	QMRF identifier (JRC Inventory): **Not Applicable**
	QMRF Title: **Carcinogenic Potency Categorization Approach (CPCA)**
	Printing Date: **October 11, 2024**

1.QSAR identifier

1.1. QSAR identifier (title):

Carcinogenic Potency Categorization Approach (CPCA)

1.2. Other related models:

None

1.3. Software coding the model:

The model does not require software for implementation; however, a free and open-source custom command-line Java application called “Featurize-Nitrosamines” for automated implementation is available for download at https://github.com/FDA/featurize-nitrosamines. Further details about the application are available in the following publication:

Kruhlak, N. L., Schmidt, M., Froetschl, R., Graber, S., Haas, B., Horne, I., Horne, S., King, S. T., Koval, I. A., Kumaran, G., Langenkamp, A., McGovern, T. J., Peryea, T., Sanh, A., Siqueira Ferreira, A., van Aerts, L., Vespa, A., Whomsley, R., 2024. Determining Recommended Acceptable Intake Limits for N-Nitrosamine Impurities in Pharmaceuticals: Development and Application of the Carcinogenic Potency Categorization Approach (CPCA). Regulatory Toxicology and Pharmacology, 150:105640.

2.General information

2.0. Abstract

The CPCA is a categorical human expert rule-based structure-activity relationship model that predicts the carcinogenic potency of an *N*-nitrosamine compound and assigns it to 1 of 5 potency categories, each with a corresponding acceptable intake (AI) limit. In this context, an AI limit is a regulated level of a nitrosamine impurity that may be present in a marketed human pharmaceutical. The model considers the number and distribution of α-hydrogens at the *N*-nitroso center, and other activating and deactivating structural features of a nitrosamine encoded as molecular fragments that affect the α-hydroxylation metabolic activation pathway implicated in carcinogenesis. The CPCA has been adopted internationally by several drug regulatory authorities as a starting point to determining recommended AI limits for nitrosamines where compound-specific empirical data are not available.

2.1.Date of QMRF:

October 11, 2024

2.2.QMRF author(s) and contact details:

Dr. Naomi L. Kruhlak

US Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, MD, 20993 USA

Naomi.Kruhlak@fda.hhs.gov

2.3.Date of QMRF update(s):

N/A

2.4.QMRF update(s):

N/A

2.5.Model developer(s) and contact details:

Nitrosamines International Technical Working Group (NITWG), Safety Subgroup, SAR Subteam

Contact through:

Dr. Naomi L. Kruhlak

US Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, MD, 20993 USA

Naomi.Kruhlak@fda.hhs.gov

2.6. Date of model development and/or publication:

Developed in July 2023. Published in regulatory guidance documents issued by various drug regulatory authorities in July and August 2023; published in the peer-reviewed scientific literature in May 2024.

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

Kruhlak, N. L., Schmidt, M., Froetschl, R., Graber, S., Haas, B., Horne, I., Horne, S., King, S. T., Koval, I. A., Kumaran, G., Langenkamp, A., McGovern, T. J., Peryea, T., Sanh, A., Siqueira Ferreira, A., van Aerts, L., Vespa, A., Whomsley, R., 2024. Determining Recommended Acceptable Intake Limits for N-Nitrosamine Impurities in Pharmaceuticals: Development and Application of the Carcinogenic Potency Categorization Approach (CPCA). Regulatory Toxicology and Pharmacology, 150:105640.

2.8. Availability of information about the model:

Model is open source. The model is described in the scientific article referenced above and in pharmaceutical regulatory guidance documents (see references below). The model does not require software for implementation; however, a free and open-source custom command-line Java application called Featurize-Nitrosamines is available for automated implementation. Featurize-Nitrosamines is described in the scientific article and is available for download at https://github.com/FDA/featurize-nitrosamines.

Regulatory Guidances:

US Food and Drug Administration: https://www.fda.gov/media/170794/download

European Medicines Agency: https://www.ema.europa.eu/en/documents/referral/nitrosamines-emea-h-a53-1490-questions-answers-marketing-authorisation-holders/applicants-chmp-opinion-article-53-regulation-ec-no-726/2004-referral-nitrosamine-impurities-human-medicinal-products_en.pdf

Health Canada: https://www.canada.ca/en/health-canada/services/drugs-health-products/compliance-enforcement/information-health-product/drugs/nitrosamine-impurities/medications-guidance.html

2.9. Availability of another QMRF for exactly the same model:

None

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Rodent, primarily rat

3.2. Endpoint:

Carcinogenic potency in a 2-year rat carcinogenicity bioassay. Carcinogenic potency is calculated as a TD_{50} value (tumorigenic dose-rate in 50% of test animals) and is then converted to an acceptable intake (AI) limit in humans.

3.3. Comment on endpoint:

Model generates a categorical prediction of carcinogenic potency by assigning a test compound to 1 of 5 potency categories (PCs) with associated AI limits in ng/day. PC 1 = 18 or 26.5 ng/day (depending on international regulatory region); PC 2 = 100 ng/day; PC 3 = 400 ng/day; PC 4 = 1500

ng/day; PC 5 = 1500 ng/day. Test compounds assigned to PC 4 are predicted to be of low carcinogenic potency but may be mutagenic in an enhanced Ames assay. Test compounds assigned to PC 5 are predicted to be of low or no carcinogenic potency and are not predicted to be DNA-reactive.

3.4. Endpoint units:

TD_{50} in mg/kg/day converted to an AI limit in ng/day by dividing the TD_{50} value by 50,000 (to adjust from a 1:2 tumor incidence in rodents to a 1:100,000 excess cancer risk in humans) and multiplying by a human body weight adjustment factor of 50 kg.

3.5. Dependent variable:

Carcinogenic potency

3.6. Experimental protocol:

As described in OECD Test Guideline 451 (https://read.oecd-ilibrary.org/environment/test-no-451-carcinogenicity-studies_9789264071186-en#page1). TD_{50} values were calculated according to methods described in the Carcinogenic Potency Database (CPDB) (see Gold et al., 2005. The Carcinogenic Potency Database (CPDB). https://files.toxplanet.com/cpdb/index.html).

3.7. Endpoint data quality and variability:

The quality of studies in the CPDB is variable, although the CPDB does utilize criteria for inclusion such as the proportion of the lifetime during which test animals were exposed. Studies of lesser quality are defined as those where one or more of the following scenarios were encountered:

< 50 animals per dose per sex; < 3 dose levels; lack of concurrent controls; intermittent dosing (< 5 days per week); dosing for less than lifetime.

The model was trained using 66 nitrosamines reported in the CPDB with harmonic mean TD_{50} values or negative carcinogenicity classifications in the rat (and one in the mouse), combined with carcinogenic potency classifications for 22 nitrosamines (10 of which also had CPDB TD_{50} values) as reported by Rao et al. (1979). Rao et al. (1979) classified compounds as strong carcinogens, carcinogens, weak carcinogens, or non-carcinogens based on dose of compound, dose rate, tumor incidence, and time taken for tumor development. Two additional compounds were classified as experimentally negative for carcinogenicity by Cross and Ponting (2021).

Structures of training chemicals, registry numbers and names were verified.

4.Defining the algorithm - OECD Principle 2

4.1. Type of model:

Human expert rule-based SAR model with categorical prediction of carcinogenic potency.

4.2. Explicit algorithm:

Human expert rule-based model implemented as a decision-tree. Model was trained by visual inspection of nitrosamine structure-activity relationship patterns, including those previously described in the published literature (see Cross and Ponting, 2021; Ponting et al., 2022; Thomas et al., 2022). Descriptor (feature) weights were manually assigned. Further refinement of feature definitions and weights was performed manually by consideration of supporting data on metabolism and chemical reactivity pathways.

4.3. Descriptors in the model:

Molecular fragment-based descriptors (features), each with associated weights corresponding to their positive or negative contribution to the prediction. Two categories of model descriptors: 1) features defining the number and position of hydrogen atoms alpha to the *N*-nitroso group (n=9), and 2) features defining substructures elsewhere on the molecule (n=14). A given test compound matches exactly one descriptor from the first category but may match zero or more descriptors from the second category.

4.4. Descriptor selection:

Initial pool of 15 descriptors was defined by visual inspection of training set structures with consideration of structure-activity relationship patterns previously described in the published literature (see Cross and Ponting, 2021; Ponting et al., 2022; Thomas et al., 2022). The set was manually expanded to 23 descriptors by consideration of supporting data on metabolism and chemical reactivity.

4.5. Algorithm and descriptor generation:

Human expert rule-based model trained by manual identification of nitrosamine structure-activity relationship patterns. Model descriptors were defined manually and generated automatically across the training set using the Featurize-Nitrosamines Java application (described in Section 1.3). Descriptors are calculated by Featurize-Nitrosamines from input structures in SMILES or SDF format using explicitly coded feature extraction methods that employ atom-type filters, bond-type filters, and both depth-first and breadth-first searches of the supplied chemical graph.

4.6. Software name and version for descriptor generation:

The model does not require software for implementation; however, if using the Featurize-Nitrosamines Java application (described in Section 1.3) for automated implementation, descriptors are calculated from input structures in SMILES or SDF format.

4.7. Chemicals/Descriptors ratio:

Number of chemicals = 81, Number of descriptors = 23. Ratio 3.52:1

5.Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Limited to *N*-nitroso compounds 1) bearing a carbon atom directly bonded to both sides of the *N*-nitroso group; 2) where the carbon atom is not directly double bonded to a heteroatom (i.e., *N*-nitrosamides, *N*-nitrosoureas, *N*-nitrosoguanidines and other related structures are excluded); and 3) where the *N*-nitroso group is not bonded to a nitrogen. No further limitations on *N*-nitrosamine structures covered by the CPCA have been implemented, reflecting the extent of available data at the time of publication.

5.2. Method used to assess the applicability domain:

Visual inspection to identify N-nitroso compound classes activated through alternate mechanistic pathways that are not adequately addressed by features encoded in the model, which is based on the α-hydroxylation pathway of metabolic activation.

5.3. Software name and version for applicability domain assessment:

None – model application does not require software.

5.4. Limits of applicability:

Limited to the three criteria listed above to reflect that these subclasses of *N*-nitroso compounds undergo different mechanisms of activation and reactivity than *N*-nitrosamines. They are, therefore, not necessarily covered by structural features encoded by the CPCA.

6.Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes, as Supplemental Data Table S1 in Kruhlak et al., 2024.

6.2. Available information for the training set:

CAS RN: Yes
Chemical Name: Yes
Smiles: Yes
Formula: No
INChI: No
MOL file: No
NanoMaterial: No

6.3. Data for each descriptor variable for the training set:

Descriptor variables are available in Supplemental Data Table S1 in Kruhlak et al., 2024.

6.4. Data for the dependent variable for the training set:

Data for the dependent variable are summarized in Supplemental Data Table S1 in Kruhlak et al., 2024. Supporting data are available in the Lhasa Carcinogenicity Database (https://carcdb.lhasalimited.org/)

6.5. Other information about the training set:

All training set data are public

6.6. Pre-processing of data before modelling:

Structural features were fingerprinted using a free and open-source custom command-line Java application called "Featurize-Nitrosamines," developed and published as supplemental data to Kruhlak et al., 2024. The full source code and documentation for Featurize-Nitrosamines is available for download at https://github.com/FDA/featurize-nitrosamines.

6.7. Statistics for goodness-of-fit:

See Kruhlak et al (2024) for full details. Specifically, goodness-of-fit statistics are presented for 32 training set compounds with categorical carcinogenicity data in Table 4 and for 58 compounds with CPDB rat TD_{50} values in Figure 6 and Supplemental Data Tables S2 and S3.

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

Not applicable – expert rule-based model

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

Not applicable – expert rule-based model

6.10. Robustness - Statistics obtained by Y-scrambling:

Not applicable – expert rule-based model

6.11. Robustness - Statistics obtained by bootstrap:

Not applicable – expert rule-based model

6.12. Robustness - Statistics obtained by other methods:

Not applicable – expert rule-based model

7.External validation - OECD Principle 4

7.1. Availability of the external validation set:

External validation was not conducted. All available data were used to train the model.

7.2. Available information for the external validation set:

Not applicable

7.3. Data for each descriptor variable for the external validation set:

Not applicable

7.4. Data for the dependent variable for the external validation set:

Not applicable

7.5. Other information about the external validation set:

Not applicable

7.6. Experimental design of test set:

Not applicable

7.7. Predictivity - Statistics obtained by external validation:

Not applicable

7.8. Predictivity - Assessment of the external validation set:

Not applicable

7.9. Comments on the external validation of the model:

Not applicable

8.Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

Metabolic activation via the α-hydroxylation pathway is generally thought to be responsible for the highly potent mutagenic and carcinogenic responses observed for many *N*-nitrosamines. Structural features included in the CPCA are those that have been identified as directly impacting the α-hydroxylation pathway and, consequently, carcinogenic potency. Therefore, the combined effect of these features is used to generate a prediction of carcinogenic potency category of a nitrosamine based on chemical structure. Other subclasses of *N*-nitroso compounds, including *N*-nitrosamide-like compounds and those with an *N*-nitroso group in an aromatic ring, are expected to be biologically active through other mechanistic pathways and are not expected to follow the α-hydroxylation pathway. Therefore, these subclasses were considered out of scope of the CPCA and structural features associated with these alternative pathways were not explicitly incorporated, although some of the features may still be predictive.

8.2. A priori or a posteriori mechanistic interpretation:

Mechanistic basis of the model was determined a priori. All training set structures and descriptors were selected to fit a pre-determined mechanism of reactivity (metabolic activation by alpha-hydroxylation)

8.3. Other information about the mechanistic interpretation:

none

9.Miscellaneous information

9.1. Comments:

None

9.2. Bibliography:

References cited in this QMRF:

Cross, K.P., Ponting, D., 2021. Developing structure-activity relationships for N-nitrosamine activity. Comput. Toxicol. 20, 100186.

Gold, L.S., Ames, B.N., Bernstein, L., Blumenthal, M., Chow, K., Da Costa, M., de Veciana, M., Eisenberg, S., Garfinkel, G.B., Haggin, T., Havender, W.R., Hooper, N.K., Levinson, R., Lopipero, P., Magaw, R., Manley, N.B., MacLeod, P.M., Peto, R., Pike, M.C., Rohrbach, L., Sawyer, C.B., Slone, T.H., Smith, M., Stern, B.R., Wong, M., 2005. The Carcinogenic Potency Database (CPDB). https://files.toxplanet.com/cpdb/index.html.

Kruhlak, N. L., Schmidt, M., Froetschl, R., Graber, S., Haas, B., Horne, I., Horne, S., King, S. T., Koval, I. A., Kumaran, G., Langenkamp, A., McGovern, T. J., Peryea, T., Sanh, A., Siqueira Ferreira, A., van Aerts, L., Vespa, A., Whomsley, R., 2024. Determining Recommended Acceptable Intake Limits for N-Nitrosamine Impurities in Pharmaceuticals: Development and Application of the Carcinogenic Potency Categorization Approach (CPCA). Regulatory Toxicology and Pharmacology (in press).

Ponting, D.J., Dobo, K.L., Kenyon, M.O., Kalgutkar. A.S., 2022. Strategies for assessing acceptable intakes for novel N-nitrosamines derived from active pharmaceutical ingredients. J. Med. Chem. 65, 15584-15607.

Rao, T.K., Young, J.A., Lijinsky, W., Epler, J.L., 1979. Mutagenicity of aliphatic nitrosamines in Salmonella typhimurium. Mutat. Res. 66, 1–7.

Thomas, R., Tennant, R.E., Oliveira, A.A.F., Ponting, D.J., 2022. What makes a potent nitrosamine? Statistical validation of expert-derived structure-activity relationships. Chem. Res. Toxicol. 35, 1997–2013.

9.3. Supporting information:

Training set(s)Test set(s)Supporting information

Training set is available as Supplemental Data Table S1 in Kruhlak et al., 2024.