



**QMRF identifier: TOXFENCE-QMRF-001a**

**QMRF Title: Mutagenicity AMES test with S9  
TOXFENCE model – version 1.0**

**Date of QMRF: 25/03/2026**

**Model Developer: RAMC Co., Ltd.**

## 1. QSAR identifier

### 1.1 QSAR identifier (title)

Mutagenicity AMES test with S9 TOXFENCE model – version 1.0

### 1.2 Other related models

Mutagenicity AMES test without S9 TOXFENCE model – version 1.0

Clastogenicity in vivo Micronucleus Test TOXFENCE model – version 1.0

Clastogenicity in vitro Chromosomal Aberration test TOXFENCE model – version 1.0

### 1.3 Software coding the model

TOXFENCE v1.0

The model is implemented in TOXFENCE, a web-based SaaS platform developed by Risk Management Consulting Co., Ltd. TOXFENCE is designed to perform QSAR-based toxicity prediction using chemical structure information as input. The software is provided as an online service without local installation and includes functions for model execution, result review, and report output. The model was implemented in a Python-based environment, the backend service is operated using FastAPI, and RDKit was used for molecular structure handling and descriptor generation.

Risk Management Consulting Co., Ltd.

<https://www.toxfence.com>

## 2 General information

### 2.1 Date of QMRF

Feb 2025

### 2.2 QMRF author(s) and contact details

[1] Organisation: Risk Assessment & Management Consulting 04156 Seoul, Korea

[2] Contact e-mail: [ramc0983@naver.com](mailto:ramc0983@naver.com)

[3] Corporate website: <https://www.ramc0983.com/>

[4] TOXFENCE web service: <https://www.toxfence.com/>

### 2.3 Date of QMRF update(s)

25/03/2026

### 2.4 QMRF update(s)

Updated by: Risk Assessment & Management Consulting Co., Ltd.

Contact: [ramc0983@naver.com](mailto:ramc0983@naver.com)

Modified field: 1.3 Software coding the model

Reason for update: The website information was revised to replace the service address with the company website address.

2.5 Model developer(s) and contact details

The model was developed by Risk Assessment & Management Consulting Co., Ltd., 04156 Seoul, Republic of Korea.

Contact e-mail: [ramc0983@naver.com](mailto:ramc0983@naver.com)

Website: <https://www.ramc0983.com/>

TOXFENCE web service: <https://www.toxfence.com/>

2.6 Date of model development and/or publication

2025

2.7 Reference(s) to main scientific papers and/or software package

[1] scikit-learn 1.5.2: <https://scikit-learn.org/stable/>

[2] xgboost 2.1.3: <https://xgboost.ai/>

[3] lightgbm 4.5.0: <https://lightgbm.readthedocs.io/en/stable/>

2.8 Availability of information about the model

The model is proprietary. The training dataset and the model are not publicly available. The model is implemented and operated through the TOXFENCE web service.

<https://www.toxfence.com/>

2.9 Availability of another QMRF for exactly the same model

Another QMRF is not available.



(주)위해관리연구소  
Risk Assessment & Management Consulting  
유해물질 위해평가/관리 컨설팅

### 3 Defining the endpoint - OECD Principle 1

#### 3.1 Species

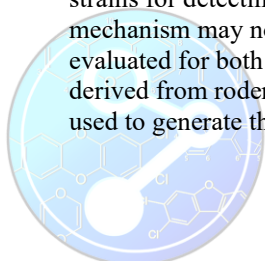
*Salmonella typhimurium*

#### 3.2 Endpoint

Bacterial reverse mutation (Ames mutagenicity) with S9 metabolic activation

#### 3.3 Comment on endpoint

Mutagenic toxicity refers to the ability of a substance to induce genetic mutations, a property that raises significant public health concerns due to its potential link to carcinogenicity and reproductive toxicity. Many mutagenic substances are suspected to be carcinogenic if a genotoxic mechanism is involved. Moreover, mutagenicity in somatic cells is concerning due to the possibility of such mutations being transmitted through germ cells, leading to hereditary genetic disorders. One of the primary methods used to assess mutagenic potential is the Ames test, a widely accepted *in vitro* assay. This bacterial reverse mutation test uses at least five strains of *S. typhimurium* to identify point mutations caused by base substitutions or frameshift mutations. The test operates on the principle that mutagenic substances can reverse mutations present in the test strains, thereby restoring the ability of the bacteria to synthesize an essential amino acid that they would otherwise be unable to produce. The test is guided by the relevant OECD Test Guideline (TG 471) and can be performed using either the original 1983 version or the updated 1997 version of the protocol. The Ames test is capable of detecting mutagenic mechanisms involving DNA base-pair substitutions and frameshift mutations, as covered by several test strains, including TA100, TA1535, TA97, TA1538, TA98, and others. In the 1997 guideline update, strains for detecting DNA cross-links were added, although mutagenic substances acting via this mechanism may not always be fully detected in the current training set [2]. The test endpoint is evaluated for both the parent compound and any metabolites formed *in vitro*, typically using an S9 mix derived from rodent liver homogenates. In some cases, liver homogenates from hamsters may also be used to generate the metabolic activation system.



Risk Assessment & Management Consulting  
유해물질 위해평가/관리 컨설팅

### 3.4 Endpoint units

Dimensionless

### 3.5 Dependent variable

The dependent variable is classified as non-mutagenic or mutagenic.

### 3.6 Experimental protocol

The bacterial reverse mutation (Ames) test with S9 metabolic activation, conducted in accordance with OECD Test Guideline 471, is an in vitro assay for assessing the mutagenic potential of chemicals using bacterial strains collectively sensitive to a broad spectrum of DNA-damaging agents.

### 3.7 Endpoint data quality and variability

Experimental data for this model were curated from studies accessible through eChemPortal and the Hazardous Substances Data Bank (HSDB). The dataset consisted of results from the bacterial reverse mutation test (Ames test) performed with metabolic activation (S9+). Only experimental records with reliability scores of 1 or 2 were retained for model development. During data curation, salts and inorganic substances were removed, duplicate records were excluded, and conflicting results reported for the same substance were resolved through expert consultation. For each retained substance, the CAS number, SMILES, and final experimental outcome classified as positive (1) or negative (0) were recorded. SMILES were standardised using a custom Python 3.10.4 script based on the RDKit library. After curation, the final dataset consisted of 4,300 organic compounds. As the data were compiled from multiple studies conducted under different experimental conditions, bacterial strains, and laboratories, some degree of inter-study and inter-laboratory variability may remain.



# (주)위해관리연구소

Risk Assessment & Management Consulting

유해물질 위해평가/관리 컨설팅

## 4 Defining the algorithm - OECD Principle 2

### 4.1 Type of model

A model in which multiple tree-based machine learning algorithms soft-vote the predicted result to select the final predicted value.

### 4.2 Explicit algorithm

The Voting model was constructed with Random Forest, XGBoost, Light GBM, and Histogram Gradient Boosting models. The hyperparameter value of each model was determined by Grid Search. Each model determines whether or not it is toxic as a value between 0 and 1, and classifies it by arithmetic average.

### 4.3 Descriptors in the model

#### 1) MolLogP:

The octanol–water partition coefficient estimated using the Wildman–Crippen method, representing the molecule’s lipophilicity and hydrophobicity. Higher values suggest enhanced membrane permeability and potential bioaccumulation.

#### 2) FpDensityMorgan:

The density of features in a molecule based on the Morgan fingerprint, calculated as the number of fingerprint bits set divided by the molecule's atom count.

#### 3) BCUT2D\_MWHI:

A 2D BCUT descriptor related to the maximum weighted hydrogen count, providing information on molecular complexity.

#### 4) BCUT2D\_LOGPHI:

A 2D BCUT descriptor related to the maximum weighted atom-based hydrophobicity (logP), representing molecular lipophilicity.

#### 5) BCUT2D\_MWLOW:

A 2D BCUT descriptor related to the minimum weighted molecular weight, reflecting molecular simplicity.

#### 6) TPSA:

The total polar surface area estimated from the topology (2D structure) of polar atoms, primarily nitrogen and oxygen with their attached hydrogens, reflecting molecular properties relevant to drug absorption, bioavailability, and blood–brain barrier penetration.

#### 7) NumHAcceptors:

The number of hydrogen bond acceptors.

#### 8) NumHDonors:

The number of hydrogen bond donors.

#### 9) PEOE\_VSA:

The volume-based surface area calculated using the PEOE (Partial Equalization of Orbital Electronegativities) method, reflecting the distribution of electrostatic potential across a molecule.

#### 10) EState\_VSA:

The volume-based surface area calculated using the Electrotopological State (EState) index, reflecting the distribution of electronic properties across the molecule.

#### 11) FractionCSP3:

The fraction of sp<sup>3</sup>-hybridized carbons in a molecule, providing information on its flexibility.

#### 12) fr\_aldehyde:

The fraction of the molecule containing an aldehyde group (–CH=O).

**13) fr\_alkyl\_halide:**

The fraction of the molecule containing an alkyl halide group ( $-C-X$ , where X is a halogen).

**14) fr\_azo:**

The fraction of the molecule containing an azo group ( $-N=N-$ ).

**15) fr\_epoxide:**

The fraction of the molecule containing an epoxide group (a three-membered ring containing oxygen).

**16) fr\_nitro:**

The fraction of the molecule containing a nitro group ( $-NO_2$ ).

**17) fr\_nitro\_ arom:**

The fraction of the molecule containing aromatic nitro substituents.

4.4 Descriptor selection

Molecular descriptors were calculated using the RDKit cheminformatics library within a custom Python 3.10.4 workflow. A total of 210 molecular descriptors were calculated, encompassing various physicochemical and structural properties of the compounds. From this comprehensive set, descriptor selection was based on SHAP (SHapley Additive exPlanations) values, prioritizing the removal of descriptors with low importance. The optimal set of 22 descriptors was finalized by comparing model performance before and after descriptor removal, selecting those that improved model performance. The selection process employed a subset evaluator that utilized 5-fold stratified cross-validation with shuffling on the training data set.

4.5 Algorithm and descriptor generation

The RDKit Descriptor Calculation module in Python 3.10.4 was utilized to compute a comprehensive set of molecular descriptors, encompassing various physicochemical and structural properties of the compounds.

4.6 Software name and version for descriptor generation

Molecular descriptors were generated using RDKit 2023.09 in Python 3.10.4.

<https://www.rdkit.org/>

<https://www.python.org/>

4.7 Chemicals/Descriptors ratio

3,400 chemicals / 22 descriptors = 156.4

## 5 Defining the applicability domain - OECD Principle 3

### 5.1 Description of the applicability domain of the model

The applicability domain is defined in the descriptor space used by the model. AD is assessed by combining Mahalanobis distance squared (MD2) and the k-nearest-neighbor (k-NN) mean distance in the standardized feature space. A query is considered inside the AD only when both criteria are met.

### 5.2 Method used to assess the applicability domain

Training descriptors are scaled using StandardScaler with optional z-clipping where applicable. In the standardized space, MD2 is computed using a LedoitWolf shrinkage covariance model and the 95th percentile of the training MD2 distribution is used as the threshold. In the same space, the k-NN mean distance is computed and the 95th percentile of the training k-NN mean-distance distribution is used as the threshold. A query is classified as inside the AD only when both thresholds are satisfied. Otherwise, it is classified as outside the AD.

### 5.3 Software name and version for applicability domain assessment

Implemented in TOXFENCE v1.0.

Applicability domain assessment was performed in the standardized descriptor space using Mahalanobis distance squared (MD2) based on a LedoitWolf shrinkage covariance model and k-nearest neighbor mean distance. A query was classified as inside the applicability domain only when both the MD2 threshold and the k-NN mean-distance threshold, defined as the 95th percentile of the corresponding training distributions, were satisfied.

### 5.4 Limits of applicability

This model is intended for organic chemicals that can be represented by valid SMILES. For salts and multi-component substances, descriptors are calculated using a parent structure obtained by removing inorganic counterions and selecting the main organic fragment. Therefore, the prediction reflects the processed parent structure rather than the full mixture or salt form. Predictions for purely inorganic substances or chemotypes that are underrepresented in the training data may have increased uncertainty and should be interpreted conservatively together with the applicability domain result.

## 6 Internal validation - OECD Principle 4

- 6.1 Availability of the training set  
No
- 6.2 Available information for the training set  
CAS RN: Yes  
Chemical Name: Yes  
Smiles: Yes  
Formula: No  
INChI: No  
MOL file: Yes
- 6.3 Data for each descriptor variable for the training set  
All
- 6.4 Data for the dependent variable for the training set  
All
- 6.5 Other information about the training set  
The training set consisted of 3,440 substances, including 1,571 positive and 1,869 negative substances. The class distribution was considered acceptable for model development, and no additional imbalance correction was applied.
- 6.6 Pre-processing of data before modelling  
Before modelling, the dataset was curated to improve the consistency and suitability of the input data. Salts and inorganic substances were removed, duplicate records were excluded, and chemical structures were verified and standardised using a custom Python script.
- 6.7 Statistics for goodness-of-fit  
Accuracy = 83%, Precision = 73%  
Recall = 88%, Specificity = 80%  
F1-score = 80%, MCC = 67%  
TP 934, TN 1354, FP 338, FN 126
- 6.8 Robustness - Statistics obtained by leave-one-out cross-validation  
Stratified 5-FOLD Average Value  
Accuracy = 85%, Precision = 77%  
Recall = 88%, F1-score = 82%  
Specificity = 0.75, MCC = 0.70  
TP 232, TN 356, FP 67, FN 33
- 6.9 Robustness - Statistics obtained by leave-many-out cross-validation
- 6.10 Robustness - Statistics obtained by Y-scrambling
- 6.11 Robustness - Statistics obtained by bootstrap
- 6.12 Robustness - Statistics obtained by other methods

## 7 External validation - OECD Principle 4

- 7.1 Availability of the external validation set  
No
- 7.2 Available information for the external validation set  
CAS RN: Yes  
Chemical Name: Yes  
Smiles: Yes  
Formula: No  
INChI: No  
MOL file: Yes
- 7.3 Data for each descriptor variable for the external validation set  
All
- 7.4 Data for the dependent variable for the external validation set  
All
- 7.5 Other information about the external validation set  
The test set is composed of 753 substances (319 positive, 434 negative)
- 7.6 Experimental design of test set  
Prior to model development, the full dataset was randomly divided into a training set and a test set at a ratio of 8:2.
- 7.7 Predictivity - Statistics obtained by external validation  
Accuracy = 76%, Precision = 66%  
Recall = 76%, Specificity = 76%  
F1-score = 70%, Sensitivity = 88%  
TP 252, TN 401, FP 67, FN 33
- 7.8 Predictivity - Assessment of the external validation set  
The external validation set was considered appropriate for assessing model predictivity because it was sufficiently large and was generated by random splitting prior to model development. The test set was assumed to be broadly representative of the descriptor space and response distribution of the training set. Therefore, the external validation results were considered suitable for evaluating the predictive performance of the model within its applicability domain.
- 7.9 Comments on the external validation of the model  
The external validation was performed using a hold-out test set obtained by random splitting of the full dataset prior to model development. Therefore, the validation can be regarded as an independent test of the final model, although the test compounds originated from the same overall dataset rather than from a completely separate external source. The reported performance should thus be interpreted as evidence of predictive ability within the chemical space represented by the dataset, while predictions for compounds outside the applicability domain or for underrepresented chemotypes should be interpreted with caution.

## 8 Providing a mechanistic interpretation - OECD Principle 5

### 8.1 Mechanistic basis of the model

The model is an ensemble classification model based on molecular descriptors encoding structural and physicochemical properties of chemicals relevant to genotoxicity. Its mechanistic basis is indirect, in that chemicals with similar descriptor patterns are assumed to have similar genotoxic potential. Because the final prediction is generated by combining multiple tree-based machine-learning algorithms, a direct one-to-one mechanistic interpretation is limited.

### 8.2 Other information about the mechanistic interpretation

The mechanistic interpretation of the model was established a posteriori, based on the interpretation of the final set of descriptors and model outputs after model development.

### 8.3 Other information about the mechanistic interpretation

No additional information is available regarding the mechanistic interpretation.

## 9 Miscellaneous information

### 9.1 Comments

No additional comments.

### 9.2 Bibliography

[1] Drug Discov Today. 2018 Aug;23(8):1538-1546. doi: 10.1016/j.drudis.2018.05.010. Epub 2018 May 8.

[2] OECD Test No. 471, Bacterial Reverse Mutation Test

### 9.3 Supporting information

Not available.