



GHS Skin Corrosion/Irritation Prediction Report

: Benzene

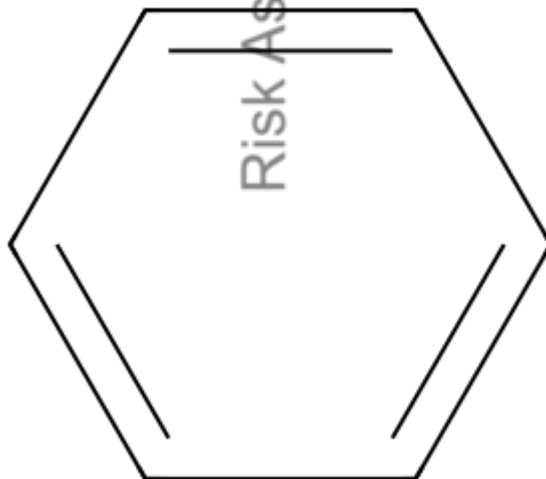
1. General Information

Date of QPRF	2026-03-23
Author	(주) 위해관리연구소
Contact	02-715-0983

2. Substance Identity

CAS Number	71-43-2
Chemical Name	Benzene
MolWeight	78.11
Structural Formula	C6H6
SMILES	c1ccccc1
InChIKey	UHOVQNZJYSORNB-UHFFFAOYSA-N
InChI	InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H

2D Struture



3. Model and Software Used

Model Used

The model implemented in the TOXFENCE QSAR platform is a supervised (Q)SAR classification model that predicts GHS skin corrosion/irritation hazards. The model was trained on a curated dataset of 4,896 compounds categorized by developmental toxicity outcomes. The prediction endpoint is a binary hazard classification (toxicant vs. non-toxicant), and this prediction reports the corresponding class for the target substance. The algorithm is gradient-boosted decision trees (XGBoost) trained on a curated dataset with experimental labels. Input features include EState indices, molecular size/topology, charge and fragment-exposed surface terms, polarity, drug-likeness, and fingerprint density. A QSAR Model Reporting Format (QMRF)/validation dossier is available for this deployment, providing detailed documentation of the training data, algorithm, applicability considerations, and validation results.

• Software Platform

TOXFENCE QSAR is a proprietary, commercially licensed QSAR software platform that provides various toxicity-prediction models, including a Skin Corrosion/Irritation model aligned with GHS/CLP. The present prediction was generated using the platform's default model settings.

• Reference to QMRF

A QMRF for the TOXFENCE Skin Corrosion/Irritation model version 1.0 is available and indicates adherence to the OECD principles for QSAR validation. No specific QMRF code is provided in this report. The model's QMRF dated 2025 documents the algorithm descriptor set and validation statistics.

• Endpoint

Skin Corrosion/Irritation. The model predicts whether a chemical is classified as skin corrosive/irritant (or not classified) under GHS/CLP, using curated labels derived from in vivo dermal studies and validated alternative (in vitro) skin corrosion/irritation tests.

4. Prediction

• Results — Skin Corrosion & Irritation Models

Result	Probability	Source
Positive	100.0%	XGBoost (Skin Corrosion/Irritation)

Read-Across Similarity Results

Result	Probability	Source
Positive	100.0%	DB

Skin Corrosion & Irritation SA Decision Tree

Structural Alert (SA) Image



Risk Assessment Management & Consulting

Overall Conclusion — Skin Corrosion/Irritation

Result	Probability	Confidence	Source	AD
Positive (Skin Corrosive/Irritant)	100.0%	High	DB	IN

Decision Rationale

Experimental DB label exists; read-across and overall conclusion follow DB value.

Basis for Probability Estimation

Model/Method	1	2	3
Model Probability			
Read-Across Similarity Probability			
Skin Corrosion/Irritation SA Decision Tree Probability			
Overall Conclusion Probability			

Model Probability

The supervised classifier's estimated probability for the positive (skin corrosion/irritation) class from molecular representations, with optional pre/post calibration.

Read-Across Similarity Probability

A similarity-based probability computed from labeled neighbors (positive/negative) in a reference set and converted via a normalized monotonic function; it is omitted when no labeled neighbors exist. When labeled neighbors exist only in one class (all positive or all negative) or when the maximum similarity is highly imbalanced, the formulation can yield probabilities near 0 or 1.

Skin Corrosion/Irritation SA Decision Tree Probability

A quantitative probability obtained by mapping the rule/decision-tree's qualitative outcome via a predefined mapping and monotonic transformation, optionally tempered to avoid extreme values.

Overall Conclusion Probability

A weighted combination of Model/Tree/Sim probabilities, with weights set and normalized based on the applicability domain (AD) and strength of evidence; when experimental labels are available, they take precedence.

Structural Similarity Components

The diagram illustrates structural similarity components through three columns of chemical structures. Each column shows a pair of molecules: a top structure with red atoms and a bottom structure with grey atoms. The first column shows a benzene ring and a cyclopentadiene ring. The second column shows a benzene ring and a cyclopentadiene ring with a different substitution pattern. The third column shows a benzene ring and a cyclopentadiene ring with a different substitution pattern. Below the structures is a table with three columns corresponding to the structures above.

Component 1	Component 2	Component 3

Similarity method (fingerprints, feature-space, and MCS ensemble)

The model quantifies analogue proximity using a multi-view ensemble that combines binary fingerprint similarities,

descriptor feature-space similarities mapped to the 0–1 range, and a maximum common substructure (MCS) ratio. The fingerprint layer uses ECFP4 Tanimoto, ECFP6 Tanimoto, RDKit topological Tanimoto. In the standardized descriptor space (StandardScaler), Mahalanobis and Euclidean distances are converted to smooth inverse-distance similarities on 0–1, and cosine similarity is rescaled to 0–1. Scaffold overlap is represented by the fraction of the query covered by the MCS. Each similarity is transformed to a percentile within the candidate set to place heterogeneous measures on a comparable scale. The final score (CombinedSim) is an equal-weight blend of the maximum percentile and the mean percentile across all similarity views, rewarding very strong evidence from any single view while also requiring overall concordance. This ensemble integrates fragment-, scaffold-, and descriptor-level evidence and aligns with analogue-based reasoning and applicability-domain assessment.

5. Model Input for Prediction

5.1 Input Structure

The prediction was based on the substance's 2D structure (SMILES and/or SDF). The structure file was imported, standardized, and featurized into 2D descriptors (and, if enabled, circular fingerprints) for use by a machine-learning classifier.

Key input details:

- Key input details.

1) Structure standardization

Where applicable, salts/solvents were handled and the largest fragment was retained; charge and aromaticity were standardized, and implicit hydrogens were processed prior to featurization.

2) Stereochemistry

A 2D representation was used. Stereochemical information may be preserved in the structure but was not explicitly modeled in 3D.

3) Tautomerism / Ionization.

When no material changes are expected under the assessment conditions, the major (default) structure was used; when relevant, tautomer/ionization normalization was applied and documented.

4) Feature generation

Descriptors were computed automatically by the software from the standardized 2D structure; no external user-supplied descriptors were used unless otherwise stated.

5.2 Model/Software Settings

Model type

A supervised machine-learning classifier with standard preprocessing was used for prediction.

AD/similarity support

Applicability Domain (AD) assessment combined data curation and outlier handling with analogue review. Analogue selection used an ensemble similarity (integrating fingerprint-, feature-space-, and MCS-based evidence) that is operated separately from the learning algorithm.

6. Considerations for Regulatory Use

6.1 Regulatory Purpose

This (Q)SAR prediction was generated to support regulatory decision-making for the skin corrosion/irritation endpoint. In practice, it is used for prioritization and internal screening to identify potentially high-risk substances early, before initiating more resource- and time-intensive testing. This approach is consistent with the intent of the UN GHS/CLP classification framework and OECD guidance, and aligns with international trends that encourage

the use of alternative approaches, such as QSAR, to reduce animal testing. Importantly, the prediction is not intended to serve as a stand-alone basis for a final conclusion; it should be interpreted as one line of evidence within a Weight-of-Evidence (WoE) framework alongside existing experimental data and analogue information.

6.2 Approach to Regulatory Interpretation

Within a Weight-of-Evidence (WoE) or Integrated Approaches to Testing and Assessment (IATA) framework, the model prediction is used together with skin corrosion/irritation–relevant structural alerts as mechanistic support. In particular, when structural alerts pertinent to skin corrosion/irritation are triggered and the model prediction is positive, confidence in a hazard concern is strengthened. This indicates that the predicted toxicity is consistent with the molecule's structural features and is therefore an important consideration for regulatory interpretation. Conversely, the absence of structural alerts or a negative model prediction does not mean the substance is “safe.” Regulatory guidance emphasizes that a lack of alerts from a (Q)SAR model or expert system is insufficient to demonstrate non-hazard. Accordingly, this prediction is interpreted conservatively and is not used as the sole basis for hazard classification. Instead, the QSAR output is treated as supporting evidence and integrated with existing test data, formulation/exposure information such as pH, buffering capacity, and concentration, read-across analogues, and other relevant information to reduce uncertainty and improve the reliability of regulatory decision-making.

6.3 Regulatory Interpretation of Result

When the substance is assessed to be within the model's Applicability Domain (AD), it lies within the chemical space on which the model was trained and is regarded as reliable, thereby supporting the validity of the prediction. In addition, if a review of the Top-5 analogues selected by the CombinedSim ensemble similarity (Tanimoto and standardized feature-space similarity transforms) shows that most are classified for skin corrosion/irritation, this provides convergent evidence supporting a positive prediction. The observation that structurally similar chemicals exhibit skin corrosive/irritant effects strengthens the plausibility that the assessed substance may present a comparable hazard. Furthermore, when skin corrosion/irritation–relevant structural alerts (SAs) are triggered, the predicted toxicity is mechanistically consistent with specific structural motifs (toxicophores), which further increases confidence in the concern signal. However, the absence of an SA must not be taken as evidence of non-hazard; because SAs can yield both false positives and false negatives, they are interpreted conservatively. In light of the foregoing, the substance may be treated as a screening-level concern, and additional information—such as targeted *in vitro* assays, refined read-across, and literature corroboration—may be recommended. Conversely, when the model predicts negative, the substance is within AD, and the empirical data for the CombinedSim Top-5 analogues are predominantly unclassified/non-irritant, the model prediction and analogue evidence are mutually consistent. If, in this situation, no skin corrosion/irritation SA is triggered or the alerts are deemed non-relevant, the substance may be considered low concern at this stage, and de-prioritization or deferral of additional testing may be reasonable. Nevertheless, if discordance is observed (e.g., the model is negative while a relevant SA is triggered, or some analogues are experimentally positive), a conservative re-evaluation is warranted. Appropriate follow-ups include re-examining or expanding the analogue set, reassessing formulation pH, buffering capacity, and concentration and the skin-exposure scenario, evaluating potential metabolites and microspecies (pH) relevance, and conducting targeted *in vitro* assays before integrating the outcome back into the WoE. Finally, a negative prediction alone must not be used to claim a test waiver. The (Q)SAR output must be interpreted together with existing relevant information and integrated into a Weight-of-Evidence (WoE) assessment for the final regulatory conclusion. Only when external evidence is sufficiently consistent with the prediction should de-prioritization or test deferral be considered.

6.4 Conclusion (Adequacy for Regulatory Conclusion)

Based on the ensemble prediction (100.0%, Confidence: High), this substance shows potential skin hazard. Classification: Not classified. Applicability Domain: IN.

Appendices

Disclaimer · Prediction Results Report

The results are estimates subject to model assumptions, input data quality, parameter choices, and applicability domain (AD) constraints; they are not experimental evidence, performance guarantees, medical/toxicological diagnoses, or regulatory advice. Uncertainty ranges and classifications reflect model behavior and may change with new data or methods. Users are solely responsible for independent verification and for any decisions made on the basis of these outputs, including additional testing and expert review where appropriate. To the extent permitted by law, the authors and provider disclaim all warranties, express or implied, and accept no liability for direct, indirect, incidental, or consequential losses arising from the use of this report. Intended use: screening, prioritization, and decision support—not as the sole basis for regulatory or safety determinations.

Risk Assessment Management & Consulting