	QMRF identifier (JRC Inventory): **Q13-410-0070**
	QMRF Title: **MultiCASE model for in vitro chromosome aberration in mammalian (CHL) cells**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

MultiCASE model for in vitro chromosome aberration in mammalian (CHL) cells

1.2.Other related models:

MultiCASE model for in vitro chromosome aberration in mammalian (CHO) cells

1.3.Software coding the model:

MultiCASE MC4PC v. 2005

http://www.multicase.com

2.General information

2.1.Date of QMRF:

18 February 2011

2.2.QMRF author(s) and contact details:

[1]Tine Ringsted Danish National Food Institute Mørkhøj Bygade 19, 2860 Søborg, Denmark
tiri@food.dtu.dk

[2]Gunde Egeskov Jensen National Food Institute Mørkhøj Bygade 19, 2860 Søborg, Denmark
gunje@food.dtu.dk

[3]Jay Russel Niemelä National Food Institute Mørkhøj Bygade 19, 2860 Søborg, Denmark
jarn@food.dtu.dk

[4]Nikolai Nikolov National Food Institute Mørkhøj Bygade 19, 2860 Søborg, Denmark
nign@food.dtu.dk

[5]Eva Bay Wedebye National Food Institute Mørkhøj Bygade 19, 2860 Søborg, Denmark
ebawe@food.dtu.dk

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

[1]Eva Bay Wedebye Danish National Food Institute Mørkhøj Bygade 19, 2860 Søborg, Denmark
ebawe@food.dtu.dk

[2]Jay Russel Niemelä Danish National Food Institute Mørkhøj Bygade 19, 2860 Søborg, Denmark
jarn@food.dtu.dk

2.6.Date of model development and/or publication:

Year of development: 2004. Year of publication: preliminary model published in 2004 (Niemelä & Wedebye, 2004).

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

[1]Niemelä J & Wedeby E (2004). Evaluation of the setubal principles for establishing the status of development and validation of (Q)SARs, Annex 4, A “global” MULTI-CASE model for in vitro chromosomal aberrations in mammalian cells. pp 113-133 in: OECD Environment Health and Safety Publications, Series on Testing and Assessment, no 49. Report from the expert group on (Quantitative) Structure-Activity Relationships ((Q)SARs) on the principles for the validation of

(Q)SARs.

[2]Klopman G (1992). Multicase, 1. A Hierarchical Computer Automated Structure Evaluation Program. Quantitative Structure-Activity Relationships 11, 176–184.

2.8.Availability of information about the model:

The training set is available in this QMRF and can be remodelled by anyone in MultiCASE or in other systems.

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

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Chinese Hamster Lung Cells

3.2.Endpoint:

4.Human Health Effects 4.10.Mutagenicity

3.3.Comment on endpoint:

The endpoint is in vitro structural chromosome aberrations, visible in light microscopy, in cultured mammalian CHL cells - a fibroblast cell line. The purpose of the in vitro chromosome aberration test is to identify agents that cause structural chromosome aberrations in cultured mammalian cells. Structural aberrations may be of two types, chromosome or chromatid. With the majority of chemical mutagens, induced aberrations are of the chromatid type, but chromosome-type aberrations also occur. Chromosome mutations and related events are the cause of many human genetic diseases and there is substantial evidence that chromosome mutations and related events causing alterations in oncogenes and tumour suppressor genes of somatic cells are involved in cancer induction in humans and experimental animals. Chromosome aberration tests in vitro have been used as an effective screen for chemicals which may have mutagenic, teratogenic, or tumourigenic potential (Ishidate, Mut. Res. 195 (1988) 151-213). The in vitro assay systems for clastogenicity 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 – as aneuploidy and polyploidy - are more easily detected.

3.4.Endpoint units:

In the Data Book (Ishidate 1988, Sofuni 1998) and in Kusakabe et al. (2002) results of the experimental studies were indicated as positive (active) or negative (inactive). MultiCASE CASE units were assigned; 10 for negatives and 45 for positives.

3.5.Dependent variable:

In vitro chromosome aberration (CHL), positive or negative.

3.6.Experimental protocol:

OECD guideline for testing of chemicals no. 473 (OECD, 1997). However, when using historical data, not all tests will have been performed in complete compliance with the newest version of the OECD test guideline. Experimental results were taken from "Data Book of Chromosomal Aberration Test In Vitro" (revised by Ishidate, 1988, and Sofuni, 1998) and the publication "Kusakabe et al 2002", so at least some of the data have been generated before the guideline from 1997.

3.7.Endpoint data quality and variability:

Endpoint data quality not specified. However, the vast majority of the training set data were taken from a single source, the “Data book of chromosomal aberration test in vitro” (Ishidate 1988 and Sofuni 1998) and supplemented with a smaller data set (Kusakabe 2002) continuing the work from the first source. The data source is very detailed and well-documented, and judged by the performance of the (Q)SAR model the quality of the data is high.

The data in the model consists of data from a preliminary model with a training set of 513 chemicals (Niemela & Wedebye, 2004) and data for 87 chemicals used for external validation (Kusakabe et al., 2002) of the preliminary model. Out of 911 substances from the Data Book (Sofuni, 1998), 513 were used to establish the preliminary model. The exclusion criteria used include inorganic status, inadequate SMILES code, etc. A decision was made to include chemicals as being positive if they were active in inducing either aberrations or polyploidy. Polyploidy is not included in the current test guideline (OECD, 1997). 87 chemicals from Kusakabe et al. (2002) completed the training set for the model.
Detailed comments on selection of training set data: Test results for a total of 901 substances are presented in the Data Book by Sofuni 1998. The chemicals were chosen for a variety of reasons, including use in foods. A number fall into the class commonly referred to as UVCB's, or chemicals that cannot be represented by a complete structure diagram and specific molecular formula. These were excluded for the obvious reason that it is impossible to model a chemical for which a structure is not available. However, we found that this is not always a totally unambiguous process, so we made the best judgement we could. Inorganic chemicals were also excluded, as our modelling platform cannot deal with them. A very small number of chemicals were excluded because we were unsure of the true identity (inconsistencies between chemical name, CAS number and structure/molecular weight that we were unable to resolve). A few stereo-isomers with conflicting results were also removed as they cannot be distinguished by SMILES notation (a computer code for 2D structures), which is required by our model system. We made a toxicological decision to include chemicals as being positive if they were active in inducing either aberrations or polyploidy. While the current test guideline does not specify testing for a length of time, which would allow polyploidy to be assessed, much of the CHL data does

and we felt the information was too valuable to lose (18 chemicals). We also decided to retain chemicals even if the test had not been performed both in the presence and absence of metabolic activation. Under current regulatory practices, metabolic activation would be a requirement for all tests. Beyond this we attempted to use the judgement of the authors in their interpretation of the final test result. This included dropping 16 of 18 chemicals that the authors considered inconclusive in repeat tests (we kept two because while they were inconclusive for polyploidy, they were clearly positive for structural aberrations). Seventy-eight chemicals were excluded because the authors considered them False Positive (only active at dose of more than 10 mM where effects could be due to osmotic pressure). As our model system cannot handle salts (e.g. sodium salts, hydrochlorides), further interpretation was necessary. In the majority of cases there was no conflict with regard to results of testing ionised or non-ionised forms. However, in certain cases there were. We decided that for some simple organic acids that were active but where the salt was clearly inactive, to consider these as being inactive in accordance with the advice, given in the OECD Guidelines and Morita et al. (Mutation Research 268, 297-305, 1992), that particularly low pH may lead to false positive predictions. We do not know if this decision is right or wrong in relation to use of results of this in vitro system for predicting in vivo effects, but it will clearly affect the performance of the model. We also made a few decisions based on additional data from the literature: vitamin B2 (Riboflavin, CAS 83-88-5) tested positive in insoluble form, but was negative in soluble form. We retained the negative results, as the mechanism for the insoluble compound appears to be physical (Kusakabe et al. 2002). After some consideration, saccharin (CAS 81-07-2) and EDTA (CAS 60-00-4) were entered as negatives, in agreement with Ashby et al. (Mutation Research 163, 63-73, 1986), even though there was conflicting information for some of the salts. Finally, about 40 chemicals having only equivocal results were excluded. This is also an arbitrary decision, but we felt that equivocal results were not likely to lead to a better training set.
In total, 513 chemicals in the training set origin from Sofuni 1998. The second data set applied was from Kusakabe et al. 2002. The data was generated over a six-year period (1991-1996) for chromosomal aberration testing of high production volume (HPV) industrial chemicals that had been conducted using Chinese hamster lung (CHL/IU) cells according to the OECD HPV testing program and the national program in Japan. Of a total of 98 substances, two were removed: dicyclopentadiene (CAS 77-73-6), because it was already in our training set, and Pigment Green No. 7 (CAS 14832-145), a copper complex that cannot be modelled in this system. On further examination of the data set, it was noticed that one substance (4-(1-Methylpropyl)phenol, CAS 99-71-8) was actually a false positive (only active at very high concentration, and ultimately judged inactive following an in vitro micronucleus test). So this substance was

removed, in addition to eight chemicals where chromosomal aberrations were induced under non-physiological culture conditions (pH<6). In total, 87 chemicals in the training set origin from Kusakabe et al. 2002.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

QSAR

Multilinear regression QSAR

Fragment based statistical system which creates a number of sub models derived by multiple linear regression. Multiple explicit algorithms operate within the MultiCASE model.

4.3.Descriptors in the model:

[1]Fragment descriptors

[2]Distance descriptors

[3]Physical descriptors

[4]Electronic descriptors

[5]Quantum mechanical descriptors

4.4.Descriptor selection:

Automated selection. See 4.5 for further details.

4.5.Algorithm and descriptor generation:

MC4PC is a fragment-based statistical model system. The methodology involves breaking down the structures of the training set into all possible fragments from 2 to 10 heavy (non-hydrogen) atoms in length. Two-dimensional distances between heavy atoms are also included in the analysis. Fragments from the entire training set are combined into gross activity categories. A structural fragment is considered as a “biophore” if it has a statistical significant association with chemicals in the active category. It is considered a “biophobe” if it has a statistically significant relation with the inactive category. Within each biophore modulators of the activity, such as substructures, molecular orbital energies and two-dimensional distance descriptors, of the biophores are identified. Statistical equations based on relevant descriptors are established within each statistical significant biophore. The program was set to maximum specificity (details available upon request).

Klopman 1992: “Fragmentation of all the compounds generates many different molecular fragments, most of them totally unrelated to the observed activity. A binomial distribution is assumed, and any considerable deviation from a random distribution of a fragment among the active and inactive classes of molecules is indicative of potential significance to the biological activity. Activating and inactivating fragments, as well as calculated values of the logarithm of the partition coefficient and the square of the logarithm of the partition coefficient are incorporated within a regression equation in a forward

stepwise manner until no significant improvement is observed between calculated and actual values. The statistical validity of each of the variables is established by application of the F-partial statistic at the 95% confidence level".

4.6.Software name and version for descriptor generation:

MC4PC v. 2005

4.7.Chemicals/Descriptors ratio:

The model uses primarily fragment descriptors, specific to a group of structurally related chemicals from the training set; therefore estimations of the number of used descriptors may be difficult.
In general, we estimate that the model uses an order of magnitude less descriptors than there are observations. It should be noted that due to MultiCASE's complex decision making scheme, over fitting is rare, compared to simpler linear models. Warnings are issued in case of statistically insufficient number of observations (total no. of chemicals in the training set), which is not the case in the present model.

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain of MultiCASE models is expressed in terms of fragments unknown to the system and statistical significance of the known biophores and biophobes. Descriptors may also be taken into account. Failure to comply with the model domain is not absolute but may be graded, depending on the number and nature of the involved fragments. Thus, different applications may define the applicability domain in different ways. The Danish QSAR group has accepted the strictest possible definition of applicability domain for its MultiCASE models, namely, only chemicals without any unknown fragments are accepted. This applicability domain definition was applied when determining the validation result.

5.2.Method used to assess the applicability domain:

Only chemicals with no warnings when predicted are within the domain. Warnings are given to chemicals with unknown fragments or/and statistical insignificance.

5.3.Software name and version for applicability domain assessment:

MC4PC v. 2005

5.4.Limits of applicability:

Discrete organics as defined by the model. See 5.2

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: No

Chemical Name: No

Smiles: Yes

Formula: No

INChI: No

MOL file: No

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

No

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

All

6.5.Other information about the training set:

600 data points: 306 negative values; 294 positive values

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

100% concordance

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

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

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

MC4PC v. 2005 five-fold 2*50% cross-validation gave:

Sensitivity: 57.8%

Specificity: 86.5%

Concordance: 74.3%

The cross-validation was done by randomly removing 50% of the training set, where the 50% contains the same ratio of positive and negatives as the training set. Then a model was created on the remaining 50% and use to predict the removed 50%, and the other way around. This was repeated five times.

6.10.Robustness - Statistics obtained by Y-scrambling:

Statistics obtained by Y-scrambling was performed for a preliminary model with a training set of 513 chemicals (not including the Kusakabe et al. 2002 data).We randomly scrambled the toxicity scores in our training set, and performed 10 cross-validations, leaving out 50% of the chemicals in each cross-validation. None of the resulting validations were statistically significant. The Chi Square value averaged 0.7126 (probability = about 0.4).For chemicals, estimated as being within the domain, concordance was 49.69%.

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

No

7.2.Available information for the external validation set:

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula: Yes

INChI: Yes

MOL file: Yes

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:

7.6.Experimental design of test set:

7.7.Predictivity - Statistics obtained by external validation:

Results for a preliminary model based on 513 chemicals with a test set of 87 chemicals from Kusakabe et al. (2002) gave for 17 positives and 45 negatives within the defined applicability domain:

Sensitivity: 58.8%
Specificity: 82.2%
Concordance: 75.8%

7.8.Predictivity - Assessment of the external validation set:

7.9.Comments on the external validation of the model:

See Kusakabe et al., 2002 and Niemelä & Wedebye, 2004. External validation was performed for the preliminary model.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

MultiCASE models identify substructures (fragments) and for each set of molecules containing a specific fragment further identifies additional parameters (modulators like e.g. logP and molecular orbital energies). Generally, many predictions may indicate modes of action that are obvious for persons with expert knowledge for the endpoint.

The exact mechanisms / MoA of the chemicals causing chromosomal aberration are not known, but it is assumed that a covalent reaction with a biological macromolecule (e.g. DNA) may be involved. A few of the more obvious biophores identified include nitroaromatics, certain PAHs and anilines.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation. The information in 8.1 may provide mechanistic interpretation.

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

The model can be used to predict in vitro chromosome aberrations in cultured CHL cells.

9.2.Bibliography:

[1]Niemelä JR, Wedebye EB, Nikolov NG, Jensen GE, Ringsted T, Ingerslev F, Tyle H & Ihlemann C (2009). The Advisory list for self-classification of dangerous substances. Danish Environmental Protection Agency. Environmental Project No.1303. www.mst.dk

[2]OECD (1997). OECD Guidelines for the Testing of Chemicals No. 473: Genetic Toxicology: In Vitro Mammalian Chromosome Aberration Test. Organisation for Economic Cooperation and Development. Paris, France.
[3]Sofuni T (1998). Data Book of Chromosomal Aberration Test In Vitro, Revised Edition. Life-Science Information Center, Tokyo, Japan.
[4]Ishidate M (1988). Data Book of Chromosomal Aberration Test In Vitro, Revised Edition. Elsevier, Amsterdam, New York, Oxford.
[5]Kusakabe H, Ymakage K, Wakuri S, Sasaki K, Nakagawa Y, Watanabe M, Hayashi M, Sufuni T, Ono H & Tanaka N (2002). Relevance of chemical structure and cytotoxicity to the induction of chromosome aberrations based on testing of 98 high production volume industrial chemicals. Mutation Research 517, 187-198.

9.3.Supporting information:

CHL_training_600.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0070/attachment/A770

Test set(s)Supporting information

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q13-410-0070

10.2.Publication date:

2013-07-02

10.3.Keywords:

MultiCASE;Danish National Food Institute;in vitro chromosome aberration;Chinese Hamster Lung Cell;CHL;

10.4.Comments:

former Q19-41-37-331