**QMRF Title**: Cholinergic model – identification of nicotinic acetylcholine receptor (nAChR) binders.

**Date**: 6th July 2022.

1. **QSAR Identifier**
   1. **QSAR Identifier (Title):** Cholinergics – nAChR binders.
   2. **Other related models:** Models for the cholinergic targets (mAChR, AChE) and mitochondrial inhibition.
   3. **Software coding the model:** Random Forest machine-learning model built in KNIME using MACCS fingerprints as the sole covariate.
2. **General Information**
   1. **Date of QMRF:** 6th July 2022.
   2. **QMRF author(s) and contact details:** Sanjeeva J. Wijeyesakere and M. Sue Marty. The Dow Chemical Company, Midland MI.
   3. **Date of QMRF update(s):** N/A.
   4. **QMRF updates(s):** N/A.
   5. **Model developers and contact details:** Sanjeeva J. Wijeyesakere and M. Sue Marty. The Dow Chemical Company, Midland MI.
   6. **Date of model development and/or publication:** January 2020.
   7. **Reference(s) to main scientific papers and/or software package:**

Wijeyesakere, S.J., Wilson, D.M., and Marty, M.S. (2020). Prediction of Cholinergic Compounds by Machine-Learning.” *Computational Toxicology.* 13: 100119. DOI: 10.1016/j.comtox.2020.100119.

* 1. **Availability of information about the model:** Data curation is detailed in the above-mentioned publication (Wijeyesakere et al., 2020; listed in section 2.7).
  2. **Availability of another QMRF for exactly the same model:** No.

1. **Defining the endpoint - OECD Principle 1**
   1. **Species:** Predictions are made for binding to human, rat and mouse nicotinic acetylcholine receptor (nAChR).
   2. **Endpoint:** Binding to the nicotinic acetylcholine receptor.
   3. **Comment on endpoint:**
   4. **Endpoint units:** The model makes binary (yes/no) predictions to identify compounds that have the potential to bind the nAChR or not.
   5. **Dependent variable:** *In vitro* data for compounds binding to the nAChR.
   6. **Experimental protocol:** *In vitro* binding data (displacement of ligand, direct binding or functional) using isolated protein or cell-based systems.
   7. **Endpoint data quality and variability:** All data were manually verified to ensure adequacy. Data curation and sources are detailed in the relevant publication (Wijeyesakere et al., 2020; listed in section 2.7).
2. **Defining the algorithm - OECD Principle 2**

**4.1 Type of model:** Machine-learning model that predicts binding potential based on structural fingerprints.

**4.2 Explicit algorithm:** Random-forest machine learning model.

**4.3 Descriptors in the model:** 2D MACCS structural fingerprints.

**4.4 Descriptor selection:** *A priori* selection.

**4.5 Algorithm and descriptor generation:** Calculated within KNIME using the CDK fingerprints node.

**4.6 Software name and version for descriptor generation:** KNIME ver. 4.4.1.

**4.7 Chemicals/Descriptors ratio:** N/A.

**5. Defining the applicability domain - OECD Principle 3**

**5.1 Description of the applicability domain of the model:** Single organic molecules (excluding organometallics) with defined 2D structures not containing As, B, Sn, Si or Zn and having molecular weights ≤ 3200 Da (based on compounds in training set).

**5.2 Method used to assess the applicability domain:** Manually defined by the model developer.

**5.3 Software name and version for applicability domain assessment:** N/A.

**5.4 Limits of applicability:** N/A.

**5. Internal validation - OECD Principle 4**

**6.1 Availability of the training set:** Yes – available in the cited publication (Wijeyesakere et al., 2020; listed in section 2.7).

**6.2 Available information for the training set:**

Chemical identifier: Yes – CAS and/or CHEMBL ID and/or Name (when available).

Smiles: Yes.

Formula: No.

INChi: No.

MOL file: No.

**6.3 Data for each descriptor variable for the training set:** MACCS fingerprints can be calculated from the provided SMILES structure.

**6.4 Data for the dependent variable for the training set:** Yes – binary (yes/no) descriptor for binding.

**6.5 Other information about the training set:** Detailed in the cited publication (Wijeyesakere et al., 2020; listed in section 2.7).

**6.6 Pre-processing of data before modelling:** Calculation of MACCS fingerprint for each structure based on its [SMILES] structure.

**6.7 Statistics for goodness-of-fit:**

**6.8 Robustness - Statistics obtained by leave-one-out cross-validation:** N/A.

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

Model built with 90% of actives and controls, with 10% being reserved as a test set (repeated 5 times):

|  |  |
| --- | --- |
| **Concordance measure** | **Mean value** |
| Sensitivity | 98.3% |
| Specificity | 94.3% |
| Positive predictive value | 97.2% |
| Negative predictive value | 96.5% |
| Balanced accuracy | 96.3% |
| Youden index | 0.926 |
| F-measure | 0.977 |
| Cohen's kappa | 0.931 |

**6.10 Robustness - Statistics obtained by Y-scrambling:** N/A.

**6.11 Robustness - Statistics obtained by bootstrap:**

Concordance statistics from production model using all curated data (out-of-bag estimates):

|  |  |
| --- | --- |
| **Concordance measure** | **Mean value** |
| Sensitivity | 98.1% |
| Specificity | 99.6% |
| Balanced accuracy | 98.9% |

**6.12 Robustness - Statistics obtained by other methods:** N/A.

**7** **External validation - OECD Principle 4**

**7.1 Availability of the external validation set:** No.

**7.2 Available information for the external validation set:**

Chemical identifier: No.

Smiles: No.

Formula: No.

INChI: No.

MOL file: No.

**7.3 Data for each descriptor variable for the external validation set:** No.

**7.4 Data for the dependent variable for the external validation set:** No.

**7.5 Other information about the external validation set:** N/A.

**7.6 Experimental design of test set:** N/A.

**7.7 Predictivity - Statistics obtained by external validation:** N/A.

**7.8 Predictivity - Assessment of the external validation set:** N/A.

**7.9 Comments on the external validation of the model:** N/A.

**8 Providing a mechanistic interpretation - OECD Principle 5**

**8.1 Mechanistic basis of the model:** Model is trained on *in vitro* binders and can be used to align novel compounds to the target [nAChR].

**8.2 A priori or a posteriori mechanistic interpretation:** The mechanistic basis of the model was developed *a priori* by manually curating *in vitro* binding data prior to developing the machine-learning model.

**8.3 Other information about the mechanistic interpretation:** All references supporting the development and mechanistic basis for the model are detailed in the references detailed above.

**9 Miscellaneous information**

**9.1 Comments:** This model can be used to screen large databases or single compounds for their potential to bind the nAChR.

**9.2 Bibliography:**

Wijeyesakere, S.J., Wilson, D.M., and Marty, M.S. (2020). Prediction of Cholinergic Compounds by Machine-Learning.” *Computational Toxicology.* 13: 100119. DOI: 10.1016/j.comtox.2020.100119.