TOXFENCE ver.1

# DART (Reproductive/Developmental Toxicity) Prediction Report

Benzene



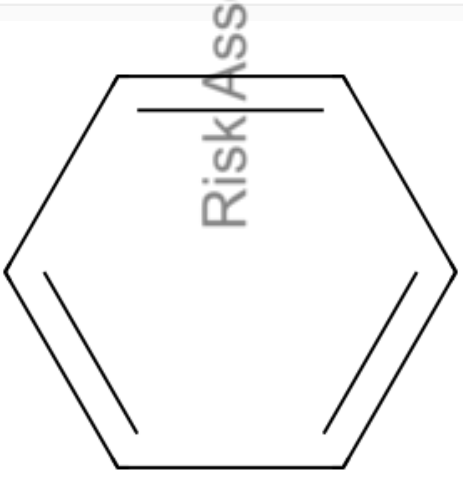
## 1. General Information

| | |
|---|---|
| **Date of QPRF** | 2026-03-23 |
| **Author** | ㈜ 위해관리연구소 |
| **Contact** | 02-715-0983 |

## 2. Substance Identity

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

**2D Structure**



## 3. Model and Software Used

**Model Used**

The DART (Developmental and Reproductive Toxicity) model (version 1.0) implemented in the TOXFENCE QSAR platform is a deep-learning (Graph Convolutional Network, GCN) (Q)SAR classification model designed to predict developmental/reproductive toxicity potential. It was trained on a curated dataset of 5,861 compounds categorized by developmental-toxicity outcomes. The endpoint is a binary hazard classification (toxicant vs. non-toxicant), and this prediction reports the corresponding class for the target substance. A QSAR Model Reporting Format (QMRF) (or model validation dossier) is available, providing detailed validation and algorithm information. According to the model's documentation/validation, classification performance (e.g., sensitivity and specificity) was evaluated, and in our deployment the decision threshold is set to balance sensitivity and specificity, thereby reducing both false negatives and false positives.

**• Software Platform**

TOXFENCE QSAR is a proprietary, commercially licensed QSAR software platform that provides various toxicity-prediction models, including a DART (Developmental and Reproductive Toxicity) model. The prediction was generated using the platform's default model settings.

**• Reference to QMRF**

TOXFENCE QMRF for DART (Developmental and Reproductive Toxicity) 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**

DART (Developmental and Reproductive Toxicity). The model predicts whether a chemical is a developmental toxicant (likely to cause adverse effects on embryo/fetal development or reproductive outcomes) or not, based on in vivo teratogenicity/developmental studies.

### • Results — DART (Reproductive/Developmental Toxicity)

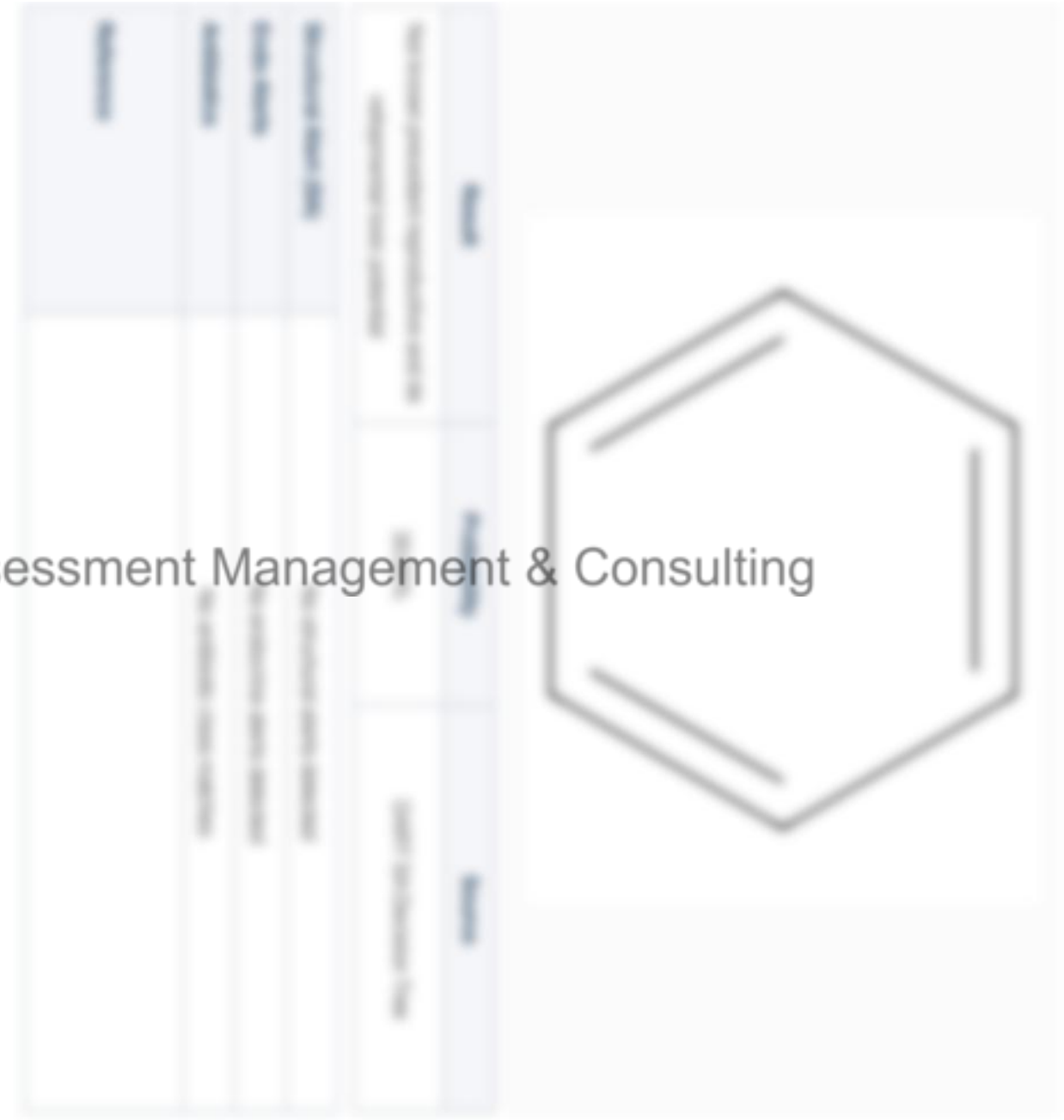
| Result | Probability | Source |
|---|---|---|
| Positive | 100.00% | GCN Model |

### Read-Across Similarity Results

| Result | Probability | Source |
|---|---|---|
| Positive | 100.00% | DB |

### DART (Reproductive/Developmental Toxicity) SA Decision Tree

**Structural Alert (SA) Image**


Risk Assessment Management & Consulting

## Overall Conclusion — DART (Reproductive/Developmental Toxicity)

| Result | Probability | Confidence | Source | AD |
|---|---|---|---|---|
| Positive | 100.00% | High | DB > Ensemble | IN |

## Decision Rationale

> An experimental label is available in the database; the final decision follows the experimental value (Decision_Source=DB).

## Basis for Probability Estimation

### Model Probability

The supervised classifier's estimated probability for the positive 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.

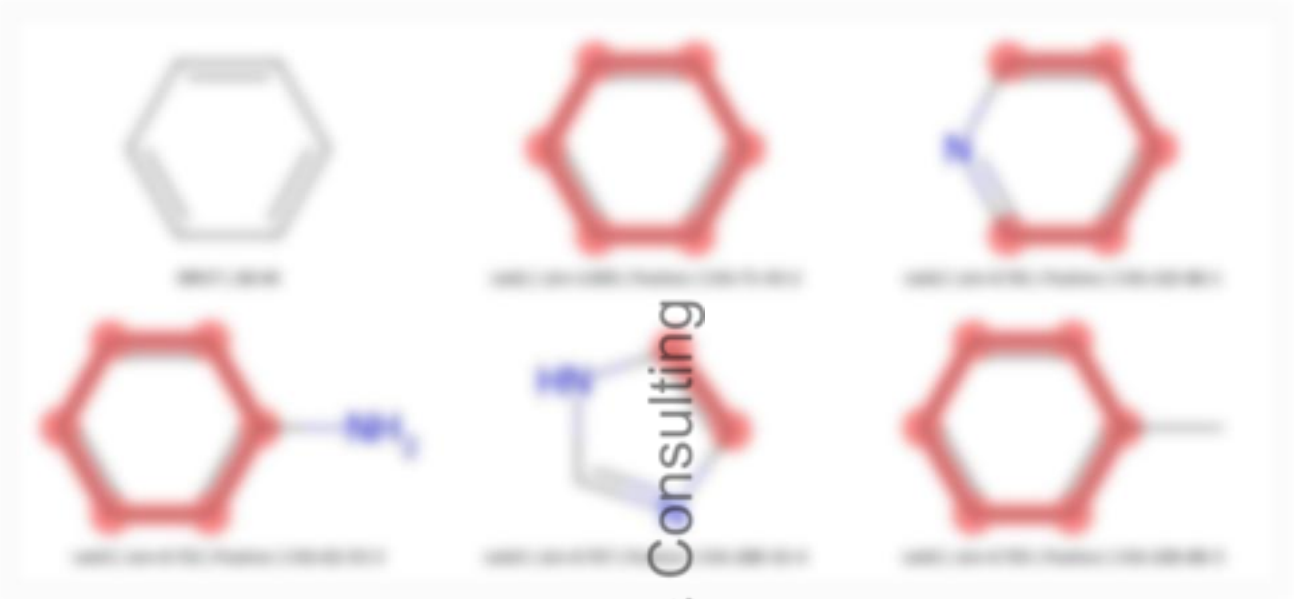
### DART (Reproductive/Developmental Toxicity) SA Decision Tree Probability

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 Similar Compounds



| Substance Name | Experimental Value | Similarity Score |
|---|---|---|
| [illegible] | Positive | [illegible] |
| [illegible] | Positive | [illegible] |
| [illegible] | Positive | [illegible] |
| [illegible] | Positive | [illegible] |
| [illegible] | Positive | [illegible] |



## Similarity method (RBF)

The model quantifies proximity between molecular feature vectors (e.g., descriptor profiles) using a Radial Basis Function (RBF) similarity. We employ the Gaussian form $S(x,x')=\exp(-\gamma \|x-x'\|^2)$, where $\gamma=1/(2\sigma^2)$ controls the width of the kernel (length scale). The score equals 1 when the vectors are identical and approaches 0 as their distance increases. Because this measure decreases smoothly with distance, it is well suited to capturing non-linear structure-activity/property relationships in QSAR/QSPR and aligns with standard kernel methods such as SVM with an RBF kernel. In this QPRF, the query compound is compared to training compounds in feature space to yield a similarity in [0,1], and was chosen to emphasize nearby compounds so that more similar training instances receive greater weight in the prediction.

# 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.**

**1) 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.

**2) 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.

**3) 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.

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

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

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

**7) 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 learni ng algorithm. |

# 6. Considerations for Regulatory Use

## 6.1 Regulatory Purpose

This (Q)SAR prediction was generated to support regulatory decision-making for the developmental and reproductive toxicity (DART) endpoint. In practice, it is used for prioritization and internal screening, helping to identify potentially high-risk substances early before committing resources and time to more intensive testing [1]. This approach aligns with international practice–amid tightening data requirements for DART under frameworks such as REACH, organizations including the OECD encourage the use of alternative methods (e.g., QSAR) to reduce animal testing. Importantly, the prediction is not intended as a stand-alone conclusion; it is interpreted as one line of evidence within a Weight-of-Evidence (WoE) framework, together with existing 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 DART-relevant structural alerts are triggered for the substance and the model prediction is positive, confidence in a hazard concern is strengthened. This indicates that the predicted toxicity aligns with structural features of the molecule, providing an important consideration for regulatory interpretation. Conversely, the absence of structural alerts or a negative model prediction should not be taken to mean the substance is "safe." Regulatory guidance emphasizes that a mere 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 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 inside the chemical space on which the model was trained and considered reliable, thereby supporting the validity of the result. If, in addition, a review of the Top-5 analogues selected by RBF-kernel similarity shows that most of them are experimentally positive for developmental/reproductive toxicity, this constitutes convergent evidence for a positive model prediction. The observation that structurally similar chemicals exhibit toxicity strengthens the plausibility that the substance under assessment may display a comparable hazard. Moreover, when DART-relevant structural alerts (SA) are triggered, the predicted toxicity is mechanistically consistent with specific structural motifs (toxicophores), which further increases confidence in a concern signal. That said, the absence of an SA must not be taken as evidence of non-hazard, and SAs can yield both false positives and false negatives; they are therefore interpreted conservatively. Considering the foregoing, the substance is treated as a screening-level concern, and additional information (e.g., targeted in vitro assays, refined read-across, and literature corroboration) may be recommended.

## 6.4 Conclusion (Adequacy for Regulatory Conclusion)

Conversely, when the model predicts negative, the substance is within AD, and the empirical data for the Top-5 analogues are predominantly non-toxic (experimentally negative), the model prediction and analogue evidence are mutually consistent. If, in this situation, no DART-relevant SA is triggered or the alerts are deemed non-relevant,

there is no mechanistic signal of concern, and the substance may be considered low concern at this stage, with de-prioritization and deferral of additional testing being reasonable options. However, if discordance is observed—e.g., the model is negative but a DART SA is triggered, or some analogues are experimentally positive—then a conservative re-evaluation is warranted. Appropriate follow-ups include (i) re-examining or expanding the analogue set, (ii) reassessing potential metabolites and microspecies (pH) relevance, and (iii) conducting targeted in vitro assays before integrating the outcome back into the WoE.

> Because this (Q)SAR prediction was generated within the model's Applicability Domain and shows concordance with analogue toxicity trends, it is 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 testing, for example by focusing subsequent study design on developmental toxicity-relevant endpoints. In short, for a positive prediction, the output serves as a credible early-warning signal that usefully informs

Finally, a negative model prediction alone must not be used to claim a test waiver. The (Q)SAR output must be interpreted together with existing relevant data and incorporated 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.

## 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.