이 간 문맥계 통해서 간으로 들어오면 다른 물질로 대사되는 데, 대부분의 약물 대사 효소 시스템(예. Cytochrome P450 약물 대사에서 가장 중요한 경로)이 간에서 매우 높은 농도로 존재한다. 즉, 약

Table 1. The equation coefficients of properties and their statistics for each principal component on the molecular descriptors

Equation Terms	PC1	PC2	PC3	Min	Max	Mean	Std. Dev
Constant	-3.085	-0.8796	-0.857	-	-	-	-
AlogP	-0.07629	0.2279	0.01741	-9.552	12.761	2.0577	2.5906
Molecular Weight	0.002445	0.001777	-0.0006761	60.055	1550.2	362.63	186.63
Number H-bond Donors	0.1832	-0.06297	-0.01853	0	17	2.251	2.6479
Number H-bond Acceptors	0.1281	-0.009962	-0.007976	0	26	5.1406	4.0914
Number Rotatable Bonds	0.05496	0.06393	-0.1124	0	43	4.8635	4.7939
Number Rings	0.1816	0.1465	0.2383	0	9	2.8313	1.8594
Number Aromatic Rings	0.08839	0.1617	0.6045	0	6	1.2912	1.1379
Molecular Fractional Polar Surface Area	1.438	-3.607	0.8894	0	0.894	0.2703	0.15437

물이 위장관에서 흡수된 후, 간문맥을 통해 간으로 들어 오면 대사되어 효과를 나타낸다. 높은 지용성, 작은 분자량을 가진 약물은 간 세포막을 쉽게 통과하여 간세포 속으로 유입이 되지만, 간세포 속에 많이 유입되는 것과 CYP 효소들에 의해 빨리 대사되는 것과는 무관하다. 대사 속도 관련 CYP 효소에 의해 분해되는 정도를 결정하는 데에 있어서, 약물의 화학구조가 지용성 또는 분자량보다 더 많이 중요하다. 지용성이 높으면 세노관을 통해 재흡수율이 높아져 콩팥에 의한 배설시키는 능력은 오히려 감소되는 경향이 있다. 일반적으로 약리학적 활성 대사산물의 대사 속도를 증가시키는 효소 유도를 통해 약물 효과의 지속 시간과 강도를 감소시킨다. 반면에 대사 효소의 억제 경우, 지속 시간과 강도를 증가시켜 잠재적으로 독성을 유발하기도 한다. 이로 인해 약물대사가 일어나는 간 부위에서 국소적인 독성 반응으로 약인성간손상(Drug Induced Liver Injury, DILI)[28,44] 될 수 있다. 게다가, 같은 약이라 해도 사람 각자에서 약물의 흡수·분포·대사·배설과 관련된 약동학 파라미터값들은 다르기 때문에, 약물대사에서 유전적 차이에 기인된 체이적 반응(Idiosyncratic effect)[45]과 약물-약물 상호작용 등이 유발된다[46].

실험적인 결과(*in vivo*)로부터 간 독성 유·무가 결정된 FDA 간독성 지식 데이터베이스(LTKB)내 2,889개 약물유사-화합물 데이터를 수집하였다. 이 중에서 간 독성 예측 모델을 개발하기 위해 다양성이 높은 training set으로, 498개 약물유사-화합물을 선별하였다(최대비 유사성 거리가 상위 18%에 해당함). 선별된 498개 약물유사-화합물의 2D-화학구조로부터 물리화학적 성질의 분자표현자를 매개함수로 하여 화학적 공간에서 분포 경향성(편향성 등)을 살펴보았다. 이는 간 독성이 있는 179개 화합물과 간 독성이 없는 319개 화합물이 위치 해있

는 구조·성질의 화학적 공간이 구분될 수 있는 지와 이를 이용하여 간 독성 유·무의 예측 정도를 확인하였다. 이때 8개 분자표현자 매개함수와 상수를 단위벡터로 하여 주 성분 분석으로, 화학공간의 새로운 좌표계(PC1, PC2, PC3)에서 상대적 위치를 결정하였다(Table 1).

Fig. 1에서 보는 바와 같이 간 독성이 있는 화합물(빨강색)과 무독성 화합물(파랑색)은 새로운 좌표계에서 명확하게 구분이 되는 경향성(편향성)이 존재하지 않았다. 간 독성이 있는 원료물질들의 화학적 공간이 근접한 영역에서 존재하는 반면에 간 독성이 없는 원료물질의 화학적 공간이 넓게 퍼져있었다(Fig. 1). 이러한 경향을 데이터프레임에서 간 독성/무독성 약물유사-화합물간의 8개 분자표현자값의 분포를 박스플롯으로 최소값, 1분위수, 중앙값, 3분위수, 이상치를 제외한 최대값, 이상치를 제외한 최소값 등으로 상세하게 비교하였다(Fig. 2). 간 독성이 있는 약물유사-화합물의 물리화학적 성질들의 분포도가 무독성 화합물보다 일정한 영역에서 존재하는 경향이 있다. 이는 자료의 퍼짐이 이상치를 제외한 최소값과 최대값 간의 차이가 적었다. 특히 물리화학적 성질들 중에서 옥탄올과 물의 분배계수로 소수성을 설명하는 AlogP의 중간 값(간 독성 1.960/ 무독성 2.112) 등은 차이가 크지 않았다. 특히, AlogP 분자표현자는 미국 환경보호청(EPA) ECOSAR(Ecological Structure Activity Relationship)에서 화학물질의 환경유해성 평가를 위해 대상 화학물질의 물리화학적 성질 특성을 설명하는 데 적용되고 있다. 반면에, 화학적 구조와 관련된 헤테로 고리 화합물, 방향족 고리 개수 등은 무독성 약물유사-화합물의 분포도가 더 넓어지는 경향이 있다. 몇몇 헤테로 고리 화합물과 방향족 고리 화합물에 결합한 작용기의 잠재적 독성에 따라 독성을 유발하기도 한다. 예를 들어 벤조디아제핀(진정제를 만들기 위해 개발된 약물) 계통은 간

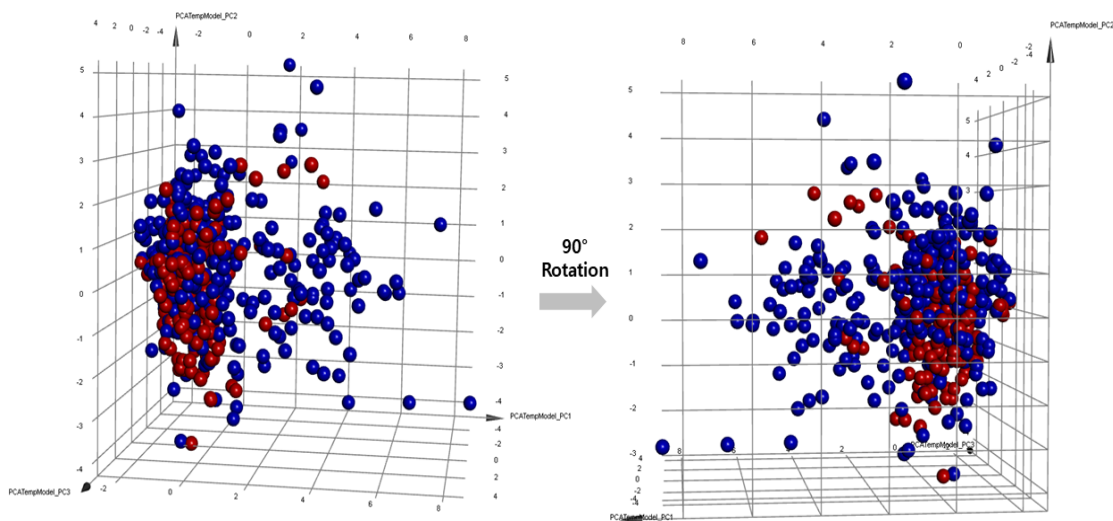


Fig. 1. Human hepatotoxicity scores (179 positives in red and 319 negatives in blue) containing the equation coefficients for each principle component on the selected 8 molecular descriptors across 498 drug-like compounds.

에서 대사를 거쳐 신체 밖으로 배출한다. 심각한 간질환 및 간경변 등을 앓고 있는 환자에게 정상인의 용법으로 투여하면, 간에서 대사되는 시간이 길어져 약인성간독성 (DILI)으로 인해 더 악화된다. 간 독성 화합물과 구조 유사성(간에서 CYP 효소에 의한 대사 관련 구조적 특성)이 물리화학적 성질(간문맥을 통과해서 간으로 수송되는 것과 관련된 소수성 AlogP, 분자량, 친수성의 극성표면적 PSA 등)의 유사성보다 간 독성/무독성을 예측하는 데에 더 중요한 결정 인자였다.

3.2 간 독성/무독성 약물의 화학구조 특성

간 독성/무독성을 가진 498개 약물유사-화합물의 구조로부터 25개 간 독성 관련된 fragment로 179개 간 독성이 있는 약물유사-화합물의 구조 특성을 제어하였다. 이때 확률기반 베이지안 QSAR 모델로 조건부 기준 값(Bayesian score) >-1.698 , 간 독성)에 따라 간 독성/무독성을 결정하였다. 더 나아가서 간 독성 예측의 정확도는 498개 약물유사-화합물에서 1개를 test set으로 결정하고 497개 데이터로 베이지안 QSAR 모델링하는 교차검증을 498번 반복 진행하여 확인하였다. 또한, 5개 검증세트(5-fold cross validation)도 실시하였다. 매번 검증할 때마다, 정확도(accuracy), 특이도(specificity, SP), 민감도(sensitivity, SE) 등의 도출된 예측 결과를 ROC(x축: 1-특이도, y축:민감도)로 하여 좌표평면에 곡선 연결하여 아래 부분 면적(Area Under the curve)

AUC로 예측력의 신뢰도를 평가하였다(AUC값이 클수록 예측력의 신뢰도는 높아짐).

즉, 498개 약물유사-화합물의 베이지안 QSAR 간 독성 예측 모델의 결과 적합성은 간 독성 있는 원료물질(true)을 간 독성 있다고 예측(positive)하고 (진양성 True positive, TP), 무독성 원료물질(true)을 무독성으로 예측(negative)하는 진음성, True negative, TN을 평가지표로 적용하였다. 이에 따라 민감도(SE)는 간 독성 원료물질을 올바르게 간 독성이 있다고 예측하는 비율(%)이기 때문에, $SE = TP / (TP + FN) \times 100\%$ 된다.

특이도(SP)는 무독성 원료물질을 무독성이라고 예측하는 비율(%)이기 때문에 $SP = TN / (TN + FP) \times 100\%$ 이다. 예측의 정확도(%)는 $(TN + TP) / (TN + TP + FN + FP) \times 100\%$ 로 계산할 수 있다. Training set으로 생성한 베이지안 모델의 정확도 90.0%보다 5개 검증세트(5-fold cross validation sets)예측 정확도가 90.6%로 약간 더 높았지만, 베이지안 모델의 신뢰도는 검증 세트에서도 유지되었다.

Table 2에서 보는 바와 같이 검증 세트에서 간 독성 유·무와 예측의 전체적인 일치율(90.6%) 나타내는 정확도와 양성 예측율을 나타내는 민감도(97.2%)를 고려할 때, 다소 예측 정확도가 높은 (ROC score 0.855, model rating 'good') 베이지안 QSAR model (Fig. 3)로 498개 약물유사-화합물의 간 독성/무독성을 예측하였다. 간 독성/무독성으로 각각 그룹화되는 약물유사-화합물에서

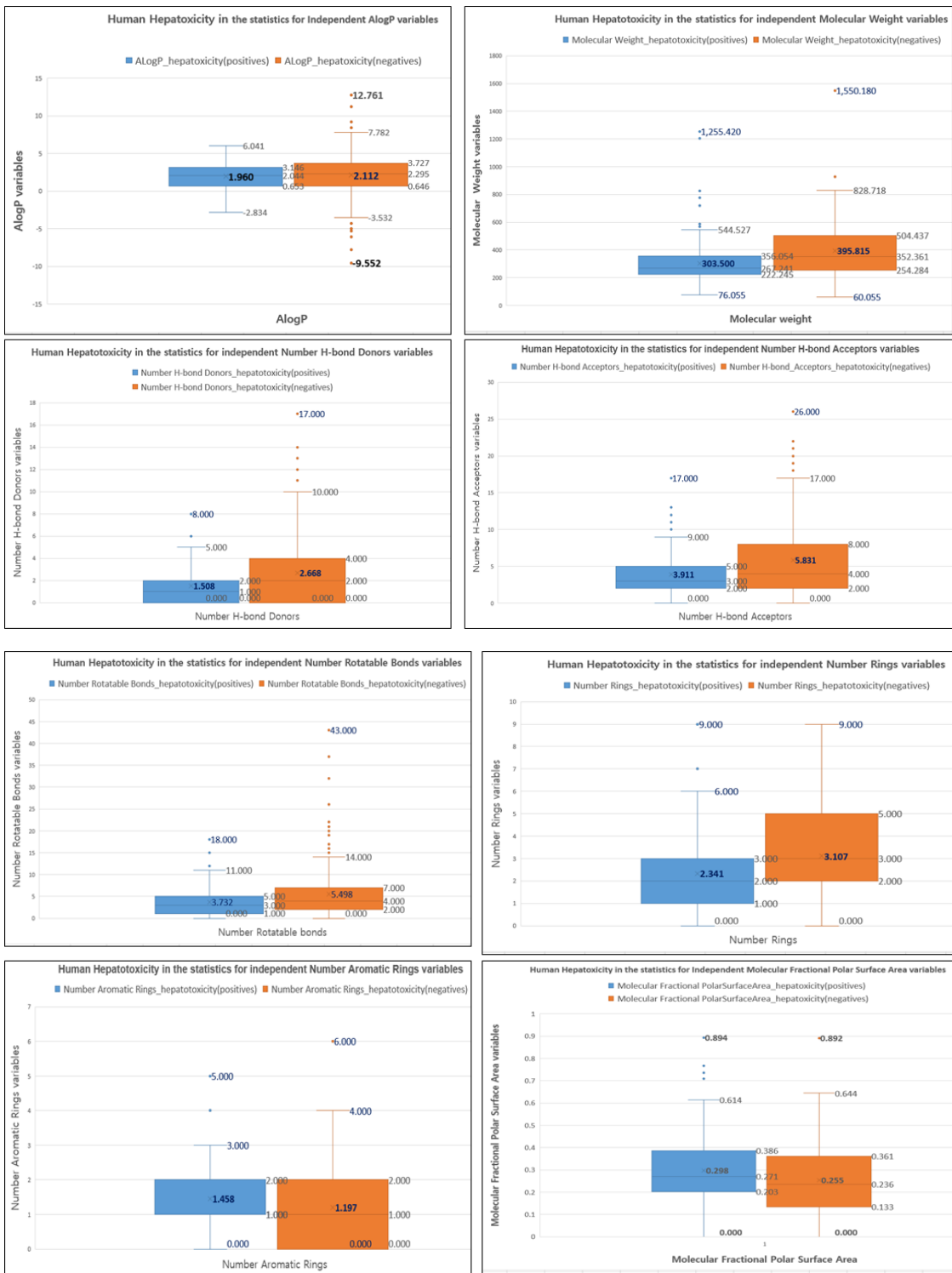


Fig. 2. A comparison of human hepatotoxicity positive (blue bars) and negative (orange bars) drug-like molecules depicted by their stats for independent variables of the selected 8 molecular descriptors.

Table 2. The confusion matrix for prediction performance of human hepatotoxicity across 498 drug-like compounds

Training set Result								
Bayesian QSAR model	ROC score	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	Concordance
	0.989	175	3	47	273	0.983	0.853	0.900
5-Fold Cross validation Result								
Bayesian QSAR model	ROC score	True Positive	False Negative	False Positive	True Negative	Sensitivity	Specificity	Concordance
	0.855	173	5	42	278	0.972	0.869	0.906

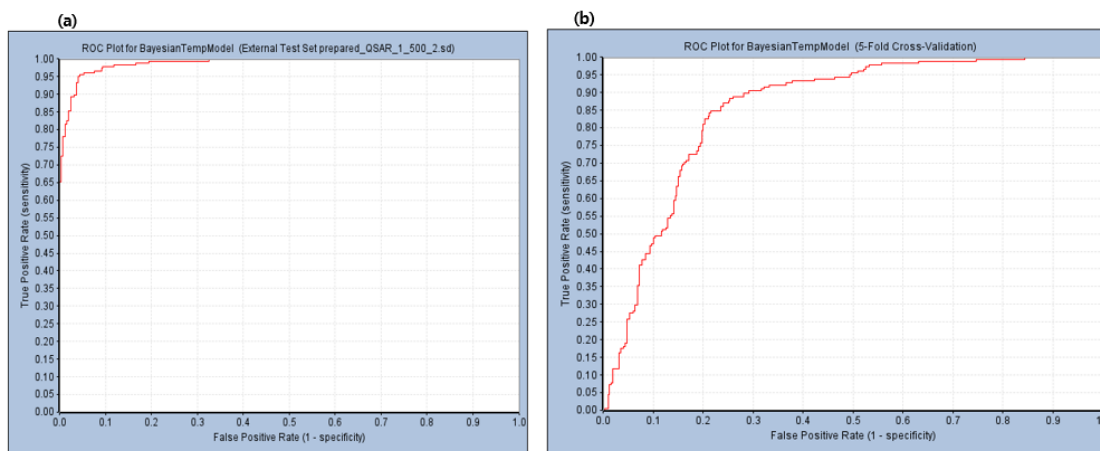


Fig. 3. The Receiver Operating Characteristic Curve for prediction performance of human hepatotoxicity within (a) training set and (b) 5-fold cross validation sets across 498 drug-like compounds.

각각 공통적으로 갖고 있는 상위 20개 substructure의 정보를 추출하였다. 이때 구조적 특성을 ECFP6 fingerprint 적용하여 substructure 특성의 유/무를 0과 1 벡터로 나타내어 분자의 구조와 특징을 표현하였다.

이는 화학구조의 특징이나 유사도를 확인하기 위해 설계된 fingerprint이다[38,39]. ECFP6 fingerprint는 주변 6개 원소들의 평균으로 substructure 구조를 표현하며, 특정 substructure는 어떤 원소와 어떤 구조로 연결이 되어있는 지에 대한 정보를 바이너리 값으로 표현하였다(지문 매개변수를 적용하여 화학구조의 특성을 추출함). 이를 통하여 간 독성의 원인이 되는 약물유사-화합물의 부분구조를 제시함으로써 간 독성 유발 잠재성에 대해 해석할 수 있었다.

간 독성 유발 잠재성을 가진 치환기를 살펴보면, 1차 아민(R-NH₂, 무독성)을 제외한 질소에 유기치환기가 결합된 2차(R₂-NH) 및 3차 아민(R₃-N), 니트릴 작용기, 방향족 아민 또는 아릴아민 등이 포함된다(Fig. 4). 간 독성의 잠재성을 가진 부분구조를 갖는 약물에서는 약인성

간독성이 유발되기도 한다(Fig. 5). 예를 들어 이소니아지드(isoniazid)는 1차 항결핵제로서 간 손상은 연령 증가와 함께 증가하는 경향성을 가졌다. 간 대사 효소의 CYP450, 2C19, 3A4 억제제로 만성 독성을 초래한다.

더 이상 판매되고 있지 않은 근육이완제인 족사졸아민(Zoxazolamine)은 간 독성으로 인해 판매가 중단되었다. 활성대사 산물 중 하나인 클로르족사존(Chlorzoxazone)은 독성이 덜 한 것으로 알려져 있다. 항정신병 약물인 페노티아진(phenothiazine)은 티아진계열의 헤테로고리화합물로서 기저핵, 시상하부 등에 존재하는 도파민 수용체를 억제하고 전뇌의 도파민 대사회전율을 증가시킨다. 페노티아진 경우, 간 손상이 심각하지 않아서 계속 임상에 쓰이는 약물이다. 이와 반대로, 할로세인(halothane)은 브롬과 플루오린, 염소원자를 포함하고 있는 전신마취제에 이용되는 흡입제로 가지는 효능이 뛰어나 간 손상의 잠재력이 있음에도 계속 쓰이는 약물이다.

할로세인처럼 그 자체가 간 독성을 발휘하지만, 대부분 간 대사 효소 CYP450에 의해 대사체로 전환(biotransform)

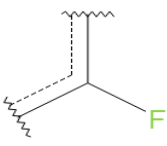
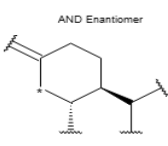
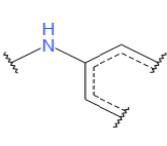
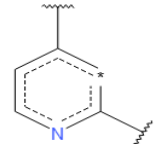
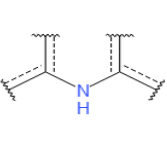
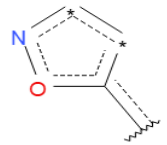
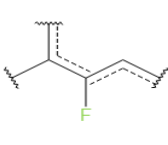
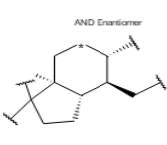
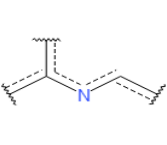
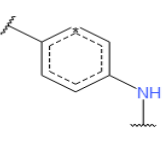
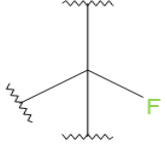
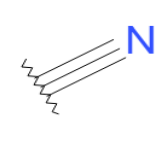
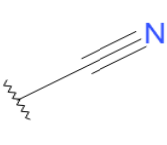
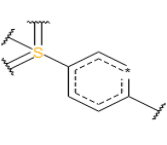
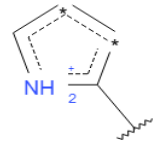
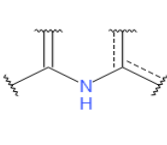
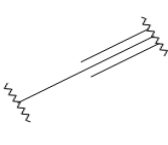
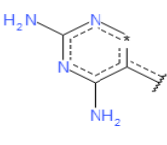
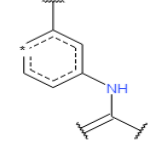
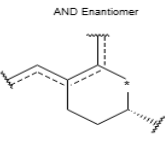
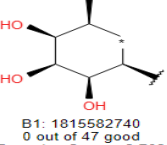
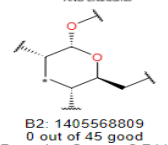
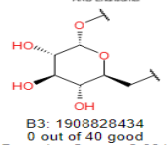
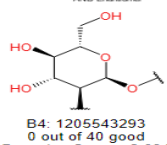
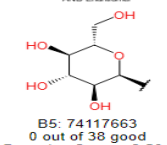
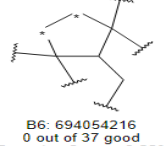
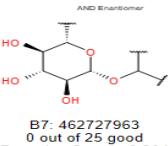
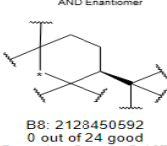
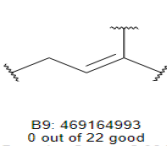
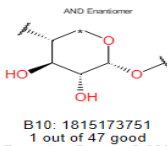
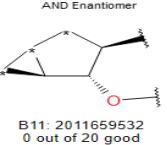
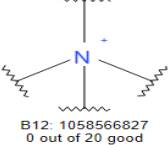
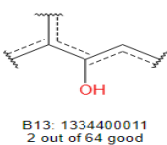
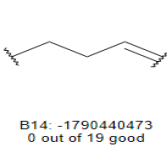
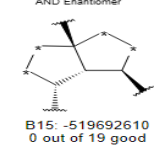
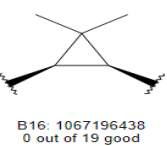
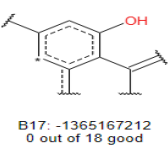
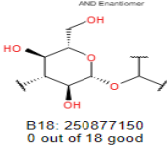
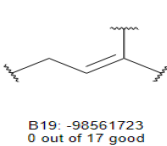
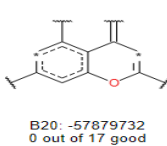
 G1: 220735655 12 out of 12 good Bayesian Score: 0.982	 AND Enantiomer G2: 151446186 9 out of 9 good Bayesian Score: 0.941	 G3: -177077903 15 out of 17 good Bayesian Score: 0.904	 G4: 1165022651 7 out of 7 good Bayesian Score: 0.899	 G5: -1242906247 7 out of 7 good Bayesian Score: 0.899
 G6: 1279696542 7 out of 7 good Bayesian Score: 0.899	 G7: -1311285389 7 out of 7 good Bayesian Score: 0.899	 AND Enantiomer G8: -1799880005 7 out of 7 good Bayesian Score: 0.899	 G9: -677309799 41 out of 51 good Bayesian Score: 0.879	 G10: 888054369 13 out of 15 good Bayesian Score: 0.875
 G11: 226796801 16 out of 19 good Bayesian Score: 0.870	 G12: -1101847286 6 out of 6 good Bayesian Score: 0.870	 G13: 1412053881 6 out of 6 good Bayesian Score: 0.870	 G14: -978131182 6 out of 6 good Bayesian Score: 0.870	 G15: 2116455019 9 out of 10 good Bayesian Score: 0.862
 G16: -1236483485 18 out of 22 good Bayesian Score: 0.854	 G17: -1114776580 14 out of 17 good Bayesian Score: 0.840	 G18: -1408385607 5 out of 5 good Bayesian Score: 0.832	 G19: 1435111106 5 out of 5 good Bayesian Score: 0.832	 AND Enantiomer G20: -1095664289 5 out of 5 good Bayesian Score: 0.832
 B1: 1815582740 0 out of 47 good Bayesian Score: -2.782	 AND Enantiomer B2: 1405568809 0 out of 45 good Bayesian Score: -2.741	 AND Enantiomer B3: 1908828434 0 out of 40 good Bayesian Score: -2.631	 AND Enantiomer B4: 1205543293 0 out of 40 good Bayesian Score: -2.631	 AND Enantiomer B5: 74117663 0 out of 38 good Bayesian Score: -2.584
 B6: 694054216 0 out of 37 good Bayesian Score: -2.559	 AND Enantiomer B7: 462727963 0 out of 25 good Bayesian Score: -2.204	 AND Enantiomer B8: 2128450592 0 out of 24 good Bayesian Score: -2.167	 B9: 469164993 0 out of 22 good Bayesian Score: -2.091	 AND Enantiomer B10: 1815173751 1 out of 47 good Bayesian Score: -2.089
 AND Enantiomer B11: 2011659532 0 out of 20 good Bayesian Score: -2.008	 B12: 1058566827 2 out of 20 good Bayesian Score: -2.008	 B13: 1334400011 2 out of 64 good Bayesian Score: -1.975	 B14: -1790440473 0 out of 19 good Bayesian Score: -1.963	 AND Enantiomer B15: -519692610 0 out of 19 good Bayesian Score: -1.963
 B16: 1067196438 0 out of 19 good Bayesian Score: -1.963	 B17: -1365167212 0 out of 18 good Bayesian Score: -1.917	 AND Enantiomer B18: 250877150 0 out of 18 good Bayesian Score: -1.917	 B19: -98561723 0 out of 17 good Bayesian Score: -1.869	 B20: -57879732 0 out of 17 good Bayesian Score: -1.869

Fig. 4. Bayesian model built with 498 drug-like molecules that features important for human hepatotoxicity active and inactive

될 때 유리기(free radical), 전자친화성산물(electrophilic intermediate), 산소 유리기(oxygen free radical)와 같은 반응성 대사산물에 의해 간 독성이 나타난다[47]. 할로알케인(haloalkane)은 유리기(unpaired electron)에 의한 간 손상의 대표적인 예이다. 막 구조를 직접 파괴, 막 지질의 불포화 지방산의 과산화, 세포 내 단백질 합성 장애, DNA나 RNA 변화 등이 초래된다. 타이레놀의 원료성분인 아세트아미노펜(acetaminophen)은 벤젠의 1번 위치에 있는 아세트아미노(R-CH₃OCNH₂)가 결합이 되어있고, 4번 위치에 하이드록시(R-OH)가 결합된 화합물이다. 해열진통제 아세트아미노펜과 같은 일반 약물도 전자 친화성 산물에 의한 세포 내 구성 성분들과 공유결합하여 세포의 구조적 또는 기능적 병변을 초래하는 손상 기전을 갖고 있다. 이와 같이 간 독성을 유발하는 substructure를 정보(Fig. 4)를 약인성간독성 약물(Fig. 5)과 연계성을 살펴보았다. Fig. 4에서 보는 바와 같이, 간 독성 유발 가능성이 높은 질소 함유 방향족 헤테로고리 화합물이 갖고 있는 피리딘(pyridine)과 피롤(pyrrole)의 substructure(G4, G15)를 갖는 이소니아지드와 족사졸아민보다 페노티아진과 아세트아미노펜의 간 독성 유발 가능성이 상대적 더 낮은 경향성으로 설명이 되었다. 또한, 탄소 원자의 플루오이드 치환기(R-F)를 갖는 할로겐화물의 substructure (G1, G7, G11) 등의 경우 할로세인처럼 간 독성이 높은 약물에도 베이저안 QSAR 모델 적합도가 잘 반영이 되었다.

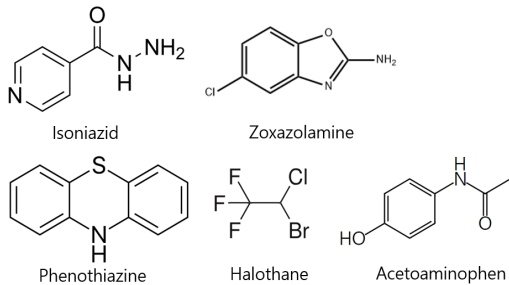


Fig. 5. Detection of positives within human hepatotoxicity ranked by frequency of representative substructure from drug classified as positive.

3.3 약물유사 화합물의 독성효과 예측

간 독성/무독성을 예측한 498개 약물유사-화합물에서 염과 무기질, 고분자 형태인 15개 원료물질을 제외한 483개 유기화합물의 독성효과를 예측하였다. 2D-화학구조로부터 substructure 매개함수와 부분구조(fragment

변화에 따른 독성효과를 TOPKAT[®] 독성모델로 분석하였다. 이때, 미국 FDA 의약품평가연구센터와 국립 독성 프로그램(NTP)의 축적된 동물독성시험자료(10,000건 이상)를 적용하였다.

이번 연구에서 적용한 원료물질의 유해용량에 대한 정보는 FDA 의약품평가연구센터에서 축적한 약물유사-화합물의 용량-반응평가(Dose-Response Assessment) 관련 최저관찰영향수준(Lowest Observed Adverse Effect Level), 또는 반수치사용량 LD₅₀ 값의 독성 종말점이다. 이를 이용한 유해인자로부터 영향이 나타날 수 있는 가능성(probability)을 평가함으로써 유해영향의 존재여부(양성/음성)로 원료물질의 안전성을 제안하였다. 이는 약물유사-화합물의 간 독성 이외, 독성 종말점 등의 데이터를 기반으로 실측 결과가 없는 원료물질에 대한 유전독성과 발암성, 발생독성 잠재성 등으로 유해영향을 예측하는 것이다.

계산된 가능성이 0.3보다 적은 경우에는 독성 발생 가능성이 적고 0.7보다 큰 경우에는 독성발생 가능성이 높은 것으로 유해영향 예측 결과를 해석하였다. 다만, 가능성이 0.3과 0.7 범위에 있는 경우에는 불확실한 영역(IND)에 존재하여 유해성 결과를 확증하는 데 추가적인 시험의 실측 결과가 필요하다. 이는 불확실성이 적은 정보를 적용하는 것과 신뢰성이 높은 과학적 자료 기반 유해성 예측 결과를 해석하기 위함이다.

TOPKAT[®] 독성모델은 유사한 구조를 가지는 화합물들은 유사한 독성효과를 나타낸다는 전제를 기반으로 한다. 483개 유기화합물의 간 독성/무독성 이외에, 경구용 급성독성, 유전독성, 발암성 등을 나타내는 화학구조의 유사성을 기계학습 방법으로 측정하고 이를 로지트릭('if and else')로 적용하였다. 인체 유해성 예측은 (1) 변이 원성과 발암성이 없는 경우에는 급성 경구독성의 종말점(LD₅₀), (2) 변이원성의 가능성이 높은 경우에는 발암성과 발생독성 가능성 예측, (3) 발암성 경우에는 발암성 원료물질이 갖고 있는 구조 특성에 대한 유사성을 기준 표본점과 분포사이의 마할라노비스 거리를 측정하는 방법을 적용하였다. 최종적으로 FDA 데이터베이스 결과와 일치도 및 최적 예측 공간(OPS)에 존재 여부를 확인 후, 예측 결과에 대한 수용 여부를 결정하였다.

만일 다른 화합물의 구조에 동일한 substructure가 있다면, 적용하는 분자표현자(독성의 결정요인) 및 분자표현자의 값(독성효과의 기여도)은 동일하게 적용된다. 즉, substructure의 fragment 위치변화에 따른 독성변화량은 고려되지 않는다. 이를 적용하여 화합물질 구조-

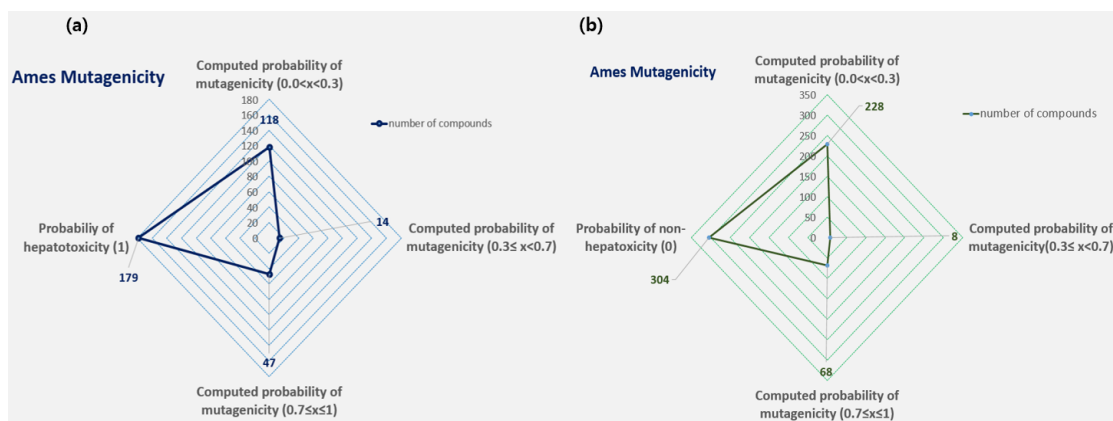


Fig. 6. Ames mutagenicity classification predicted by TOPKAT® on 483 human hepatotoxicity prediction results of drug-like compound

독성 상관관계 예측에 있어, 적용할 수 있는 구조 유사체 범위는 최적 예측 공간(OPS)에 있는 것으로 모델의 범주를 설정하였다.

TOPKAT® 변이원성에 대한 예측 모델은 미국 환경보호청 EPA의 GeneTox 프로토콜에 따라, 살모넬라 티피뮤리움(*Salmonella Typhimurium*) 5종 균주 TA100, TA1535, TA1537, TA1538, TA98 복귀돌연변이 분석(histidine reversion assay) 결과를 적용하였다. 483개 화합물의 (1) 변이원성에 대한 예측은 Ames Mutagenicity 3.1를 적용하여 유전적 돌연변이를 일으킬 수 있는 화합물의 구조 특성으로 예측 결과를 해석하였다.

평가대상 화합물의 변이원성 가능성이 0.0과 0.3 범위에 있는 경우(non-mutagen)에는 돌연변이를 유발하지 않는다. 간 독성이 있는 179개 약물유사-화합물에서 118개 원료물질은 변이원성이 검출되지 않았다(Fig. 6(a)). 반면에, 47개 원료물질은 변이원성에 대한 가능성(0.7 이상) 있는 substructure를 가졌다. 게다가 34개 원료물질은 가능성(probability) 1로 최적예측 공간(OPS)에서 변이원성을 양성으로 예측하였다. 반면에, 304개 약물유사 화합물에서 간 독성은 없지만, 변이원성의 가능성을 갖는 substructure가 있는 68개 원료물질을 추가적으로 검출하였다(Fig. 6(b)). 483개 약물유사 화합물에서 115개 변이원성으로 예측된 원료물질과 346개 비변이원성 화합물질의 substructure 화학구조 특성을 비교하였다. 놀랍게도 동물시험자료가 축적된 FDA 데이터베이스 내에 존재하는 9개 변이원성 화합물질(9/115, 7.82%)에서 기준 구조적인 특징(structural alert)로 방향족에 니트로기, 아민, 할로젠족 작용기가 치환되

었을 때, 페닐기에 하이드록시기(R-OH)가 결합된 화합물을 변이원성 화합물질로 평가되었다(Fig. 7). 특히, 방향족 화합물에서 전자를 끌어당기는 니트로기(R-NO₂) 작용기가 공액결합(conjugation)으로 결합세기가 강해지고 안정화되면서 변이원성을 증가시킬 수 있다.

다카바진(decarbazine)은 흑색종과 림프종의 치료에 사용되는 화학요법의 약물이다. 작용기전은 간 효소에 의해 활성화되어 알킬화 화합물인 MTIC(Monomethyl Triazeno Imidazole Carboxamide)를 형성하는 데 DNA의 메틸화, 변형, 교차결합을 유발해서 DNA, RNA, 단백질 합성을 억제하는 변이원성과 간 독성을 가지고 있다. 변이원성 가능성이 1 값으로 실험적으로 측정된 Ames test 결과와 변이원성에 대해 양성인 것으로 일치하였다.

또한, 3개 이상 고리화합물에서도 변이원성이 높아지는 경향이 있다. 예를 들어, 디하이드로퀘르세틴으로 알려진 탁시폴린(toxifolin)은 플라보노이드 하위클래스 플라보놀에 속하는 데 시베리아 낙엽송, 밀크시슬과 같은 식물에서 추출되는 성분이다. 탁시폴린 수화물의 경우에는 물질안전보건자료(MSDS, Material Safety Data Sheets)에 대사활성계 유무와 상관없이 유전독성은 음성이다. 킬레이트 형성작용이 있는 플라보노이드나 카테킨 및 벤젠환에 2개 이상 수산기가 결합된 폴리페놀이 *in vitro* 내에서 산화저밀도지단백질(oxidized LDL) 산화를 억제하고, 항산화성에 기인하며 활성산소가 관여하는 질병을 감소시키는 것으로 알려져 있다. 탁시폴린의 기존 시험자료에 대한 변이원성 예측 결과는 일치하지 않았다. 이는 TOPKAT® 변이원성 모델에서 특정 분자의 fragment 독성에 대한 기여도를 고려할 때, 생물학적인

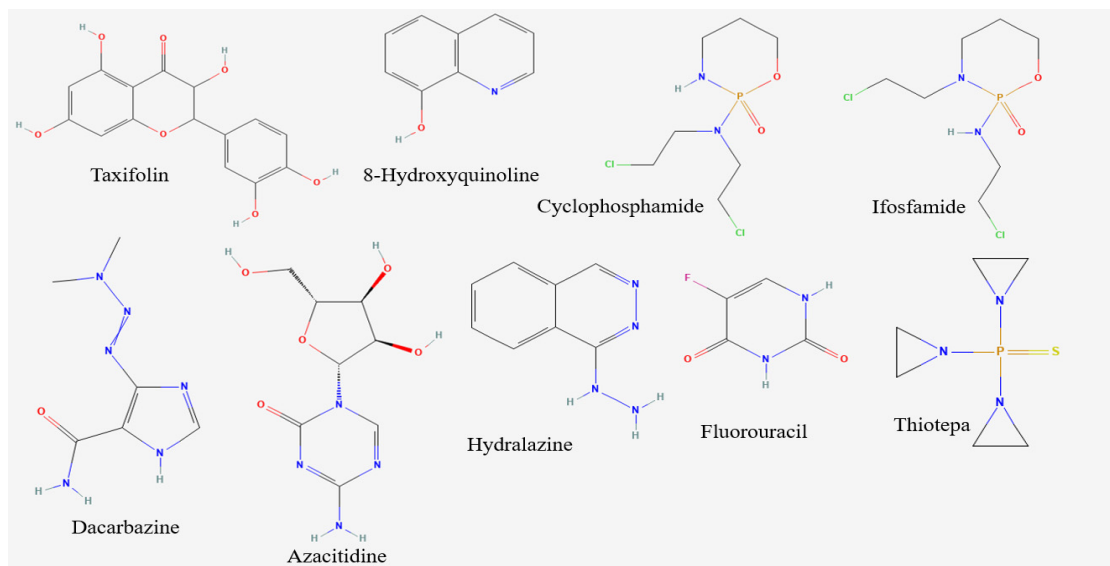


Fig. 7. Nine mutagen compounds as positive predicted by TOPKAT® with high computed probability values in the hypothesized assessment.

측면을(작용기 위치 및 개수, 작용기-작용기 상호작용에 의한 효과 상쇄·증가, 수용체와 전자적 작용 등) 충분히 고려하지 못했기 때문일 것이다. 즉, 약물유사-화합물의 작용점으로 수송과 관련된 화학구조 특성 이외에 분자량, 크기, 모양, 구조 대칭성, 소수성뿐만 아니라, 변이원성(복귀돌연변이)의 다른 결정요인이 작용했기 때문이다. 최종적으로, 생체 매개변수와 근접하지 않은 타시폴린에 대한 변이원성의 예측 결과를 수용하지 않았다.

TOPKAT® 발암성 예측 모델은 FDA 의약품평가조사센터(CDER)에서 설치류 대상(male mouse, female mouse) 발암성 평가 데이터에 기반으로 하였다. 원료물질의 2D-화학구조로부터 선형판별분석(LDA, Linear Discrimination Analysis)을 통해 발암성/비발암성 화학물질을 분류한다. 483개 약물유사-화합물의 2D-화학구조 표현자를 하나의 벡터 선으로 투영하였다. 즉, 특정 화합물 표현자 공간에서 클래스 분리를 최대화하는 축을 찾기 위해 클래스 간 분산(between class scatter)과 클래스 내부 분산(within-class scatter)의 비율을 최대화하는 방식으로 차원을 축소한다. 결과적으로 클래스 내의 분산이 최소화 되게 하고, 클래스 간의 거리가 최대가 되게 하는 벡터를 찾아 각 데이터 포인트를 투영시키는 것이다. 기존 FDA 설치류 발암성 평가 데이터의 정보를 적용하여 발암성/비발암성 클래스 간의 화학구조 분별 정보를 최대한 유지하면서 차원을 축소시켜 분석대상 원료물질에 대하여 발암독성 값을 예측하게 된다. 다음 단

계로 최적예측 공간(OPS)에 존재하는 지와 CDER 실험적 발암성 데이터에 있는 화학물질들과 유사성 거리를 계산하여 발암성 가능성을 예측하였다(유사성 거리가 짧을수록 발암성은 높아진다). 이때 가능성이 0.7이상 ($0.7 \leq x \leq 1$)이면, 발암물질로 간주하였지만, 가능성이 0.3 미만인 경우에는 비발암성 물질로 평가하였다. 다만, 가능성이 0.3이상 0.7 미만일 경우에는 불확실성 영역 내에 있어서 예측 결과를 수용하지 않았다.

각 실험동물 품종 간 투여경로, 간 독성 동태의 차이가 존재하기 때문에, 투여경로를 고려하지 않고 실험적 발암성/비발암성 결과가 있는 원료물질을 우선적 선별하였다. 수컷 쥐에서 발암 가능성이 높은 것으로 예측된 190개 원료물질 중에서 27개 원료물질이 기존 CDER 실험적 발암성 데이터에 포함이 되었다(27/190, 14.2%). 반면에, 암컷 쥐에서 발암 가능성이 높은 것으로 예측된 197개 원료물질 중에서 23개 원료물질이 기존 실험데이터에 포함이 되었다(23/197, 11.67%). 이번 연구에서 TOPKAT® 발암성 예측 모델은 실험동물의 성별에 대한 독성효과의 차이를 별도로 고려하지 않았기 때문에, 암수 모두에서 발암성이 있고, 또한 기존 CDER 실험적 발암성 데이터에 포함이 되어 있는 19개 원료물질을 선별하였다. 선별된 원료물질의 발암성 기준은 10개 비유전독성 발암물질과 9개 DNA를 손상시키는 돌연변이 유발성에 의한 유전독성이 있는 발암물질로 분류할 수 있다 (Table 3).

Table 3. Candidate chemicals with carcinogenicity.

CAS No.	Synonyms	Molecular formula	Prediction
80-08-0	dapsone	C ₁₂ H ₁₂ N ₂ O ₂ S	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 0.008 - Computed Probability of Carcinogenicity :1.000 - Computed Rat Oral LD₅₀ : 2.4g/kg - Computed Maximum Tolerated Dose (feed/water) : 171.6mg/kg - Computed Probability of Developmental Toxicity Potential : 0.812
86-54-4	Hydralazine	C ₈ H ₈ N ₄	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 1.000 - Computed Probability of Carcinogenicity :0.995 - Computed Rat Oral LD₅₀ : 181.7mg/kg - Computed Maximum Tolerated Dose (feed/water) : 20.0mg/kg - Computed Probability of Developmental Toxicity Potential : 0.072
50-18-0	Cyclophosphamide	C ₁₇ H ₁₅ ClN ₂ O ₂ P	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 0.921 - Computed Probability of Carcinogenicity : 1.000 - Computed Rat Oral LD₅₀ : 82.3mg/kg - Computed Maximum Tolerated Dose (feed/water) : 3.5mg/kg - Computed Probability of Developmental Toxicity Potential : 0.204
126-07-8	Griseofulvin	C ₁₇ H ₁₇ ClO ₆	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 0.000 - Computed Probability of Carcinogenicity : 1.000 - Computed Rat Oral LD₅₀ : 289.3mg/kg - Computed Maximum Tolerated Dose (feed/water) : 51.7mg/kg - Computed Probability of Developmental Toxicity Potential : 0.005
50-28-2	Estradiol	C ₁₈ H ₂₄ O ₂	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 1.000 - Computed Probability of Carcinogenicity : 0.910 - Computed Rat Oral LD₅₀ : 180.9mg/kg - Computed Maximum Tolerated Dose (feed/water) : 93.8mg/kg - Computed Probability of Developmental Toxicity Potential : 0.985
671-16-9	Procarbazine	C ₁₂ H ₁₉ N ₃ O	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 0.000 - Computed Probability of Carcinogenicity : 0.994 - Computed Rat Oral LD₅₀ : 411.6mg/kg - Computed Maximum Tolerated Dose (feed/water) : 2.4mg/kg - Computed Probability of Developmental Toxicity Potential : 1.000
52-01-7	Spironolactone	C ₂₄ H ₃₂ O ₄ S	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 0.005 - Computed Probability of Carcinogenicity : 1.000 - Computed Rat Oral LD₅₀ : 10.0g/kg - Computed Maximum Tolerated Dose (feed/water) : 338.1mg/kg - Computed Probability of Developmental Toxicity Potential : 0.000
637-07-0	Clofibrate	C ₁₂ H ₁₅ ClO ₃	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 0.009 - Computed Probability of Carcinogenicity : 0.994 - Computed Rat Oral LD₅₀ : 2.1g/kg - Computed Maximum Tolerated Dose (feed/water) : 167.4mg/kg - Computed Probability of Developmental Toxicity Potential : 0.022
4342-03-4	Decarbazine	C ₆ H ₁₀ N ₆ O	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 1.000 - Computed Probability of Carcinogenicity : 1.000 - Computed Rat Oral LD₅₀ : 2.6g/kg - Computed Maximum Tolerated Dose (feed/water) : 8.6mg/kg - Computed Probability of Developmental Toxicity Potential : 1.000
60-56-0	Methimazole	C ₄ H ₆ N ₂ S	<ul style="list-style-type: none"> - Computed Probability of Mutagenicity : 1.000 - Computed Probability of Carcinogenicity : 1.000 - Computed Rat Oral LD₅₀ : 202.7mg/kg - Computed Maximum Tolerated Dose (feed/water) : 49.0mg/kg - Computed Probability of Developmental Toxicity Potential : 0.000

CAS No.	Synonyms	Molecular formula	Prediction
298-81-7	Methoxsalen	C ₁₂ H ₈ O ₄	- Computed Probability of Mutagenicity : 1.000 - Computed Probability of Carcinogenicity : 0.997 - Computed Rat Oral LD ₅₀ : 765.1mg/kg - Computed Maximum Tolerated Dose (feed/water) : 39.7mg/kg - Computed Probability of Developmental Toxicity Potential : 0.000
297-76-7	Ethinodiol diacetate	C ₂₄ H ₃₂ O ₄	- Computed Probability of Mutagenicity : 0.000 - Computed Probability of Carcinogenicity : 0.999 - Computed Rat Oral LD ₅₀ : 10g/kg - Computed Maximum Tolerated Dose (feed/water) : 8.0mg/kg - Computed Probability of Developmental Toxicity Potential : 0.000
103-90-2	Acetaminophen	C ₈ H ₉ NO ₂	- Computed Probability of Mutagenicity : 0.000 - Computed Probability of Carcinogenicity : 0.999 - Computed Rat Oral LD ₅₀ : 1.5g/kg - Computed Maximum Tolerated Dose (feed/water) : 256.5mg/kg - Computed Probability of Developmental Toxicity Potential : 0.000
90357-06-5	Bicalutamide	C ₁₈ H ₁₄ F ₄ N ₂ O ₄ S	- Computed Probability of Mutagenicity : 0.000 - Computed Probability of Carcinogenicity : 0.888 - Computed Rat Oral LD ₅₀ : 173.6mg/kg - Computed Maximum Tolerated Dose (feed/water) : 12.9mg/kg - Computed Probability of Developmental Toxicity Potential : 1.000
36791-04-5	Ribavirin	C ₈ H ₁₂ N ₄ O ₅	- Computed Probability of Mutagenicity : 1.000 - Computed Probability of Carcinogenicity : 0.995 - Computed Rat Oral LD ₅₀ : 2.4g/kg - Computed Maximum Tolerated Dose (feed/water) : 98.8mg/kg - Computed Probability of Developmental Toxicity Potential : 1.000
298-46-4	Carbamazepine	C ₁₅ H ₁₂ N ₂ O	- Computed Probability of Mutagenicity : 0.000 - Computed Probability of Carcinogenicity : 0.996 - Computed Rat Oral LD ₅₀ : 2.6g/kg - Computed Maximum Tolerated Dose (feed/water) : 39.6mg/kg - Computed Probability of Developmental Toxicity Potential : 1.000
68-22-4	Norethindrone	C ₂₀ H ₂₆ O ₂	- Computed Probability of Mutagenicity : 0.000 - Computed Probability of Carcinogenicity : 0.978 - Computed Rat Oral LD ₅₀ : 5.1g/kg - Computed Maximum Tolerated Dose (feed/water) : 296.6.1mg/kg - Computed Probability of Developmental Toxicity Potential : 0.000

원인물질에 의한 발암유전자(예. 단백질활성화효소, G단백질과 조절자, 전사촉진인자 등)의 특정부위가 손상 되어 복제과정에서 돌연변이나 변이세포 형성되면서 발 암성을 나타낸다. 대사물질이 불안정하고 화학반응성이 큰 경우 단백질/핵산 등 고분자 화학물질과 결합하여 암 이 유발될 수도 있다. 발암물질을 분류할 때 주로 활성분 체, 변이원성물질 전구체, 에폭사이드, 방향족 아민, 아 조화합물(R-N=N-R), 다환방향족 탄화수소, N-니트로 소 화합물 등이 발암성의 화학구조에 따른다. 그 이외에 노스에티스테론(norethisterone)은 에스트로겐의 특징 의 일부를 가지고 있어 유방암 세포증식을 자극하는 잠 재적 유방암 위험인자로 여겨질 수 있다. 카르바마제핀 (Carbamazepine)은 모노 히드록시 유도체 발암성 연구 에서 암컷 생식기관(자궁 경부 및 질) 과립세포 종양발생

이 관찰되었다.

리바비린(ribavirin)은 항바이러스제로 주로 C형 간 염치료에 쓰이는 약물인데, 캐나다 유해물질 정보시스템 (WHMIS)에 생식독성이 있는 약물로 등록되어 있다. 위 와 같은 약물에 대해 미국병원약사회(ASHP)에서 항암 제, 항바이러스제, 호르몬제(에스트로겐 등), 면역억제제, 전립선치료제 일부를 위해약물(hazardous drug)로 분 류하였다. 스피노놀락톤(Spironolactone)은 체내 알도 스테론과 수용체간의 결합을 억제하여 체내 칼륨이온 보 존성 이뇨제이다. 메티마졸(methimazole)은 갑상샘 기 능항진증을 치료하기 위해 사용되는 약물이다. 스피노놀 락톤과 메티마졸은 임신부나 모유수유 중인 여성에게 위 해약물 범주에 속한다.

게다가, 발암성 약물은 주로 암 질환과 자가면역 질환

등 희귀난치성 질병에 사용되는 약제로 발암 위험의 문제를 국제암연구소(IARC, International Agency for Research on Cancer)에서 제시한 바 있다. 시클로포스파마이드(Cyclophosphamide)는 사용군에서 비사용군에 비하여 혈액중양, 방광암 발생위험이 증가한다(Fig. 7, Table 3). 특히, 메톡살렌(Methoxalen)은 건선, 습진, 백반증 및 일부 피부립프증을 치료하는 데 사용되는 약물이지만, 메톡살렌은 광활성화 이후에 유전독성으로 피부암 발암성에 포함이 되어있다. 항암제에 의한 노출은 암세포뿐만 아니라, 정상세포에도 영향을 미쳐 피부 및 점막자극, 발암성 등 장기적 독성을 나타낼 수 있기 때문이다.

비칼루타미드(bicalutamide)의 경우 마우스에는 항안드로겐 활성에 기인하는 양성중양 효과와 간세포중양의 발생이 보고되었으나, 사람에서는 유전독성 발암성을 시사하는 중양효과는 없었다(중간의 차이). 유럽의약품안전청(EMA)에 따르면, 아세트아미노펜(acetaminophen)의 경우에서 원료물질 자체보다는 생산과정에서 형성된 발암추정 가능물질(유전독성 발암물질)에 4-클로로아닐린(4-chloroaniline)이 검출된 적이 있다. 이번 연구에서 발암성으로 검출된 약물은 주로 암 질환과 자가면역질환 등 희귀난치성 질병에 사용되는 약제이기 때문에, 위해성은 있지만 완전히 사용을 중단하기는 어렵다. 일부 약물의 경우 더 안전하거나 생존에 이득이 있는 약제들로 대체되고 있다.

Table 3에서 보는 바와 같이, 장기간 간 독성과 유전독성, 발생독성 잠재성이 높아질 때 발암원성도 높아진다. 에스트라디올, 다카바진, 리바비린 약물이 해당된다. 한 예제로, 성호르몬인 스테로이드 약물인 에스트라디올(Estradiol)은 성장촉진 효과를 발휘하는 생리호르몬제로 알려져 있다. 성장촉진용 호르몬제는 일반적으로 물질대사과정 전반에 작용하여 체내 질소 총량을 증대시키고, 단백질 및 근육형성을 증대시키며 적혈구 생성을 활성화하여 성장을 촉진한다. 그러나, 에스트라디올의 산화적 대사산물로 인해, 염색체 이수배성(aneuploidy), 염색체 변형, 세포질 미세소기관 파괴, DNA 단일가닥절단, 유리기 라디칼 증가, 지질과 산화물 증가에 의해 유전독성을 유발한다. 에스트라디올은 국제암연구소에서 인간에 대한 충분한 근거를 가진 발암물질에 포함이 되어 있다. 게다가, 남성과 여성에게 불임을 초래할 수 있는 생식독성도 있는 것으로 알려져 있다.

위와 같이, TOPKAT[®] 독성모델의 범주에 속하는 화학구조를 갖는 약물유사-화합물 경우에는 변이원성, 발

암성, 발생독성 잠재성이 높은 원료물질에 대해서는 위해약물의 정보와 일치율이 높은 정확도로 독성효과의 예측 결과를 해석할 수 있었다.

4. 결론

의약품등을 통하여 체내에 유입되는 화합물 형태의 원료물질은 위장관 등을 통해 체내로 흡수된다. 약물들은 체내에서 대사체 중간체로 변화되어 배설되기도 하지만, 신장의 세포막을 통하여 확산이 되거나 재흡수 될 수 있다. 지용성 약물들은 소포체의 지용성 막에 비선택적으로 작용하여 효소들을 유도하기 때문에, 약물-약물 상호작용에 의해 간에서 약물 대사효소(CYP P450등)의 활성을 낮춰 해독과정을 방해할 수 있다.

생화학적 반응으로 모체약물(parent drug)은 산화, 환원 및 가수분해되어서 수용성 작용기인 하이드록시기(R-OH), 1차-아민기(R-NH₂), 티올(R-SH), 카르복실기(R-COOH)를 갖게 된다. 특히, 모체약물의 90% 산화반응은 사이토크롬 CYP P450 대사효소에 의해 불용성인 약물들이 수산화되어 수용성 유기 대사 중간체를 생성한다. 이러한 약물의 생화학 반응은 간에서 이뤄지며, 해독과정(detoxication)이라 할 수 있다. 약물의 대사효소는 모체약물의 구조적 특징(structural alert)을 인식하여 해독과정의 활성 프로파일을 나타내기 때문에, 간 독성/무독성을 예측하는 데, 결정인자는 모체약물이 갖고 있는 화학구조이다.

이번 연구에서 FDA 간 독성 지식 베이스(LTKB)에 축적된 간 독성 유-무(독성 1, 무독성 0)가 결정된 2,889개 데이터 목록 중에서 다양성이 높은 498개 약물유사-화합물을 대상으로 확률기반 베이지안 QSAR 모델로 간 독성/무독성을 예측하였다. 이때 179개 간 독성/319개 무독성을 가진 498개 약물유사-화합물의 2D-구조로부터 25개 간 독성과 관련된 fragment 분자표현자로 수치화하여 베이지안 QSAR 모델로 조건부 기준값(Bayesian score) >-1.698 , 간 독성)에 따라 간 독성 유-무를 결정하였다. 베이지안 QSAR 모델링 과정에서 교차검증과 5개 검증세트(5-fold cross validation)로 신뢰도를 평가하였을 때, 전체적인 일치율 90.6%와 양성예측율을 나타내는 민감도 97.2%를 고려할 때 예측 정확도가 높았다. 이뿐만 아니라, 319개 간 독성이 없는 약물유사-화합물의 공통적 fragment 정보를 살펴보았을 때, 간 대사 효소들에 의해 친수성 형태로 대사되어 체외로 배출되는 작용

기들은 86.9% 특이도로 무독성 원료물질로 예측되었다.

그러나 일부 약물은 체내에서 생화학적 반응을 거친 후, 활성을 가지게 되며 모체약물보다 반응성이 강한 대사중간체로 전환되어 간 독성 및 발암성을 유발한다. 179 개 간 독성을 갖는 약물유사-화합물에서도, CYP P450 대사 효소에 의한 해독을 방해하는 fragment를 갖는 나이트로소아민(Nitrosamine, $R^1-N-R^2-N=O$)의 2차 및 3차 아민과 아질산염이 산성 조건에서 나이트로소화($R^2-N-N=O$) 반응에 의해 나이트로사만류를 형성시킨다. 그래서, 동종효소 CYP P450 2B에 의하여 대사되어 반응성이 강하고 독성이 강한 중간체를 형성하여 DNA 및 단백질과 공유결합으로 세포독성과 발암성을 유발한다. 예를 들어, 항결핵제인 이소니아지드(isoniazid), 피리딘(pyridine), 피라진(pyrazine) 방향족 질소-헤테로고리 fragment를 갖는 약물유사-화합물은 CYP P450 2E1 효소에 의해 대사된다. 이 대사과정 중에서 반응성이 강한 중간체가 형성되어 DNA 및 단백질에 결합하여 변이원성(돌연변이)을 나타낸다. 또한, 간 독성 및 발암성에 연관성이 있다. 전신마취제인 할로세인(halothane)의 경우 CYP P450 2E1에 산화되어 트리플루오로아세틸(trifluoroacetyl) 유리기(free radical)를 형성하여 간 조직의 단백질과 결합한다. 이로부터 신생항원(neoantigen) 형성되어 간 독성을 나타낸다.

베이스안 QSAR 간독성 모델은 약물유사-화합물의 구조 특성 및 작용점으로 수송과 관련된 물리화학적 성질만을 고려하였지만, 간 대사 효소들의 기질 특성이 잘 반영되어 정확도가 높았다. 그러나, 간 대사 효소들의 작용기전을 고려하지 않아서 위양성률(1-특이도)도 존재했는데, 특히, 해열진통제인 아세트아미노펜은 CYP P450 2E1에 의해 산화되어 반응성이 큰 N-acetylbenzoquinone imine (NAPQI) 대사중간체를 생성하여 간 독성과 발암 가능성이 있다고 예측하였다. 하지만, 고용량 투여 시 간 손상을 일으켜 간 기능장애 우려가 있다. 아세트아미노펜의 발암 가능성의 예측결과는 과학적 근거자료로 위해성 평가가 더 필요하여 수용하지 않았다.

약물대사에 영향을 미치는 내·외적 요소인 종(species)에 따른 차이, 나이와 성별에 따른 대사 기능의 차이, 환경적 차이에 따른 개인의 유전적 요인, 다른 약물 상호작용은 간 독성 예측 모델인 베이지안 QSAR 모델의 범주에 포함되지 않았다. 따라서, 정량적인 구조-독성관계(QSAR)를 예측하는 데, 모델의 신뢰도는 기반이 되는 실험적 실측치(training set)의 일관성과 모델의 주요 결정인자(간 독성 경우 모체약물의 화학구조)에서 설명이 가

능한 모델의 범주(category)라고 할 수 있다. 이는 주어진 데이터를 학습하여 새로운 사례에 대한 확률기반 베이지안 모델 성능을 일반화(generalization)하거나 TOPKAT[®] 독성모델처럼 로지스틱 선형판별분석은 종속변수의 값을 직접 예측하기보다는 두 범주(발암성/비발암성 등) 중 특정한 범주에 속할 확률을 기계학습으로 모델링하는 접근이기 때문이다. 결과적으로, 분석대상 데이터 종류에 따라 모델의 선택, 모델 적합, 해석 등 적용 방법이 달라져야 한다.

종에 따른 약물의 대사경로와 대사 속도에 많은 차이는 유전적으로 대사에 관여하는 효소의 기능 및 발현이 다르기 때문이다. 내외적 요인으로 간 독성 효과를 해석하는 데에는 많은 요인을 고려해야 한다. 그래서 다음 단계의 연구에서는 간 독성 예측 베이지안 QSAR 모델로부터 간 독성/무독성을 예측한 후, 약인성간손상의 손상기전을 초기에 발견하고 진단 할 수 있는 바이오마커 정보 [48]를 도입하여 해석하고자 한다. 가능한 바이오마커에는 글루타메이트 탈수소효소, 케라틴18, 소르비톨탈수소효소, 글루타치온 S-전이효소, 담즙산, 사이토크롬 CYP P450 대사 효소, 지방산 결합단백질, miR-122등 유전자 진단과 간 독성 예측 QSAR 모델에 오픈믹스 기술을 적용하여 생체 매개 함수를 접목하고자 한다.

References

- [1] S. Ekins, "Progress in computational toxicology", *Journal of Pharmacological and Toxicological Methods*, Vol.69, No.2, pp.115-140, Dec. 2013. DOI: <https://doi.org/10.1016/j.vascn.2013.12.003>
- [2] P. Pound, M. B. Bracken, "Is animal research sufficiently evidence based to be a cornerstone of biomedical research?", *British Medical Journal*, Vol.348, pp.g3387, May 2014. DOI: <https://doi.org/10.1136/bmi.g3387>
- [3] M. Clark, T. S. Hartmann, "A big data approach to the concordance of toxicity of pharmaceutical in animals and humans", *Regulatory Toxicology and Pharmacology*, Vol.96, pp.94-105, Jul. 2018. DOI: <https://doi.org/10.1016/j.yrtph.2018.04.018>
- [4] C. Rovida, T. Hartung, "Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals- a report by the transatlantic think for toxicology(tt(4))", *ALTEX*, Vol.26, No.3, pp.187-208, 2009.
- [5] M. Vinken, "3Rs toxicity testing and disease modeling, projects in the european horizon 2020 research and innovation program", *Experimental and Clinical*

- sciences Journal*, Vol.19, pp.775-784, June 2020.
DOI: <https://doi.org/10.17179/excli2020-2450>
- [6] M. J. Mone, G. Pallocca, S. E. Escher, T. Exner, M. Herzler, S. H. Bennekou, H. Kamp, E. D. Kroese, M. Leist, T. S. Hartman, B. van de Water, "Setting the stage for next-generation risk assessment with non-animal approaches: the EU-ToxRisk project experience", *Archives of Toxicology*, Vol.94, pp.3581-3592, Sep. 2020.
DOI: <https://doi.org/10.1007/s00204-020-02866-4>
- [7] M. Vinken, E. Benfenati, F. Busquest, J. Castell, D. A. Clevert, T. M. de Kok, H. Dirven, E. Fritsche, L. Geris, R. Gozalbes, T. Hartung, D. Jennen, R. Jover, H. Kandarova, N. Kramer, C. Krul, T. Luechtefeld, R. Masereeuw, E. Roggen, S. Schaller, T. Vanhaecke, C. Yang, A. H. Piersma, "Safer chemicals using less animals: kick-off of the European ONTOX project", *Toxicology*, Vol. 458, pp.152846, June, 2021.
DOI: <https://doi.org/10.1016/j.tox.2021.152846>
- [8] R. de Kanter, M. Monshouwer, D. K. F. Meijer, G. M. M. Groothuis, "Precision-cut organ slices as a tool to study toxicity and metabolism of xenobiotics with special reference to non-hepatic tissues", *Current Drug metabolism*, Vol.3, No.1, pp.39-59, Feb. 2002.
DOI: <https://doi.org/10.2174/1389200023338071>
- [9] G. Pallocca, M. J. Moné, H. Kamp, M. Luijten, B. V. de Water, M. Leist, "Next-generation risk assessment of chemicals-rolling out a human-centric testing strategy to drive 3R implementation: The RISK-HUNT3R project perspective", *ALTEX*, Vol.39, No.3, pp.419-426, Apr. 2022
DOI: <https://doi.org/10.14573/altex.2204051>
- [10] A. M. Richard, R. Huang, S. Waidyanatha, P. Shinn, B. J. Colins, I. Thillainadarajah, C. M. Grulke, A. J. Williams, R. R. Lougee, R. S. Judson, K. A. Houck, M. Shobair, C. Yang, J. F. Rathman, A. Yasgar, S. C. Paules, J. R. Bucher, C. F. Austin, R. J. Kavlock, R. R. Tice, "The Tox21 10K compound library: collaborative chemistry advancing toxicology", *Chemical Research in Toxicology*, Vol.34, No.2, pp.189-216, Feb. 2021.
DOI: <https://doi.org/10.1021/acs.chemrestox.0c00264>
- [11] J. Jaeseong, K. Donghyeon, C. Jinhee, "Application of ToxCast/Tox21 data for toxicity mechanism-based evaluation and prioritization of environmental chemicals: perspective and limitations", *Toxicology in Vitro*, Vol.84, 105451, Oct 2022.
DOI: <https://doi.org/10.1016/j.tiv.2022.105451>
- [12] FDA(2021), "Advancing New Alternative Methodologies at FDA"
<https://www.fda.gov/media/144891>
- [13] J. Jiang, C. D. Pieterman, G. Ertaylan, R. L. M. Peeters, T. M. C. M. de Kok, "The application of omics-based human liver platforms for investigating the mechanism of drug-induced hepatotoxicity in vitro", *Archives of Toxicology*, Vol.93, pp.3067-3098, Aug. 2019.
DOI: <https://doi.org/10.1007/s00204-019-02585-5>
- [14] C. H. Beckwitt, A. M. Clark, S. Wheeler, D. L. Taylor, D. B. Stolz, L. Griffith, A. Wells, "Liver 'organ on a chip'", *Experimental Cell Research*, Vol.363, No.1, pp.15-25, Feb. 2018.
DOI: <https://doi.org/10.1016/j.yexcr.2017.12.023>
- [15] M. Karabicici, S. Akbari, O. Ertern, M. Gumustekin, E. Erdal, "Human liver organoid models for assessment of drug toxicity at the preclinical stage", *Endocr Metab Immune Disord Drug Targets*, Vol.23, No.14, pp. 1713-1724, June, 2023.
DOI: <https://doi.org/10.2174/1871530323666230411100121>
- [16] M. Yoshizawa, M. Sugimoto, M. Tanaka, Y. Sakai, M. Nishikawa, "Computational simulation of liver fibrosis dynamics", *Scientific Reports*, Vol.12, No.1, pp.14112, Aug. 2022.
DOI: <https://doi.org/10.1038/s41598-022-18123-w>
- [17] E. Asilar, J. Hemmerich, G. F. Ecker, "Image based liver toxicity prediction", *Journal of Chemical Information and Modeling*, Vol.60, No.3, pp.1111-1121, Mar. 2020.
DOI: <https://doi.org/10.1021/acs.icim.9b00713>
- [18] P. Rana, M. D. Aleo, X. Wen, S. Kogut, "Hepatotoxicity reports in the FDA adverse event reporting system database: A comparison of drugs that cause injury via mitochondrial or other mechanisms", *Acta Pharmaceutica Sinica B*, Vol.11, No.12, pp.3857-3868, May 2021.
DOI: <https://doi.org/10.1016/j.apsb.2021.05.028>
- [19] S. Babai, L. Auclert, H. L. Louët, "Safety data and withdrawal of hepatotoxic drugs", *Thérapie*, Vol.76 No.6, pp.715-723, Nov-Dec. 2021.
DOI: <https://doi.org/10.1016/j.therap.2018.02.004>
- [20] S. D. Dimitrov, R. Diderich, T. Sobanski, T. S. Pavlov, G. V. Chankov, Y. H. Karakolev, S. G. Temelkov, R. A. Vasilev, K. D. Gerova, C. D. Kuseva, N. D. Todorova, A. M. Mehmed, M. Rasenberg, O. G. Mekenyan, "QSAR Toolbox-workflow and major functionalities", *SAR and QSAR in Environmental Research*, Vol. 27, No.3, pp.203-219, Mar 2016.
DOI: <https://doi.org/10.1080/1062936X.2015.1136680>
- [21] E. Mombelli, P. Pandard, "Evaluation of the OECD QSAR toolbox automatic workflow for the prediction of the acute toxicity of organic chemicals to fathead minnow", *Regulatory Toxicology and Pharmacology*, Vol.122, pp.104893, June 2021.
DOI: <https://doi.org/10.1016/j.yrtph.2021.104893>
- [22] C. Johnson, L. T. Anger, R. Benigni, D. Bower, F. Bringezu, K. M. Crofton, M. T. D. Cronin, K. P. Cross, M. Dettwiler, M. Frericks, F. Melnikov, S. Miler, D. W. Roberts, D. S. A. Roncaglioni, E. L. Piparo, R. R. Tice, C. Zwickl, G. J. Myatt, "Evaluating confidence in toxicity assessment based on experimental data and *in silico* predictions", *Computational Toxicology*, Vol. 21, pp.100204, Nov. 2021.
DOI: <https://doi.org/10.1016/j.comtox.2021.100204>
- [23] C. M. Ellison, J. C. Madden, M. T. D. Cronin, S. J. Enoch, "Investigation of the verhaar scheme for predicting acute aquatic toxicity: Improving predictions obtained

- from Toxtree ver.2.6”, *Chemosphere*, Vol. 139, pp.146-154, Nov. 2015.
DOI: <https://doi.org/10.1016/j.chemosphere.2015.06.009>
- [24] U. S. EPA(2020), “User’s guide for T.E.S.T(version 5.1) (Toxicity Estimation Software Tool): A program to estimate toxicity from molecular structure”
<https://www.epa.gov/sites/default/files/2016-05/documents/600r16058.pdf>
- [25] R. Dhingra, M. Malhotra, V. Sharma, T. R. Bhardwaj, N. Dhingra, “Exploration of novel 5 α -reductase inhibitor for benign prostatic hyperplasia by 2D/3D QSAR, cytotoxicity Pre-ADME and docking studies”, *Current Topics in Medicinal Chemistry*, Vol. 18, No.32, pp.2816-2834, 2018.
DOI: <https://doi.org/10.2174/1568026619666190119145959>
- [26] M. Hassan, J. P. Bielawski, J. C. Hempel, M. Waldman, “Optimization and visualization of molecular diversity of combinatorial libraries”, *Molecular Diversity*, Vol.1-2, pp.64-74, Oct. 1996.
DOI: <https://doi.org/10.1007/BF01718702>
- [27] H. Hong, S. Thakkar, M. Chen, W. Tong, “Development of decision forest models for prediction of drug-induced liver injury in humans using a large set of FDA-approved drugs”, *Scientific Reports*, Vol.7, No.1, pp.17311, Dec. 2017.
DOI: <https://doi.org/10.1038/s41598-017-17701-7>
- [28] S. Thakkar, T. Li, Z. Lu, L. Wu, R. Roberts, W. Tong, “Drug-induced liver injury severity and toxicity (DILIst): binary classification of 1279 drugs by human hepatotoxicity”, *Drug Discovery Today*, Vol.25, No.1, pp.201-208, Jan. 2020.
DOI: <https://doi.org/10.1016/j.drudis.2019.09.022>
- [29] J. H. Hoofnagle, J. Serrano, J. E. Knoben, V. J. Navarro, “LiverTox: a website on drug-induced liver injury”, *Hepatology*, Vol.57, No.3, pp.873-874, Mar. 2013.
DOI: <https://doi.org/10.1002/hep.26175>
- [30] S. Thakkar, M. Chen, H. Fang, Z. Liu, R. Roberts, Weida Tong, “The liver Toxicity Knowledge Base (LTKB) and drug-induced liver injury (DILI) classification for assessment of human liver injury”, *Expert Review of Gastroenterology & Hepatology*, Vol.12, No.1, pp.31-38, Jan. 2018.
DOI: <https://doi.org/10.1080/17474124.2018.1383154>
- [31] S. N. Ilemo, D. Barth, O. David, F. Quessette, M. A. Weisser, D. Water, “Improving graphs of cycles approach to structural similarity of molecules”, *PLOS One*, Vol.14, No.12, pp.e0226680, Dec. 2019.
DOI: <https://doi.org/10.1371/journal.pone.0226680>
- [32] R. L. M. Robinson, D. Geatches, C. Morris, R. Mackenzie, A. G. P. Maloney, K. J. Roberts, A. Moldovan, E. Chow, K. Pencheva, D. Ramesh, 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, Nov. 2019.
DOI: <https://doi.org/10.1021/acs.jcim.9b00601>
- [33] Dassault Systems BIOVIA Discovery Studio modeling environment release 22.1, San Diego 2022.
<https://biovia-discovery-studio-2022-client.software.informer.com>
- [34] W. J. Egan, K. M. Merz Jr, J. J. Baldwin, “Prediction of drug absorption using multivariate statistics”, *Journal of Medicinal Chemistry*, Vol.43, No.21, pp.3867-3877, Oct. 2000.
DOI: <https://doi.org/10.1021/jm000292e>
- [35] A. Cheng, K. M. Merz Jr, “Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationship”, *Journal of Medicinal Chemistry*, Vol.46, No.17, pp.3572-3580, Aug. 2003.
DOI: <https://doi.org/10.1021/jm020266b>
- [36] A. Cheng, S. L. Dixon, “in silico models for the prediction of does-dependent human hepatotoxicity”, *Journal of Computer-Aided Molecular Design*, Vol.17, No.12, pp.811-823, Dec. 2003.
DOI: <https://doi.org/10.1023/b:jcam.0000021834.50768.c6>
- [37] X. Xia, E. G. Maliski, P. Gallant, D. Rogers, “Classification of kinase inhibitors using a Bayesian model”, *Journal of Medicinal Chemistry*, Vol.47, No.18, pp.4463-4470, Aug. 2004.
DOI: <https://doi.org/10.1021/jm0303195>
- [38] M. Marzo, E. Benfenati, “Classification of a native Bayesian fingerprint model to predict reproductive toxicity”, SAR and QSAR in Environmental Research, Vol.29, No.8, pp.631-645, Aug. 2018.
DOI: <https://doi.org/10.1080/1062936X.2018.1499125>
- [39] 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>
- [40] B. Bhatarai, D. M. Wilson, A. K. Parks, E. W. Carney, P. J. Spencer, “Evaluation of TOPKAT, Toxtree, and Derek Nexus in silico models for ocular irritation and development of a knowledge-based fragment to improve the prediction of severe irritation”, *Chemical Research in Toxicology*, Vol. 29, No.5, pp.810-822, May. 2016.
DOI: <https://doi.org/10.1021/acs.chemrestox.5b00531>
- [41] M. S. Elnaggar, A. M. Elissawy, F. S. Youssef, M. Kicsák, T. Kurtán, A. N. B. Singab, R. Kalscheuer, “Austalide derivative from marine-derived *Aspergillus* sp. and evaluations of its cytotoxic and ADME/TOPKAT properties”, *Royal Society of Chemistry advances*, Vol.13, No.24, pp.16480-16487, June 2023.
DOI: <https://doi.org/10.1039/d3ra02632a>
- [42] K. Roy, I. Mitra, “Electrotopological state atom (E-state) index in drug design, QSAR, property prediction and toxicity assessment”, *Current Computer Aided Drug Design*, Vol.8, No.2, pp. 135-158, Jun. 2012.
DOI: <https://doi.org/10.2174/157340912800492366>

- [43] M. J. Prival, "Evaluation of the TOPKAT system for predicting the carcinogenicity of chemicals", *Environmental and Molecular Mutagenesis*, Vol.37, No.1, pp.55-69, 2001
DOI: [https://doi.org/10.1002/1098-2280\(2001\)37:1<55::aid-em1006>3.0.co;2-5](https://doi.org/10.1002/1098-2280(2001)37:1<55::aid-em1006>3.0.co;2-5)
- [44] A. Iorga, L. Dara, N. Kaplowitz, "Drug-induced liver injury: cascade of events leading to cell death, apoptosis or necrosis", *International Journal of Molecular Sciences*, Vol.18, No.5, pp.1018, May 2017.
DOI: <https://doi.org/10.3390/ijms18051018>
- [45] T. Sahota, M. Danhof, O. D. Pasqua, "Pharmacology-based toxicity assessment: towards quantitative risk prediction in humans", *Mutagenesis*, Vol.31, No.3, pp.359-374, May. 2016.
DOI: <https://doi.org/10.1093/mutage/gev081>
- [46] FDA(2020) "Clinical drug interaction studies-cytochrome P450 enzyme and transporter mediated drug interactions guidance for industry"
<https://www.fda.gov/media/134581>
- [47] H. K. Björnsson, E. S. Björnsson, "Drug-induced liver injury: Pathogenesis, epidemiology, clinical features, and practical management", *European Journal of Internal Medicine*, Vol.97, pp.26-31, Mar. 2022.
DOI: <https://doi.org/10.1016/j.ejim.2021.10.035>
- [48] S. Fu, D.Wu, W. J. J. Long, C. Jia, T. Zhou, "Molecular biomarkers in drug-induced liver injury: challenges and future perspectives", *Frontiers in Pharmacology*, Vol.10, pp.1667, Jan. 2020.
DOI: <https://doi.org/10.3389/fphar.2019.01667>

채 한 화(Han-Ha Chai)

[정회원]



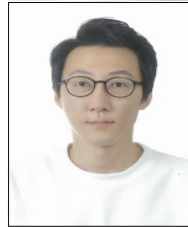
- 2010년 11월 ~ 현재 : 농촌진흥청 국립축산과학원 농업연구사

<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

진 영 배(Yeung Bae Jin)

[정회원]



- 2020년 9월 ~ 현재 : 경상국립대학교 수의과대학 실험동물의학실 협실
- 2017년 11월 ~ 2020년 8월 : 한국 생명공학연구원 국가영장류 센터장

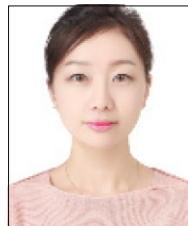
- 2014년 7월 ~ 2020년 8월 : 한국 생명공학연구원 국가 영장류센터 선임연구원

<관심분야>

독성병리학, 실험동물 이용 질환모델 개발

임 다 정(Dajeong Lim)

[정회원]



- 2023년 9월 ~ 현재 : 충남대학교 동물바이오시스템과학과 조교수
- 2022년 1월 ~ 2023년 9월 : 대통령 소속 농어업·농어촌특별위원회
- 2007년 11월 ~ 2023년 9월 : 농촌진흥청 국립축산과학원 농업 연구관

<관심분야>

가축 집단 유전체, 생명정보 적용 유전자 선발