**External Validation of STopTox as an Alternative to Animal Testing for Toxicological Assessments**

Ricardo Scheufen Tieghi#a, Cleber C. Melo-Filho#a, Holli-Joi Martina, José Teófilo Moreira-Filhob, Tripp LaPrattc, Dave Allen*c*, Judy Strickland*c*, Alexander Tropsha\*a,d, Nicole Kleinstreuer\*b, Eugene N. Muratov\*a.

# These authors contributed equally.

*a UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA.*

*b National Toxicology Program Interagency Center for Evaluation of Alternative Toxicological Methods (NICEATM), Research Triangle Park, NC, 27711, USA.*

*c Inotiv-RTP, Morrisville, 27560, NC, United States*

*d Predictive, LLC, Raleigh, 27614, NC, United States*

\*To whom correspondence and materials requests should be addressed: 100K Beard Hall, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA; Telephone: (919) 966-2955; FAX: (919) 966-0204; E-mail: [nicole.kleinstreuer@nih.gov](mailto:nicole.kleinstreuer@nih.gov), [alex\_tropsha@unc.edu](mailto:alex_tropsha@unc.edu), [murik@email.unc.edu](mailto:murik@email.unc.edu).

**Competing Interests**

AT and ENM are co-founders of Predictive, LLC, which develops novel alternative methodologies and software for toxicity prediction. All the other authors declare no conflicts.

**Data Collection Protocol**

**Integrated Chemical Environment (ICE):**

1. ICE contains highly curated and verified datasets for various specific toxicity endpoints.
2. The datasets can be downloaded by accesing the link:
   1. (<https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION>, accessed June 2024)
3. Datasets of interest were directly downloaded as xlsx files from their “datasets tab”. All original ICE files are also available in the folder: “ICE Datasets”. File descriptions:
   1. Oral Toxicity – 9110 chemicals for 4 endpoints
   2. Acute Dermal Toxicity – 275 chemicals for 3 endpoints
   3. Acute Inhalation Toxicity – 1781 chemicals for 3 endpoints.
   4. Eye Irriation Toxicity – 454 chemicals for 7 endpoints
   5. Skin irritation/Corrosion – 564 chemicals for 16 endpoints
   6. Skin Sensitization – 1956 chemicals for 26 endpoints.
4. All datasets were filtered to include only compounds tested *in vivo* using the same OECD Test Guidelines the original STopTox was designed to predict. Specifically,
   1. Eye Irritation/Corrosion
      1. OECD TG 405
   2. Skin Irritation/Corrosion
      1. OECD TG 404
   3. Skin Sensitization
      1. OECD TG 429
   4. Acute Oral
      1. OECD TG 401, 420, 423, 425
   5. Acute Dermal
      1. OECD TG 402
   6. Acute Inhalation
      1. OECD TG 403
5. Duplicate chemicals or compounds missing SMILES were removed from all datasets.

**ChemID Plus / PubChem**

ChemIDplus is a database that contains about 350,000 chemical records. Access to the database is provided by Specialized Information Services (SIS) of the U.S. National Library of Medicine (NLM). In 2022, ChemID Plus was integrated into PubChem.

**Key Reference Studies.**

Specific literature sources were identified that identified *in vivo* compound testing in alignment with OECD test guidelines for each endpoint. Compounds from these studies were included for analysis.

1. Skin irritation/Corrosion:
   1. Rooney et al. 20211 and Lewis, R.J. Sr. Sax's Dangerous Properties of Industrial Materials2
2. Eye Irritation/Corrosion
   1. 118 compounds tested for eye irritation/corrosion by Yamaguchi et al.3