TOXFENCE ver.1

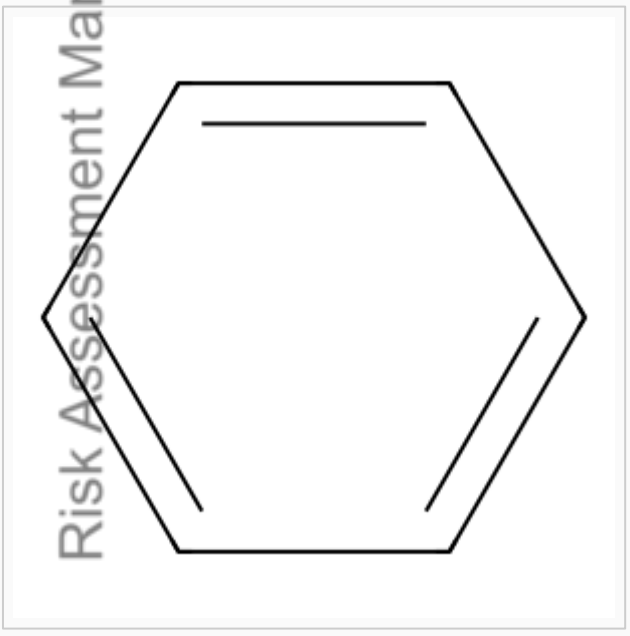
Genotoxicity Prediction Report (AMES, In Vitro CA, In Vivo MN)

1. General Information

Author: ㈜ 위해관리연구소

Contact: 02-715-0983

2. Substance Identity

CAS Number	71-43-2
Chemical Name	Benzene
Structural Formula	C6H6
SMILES	`c1ccccc1`
Molecular Weight	78.11
2D Structure	

3. Model and Software Used

Model Used: TOXFENCE QSAR Platform Genotoxicity Model version 1.0 is a machine learning based QSAR binary classification model designed to predict potential hazard across genotoxicity test endpoints, including AMES with S9 positive and S9 negative, in vitro chromosomal aberration, and in vivo micronucleus. The platform builds independent ensemble models for each endpoint. For each endpoint, an ensemble of four algorithms Random Forest, XGBoost, LightGBM, and Histogram based Gradient Boosting is combined using a soft voting approach based on predicted probabilities to output both the prediction probability and a binary class result positive or negative. In addition, a Comprehensive module uses the endpoint level prediction results as inputs to provide an overall genotoxicity conclusion for the target substance.

A QSAR Model Reporting Format dossier is available and includes model development and validation information, including reported classification performance such as sensitivity and specificity. In the deployed environment, the decision threshold is configured to balance sensitivity and specificity.

Software Platform: TOXFENCE QSAR (version 1.0) is a proprietary, commercially licensed QSAR software platform that provides multiple genotoxicity prediction models, including endpoint specific models for AMES, in vitro chromosomal aberration, and in vivo micronucleus. The predictions were generated using the platform's default model settings.

Reference to QMRF: TOXFENCE QMRF for the Genotoxicity model (version 1.0), indicating the model's adherence to OECD principles. No specific QMRF code was provided in this report, but the model's QMRF (dated 2025) documents its algorithm, descriptors, and validation statistics.

4. Prediction

- **Endpoint:** Genotoxicity. The model predicts whether a chemical is genotoxic (positive) or non genotoxic (negative) based on genotoxicity test endpoints, including AMES, in vitro chromosomal aberration, and in vivo micronucleus assays.
- **Results — Genotoxicity (AMES, In Vitro CA, In Vivo MN)**

Prediction Results

OECD TG	Probablility	Result	Others	AD Flag / Confidence
TOXFENCE AMES_S9+	1.000	Negative	Test organisms: Salmonella typhimurium Test type: Bacterial Reverse Mutation Assay Prediction approach: QSAR	IN / High
TOXFENCE AMES_S9-	1.000	Negative	Test organisms: Salmonella typhimurium Test type: Bacterial Reverse Mutation Assay Prediction approach: QSAR	IN / High
TOXFENCE In vitro CA	1.000	Positive	Test organisms: Mammalian cell lines according to the OECD TG 473 Test type: in vitro Chromosomal Aberration test Prediction approach: QSAR	IN / High

OECD TG	Probablility	Result	Others	AD Flag / Confidence
TOXFENCE In vivo MN	1.000	Positive	Test organisms: Mice, rats, or other mammalian species Test type: in vivo Micronucleus Test Prediction approach: QSAR	IN / High

Comprehensive Result

Result	AD Flag	Confidence

Structural Alert (SA) Comment

[illegible]

Basis for Probability Estimation

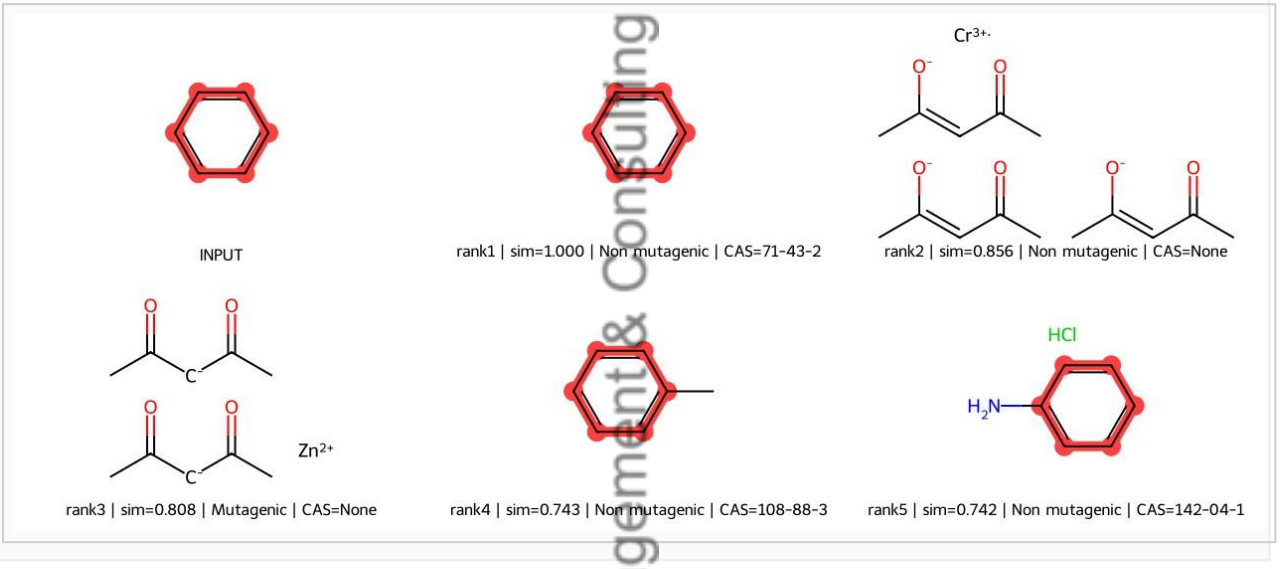
Model Probablility: The positive class probability is estimated using numerical molecular features calculated from the chemical structure. As defined in the model documentation, the input features are composed of structure based molecular descriptors. The machine learning model uses these descriptors to generate, for each genotoxicity endpoint, the probability that the target substance is classified as positive.

Genotoxicity SA Decision Tree (AMES, In Vitro CA, In Vivo MN) **Probablility:** 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.

Structural Similar Compounds

Similarity method: The Mahalanobis distance (MD) serves as an extension of the Euclidean distance, incorporating correlations between the X properties. This metric is determined by measuring the distance of a data point from the center of the training dataset. A higher MD value indicates lower reliability in the corresponding prediction.

Bacterial Reverse Mutation Test with S9 mixture (AMES +S9)



Substance Name	Experimental Value	Similarity Score
BENZENE	Non mutagenic	1.000
CHROMIUM ACETYLACETONATE	Non mutagenic	0.856
ZINC 2,4-PENTANEDIONE COMPLEX	Mutagenic	0.808
TOLUENE	Non mutagenic	0.743
ANILINE HYDROCHLORIDE	Non mutagenic	0.742

Bacterial Reverse Mutation Test without S9 mixture (AMES -S9)

In Vitro Mammalian Chromosomal Aberration Test (TG473 CA)

Substance Name	Experimental Value	Similarity Score
Benzaldehyde	Genotoxic	0.735
1,2-xylene	Non genotoxic	0.732

Mammalian Erythrocyte Micronucleus Test (TG474 MN)

Substance Name	Experimental Value	Similarity Score
Benzene	Genotoxic	1.000
aniline	Genotoxic	0.748
Chlorobenzene	Genotoxic	0.735
Acetobenzone	Non genotoxic	0.717
Phenol	Genotoxic	0.716

5. Model Input for Prediction

5.1 Input Structure

The prediction was based on the substance's 2D structure. The structure file (SMILES string and/or SDF) was imported into the software and converted into a molecular graph composed of nodes (atoms) and edges (bonds). Key input details:

Structure standardization. Where applicable, salts/solvents were removed and the largest fragment was retained; charges were normalized; aromaticity was standardized; implicit hydrogens were handled prior to graph generation.

Stereochemistry. The prediction used a 2D graph. Stereochemical information was either ignored or preserved without explicit 3D treatment, depending on model settings. If the substance has no stereocenters/geometric isomers, no special handling was required.

Tautomerism / Ionization. When no material tautomerism or ionization is expected under the assessment conditions, the default (major) structure was used. If relevant, the major microspecies (e.g., at pH 7.4) or the tool's tautomer normalization rules were applied and documented.

Node features. Atom type, atomic degree, total hydrogen count, hybridization, formal charge, number of radical electrons, aromaticity, and (if enabled) SMARTS-based toxicophore alerts.

Edge features. One-hot encodings of bond order/type (single, double, triple, aromatic) and bond direction where applicable.

Coordinates. No 3D coordinates or conformer generation were used; a geometry-independent 2D representation was employed.

Descriptor/embedding calculation. From the graph input, descriptors/embeddings were computed automatically by the software; no user-supplied external descriptors were used unless otherwise stated.

5.2 Model/Software Settings

Custom Settings: None – prediction generated with the platform's default model settings.

Comments on Settings: No metabolism simulation or additional expert options enabled. Similarity (RBF) was used only for analogue selection/AD support and is separate from the model's learning algorithm.

6. Considerations for Regulatory Use

6.1 Regulatory Purpose

This (Q)SAR prediction was generated to support regulatory decision making for the genotoxicity endpoint. In practice, it is used for prioritization and internal screening to help identify potentially high risk substances early, before committing resources and time to more intensive testing [1]. This approach aligns with international practice. As data requirements for genotoxicity continue to increase under regulatory frameworks such as REACH, organizations including the OECD encourage the use of alternative approaches such as QSAR to reduce animal testing. Importantly, the prediction is not intended as a stand alone conclusion. It should be interpreted as one line of evidence within a Weight of Evidence framework, together with available experimental data and analogue toxicity 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 structural alert SA information as mechanistic support. In particular, when genotoxicity relevant structural alerts are triggered for the substance and the model prediction is positive, confidence in a genotoxicity concern is strengthened. This indicates that the predicted outcome is consistent with structural features associated with genotoxic mechanisms, which is an important consideration for regulatory interpretation. Conversely, the absence of structural alerts or a negative model prediction should not be interpreted as evidence that the substance is safe. Regulatory guidance emphasizes that a lack of alerts from a QSAR model or expert system alone 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 to be integrated with other information sources, thereby reducing uncertainty and strengthening 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 falls within the chemical space on which the model was trained. The prediction can therefore be interpreted as relatively more reliable, supporting the validity of the result. In addition, when the Top 5 analogues selected using the Mahalanobis distance (MD) are reviewed and most of them are experimentally positive in genotoxicity assays such as AMES, in vitro chromosomal aberration (CA), and in vivo micronucleus (MN), this provides convergent evidence supporting a positive model prediction. The observation of genotoxicity among structurally similar chemicals or among chemicals close in feature space strengthens the plausibility that the substance under assessment may exhibit a comparable hazard. Moreover, when genotoxicity relevant structural alerts (SA) are triggered, the predicted hazard can be interpreted as mechanistically consistent with specific structural motifs, further increasing confidence in a concern signal. However, the absence of an SA must not be interpreted as evidence of non hazard, and SAs may generate both false positives and false negatives. They should therefore be interpreted conservatively. Taken together, the substance is treated as a screening level genotoxicity concern, and additional information may be recommended.

Conversely, when the model predicts negative, the substance is within the AD, and the experimental data for the MD based Top 5 analogues are predominantly genotoxicity negative, the model prediction and analogue evidence are mutually consistent. If, under these conditions, no genotoxicity relevant SA is triggered or the alerts are deemed non relevant, there is no clear mechanistic concern signal at this stage. The substance may therefore be considered low concern, and de prioritization and deferral of additional testing can be reasonable options. However, if discordance is observed, for example the model is negative but a genotoxicity SA is triggered, or some analogues are experimentally positive, a conservative re evaluation is warranted. Appropriate follow ups may include (i) re examining or expanding the analogue set, (ii) assessing the relevance of potential metabolites and microspecies or the possibility of reactive intermediates, and (iii) conducting genotoxicity testing and integrating the results into the WoE framework. In addition, if the MD value is high, the prediction may be less reliable. In such

cases, strengthening analogue evidence or generating additional experimental data should be considered as a priority.

Finally, a negative model prediction alone must not be used to justify a test waiver. The QSAR output should be interpreted together with relevant existing data and analogue toxicity information and incorporated into a Weight of Evidence assessment for the final regulatory conclusion. De prioritization or test deferral should be considered only when external evidence is sufficiently consistent with the prediction.

6.4 Conclusion (Adequacy for Regulatory Conclusion)

Positive prediction, AD in

Positive case. When the (Q)SAR prediction is generated within the model's Applicability Domain and is concordant with analogue toxicity trends, it can be considered adequate as supportive evidence for regulatory decision making. The result may be used to flag the substance for priority management and to guide follow up actions. For example, subsequent evaluation or testing can focus on genotoxicity relevant endpoints such as AMES, in vitro chromosomal aberration, and in vivo micronucleus. The prediction is not intended as a stand alone conclusion and should be interpreted conservatively within a Weight of Evidence framework together with available experimental data and analogue information.

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.