'Relevant' versus 'Similar' Structural Analogs to Support Genotoxicity (Q)SAR Predictions

Suman Chakravarti[a], Sai Radha Mani Alla[a]

GTA 2024

Background

- **Structural analogs** with experimental data are important for expert review of (Q)SAR model evaluations.
- Analogs help **clarify uncertain (Q)SAR outcomes**, making inconclusive or contradictory evaluations more definitive.
- Finding analogs is **challenging** when query compounds have multiple alerts of varied significance, a common scenario in statistical model evaluations.
- **Conventional similarity** metrics may yield structurally similar analogs but often **lack relevance** by failing to identify analogs with query compound's alerts.
- Here we introduce a novel approach for finding analogs, employing a sophisticated fingerprinting technique that considers **both specific alerts** and **general structural features**.

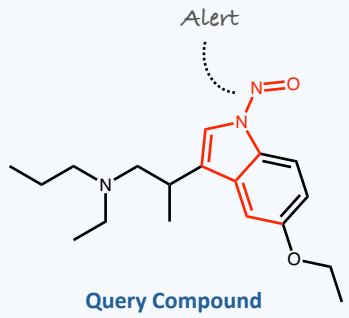
Methodology

- **Fingerprint Construction:** The fingerprint for a compound consists of two parts:
  - **Alert Part** – Contains importance weights of toxicity alerts identified in the compound, specific to the end point. As many elements as alerts in the model.
  - **Fragment Part** – Indicates presence of molecular fragments from the whole structure. Usually of fixed size, e.g., 256, 512 or 1024.
- **Relevance Metric:** Cosine similarity function was used to compute relevance.
- **Software:** CASE Ultra, 1.9.1.2 beta

Conclusions

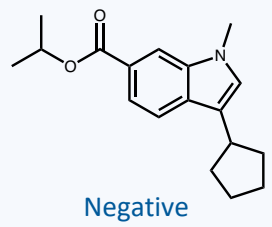
- This refined method **improves** the discovery of suitable and meaningful analogs.
- Offers a more sophisticated analysis of structural features **relevant to toxicity**.
- It is **versatile**, applicable beyond genotoxicity, compatible with both statistical and expert knowledge-based (Q)SAR models.

Relevant vs Similar Analogs

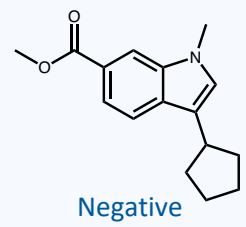


Query Compound
Positive QSAR Prediction
Ames Mutagenicity

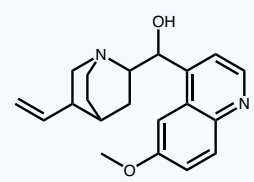
Similar Analogs



Negative
Similarity: 0.451



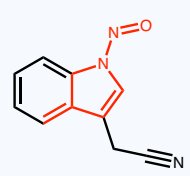
Negative
Similarity: 0.446



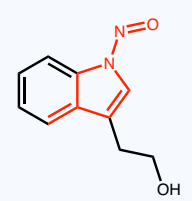
Positive
Similarity: 0.435

- Identified using **traditional similarity**
- Causality absent
- Low similarity
- **Not meaningful**

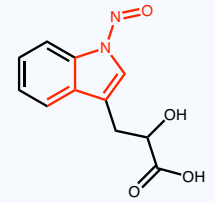
Relevant Analogs



Positive
Relevance: 0.822



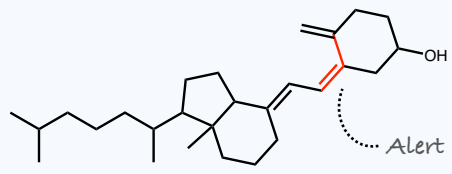
Positive
Relevance: 0.809



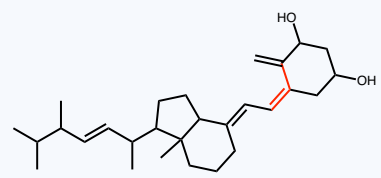
Positive
Relevance: 0.805

- Identified using the **enhanced method**
- Analogs with causality
- High relevance
- **Meaningful**

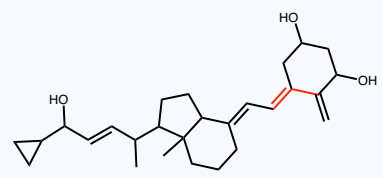
Downgrading Positive Prediction Calls to Negative



Query Compound
Positive QSAR Prediction
Mouse Lymphoma, Unactivated



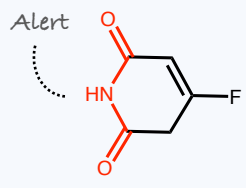
Negative
Relevance: 0.695



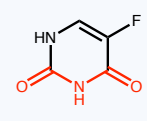
Negative
Relevance: 0.569

Negative Decision

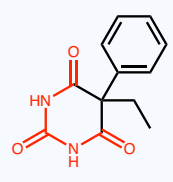
Resolving Inconclusive Prediction Calls



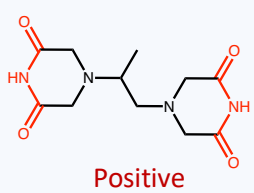
Query Compound
Inconclusive QSAR Prediction
Mouse Micronucleus, in vivo



Positive
Relevance: 0.808



Positive
Relevance: 0.737



Positive
Relevance: 0.730

Conclusive Positive Decision

[a]MultiCASE Inc., Mayfield Heights, OH, USA. Emails: chakravarti@multicase.com, srmalla@multicase.com