GT_EXPERT

	Element	Explanation
1.	**QSAR identifier**	
1.1.	QSAR identifier (title)	GT_Expert Expert rule-based model for chemical-induced mutagenicity version 1.9.2.0.18473.500
1.2	Other related models	GT1_BMUT OECD 471 bacterial mutagenicity model version 1.9.2.0.15546.500
1.3.	Software coding the model	**Name** CASE Ultra Version 1.9.2.0 **URL** http://www.multicase.com/case-ultra **Description** QSAR based bioactivity and toxicity prediction software **Contact** sales@multicase.com, MultiCASE Inc, 5885 Landerbrook Dr. #210 Mayfield Heights, OH 44124 USA www.multicase.com
2.	**General information**	
2.0	Abstract	GT_Expert is an Expert rule-based model for chemical-induced mutagenicity.
2.1.	Date of QMRF	June 13, 2022
2.2.	QMRF author(s) and contact details	**Name** Dr Roustem D Saiakhov **Affiliation** MultiCASE Inc., **Contact** +1-440-565-7221 **URL** www.multicase.com **Email** saiakhov@multicase.com **Name** Mounika Girireddy **Affiliation** MultiCASE Inc., **Contact** +1-440-565-7221 **URL** www.multicase.com **Email** girireddy@multicase.com
2.3.	Date of QMRF update(s)	October 10, 2024
2.4.	QMRF update(s)	Mounika Girireddy, entire document.
2.5.	Model developer(s) and contact details	**Name** Model was constructed at MultiCASE Inc. **Affiliation** MultiCASE Inc., **Contact** +1-440-565-7221 **URL** www.multicase.com **Email** sales@multicase.com

2.6.	Date of model development and/or publication	The original GT_EXPERT model was developed in 2015; Published in 2015; Updated in 2020, 2022 and 2024.
2.7.	Reference(s) to main scientific papers and/or software package	1. Chakravarti SK, Saiakhov RD. MultiCASE Platform for In Silico Toxicology. https://pubmed.ncbi.nlm.nih.gov/35188644/ 2. Catrin Hasselgren, Joel Bercu, Alex Cayley, Kevin Cross, Susanne Glowienke, Naomi Kruhlak, Wolfgang Muster, John Nicolette 8, M Vijayaraj Reddy, Roustem Saiakhov, Krista Dobo Management of pharmaceutical ICH M7 (Q)SAR predictions - The impact of model updates. https://pubmed.ncbi.nlm.nih.gov/33058939/ 3. Honma M, Kitazawa A, Cayley A, Williams RV, Barber C, Hanser T, Saiakhov R, Chakravarti S, Myatt GJ, Cross KP, Benfenati E, Raitano G, Mekenyan O, Petkov P, Bossa C, Benigni R, Battistelli CL, Giuliani A, Tcheremenskaia O, DeMeo C, Norinder U, Koga H, Jose C, Jeliazkova N, Kochev N, Paskaleva V, Yang C, Daga PR, Clark RD, Rathman J. Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. https://pubmed.ncbi.nlm.nih.gov/30357358/
2.8.	Availability of information about the model	Model is commercial. Although the training set is not publicly available, Information about the non-proprietary training set chemicals, assay conditions and details, information about the alerts are available through CASE Ultra interface. For any other specific details contact: sales@multicase.com, MultiCASE Inc. 5885 Landerbrook Dr. #210 Mayfield Heights, OH 44124 USA. Phone: +1-440-565-7221.
2.9.	Availability of another QMRF for exactly the same model	None
3	**Defining the endpoint - OECD Principle 1: "A DEFINED ENDPOINT"**	**PRINCIPLE 1: "A DEFINED ENDPOINT". ENDPOINT refers to any physicochemical, biological, or environmental property/activity/effect that can be measured and therefore modelled. The intent of PRINCIPLE 1 (a (Q)SAR should be associated with a defined endpoint) is to ensure clarity in the endpoint being predicted by a given model, since a given endpoint could be determined by different experimental protocols and under different experimental conditions. It is therefore important to identify the experimental system and test conditions that is being modelled by the Q)SAR.**
3.1.	Species	Salmonella and E.Coli composite: Combination of results from the S. typhimurium histidine reversion gene mutation test using tester strains TA97, TA98, TA100, TA102, TA1535, TA1537 and E.Coli per OECD471.

3.2.	Endpoint	**Group** Bacterial Mutagenicity **Name** Salmonella and EColi Mutagenicity **Protocol** As described in Alexander Sedykh, Suman Chakravarti, and Roustem Saiakhov. MultiCASE rule-based expert system for mutagenicity prediction: creating and validating. Genetic Toxicity Association Annual Meeting, 2015, Delaware. Poster presentation.
3.3	Comment on endpoint	As per OECD 471 guideline.
3.4.	Endpoint units	Binary score
3.5.	Dependent variable	Negative 0, Positive 1
3.6.	Experimental protocol	As described in Alexander Sedykh, Suman Chakravarti, and Roustem Saiakhov. MultiCASE rule-based expert system for mutagenicity prediction: creating and validating. Genetic Toxicity Association Annual Meeting, 2015, Delaware. Poster presentation.
3.7.	Endpoint data quality and variability	As described in Alexander Sedykh, Suman Chakravarti, and Roustem Saiakhov. MultiCASE rule-based expert system for mutagenicity prediction: creating and validating. Genetic Toxicity Association Annual Meeting, 2015, Delaware. Poster presentation.
4	**Defining the algorithm - OECD Principle 2 : "AN UNAMBIGUOUS ALGORITHM"**	**PRINCIPLE 2: "AN UNAMBIGUOUS ALGORITHM". The (Q)SAR estimate of an endpoint is the result of applying an ALGORITHM to a set of structural parameters which describe the chemical structure. The intent of PRINCIPLE 2 (a (Q)SAR should be associated with an unambiguous algorithm) is to ensure transparency in the model algorithm that generates predictions of an endpoint from information on chemical structure and/or physicochemical properties. In this context, algorithm refers to any mathematical equation, decision rule or output approach.**
4.1.	Type of model	Expert rule based QSAR model.
4.2.	Explicit algorithm	**Definition** Rule based QSAR. **Description** Set of rules, based on substructures identified by experts as related to bacterial mutagenicity.
4.3.	Descriptors in the model	**Name** Structural Fragments. **Units** Probability **Description** Each expert rule (i.e., structural alert) has its mutagenic potential (probability, %) value assigned based on the reference database of bacterial mutagenicity outcomes (n=18473).
4.4.	Descriptor selection	GT_EXPERT, is a rule-based system of bacterial mutagenicity, which consists of 245 structural alerts, of which 81 represent general mechanisms and the remainder accompanies them as refining factors. These expert alerts were collected from published reviews and scientific studies and were

		refined and benchmarked on a reference set of over 21,000 chemicals with known bacterial mutagenicity outcomes.
4.5.	Algorithm and descriptor generation	Expert rules were grouped by generic mechanisms. Each such group could comprise several, more specific alerts (activating or deactivating factors).
4.6.	Software name and version for descriptor generation	**Name** CASE Ultra Version 1.9.2.0 **URL** http://www.multicase.com/case-ultra **Description** QSAR based bioactivity and toxicity prediction software. **Contact** sales@multicase.com, MultiCASE Inc, 5885 Landerbrook Dr. #210 Mayfield Heights, OH 44124 USA www.multicase.com
4.7.	Chemicals/Descriptors ratio	Number of Chemicals = 18473 (7316 positives/11157 negatives) Number of Descriptors = 245
5	**Defining the applicability domain - OECD Principle 3: "A DEFINED DOMAIN OF APPLICABILITY"**	**PRINCIPLE 3: "A DEFINED DOMAIN OF APPLICABILITY". APPLICABILITY DOMAIN refers to the response and chemical structure space in which the model makes predictions with a given reliability. Ideally the applicability domain should express the structural, physicochemical and response space of the model. The CHEMICAL STRUCTURE (x variable) space can be expressed by information on physicochemical properties and/or structural fragments. The RESPONSE (y variable) can be any physicochemical, biological or environmental effect that is being predicted. According to PRINCIPLE 3 a (Q)SAR should be associated with a defined domain of applicability. Section 5 can be repeated (e.g., 5.a, 5.b, 5.c, etc) as many times as necessary if more than one method has been used to assess the applicability domain.**
5.1.	Description of the applicability domain of the model	Applicability domain based on the fragment dictionary, created for every model.
5.2.	Method used to assess the applicability domain	The CASE Ultra program evaluates automatically whether a tested molecule is within the domain of applicability of the module it is tested with. Because of the complexity of the system it is rather difficult to estimate if the tested chemical belongs to the domain of applicability before the actual test takes place. However, since all the structures, names and CAS numbers (if available) are listed for every training set, the user can assess if the chemicals-to-be-tested belong to the chemical space which was defined by the learning set before performing the CASE Ultra assessment. In the prediction mode CASE Ultra evaluates the potentially relevant effect of all parts of the tested molecule. If any fragment of the tested molecule does not exist in any of the compounds of the learning set or exists within a different environment, then a warning is issued advising the user that the molecule is somewhat out of the domain of applicability of the model. In addition, the program uses the range of descriptor values in training set with ±30% confidence. Descriptor values must fall between maximal and minimal descriptor values of training set ±30%.

5.3.	Software name and version for applicability domain assessment	**Name** CASE Ultra Version 1.9.2.0 **URL** http://www.multicase.com/case-ultra **Description** QSAR expert system for in-silico prediction of toxicity and bioactivity of chemicals. **Contact** sales@multicase.com, MultiCASE Inc, 5885 Landerbrook Dr. #210 Mayfield Heights, OH 44124 USA www.multicase.com
5.4.	Limits of applicability	Inorganic compounds, mixtures and large biomolecules are not covered.
6	**Defining goodness-of-fit and robustness (internal validation) – OECD Principle 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY"**	**PRINCIPLE 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY". PRINCIPLE 4 expresses the need to perform validation to establish the performance of the model. GOODNESS-OF-FIT and ROBUSTNESS refer to the internal model performance.**
6.1.	Availability of the training set	Training set data and associated references can be seen in CASE Ultra interface under "Display Training Set Chemicals".
6.2.	Available information for the training set	a) Chemical names (common names and/or IUPAC names) b) CAS c) SMILES d) Mol
6.3.	Data for each descriptor variable for the training set	Some data available.
6.4.	Data for the dependent variable for the training set	Some data available.
6.5.	Other information about the training set	The model is based on a reference database of 18473 unique chemicals with overall bacterial mutagenicity outcomes harmonized from multiple sources • There were 7316 positive and 11157 negative activity entries. • Data sources included Japan NIH, NTP, CCRIS, GENETOX, and RTECS databases, as well as datasets from US FDA.
6.6.	Pre-processing of data before modelling	Extensive data curation and verification name-registry number-structure relationship.
6.7.	Statistics for goodness-of-fit	Sensitivity 82.6% Specificity 88.3% Positive predictivity 82.3% Negative predictivity 88.5%, Coverage 94.8% AUC 86.0%, Classification cut-off 0.50
6.8.	Robustness - Statistics obtained by leave-one-out cross-validation	Not performed.
6.9.	Robustness - Statistics obtained by leave-many-out cross-validation	Not performed.

6.10.	Robustness - Statistics obtained by Y-scrambling	Not performed.
6.11.	Robustness - Statistics obtained by bootstrap	Not performed.
6.12.	Robustness - Statistics obtained by other methods	Not performed.
7	**Defining predictivity (external validation) – OECD Principle 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY"**	**PRINCIPLE 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY". PRINCIPLE 4 expresses the need to perform validation to establish the performance of the model. PREDICTIVITY refers to the external model validation. Section 7 can be repeated (e.g., 7.a, 7.b, 7.c, etc) as many times as necessary if more validation studies need to be reported in the QMRF.**
7.1.	Availability of the external validation set	Not available.
7.2.	Available information for the external validation set	a) Chemical names (common names and/or IUPAC names) b) CAS c) SMILES
7.3.	Data for each descriptor variable for the external validation set	Some data available.
7.4.	Data for the dependent variable for the external validation set	Some data available.
7.5.	Other information about the external validation set	The test set is 1323 chemicals randomly taken out from the internal mutagenicity database.
7.6.	Experimental design of test set	1323 chemicals (308 Positives, 1015 Negatives).
7.7.	Predictivity - Statistics obtained by external validation	Sensitivity 75.0% Specificity 86.0% Positive predictivity 82.0% Negative predictivity 79.0%, Coverage 89.0% AUC 0.8113, Classification cut-off 0.50
7.8.	Predictivity - Assessment of the external validation set	This external validation set is well balanced and sufficiently represents the structural domain. However, the experimental design of the set and labeling is questionable for some compounds and contradicts the regulatory practice.
7.9.	Comments on the external validation of the model	Not available.
8	**Providing a mechanistic interpretation - OECD Principle 5: "A MECHANISTIC INTERPRETATION, IF POSSIBLE"**	**PRINCIPLE 5: "A MECHANISTIC INTERPRETATION, IF POSSIBLE". According to PRINCIPLE 5, a (Q)SAR should be associated with a mechanistic interpretation, if possible.**

8.1.	Mechanistic basis of the model	These expert alerts were collected from published reviews and scientific studies and were refined and benchmarked on a reference set of over 21,000 chemicals with known bacterial mutagenicity outcomes. Expert rules were grouped by generic mechanisms. Each such group could comprise several, more specific alerts (activating or deactivating factors).
8.2.	A priori or a posteriori mechanistic interpretation	None.
8.3.	Other information about the mechanistic interpretation	None.
9	**Miscellaneous information**	
9.1.	Comments	Model is compatible with all 1.9.x.x versions of CASE Ultra.
9.2.	Bibliography	1. Optimizing predictive performance of CASE Ultra expert system models using the applicability domains of individual toxicity alerts; Chakravarti, S.K., Saiakhov, R.D. and Klopman, G., Journal of Chemical Information and Modeling, 2012, 52, 2609-2618. DOI: 10.1021/ci300111r. 2. Effectiveness of CASE Ultra Expert System in Evaluating Adverse Effects of Drugs; Saiakhov, R.D., Chakravarti, S.K. and Klopman, G.; Molecular Informatics, 2012, 32, 87-97. DOI : 10.1002/minf.201200081. 3. Computing similarity between structural environments of mutagenicity alerts, Chakravarti, S.K.,Saiakhov, R. D.; Mutagenesis,October20,2018,DOI: https://doi.org/10.1093/mutage/gey032 4. Kim M.T., Sedykh A., Chakravarti S.K., Saiakhov R.D., Zhu H. Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches. https://www.ncbi.nlm.nih.gov/pubmed/24306326
9.3	Supporting information	Not available.