Leadscope Enterprise version of commercial CASE Ultra model A61 for Chromosome Aberrations in Chinese Hamster Ovary cells *in vitro* (NTP data)

1. QSAR identifier

1.1 QSAR identifier (title)

Leadscope Enterprise version of commercial CASE Ultra model A61 for Chromosome Aberrations in Chinese Hamster Ovary cells *in vitro* (NTP data), Danish QSAR Group at DTU Food.

1.2 Other related models

MultiCASE CASE Ultra commercial model A61 for Chromosome Aberrations in Chinese Hamster Ovary cells *in vitro* (NTP data), Danish QSAR Group at DTU Food.

SciMatics SciQSAR version of commercial CASE Ultra A61 for Chromosome Aberrations in Chinese Hamster Ovary cells *in vitro* (NTP data), Danish QSAR Group at DTU Food.

1.3. Software coding the model

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

2. General information

2.1 Date of QMRF

January 2015.

2.2 QMRF author(s) and contact details

QSAR Group at DTU Food;

Danish National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Marianne Dybdahl;

National Food Institute at the Technical University of Denmark;

Sine Abildgaard Rosenberg;

National Food Institute at the Technical University of Denmark;

2.3 Date of QMRF update(s)

2.4 QMRF update(s)

2.5 Model developer(s) and contact details

MultiCASE Inc.;

23811 Chagrin Blvd Ste 305, Beachwood, OH, 44122, USA;

www.multicase.com

MultiCASE Inc. has kindly given their permission that remodelling of their training set for the commercial A61 model in Leadscope was performed by:

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Danish QSAR Group at DTU Food;

National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

2.6 Date of model development and/or publication

January 2014.

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

Roberts, G., Myatt, G. J., Johnson, W. P., Cross, K. P., and Blower, P. E. J. (2000) LeadScope: Software for Exploring Large Sets of Screening Data. *Chem. Inf. Comput. Sci.*, 40, 1302-1314.

Cross, K.P., Myatt, G., Yang, C., Fligner, M.A., Verducci, J.S., and Blower, P.E. Jr. (2003) Finding Discriminating Structural Features by Reassembling Common Building Blocks. *J. Med. Chem.*, *46,* 4770-4775.

Valerio, L. G., Yang, C., Arvidson, K. B., and Kruhlak, N. L. (2010) A structural feature-based computational approach for toxicology predictions. *Expert Opin. Drug Metab. Toxicol.*, 6:4, 505-518.

2.8 Availability of information about the model

The training set is proprietary and commercially available from MultiCASE Inc. It was originally compiled by MultiCASE Inc. and used to train the commercial MultiCASE A61 model. The Danish QSAR Group bought this model from MultiCASE Inc. in 1999. Permission to remodel the training set in Leadscope was kindly granted by MultiCASE Inc. The model algorithm is proprietary from commercial software.

2.9 Availability of another QMRF for exactly the same model

3. Defining the endpoint

3.1 Species

Chinese Hamster (Chinese Hamster Ovary (CHO) cells).

3.2 Endpoint

QMRF 4.10. Mutagenicity

OECD 473 In Vitro Mammalian Chromosome Aberration Test

3.3 Comment on endpoint

The chromosome aberration test using cultured mammalian cells is one of the sensitive methods to predict environmental mutagens and/or carcinogens, and is a complementary test to the *Salmonella typhimurium* mutagenicity assay.The purpose of the *in vitro* chromosome aberration test is to identify agents that cause structural chromosome aberrations in cultured mammalian cells arrested in metaphase. The structural aberrations detected may be of two types, chromosome (i.e. breakage, or breakage and reunion, of both chromatids at an identical site) or chromatid (i.e. breakage of single chromatids or breakage and reunion between chromatids). With the majority of chemical mutagens, induced aberrations are of the chromatid type, but chromosome-type aberrations also occur.
Chromosome aberrations and related events are the cause of many human genetic diseases and there is substantial evidence that chromosome aberrations and related events causing alterations in oncogenes and tumor suppressor genes of somatic cells are involved in cancer induction in humans and experimental animals. Chromosome aberration *in vitro* tests have been used as an effective screen for chemicals which may have mutagenic, teratogenic, or tumorigenic potential. The *in vitro* assay systems for clastogenicity (i.e. any process resulting in the breakage of chromosomes or the loss or rearrangement of pieces of chromosomes) testing have certain advantages over *in vivo* systems such as, cells of human origin can be used if desired, a chemical can be tested for both direct effect and in the presence of metabolic activating systems, active but short-lived metabolites can be more easily detected, tests can be repeated with the same or different cell types under the same experimental conditions, and numerical aberrations – such as aneuploidy (i.e. abnormal number of chromosomes) and polyploidy (i.e. more than two paired (homologous) sets of chromosomes) - are more easily detected.
The mammalian chromosome aberration *in vitro* test is used to screen for possible mammalian mutagens and carcinogens. Many compounds that are positive in this test are mammalian carcinogens; however, there is not a perfect correlation between this test and carcinogenicity. Correlation is dependent on chemical class and there is increasing evidence that there are carcinogens that are not detected by this test because they appear to act through mechanisms other than direct DNA damage (i.e., non-genotoxic carcinogens). I addition, it is important to be aware that positive results may arise from changes in pH, osmolality or high levels of cytotoxicity and do not reflect intrinsic mutagenicity (OECD guideline 473, 1997).
In the fact sheet for the A61 model (personal communication with MultiCASE in 2001), MultiCASE Inc. refers to two publications of the MultiCASE A61 model (Rosenkranz *et al.* 1990, Liu *et al.* 1997). The compilation of the training set has been described by Rosenkranz *et al.* (1990) and the training set consists of chromosome aberration results from US National Toxicology Program (NTP) (Galloway *et al.* 1985, 1987, Gulati *et al.* 1989, Loveday *et al.* 1989).

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Chromosome aberrations in Chinese Hamster Ovary cells, positive or negative.

3.6 Experimental protocol

The training set data were compiled by MultiCASE Inc. from US NTP and consist of results from the chromosome aberration test in Chinese Hamster Ovary (CHO) Cells. The experimental protocol for this test is described in OECD guideline 473 (1997). Briefly, the cell cultures (a variety of cell lines can be used, e.g. CHO cells) are exposed to the test substance both with and without metabolic activation. At predetermined intervals after exposure of cell cultures to the test substance, they are treated with a metaphase-arresting substance, harvested, stained and metaphase cells are analysed microscopically for the presence of chromosome aberrations.

3.7 Endpoint data quality and variability

As the training set is commercial by MultiCASE Inc. the quality and variability of the data used is unknown.

4. Defining the algorithm

4.1 Type of model

A categorical (Q)SAR model based on structural features and numeric molecular descriptors.

4.2 Explicit algorithm

This is a categorical (Q)SAR model made by use of partial logistic regression (PLR). The specific implementation is proprietary within the Leadscope software.

4.3 Descriptors in the model

structural features,

aLogP,

polar surface area,

number of hydrogen bond donors,

Lipinski score,

number of rotational bonds,

parent atom count,

parent molecular weight,

number of hydrogen bond acceptors

4.4 Descriptor selection

Leadscope Predictive Data Miner is a software program for systematic sub-structural analysis of a chemical using predefined structural features stored in a template library, training set-dependent generated structural features (scaffolds) and calculated molecular descriptors. The feature library contains approximately 27,000 pre-defined structural features and the structural features chosen for the library are motivated by those typically found in small molecules: aromatics, heterocycles, spacer groups, simple substituents. Leadscope allows for the generation of training set-dependent structural features (scaffold generation), and these features can be added to the pre-defined structural features from the library and be included in the descriptor selection process. It is possible in Leadscope to remove redundant structural features before the descriptor selection process and only use the remaining features in the descriptor selection process. Besides the structural features Leadscope also calculates eight molecular descriptors for

each training set structure: the octanol/water partition coefficient (alogP), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), Lipinski score, atom count, parent compound molecular weight, polar surface area (PSA) and rotatable bonds. These eight molecular descriptors are also included in the descriptor selection process.

Leadscope has a default automatic descriptor selection procedure. This procedure selects the top 30% of the descriptors (structural features and molecular descriptors) according to X^2-test for a binary variable, or the top and bottom 15% descriptors according to *t*-test for a continuous variable. Leadscope treats numeric property data as ordinal categorical data. If the input data is continuous such as IC_{50} or cLogP data, the user can determine how values are assigned to categories: the number of categories and the cut-off values between categories. (Roberts *et al.*2000).

When developing this model, intermediate models with application of different modelling approaches in Leadscope were tried:

1. 'Single model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors for descriptor selection.
2. 'Single model' using both the Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.
3. 'Single model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation), with subsequent removal of redundant structural features, and calculated molecular descriptors for descriptor selection.
4. 'Composite model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors in the descriptor selection.
5. 'Composite model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.

Based on model performance as measured by a preliminary cross-validation the model developed using approach number 2. was chosen.

For this model scaffolds were generated by Leadscope for the training set structures and added to the Leadscope library of structural features. Descriptors were then automatically selected among the structural features and the eight molecular descriptors.

4.5 Algorithm and descriptor generation

For descriptor generation see 4.4.

After selection of descriptors the Leadscope Predictive Data Miner program performs partial least squares (PLS) regression for a continuous response variable, or partial logistic regression (PLR) for a binary response variable, to build a predictive model. By default the Predictive Data Miner performs leave-one-out or leave-groups-out (in the latter case, the user can specify any number of repetitions and percentage of structures left out) cross-validation on the training set depending on the size of the training set. In the cross-validation

made by Leadscope the descriptors selected for the 'mother model' are used when building the validation submodels and they therefore have a tendency to be overfittet and give overoptimistic validation results.

In this model, because of the categorical outcome in the response variable, PLR was used to build the predictive model. For this model 172 descriptors were selected to build the model. These include 8 Leadscope calculated molecular descriptors, 127 hierarchy features, 4 dynamic features and 33 scaffolds. The 172 descriptors were distributed on 1 PLS factor.

4.6 Software name and version for descriptor generation

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

4.7 Descriptors/chemicals ratio

In this model 172 descriptors were used and distributed on 1 PLS factor. The training set consists of 229 compounds. The descriptor/chemical ratio is 1:1.3 (172:229).

5. Defining Applicability Domain

5.1 Description of the applicability domain of the model

The definition of the applicability domain consists of two components; the definition of a structural domain in Leadscope and the in-house further probability refinement algorithm on the output from Leadscope to reach the final applicability domain call.

1. Leadscope

For assessing if a test compound is within the structural applicability domain of a given model Leadscope examines whether the test compound bears enough structural resemblance to the training set compounds used for building the model (i.e. a structural domain analysis). This is done by calculating the distance between the test compound and all compounds in the training set (distance = 1 - similarity). The similarity score is based on the Tanimoto method. The number of neighbours is defined as the number of compounds in the training set that have a distance equal to or smaller than 0.7 with respect to the test compound. The higher the number of neighbours, the more reliable the prediction for the test compound. Statistics of the distances are also calculated. Effectively no predictions are made for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated in Leadscope.

2. The Danish QSAR group

In addition to the general Leadscope structural applicability domain definition the Danish QSAR group has applied a further requirement to the applicability domain of the model. That is only positive predictions with a probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted. Predictions within the structural applicability domain but with probability between 0.5 to 0.7 or 0.3 to 0.5 are defined as positives out of applicability domain and negatives out of applicability domain, respectively. When these predictions are wed out the performance of the model in general increases at the expense of reduced model coverage.

5.2 Method used to assess the applicability domain

Leadscope does not generate predictions for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated.

Only positive predictions with probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted.

5.3 Software name and version for applicability domain assessment

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

5.4 Limits of applicability

The Danish QSAR group applies an overall definition of structures acceptable for QSAR processing which is applicable for all the in-house QSAR software, i.e. not only CASE Ultra. According to this definition accepted

structures are organic substances with an unambiguous structure, i.e. so-called discrete organics defined as: organic compounds with a defined two dimensional (2D) structure containing at least two carbon atoms, only certain atoms (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, and I), and not mixtures with two or more 'big components' when analyzed for ionic bonds (for a number of small known organic ions assumed not to affect toxicity the 'parent molecule' is accepted). Calculation 2D structures (SMILES and/or SDF) are generated by stripping off ions (of the accepted list given above). Thus, all the training set and prediction set chemicals are used in their non-ionized form. See 5.1 for further applicability domain definition.

6. Internal validation

6.1 Availability of the training set

No

6.2 Available information for the training set

SMILES

6.3 Data for each descriptor variable for the training set

No

6.4 Data for the dependent variable for the training set

No

6.5 Other information about the training set

229 compounds are in the training set: 95 positives and 134 negatives.

6.6 Pre-processing of data before modelling

As the training set is commercial by MultiCASE Inc. the pre-processing of data is unknown.

Only structures acceptable for Leadscope were used in the final training set. That is only discrete organic chemicals as described in 5.4 were used. In case of replicate structures, one of the replicates was kept if all the compounds had the same activity and all were removed if they had different activity. No further structures accepted by the software were eliminated (i.e. outliers).

6.7 Statistics for goodness-of-fit

Not performed.

6.8 Robustness – Statistics obtained by leave-one-out cross-validation

Not performed. (It is not a preferred measurement for evaluating large models).

6.9 Robustness – Statistics obtained by leave-many-out cross-validation

A five times two-fold 50 % cross-validation was performed. This was done by randomly removing 50% of the full training set used to make the “mother model”, where the 50% contains the same ratio of positive and negatives as the full training set. A new model (validation submodel) was created on the remaining 50% using the same settings in Leadscope but with no information from the “mother model” regarding descriptor selection etc. The validation submodel was applied to predict the removed 50% (within the defined applicability domain for the submodel). Likewise, a validation submodel was made on the removed 50% of the training set and this model was used to predict the other 50% (within the defined applicability domain for this submodel). This procedure was repeated five times.

Predictions within the defined applicability domain for the ten validation submodels were pooled and Cooper’s statistics calculated. This gave the following results for the 44.0 % (504/(5*229)) of the predictions which were within the applicability domain:

- Sensitivity (true positives / (true positives + false negatives)): 54.1%
- Specificity (true negatives / (true negatives + false positives)): 79.3%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 68.8%

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. External validation

7.1 Availability of the external validation set

7.2 Available information for the external validation set

7.3 Data for each descriptor variable for the external validation set

7.4 Data for the dependent variable for the external validation set

7.5 Other information about the training set

7.6 Experimental design of test set

7.7 Predictivity – Statistics obtained by external validation

7.8 Predictivity – Assessment of the external validation set

7.9 Comments on the external validation of the model

External validation has not been performed for this model.

8. Mechanistic interpretation

8.1 Mechanistic basis of the model

The global model identifies structural features and molecular descriptors which in the model development was found to be statistically significant associated with effect. Many predictions may indicate modes of action that are obvious for persons with expert knowledge for the endpoint.

8.2 A priori or posteriori mechanistic interpretation

A posteriori mechanistic interpretation. The identified structural features and molecular descriptors may provide basis for mechanistic interpretation.

8.3 Other information about the mechanistic interpretation

9. Miscellaneous information

9.1 Comments

The model can be used to predict results for the chromosome aberration test in Chinese Hamster Ovary (CHO) cells *in vitro*.

9.2 Bibliography

Ishidate, M. Jr., Miura, K.F., and Sofuni, T. (1998) Chromosome aberration assays in genetic toxicology testing in vitro. *Mutation Research*, 404, 167–172.

Galloway, S.M., Bloom, A.D., Resnick, M., Margolin, B.H., Nakamura, F., Archer, P. and Zeiger, E. (1985) Development of a standard protocol for *in vitro* cytogenetic testing with Chinese hamster ovary cells: Comparison of results for 22 compounds in two laboratories. *Environ. Mutagenesis,* 7, 1-51.

Galloway, S.M., Armstrong, M.J., Reuben, C., Colman, S., Brown, B., Cannon, C, Bloom, A.D., Nakamura, F., Ahmed, M., Duk, S., Rimpo, J., Margolin, B.H., Resnick, M.A., Anderson, B. and Zeiger, E. (1987) Chromosome aberrations and sister chromatid exchanges in Chinese hamster ovary cells. Evaluation of 108 chemicals. *Environ. Mol. Mutagenesis,* 10 (Suppl. 10), 1-175.

Gulati, D.K., Witt, K., Anderson, B., Zeiger, E. and Shelby, M.D. (1989) Chromosome aberration and sister chromatid exchange tests in Chinese hamster ovary cells *in vitro.* III: Results with 27 chemicals. *Environ. Mol. Mutagenesis*, 13, 133-193.

Liu, M., Grant, S.G., Macina, O.T., Klopman, G., and Rosenkranz, H.S. (1997) Structural and mechanistic bases for the induction of mitotic chromosomal loss and duplication ('malsegregation') in the yeast Saccharomyces cerevisiae: relevance to human carcinogenesis and developmental toxicology. *Mutation Research,* 374, 209-231.

Loveday, K.S., Lugo, M.H., Resnick, M.A., Anderson, B.E. and Zeiger, E. (1989) Chromosome aberrations and sister chromatid exchanges in Chinese hamster ovary cells in vitro: II. Results with 20 chemicals. *Environ. Mol. Mutagenesis,* 13, 60-94.

OECD guideline 473 (1997) OECD Guidelines for the Testing of Chemicals No. 473, *In Vitro* Mammalian Chromosome Aberration Test. Organisation for Economic Cooperation and Development; Paris, France. Available online at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Rosenkranz, H.S., Ennever, F.K., and Klopman, G. (1990) Relationship between carcinogenicity in rodents and the induction of sister chromatid exchanges and chromosomal aberrations in Chinese hamster ovary cells. *Mutagenesis,* 5, 559-571.

9.3 Supporting information