



ORIGINAL ARTICLES

농약의 GHS 분류를 위한 QSAR 기반 독성 예측 및 H코드 도출

이은지¹ · 김시은^{2,3} · 박재현^{2,3} · 조유미⁴ · 박수진⁴ · 신현길^{2,3*}

¹국가독성과학연구소 첨단예측연구본부 생체신호연구센터, ²국가독성과학연구소 첨단예측연구본부 예측모델연구센터
³UST 과학기술연합대학원대학교 인체 및 환경 독성학, ⁴농촌진흥청 국립농업과학원 농산물안전성부 독성위해성평가과

QSAR Model for Toxicity Prediction and H Code Assignment for GHS Classification of Pesticides

Eun Ji Lee¹, Si Eun Kim^{2,3}, Jaehyeon Park^{2,3}, Youmi Jo⁴, SooJin Park⁴, and Hyun Kil Shin^{2,3*}¹Biological Signal Research Center, Advanced Prediction Research Department²Prediction Model Research Center, Advanced Prediction Research Department, Korea Institute of Toxicology³Human and Environmental Toxicology, University of Science and Technology⁴Toxicity and Risk Assessment Division, Department of Agro-Food Safety & Crop Protection, National Institute of Agricultural Science, Rural Development Administration

(Received on August 25, 2025. Revised on September 16, 2025. Accepted on September 25, 2025)

Abstract The Globally Harmonized System (GHS) for chemical safety assessment uses hazard statement codes (H-codes) to standardize the classification of chemical hazards. H-codes are primarily determined based on data obtained from human or animal experiments; therefore, there are limitations in the rapid assessment of new chemicals and the application of alternative testing methods to assign H codes. To overcome these limitations, this study proposes a Quantitative Structure-Activity Relationship (QSAR) model approach for assigning H-codes. Based on a chemical's molecular structure, the QSAR model predicts chemical toxicity. Although QSAR models offer a low-cost, high-efficiency way to assess the hazards of a large number of chemicals, they are restricted by predictive uncertainty and limited applicability domain. Therefore, this study explores an approach based on reliable predictions by excluding results that fall outside the QSAR model's applicability domain. This research introduces an algorithm that integrates QSAR predictions with GHS classification criteria to derive H-codes. Through the exclusion of certain H-codes that are difficult to determine with prediction values alone according to the definition of the endpoint, this analysis specifically focused on a limited set of H-codes that can be obtained through the QSAR model. When this algorithm was applied to the active ingredients of two pesticides (Novaluron and Docusate) with sparse hazard information, Novaluron showed that the QSAR-derived H-code was consistent with its actual H-code. In case of Docusate, a potential hazard was predicted by the model; however, no toxicity evidence was found in the PubChem database. This outcome demonstrates both the utility and the inherent limitations of QSAR predictions.

Key words: Applicability domain, *in silico*, GHS classification

서론

매년 수많은 새로운 화학물질이 개발되고 있으며, 사람의 건강과 환경 보호를 위해 이들의 안전성 평가가 필수적이다. GHS(Globally Harmonized System of Classification and

Labelling of Chemicals)는 화학물질의 유해성과 위험성을 통일된 기준을 바탕으로 한 분류 체계이다(Winder et al., 2005). GHS는 동일한 형태의 경고표지를 통해 화학물질의 위험성을 표준화된 방식으로 전달한다. 특히 GHS에서 정의하는 H 코드(Hazard statement codes)는 화학물질의 특정 위험성을 나타내는 핵심 지표로, 화학물질의 안전한 취급과 관리를 위한 기본 정보를 제공한다. H 코드는 H200(물리적 위험성), H300(건강 유해성), H400(환경 유해성)으로 구분

*Corresponding author
E-mail: hyunkil.shin@kitox.re.kr

되며, 각각의 코드는 특정한 위험성 유형과 정도를 나타낸다(UN, 2023). H 코드는 인체에 대한 유해성 자료 혹은 동물 시험 결과를 기준으로 결정된다. 따라서 H 코드 결정 과정에는 대체시험법 적용이 어렵다는 한계가 존재한다.

QSAR(Quantitative Structure-Activity Relationship)는 화학 물질의 분자 구조와 생물학적 활성 간의 정량적 관계를 수학적 모델로 표현하는 방법론이다(Li et al., 2025). 화학물질의 구조적 특성으로부터 독성학적 종말점을 예측함으로써, 실험적 검증 이전에 화학물질의 잠재적 위험성을 평가하는데 활용된다. QSAR 모델은 비교적 짧은 시간과 낮은 비용으로 대량의 화학물질에 대한 위험성 평가가 가능하다는 장점이 있지만, 예측의 불확실성과 예측 결과에 대한 해석이 불가능한 경우가 빈번하게 발생해서 대체시험법으로 활용하기에는 한계가 있다(Madden et al., 2020). QSAR 모델의 예측 결과를 활용하기 위해서는 적용범위(applicability domain) 분석을 통해 예측 결과의 신뢰도를 파악하는 것이 필수적으로 요구된다. 적용범위란 QSAR 모델이 신뢰할 만한 예측을 제공할 수 있는 화학물질의 구조 분포를 의미하며, 이는 모델 개발에 사용된 훈련 데이터셋의 구조 다양성에 의해 결정된다(Vasilev and Atanasova, 2025). 예측 대상 화학물질이 모델의 적용범위를 벗어날 경우 예측 불확실성이 크게 증가하므로, 예측 값을 근거로 의사 결정을 하기 전에 적용범위 분석을 통해 신뢰할 수 있는 예측 결과를 얻었는지 평가하는 것이 중요하다. 특히 규제 목적으로 QSAR 모델을 활용할 경우에는 QSAR assessment framework에서 제시하는 원칙에 따라 예측결과의 투명성과 명확성을 확보해야 한다(OECD, 2023).

지금까지 QSAR를 이용한 독성 예측에 관한 연구는 주로 특정 독성학적 종말점에 대한 개별적인 모델 개발에 초점이 맞춰져 있었다. 일부 연구에서는 다중 분류(multi-class classification) 접근법을 통해 여러 독성 카테고리를 동시에 예측하려는 시도가 있었으나, H 코드와 같은 표준화된 위험성 분류 체계와 직접적으로 연결되는 연구는 부족한 실정이다. 기존 연구의 주요 한계점으로는 다음과 같은 것들이 있다. 첫째, 대부분의 연구가 단일 독성 종말점에 국한되어 있어 통합적인 위험성 평가가 어렵다. 둘째, 모델의 예측 성능이 화학물질의 구조적 다양성과 데이터의 질에 크게 의존한다. 셋째, 예측 결과의 해석 가능성과 규제 당국의 수용성 측면에서 보완이 필요하다(Cherkasov et al., 2014).

본 연구에서는 화학물질의 분자 구조 정보만을 활용한 QSAR 모델의 예측값을 바탕으로 H 코드를 분류하는 작업을 수행했다. 일반적으로 H 코드는 실험값을 통해 분류되지만, 본 연구에서는 예측 값을 보수적으로 활용하여 도출된 H코드가 의미를 가질 수 있는지 확인하고자 했다. QSAR 연구 대부분은 모델을 개발하고 검증하는데 집중되어 있다. 규제 관점에서 QSAR 모델을 활용하는 방법도 연구가 필요한 중

요한 부분이기 때문에 본 연구에서는 QSAR 모델을 통해 H 코드를 산출 가능한지 실험해 봄으로써 QSAR의 규제 활용성을 보완하고자 했다. QSAR 모델은 실험값이 없는 경우 사전 예측에 활용된다. QSAR 모델이 사용되는 상황을 가정해서 유해성 정보가 부족한 물질 리스트를 확보하고 예측을 수행하는 것으로 실험을 설계했으며, 수집한 물질 리스트 중에서 농약 원제는 2종(Novaluron, Docusate) 확인되었다. 농약 원제의 QSAR 예측 값을 바탕으로 H코드를 도출하고 이를 검증함으로써 QSAR 모델의 규제적 활용 가능성을 검토했다.

재료 및 방법

유해성 데이터가 희소한 화합물 리스트

ECHA와 EPA 데이터베이스(ECHA, 2023; EPA, 2024)를 활용하여 유해성 정보가 부족해 예측으로 유해성 확인이 필요한 화합물 목록을 확보하였다. 이 목록에서 무기화합물과 혼합물은 제외했는데, 이는 QSAR 모델이 단일 유기 화합물을 대상으로 개발되었기 때문이다(Cherkasov et al., 2014). 또한 QSAR 모델을 사용하려면 구조 정보를 SMILES 또는 MOL 파일 형태로 확보해야 하므로, 구조 식별이 불가능한 화합물도 목록에서 제외하였다. 그 결과, 유기화합물인면서 구조 정보가 명확한 418종의 화합물 목록을 구축하였다. 해당 리스트를 농약 원제 구조와 비교한 결과, 농약 2종(Novaluron, Docusate)이 포함되어 있는 것으로 확인되었다.

QSAR 예측

VEGA는 IRFMN(Instituto di Ricerche Farmacologie Mario Negri)연구소에서 개발한 QSAR 프로그램이다(IRFMN, 2024). 본 연구에서는 VEGA 1.2.4 버전을 이용해서 예측을 수행했다. 화학물질의 등록에 요구되는 실험 값 중에서 VEGA 1.2.4를 통해 예측할 수 있는 종말점은 총 30개가 있었다. 30개의 종말점을 예측할 수 있는 모델들은 전체 72개로 확인되었다(Table 1).

적용범위 분석

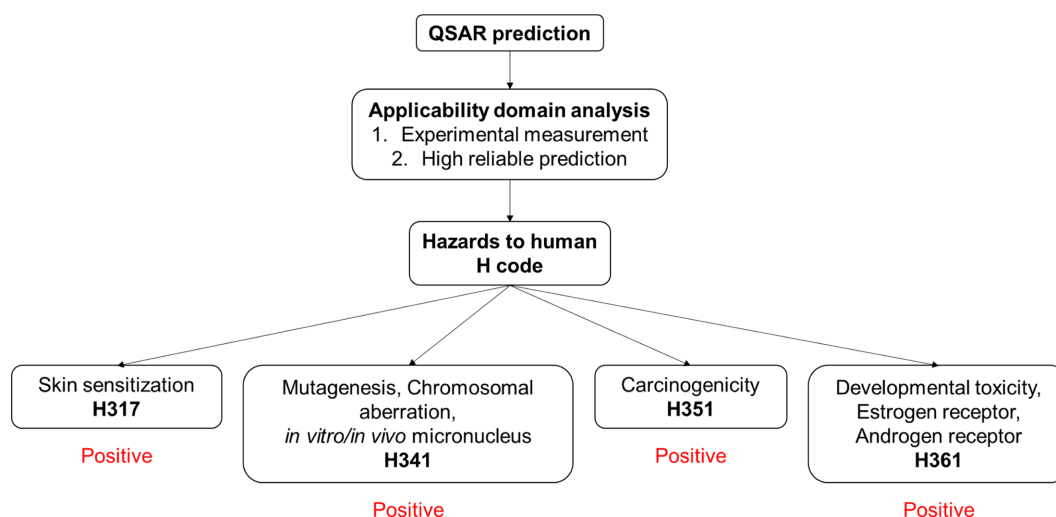
VEGA에서는 적용범위를 4가지로 분류하고 있다: EXPERIMENT(훈련데이터에 이미 실험값 있음), GOOD(적용범위에 들어올 가능성이 높음), MODERATE(적용범위 내에 있는지 확실하지 않음), LOW(적용범위에 들어오지 않음). 4가지 중에서 EXPERIMENT, GOOD인 경우에만 믿을 수 있는 값을 제시하고 있다고 판단하여 H 코드 도출에 활용하였다.

QSAR 기반 H code 도출

QSAR 모델의 예측 결과는 실험을 통해 얻은 결과에 비해

Table 1. List of QSAR models available in VEGA with endpoints required for registration

Endpoint category	Endpoint	Models	Descriptions
Mutagenicity, Carcinogenicity	Ames test	6	Salmonella typhimurium
	Chromosomal aberration	1	Chinese hamsters
	Micronucleus activity	2	in vitro / in vivo
Acute toxicity	Carcinogenicity	10	Oral / inhalation
	LD ₅₀	1	Rat
Developmental toxicity	Developmental/reproductive Toxicity	2	Homo sapiens
	Zebrafish embryo AC ₅₀	1	Danio rerio
Sensitization	Binding assay	6	Endocrine disruptor, Estrogen receptor, Androgen receptor, Thyroid receptor
Irritation	Skin sensitization	6	CBA mice
Chronic toxicity (repeated toxicity)	Skin/eye irritation	6	No QMRF documents were found
	NOAEL	2	Tested in rat or other rodents
Bioaccumulation	LOAEL	1	Q No QMRF documents were found
	BCF	4	Fish (Oncorhynchus mykiss, Cyprinus Carpio, and salmonids)
Acute daphnia toxicity	EC ₅₀	4	Daphnia Magna
Chronic daphnia toxicity	NOEC	1	Daphnia Magna
Acute fish toxicity	LC ₅₀	8	Guppy (Poecilia reticulata), rainbow trout, fathead minnow (pimephales promelas), Medaka (Oryzias latipes)
Chronic fish toxicity	NOEC	1	Medaka (Oryzias latipes)
Acute algae toxicity	EC ₅₀	3	Rhapidocelis subcapitata
Chronic algae toxicity	NOEC	1	Rhapidocelis subcapitata
Partition coefficient	Octanol/water	3	logP
Solubility	Water	1	-LOG ₁₀ (mol/L)
Biodegradability	Ready biodegradability	1	OECD 301C modified MITI (I) Test
	Hydrolysis as a function of pH	1	OECD 111

**Fig. 1.** Algorithm to derive human hazard H code. Only four H codes can be reliably obtained from QSAR predictions such as skin sensitization (H317), mutagenicity (H341), carcinogenicity (H351), and developmental toxicity (H361).

정확도가 낮다는 점을 감안해서 신뢰도가 높은 경우만을 선별하여 H코드 산출을 진행하였다. 분자 구조로부터 예측

값을 얻었더라도, 해당 화합물이 모델의 적용범위 밖에 있는 것으로 확인되면, 해당 예측 값을 H코드 산출에 사용하지

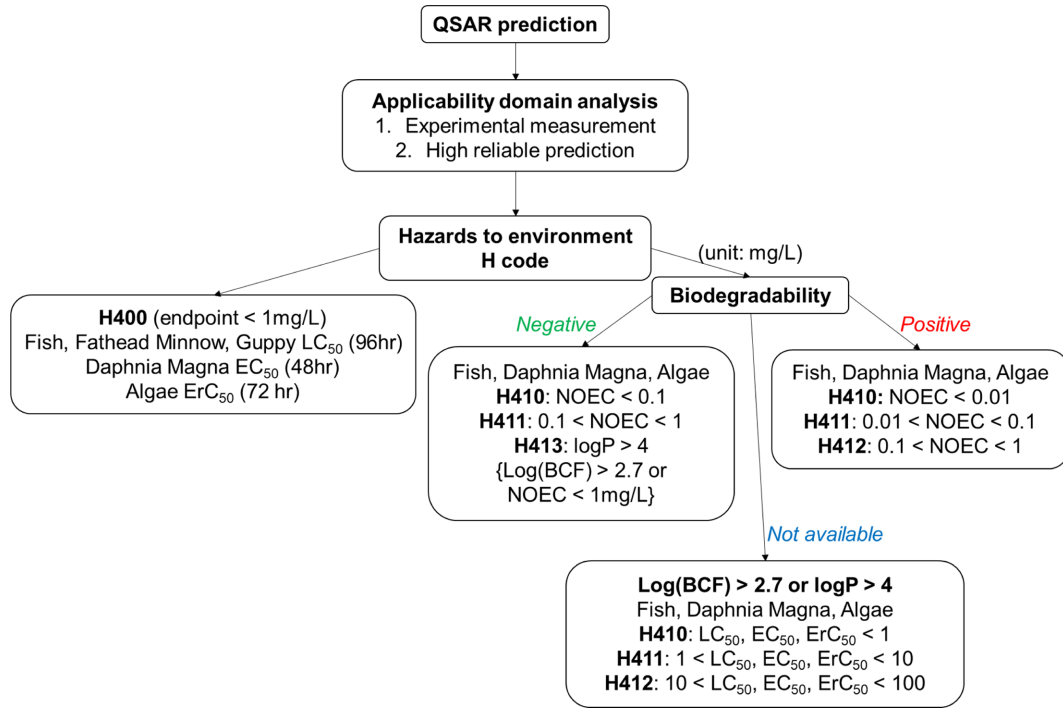


Fig. 2. Algorithm to derive environmental hazard H code. In total, five H codes were can be obtained from QSAR predictions such as H400, H410, H411, H412, and H413.

않았다. 예측을 통해 얻은 H 코드는 PubChem 데이터베이스에서 확인된 H 코드와 비교함으로써 결과의 정확성을 검증했다. 인체 유해성과 환경 유해성 각각에 대해 H 코드를 도출하기 위한 알고리즘을 정의했는데 인체유해성(Fig. 1)과 환경유해성(Fig. 2)에 대한 H 코드 분류 기준이 각각 상이하기 때문에, 두 영역에 대해 알고리즘을 별도로 정리했다.

결과 및 고찰

QSAR 예측 결과

QSAR 모델로 예측을 수행했을 때, 입력한 구조가 모델의 적용범위에 들어오면 예측 결과를 신뢰할 수 있다. OECD QSAR validation guideline에서는 QSAR 모델 개발 시 지켜야할 5가지 원칙을 제시하고 있는데, 3번째 원칙이 적용 범위 분석에 대한 내용이다(OECD, 2014). QSAR 모델은 훈련데이터의 구조 패턴과 활성 값 간의 연관 관계를 학습하게 된다. 그래서 QSAR 모델의 예측은 훈련데이터와 비슷한 경우 신뢰도가 높다. 모델의 적용범위는 훈련데이터의 구조 공간(chemical space)과 밀접한 연관이 있기 때문에 훈련데이터의 분포를 바탕으로 적용범위 분석을 하게 된다. QSAR 예측 모델의 규제 활용을 위해 발간된 가이드문서인 QSAR assessment framework에서도 적용범위의 중요성을 강조하고 있다(OECD, 2023). 신뢰할 수 있는 예측 값인지 분석하기 위해서는 입력한 분자구조가 모델의 적용범위에 포함되는지가 중요하기 때문이다. 적용범위를 벗어나는 예

측 결과는 신뢰할 수 없기 때문에 적용범위 내에 들어오는 예측 결과를 기반으로 H 코드를 도출하였다. 유해성 정보가 없는 물질들은 대부분의 예측값에서 적용 범위 밖에 있는 것으로 확인되었다. VEGA는 적용 범위 분석 시 다음 기준을 활용하여 평가한다. 1) 입력 물질과 훈련 데이터 간의 구조적 유사성. 2) 유사 물질의 예측 정확도. 3) 유사 물질의 실험값과 입력 물질의 예측값 간의 일치도. 4) 입력 물질의 구조 패턴이 훈련 데이터에 존재하는지 여부. 5) 훈련 데이터의 기술자(descriptor) 값 범위 내에 입력 물질의 기술자 값이 포함되는지. 이 다섯 가지 기준은 결국 구조적 유사성을 다양한 방식으로 확인하는 과정이라 할 수 있다. 따라서 적용 범위 밖에 있다는 것은 입력된 물질의 구조가 모델 훈련에 사용된 물질의 구조와 유사하지 않음을 의미한다. 유해성 정보가 없는 물질들은 단순히 실험값이 없을 뿐만 아니라, 분자 구조의 분포도 달라 예측이 어려운 것으로 확인되었다. QSAR 모델은 실험 정보가 없는 물질의 실험값을 예측하는 용도로 사용된다(Fig. 3). 이때 모델 학습에 사용된 물질과 구조적 유사성이 높은 경우에만 예측 정확도가 보장된다. 즉 새로운 구조 예측은 가능하지만, 훈련 데이터와 지나치게 상이한 구조를 입력하는 경우에는 예측 값이 틀릴 가능성이 크게 높아진다. 그렇기 때문에 QSAR 모델을 활용하기 전에는 적용범위 분석이 필수적으로 요구된다. QSAR 모델에서 얻은 예측 값이 큰 오차를 가질 경우 모델에 입력한 분자구조가 모델의 적용범위 밖에 있기 때문에 오차가 발생할 가능성이 높다. 그래서 예측한 값이 얼마나 큰 오차

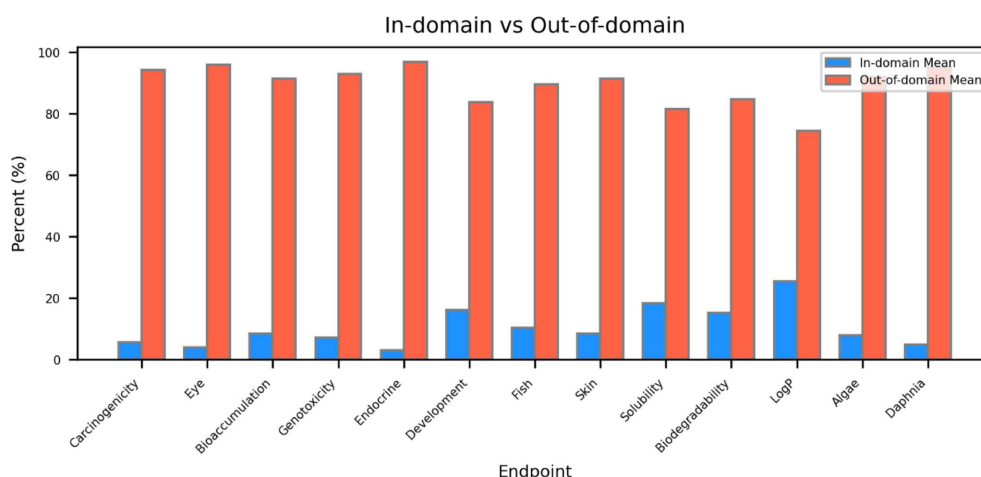


Fig. 3. Applicability domain analysis results on the collected molecules. Toxicity information of the compounds in the dataset was not available in public database. These molecules were also absent in the QSAR model training dataset; therefore, most of the molecules were out of the applicability domain.

Table 2. Human hazard H code for each endpoint and possibility of using QSAR predictions to derive H codes

Endpoint	H code	Possibility O/X (description)
Oral acute toxicity	H300, H301, H302, H303	O
Aspiration hazard	H304, H305	X (Human data required)
Acute dermal toxicity	H310, H311, H312	X (No model available due to lack of data)
Skin corrosion/irritation	H314, H315	X (Detailed information required such as severity of damage, exposure, exposure time, observation time)
Skin sensitization	H317	O (QSAR model based on OECD TG429 data required)
Eye corrosion/irritation	H318, H319, H320	X (Detailed information on the number of affected animals and the severity of damage)
Acute inhalation toxicity	H330, H331, H332	X (No model available due to lack of data)
Respiratory sensitization	H334, H335 (Respiratory irritation), H336 (Anesthetic effect)	X (No model available due to lack of data)
Germ cell mutagenicity	H340, H341	H340 Unfeasible (epidemiological survey and germ cell data required), H341 feasible with somatic cell data
Carcinogenicity	H350, H351	H350 Unfeasible (human data required), H351 Feasible with animal data
Reproductive toxicity	H360, H361, H362	H360 Unfeasible (human data required), H361 Feasible with animal data, H362 Unfeasible (first- and second-generation data required)
Specific target organ toxicity	H370, H371, H372, H373	X (human data and the number of exposures required)
Water solubility	H413	
Octanol/water partition coefficient		
Biodegradability		
Hydrolysis		
Bioaccumulation (aquatic species)	H410, H411, H412, H413	O (Acute and chronic aquatic toxicity)
Algae (acute, chronic)		
Daphnia (acute, chronic)		
Fish (acute, chronic)		

를 갖는지 확인하기 전에 입력한 물질이 적용 범위 내에 들어오는지 여부를 사전에 분석을 해야 QSAR 모델의 정확도를 올바르게 평가할 수 있다.

QSAR 기반 H 코드 도출

GHS 분류에서는 건강 유해성과 환경 유해성의 심각성에 따라 H코드를 부여하게 된다. 건강 유해성에서 분류하는 중

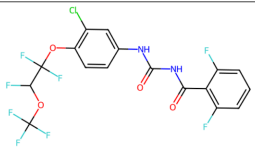
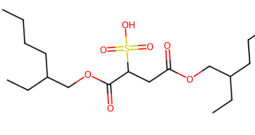
Molecular structure of pesticides (Name of pesticide and CAS No.)	H-code (PubChem)	Predicted H-code
 Novaluron (116714-46-6)	H400, H410	H400, H410
 Docusate (10041-19-7)	No evidence	H351, H361

Fig. 4. H code prediction results on two pesticides. In case of Novaluron, available H codes from PubChem were matched with the predicted H codes. On the contrary, predicted H codes for Docusate cannot be evaluated due to absence of evidence on human hazard according to PubChem.

말점은 급성독성, 피부 부식성/자극성, 심한 눈 손상/자극성, 호흡기 또는 피부 과민성, 생식세포 변이원성, 발암성, 생식독성, 특정 표적장기 독성-1회 노출, 특정 표적장기 독성-반복 노출, 흡인 유해성이다. 환경 유해성은 수생환경유해성에 따라 급성1, 만성1, 2, 3, 4로 나뉜다. 급성과 만성을 나눌 때는 어류, 갑각류, 조류 혹은 기타 수생식물에 대한 실험 값 중 하나를 이용해서 유해성을 분류하게 된다. GHS 분류는 실험 값을 기준으로 하게 되어 있다. 일반적으로 예측 값으로부터 H코드를 도출하지는 않는다. 하지만 이번 연구에서는 QSAR 예측 값으로부터 얻어낸 H코드가 활용 가치가 있는지 검증하는 차원에서 예측 결과로부터 H코드를 도출하였다. H코드를 도출하는 기준이 사람에게 대한 데이터를 요구하거나 동물 시험에서 관찰되는 증상을 바탕으로 결정되는 경우에는 QSAR 모델을 통해 H 코드를 도출하는 것이 불가능했다. 이에 따라 먼저 QSAR 모델의 종말점이 갖는 의미를 분석(Table 1) 하고 예측된 종말점으로부터 확보할 수 있는 H 코드를 정리하는 과정을 수행하였다(Table 2). 이렇게 도출된 제한적인 H 코드 중에서도, QSAR 예측 결과가 적용 범위 내에 포함되는 경우만을 선별하여 H 코드를 선별함으로써, 상당히 보수적인 접근 방식을 적용하였다. 농약 2종에 대한 예측 결과로부터 H 코드를 확인한 결과, Novaluron에서는 PubChem 데이터베이스에 등록된 H 코드와 동일한 결과가 예측되었다. 반면, Docusate의 경우는 QSAR예측 값에서 H코드가 도출되었지만, PubChem데이터베이스에서는 해당 독성에 대한 근거가 없는 것으로 확인되었다(Fig. 4).

QSAR 모델은 크게 회귀 모델과 분류 모델로 구분된다. 회귀 모델은 실험을 통해 측정된 값을 정량적으로 예측하는 모델이다. 회귀 모델의 예시로는 LD₅₀을 예측하는 모델이 있다. 분자 구조를 입력하면 그 물질의 LD₅₀ 값(mg/kg/day)을 예측해준다. 반면에 분류 모델은 실험 결과를 정성적으

로 나누는 방법이다. 분류 모델에 분자 구조를 입력하면 ‘독성 있음(positive)’ 혹은 ‘독성 없음(negative)’으로 결과값이 출력된다. H코드를 부여하는 작업 자체는 분류 모델의 영역으로 볼 수 있다. 본 연구에서는 H 코드의 근거가 되는 실험 값을 QSAR 모델로 예측한 후, 해당 예측 값을 바탕으로 H코드를 도출하였다. 즉, 최종적으로 확인된 H코드는 모델에서 직접 산출된 결과 값이 아니라, 예측 값을 기반으로 분류 기준에 따라 도출된 것이다. H코드를 예측하는 작업은 회귀 모델에서 얻은 예측 값을 활용하는 방식보다는, 분류 모델을 통해 직접 H 코드의 유무를 판단하는 접근이 더 높은 정확도를 제공할 가능성이 높아 보인다.

감사의 글

본 연구는 농촌진흥청 국립농업과학원 농업과학기술 연구개발사업(과제번호: RS-2024-00400007)의 지원에 의해 수행되었습니다.

Author Information and Contributions

Eun Ji Lee, Biomimetic Research Center, Korea Institute of Toxicology, Senior Researcher, <https://orcid.org/0009-0002-8048-6655>, Writing-original draft and conducting the experiment.

Si Eun Kim, Human and Environmental Toxicology, University of Science and Technology, <https://orcid.org/0009-0001-9999-4855>, Researcher, Data analysis and visualization.

Jaehyeon Park, Human and Environmental Toxicology, University of Science and Technology, <https://orcid.org/0009-0004-7459-809X>, Researcher, Data analysis and Review.

Youmi Jo, Toxicity and Risk Assessment Division, National Institute of Agricultural Sciences, <https://orcid.org/0009-0008-6733-7181>, Senior Researcher, Data analysis and Review.

Soojin Park, Toxicity and Risk Assessment Division, National Institute of Agricultural Sciences, <https://orcid.org/0000-0002-2522-3185>, Senior Researcher, Literature review and investigation.

Hyun Kil Shin, Prediction Model Research Center, Korea Institute of Toxicology, Senior Researcher, <https://orcid.org/0000-0003-3665-0841>, Conceptualization, study design and Project administration.*

이해상충관계

본 논문의 저자들은 이해상충관계가 없습니다.

Literature Cited

- Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, *et al.*, 2014. QSAR modeling: where have you been? where are you going to? *J. Med. Chem.* 57(12):4977-5010.
- ECHA, 2023. ECHA chemicals database. <https://echa.europa.eu/information-on-chemicals>
- EPA, 2024. CompTox chemicals dashboard v2.5.3. <https://comptox.epa.gov/dashboard/>
- IRFMN, 2024. VEGA QSAR v.1.2.4. <https://www.vegahub.eu/portfolio-item/vega-qsar/>
- Li J, Zhao T, Yang Q, Du S and Xu L, 2025. A review of quantitative structure-activity relationship: The development and current status of data sets, molecular descriptors and mathematical models. *Chemometrics Intellig. Lab. Syst.* 256:105278.
- Madden JC, Enoch SJ, Paini A and Cronin MTD, 2020. A review of in silico tools as alternatives to animal testing: principles, resources and applications. *Alternatives to Laboratory Animals.* 48(4):146-172.
- OECD, 2014. Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. In.
- OECD, 2023. (Q)SAR assessment framework: guidance for the regulatory assessment of (quantitative) structure activity relationship models, predictions, and results based on multiple predictions.
- UN, 2023. Globally harmonized system of classification and labelling of chemicals (GHS) 10th revised edition. In.
- Vasilev B and Atanasova M, 2025. A (comprehensive) review of the application of quantitative structure-activity relationship (QSAR) in the prediction of new compounds with anti-breast cancer activity. In, *Applied Sciences.*
- Winder C, Azzi R and Wagner D, 2005. The development of the globally harmonized system (GHS) of classification and labelling of hazardous chemicals. *J. Hazard. Mater.* 125(1): 29-44.

● ● 농약의 GHS 분류를 위한 QSAR 기반 독성 예측 및 H코드 도출

이은지¹ · 김시은^{2,3} · 박재현^{2,3} · 조유미⁴ · 박수진⁴ · 신현길^{2,3*}

¹국가독성과학연구소 첨단예측연구본부 생체신호연구센터, ²국가독성과학연구소 첨단예측연구본부 예측모델연구센터
³UST 과학기술연합대학원대학교 인체 및 환경 독성학, ⁴농촌진흥청 국립농업과학원 농산물안전성부 독성위해성평가과

요약 화학물질 분류 표기 세계조화시스템(GHS)은 화학물질의 안전성을 평가하기 위한 분류 체계로 화학적 위험성을 표준화한 H 코드를 사용해서 화학물질을 분류한다. H 코드는 주로 인체 또는 동물 실험에서 얻은 데이터를 기반으로 결정된다. 현재 기준에서는 대체시험법 결과로 H 코드를 지정하기에는 어려움이 있다. 본 연구는 QSAR 예측을 GHS 분류 기준과 통합하여 H-코드를 도출하는 알고리즘을 소개한다. QSAR 모델은 화학물질의 분자 구조를 기반으로 물질의 독성을 예측한다. QSAR 모델은 저비용으로 다수의 화학물질 위험성을 예측할 수 있는 방법이지만, 모델의 적용 범위를 벗어나는 경우 예측 결과의 정확도가 현저히 낮아지는 문제가 있다. 그래서 QSAR 모델의 적용 범위에 포함되는 신뢰할 수 있는 예측 값을 기반으로 H 코드를 도출했다. H 코드 도출 기준에 의거하여 QSAR 예측 값만으로는 결정할 수 없는 경우는 본 연구에서 제외했다. 제안한 알고리즘을 유해성 정보가 부족한 두 가지 농약 원제(Novaluron과 Docusate)에 적용한 결과 Novaluron의 경우 QSAR로 도출된 H 코드가 실제 H 코드와 일치하는 것으로 나타난다. Docusate의 경우 예측 결과로 H 코드가 도출되었지만 PubChem 데이터베이스에서 검증한 결과 H 코드가 확인되지는 않았다. 이번 분석 결과는 QSAR 예측의 유용성과 내재된 한계를 동시에 보여주고 있다.

색인어: 적용 범위, 인실리코, 화학물질 분류 표기 세계조화시스템

● ●