MultiCASE CASE Ultra model for Androgen Receptor (AR) antagonism (human vector) *in vitro*

1. QSAR identifier

1.1 QSAR identifier (title)

MultiCASE CASE Ultra model for Androgen Receptor (AR) antagonism (human vector) *in vitro*, Danish QSAR Group at DTU Food.

1.2 Other related models

Leadscope Enterprise model for Androgen Receptor (AR) antagonism (human vector) *in vitro*, Danish QSAR Group at DTU Food.

SciMatics SciQSAR model for Androgen Receptor (AR) antagonism (human vector) *in vitro*, Danish QSAR Group at DTU Food.

1.3. Software coding the model

MultiCASE CASE Ultra 1.4.6.6 64-bit.

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

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Gunde Egeskov Jensen;

National Food Institute at the Technical University of Denmark;

Eva Bay Wedebye;

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

Klopman, G. (1992) MULTICASE 1. A Hierarchical Computer Automated Structure Evaluation Program. *Quant. Struct.-Act. Relat.,* 11*,* 176 - 184.

Chakravarti, S.K., Saiakhov, R.D., and Klopman, G. (2012) Optimizing Predictive Performance of CASE Ultra Expert System Models Using the Applicability Domains of Individual Toxicity Alerts. *J. Chem. Inf. Model*., 52, 2609 −2618.

Saiakhov, R.D., Chakravarti, S.K., and Klopman, G. (2013) Effectiveness of CASE Ultra Expert System in Evaluating Adverse Effects of Drugs. *Mol. Inf*., 32, 87 – 97.

2.8 Availability of information about the model

The training set is non-proprietary and is composed of experimental data from our own laboratory and additional data from the literature (see references under 9.2). 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

Human (human androgen receptor in Chinese Hamster Ovary (CHO) cells).

3.2 Endpoint

QMRF 4. Human Health Effects

QMRF 4.18.c. Endocrine Activity. Other (human Androgen Receptor antagonism in a reporter gene assay)

3.3 Comment on endpoint

There is increasing evidence that a variety of environmental chemicals have the potential to disrupt the endocrine system by mimicking or inhibiting endogenous hormones such as estrogens and androgens. These endocrine disrupting chemicals (EDCs) may adversely affect development and/or reproductive function.

Among the many biological mechanisms that can result in endocrine disruption, one important is the expression of an antiandrogenic response. Chemicals with antiandrogenic activity counteract the effect of the male sex steroid hormones either by affecting their synthesis or metabolism or by blocking the effects of androgens. Androgens such as testosterone and dihydrotestosterone play a crucial role at several stages of male development and in the maintenance of the male phenotype. The development of the male phenotype during gestation is totally dependent on the action of androgens, and interference with the androgen receptor (AR) at this point of development is hypothesized as being linked to the increased frequency of male reproductive disorders such as testicular dysgenesis syndrome. Blocking of the androgen action may be exerted by antagonism of the AR, that is, by direct interaction of a chemical with AR.

The AR is a member of the nuclear receptor superfamily. Upon ligand binding to the AR in the cytoplasm the receptor undergoes a conformational change and the receptor-ligand dimer is transported to the nucleus where it binds to an androgen response element (RE) on the DNA. This binding modulates the transcription of target genes downstream the RE. The structural diversity of chemicals, which can bind to and affect the activity of AR is very broad. *In vivo* assays for the detection of antiandrogenic action are time-consuming, costly, and labour intensive, which makes them impractical for routine screening and testing of a large number of chemicals. Although *in vitro* data for AR antagonism alone are not sufficient to characterize a compound as an EDC, information on the ability of a chemical to antagonize AR *in vitro* provides an important piece of information for priority setting of chemicals for the more elaborate *in vivo* assays.
For this model training set data originates from reporter gene assays using hAR plasmid transfected Chinese Hamster Ovary (CHO) cells. The training set consists of data from our own laboratory (Vinggaard *et al.* 2008) and data compiled from the literature.

3.4 Endpoint units

CASE units, 45 for positives and 10 for negatives.

3.5 Dependent variable

Human Androgen Receptor (hAR) antagonism *in vitro*, positive or negative.

3.6 Experimental protocol

The experimental protocol for the data obtained in our own laboratory can be found in Vinggaard *et al.* (2008). Briefly, Chinese Hamster Ovary (CHO) cells were transfected with a plasmid containing a gene coding for the human androgen receptor (AR) and a plasmid containing a gene coding for the reporter enzyme Luciferase. The synthetic androgen, R1881, responsible for AR activity, was added, and the response of 0.1 nM R1881 was set to 100%. Chemicals were tested at various concentrations and data was related to the response of 0.1 nM R1881. Cytotoxicity was determined in parallel using CHO cells transfected with a plasmid containing a gene coding for a constitutively active AR lacking the ligand binding domain. The IC_{25}, defined as the concentration of the test compound that caused a 25% inhibition of the luciferase activity induced by R1881, was calculated for each compound.

For the data obtained from the literature different experimental protocols have been employed. We therefore refer to the references in 9.2 for a specific description of the different protocols.

All the AR antagonism data was separated in to two groups: chemicals reaching an IC_{25} at non-cytotoxic concentration ≤10 μM were defined as positives, and chemicals with IC25 > 10 µM or showing no activity were defined as negatives.

3.7 Endpoint data quality and variability

The dataset from our own laboratory is expected to have low data variability. Because multiple different experimental protocols were used for the data obtained from the literature a certain degree of interlaboratory variability in the data is expected. Jensen *et al.* (2012) compared data where different laboratories had tested the same substances and found an agreement of 83% (29/35) in one case and 91% (40/44) in another. Some chemicals were excluded from the training set due to significant discrepancies between data from different sources without other supporting data.

4. Defining the algorithm

4.1 Type of model

A categorical (Q)SAR model based on structural fragments and calculated molecular descriptors.

4.2 Explicit algorithm

This is a categorical (Q)SAR model composed of multiple local (Q)SARs made by use of stepwise regression. The specific implementation is proprietary within the MultiCASE CASE Ultra software.

4.3 Descriptors in the model

Fragment descriptors,

Distance descriptors,

Physical descriptors,

Electronic descriptors,

Quantum mechanical descriptors

4.4 Descriptor selection

Automated hierarchical selection (see 4.5).

4.5 Algorithm and descriptor generation

MultiCASE CASE Ultra is an artificial intelligence (AI) based computer program with the ability to learn from existing data and is the successor to the program MultiCASE MC4PC. The system can handle large and diverse sets of chemical structures to produce so-called global (Q)SAR models, which are in reality series of local (Q)SAR models. Upon prediction of a query structure by a given model one or more of these local models, as well as global relationships if these are identified, can be involved if relevant for the query structure. The CASE Ultra algorithm is mainly built on the MCASE methodology (Klopman 1992) and was released in a first version in 2011 (Chakravarti *et al.* 2012, Saiakhov *et al.* 2013).

CASE Ultra 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. The fragment generation procedure produces simple linear chains of varying lengths and branched fragments as well as complex substructures generated by combining the simple fragments.

A structural fragment is considered as a positive alert if it has a statistical significant association with chemicals in the active category. It is considered a deactivating alert if it has a statistically significant relation with the inactive category.

Once final lists of positive and deactivating alerts are identified, CASE Ultra attempts to build local (Q)SARs for each alert in order to explain the variation in activity within the training set chemicals covered by that alert. The program calculates multiple molecular descriptors from the chemical structure such as molecular orbital energies and two-dimensional distance descriptors. A stepwise regression method is used to build the local (Q)SARs based on these molecular descriptors. For each step a new descriptor (modulator) is added if the addition is statistically significant and increases the cross-validated R2 (the internal performance) of the model. The number of descriptors in each local model is never allowed to exceed one fifth of the number of training set chemicals covered by that alert. If the final regression model for the alert does not satisfy certain criteria ($R2 \geq 0.6$ and $Q2 \geq 0.5$) it is rejected. Therefore, not all alerts will necessarily have a local (Q)SAR.

The collection of positive and deactivating alerts with or without a local (Q)SAR constitutes a global (Q)SAR model for a particular endpoint and can be used for predicting the activity of a test chemical.

More detailed information about the algorithm can be found in Chakravarti *et al.* (2012), Saiakhov *et al.* (2013).

4.6 Software name and version for descriptor generation

MultiCASE CASE Ultra 1.4.6.6 64-bit.

4.7 Descriptors/chemicals ratio

The program primarily uses fragment descriptors specific to a group of structurally related chemicals from the training set. Therefore estimation of the number of descriptors used in a specific model, which is a collection of local models as explained under 4.5, may be difficult. In general, we estimate that the model uses an order of magnitude less descriptors than there are observations. The number of descriptors in each local (Q)SAR model is never allowed to exceed one fifth of the number of training set chemicals covered by that alert (Saiakhov *et al.* 2013).

It should be noted that due to CASE Ultra's complex decision making scheme overfitting is rare compared to simpler linear models. Warnings are issued in case of statistically insufficient overall number of observations to produce a model, which is not the case in the present model.

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 in CASE Ultra and the in-house further refinement algorithm on the output from CASE Ultra to reach the final applicability domain call.

1. CASE Ultra

CASE Ultra recognizes unknown structural fragments in test chemicals that are not found in the training data and lists these in the output for a prediction. Fragments this way impose a type of global applicability domain for the overall model. The presence of more than three unknown structural fragments in the test chemical results in an 'out of domain' call in the program. (Chakravarti *et al.*2012, Saiakhov *et al.*2013).

For each structural alert, CASE Ultra uses the concept of so-called domain adherences and statistical significance.

The domain adherence for an alert in a query chemical depends on the similarity of the chemical space around the alert in the query chemical compared to the chemical space (in terms of frequencies of occurrences of statistically relevant fragments) of the training set chemicals used to derive the alert. The domain adherence value (between zero and one) is the ratio of the sum of the squared frequency of occurrence values of the subset of the fragments that are present in the test chemical and sum of the squared frequency of occurrence of all the fragments that constitute the domain of the alert in question. The more fragments of the domain of the alert in the test chemical the closer the domain adherence value is to 1. The value will never be zero as the alert itself is part of the alerts domain.

Furthermore, all alerts come with a measure of its statistical significance, and this depends on the number of chemicals in the training set which contained the alert and the prevalence within these of actives and inactives. (Chakravarti *et al.*2012).

2. In-house refinement algorithm to reach the final applicability domain call

The Danish QSAR group has applied a stricter definition of applicability domain for its MultiCASE CASE Ultra models.

An optimization procedure based on preliminary cross-validation is applied to further restrict the applicability domain for the whole model based on non-linear requirements for domain adherence and statistical significance, giving the following primary thresholds:

Domain adherence = 0.50 and significance = 70%.

Any positive prediction is required to contain at least one valid positive alert, namely an alert with statistical significance and domain adherence exceeding thresholds defined for the specific model.

The positive predictions for chemicals which only contain invalid positive alerts are considered 'out of domain' (in CASE Ultra these chemicals are predicted to be inactive).

Furthermore, only query chemicals with no unknown structural fragments are considered within the applicability domain, except for chemicals predicted 'positive', where one unknown fragment is accepted. Also no significant positive alerts are accepted for an inactive prediction.

5.2 Method used to assess the applicability domain

The applicability domain is assessed in terms of the output from CASE Ultra with the Danish QSAR group's further refinement algorithm on top as described in 5.1.

Because of the complexity of the system (see 5.1), the assessment of whether a test chemical is within the applicability domain of the model requires predicting the chemical with the specific model, and the application of the Danish QSAR group model-specific thresholds for domain adherence and significance.

This applicability domain was also applied when determining the results from the cross-validations (6.9).

5.3 Software name and version for applicability domain assessment

MultiCASE CASE Ultra 1.4.6.6 64-bit.

5.4 Limits of applicability

All structures are run through the DataKurator feature within CASE Ultra to check for compatibility with the program. Furthermore, 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). Structures with less than two carbon atoms or containing atoms not in the list above (e.g. heavy metals) are rendered out as not acceptable for further QSAR processing. Calculation 2D structures (SMILES and/or SDF) are generated by stripping off accepted organic and inorganic ions. 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

Yes

6.2 Available information for the training set

CAS

SMILES

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

874 compounds are in the training set: 231 positives and 643 negatives.

6.6 Pre-processing of data before modelling

Only structures acceptable for CASE Ultra 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”, thereby splitting the full training set into two subsets A and B, each containing the same ratio of positives to negatives as the full training set. A new model (validation sub-model) was created on subset A without using any information from the “mother model” (regarding e.g. descriptor selection etc.). The validation sub-model was applied to predict subset B (within the CASE Ultra applicability domain for the validation sub-model and the in-house further refinement algorithm for the full model). Likewise, a validation sub-model was made on subset B and this model was used to predict subset A (within the CASE Ultra applicability domain for the validation sub-model and the in-house further refinement algorithm for the full model). This procedure was repeated five times.

Predictions within the defined applicability domain for the ten validation sub-models were pooled and Cooper’s statistics calculated. This gave the following results for the 69.0% (3014/(5*874)) of the predictions which were within the applicability domain:

- Sensitivity (true positives / (true positives + false negatives)): 53.1%
- Specificity (true negatives / (true negatives + false positives)): 89.8%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 79.4%

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

External validation has not been performed for this model.

However, an external validation was performed on an earlier version of this model. That version was based on a training set of 523 chemicals and made in MultiCASE MC4PC software, a predecessor to MultiCASE CASE Ultra. The experimental results for the test set were obtained in our own laboratory. The test set comprised 96 chemicals with 14 experimentally tested positive chemicals and 82 negative chemicals. The internal cross validated QSAR model was "closed" and used to select 102 chemicals within the domain for external validation. The selection of chemicals was done according to the following criteria: 1) Only EINECS chemicals (European Inventory of Existing Commercial Chemical Substances), approximately 47000 chemicals, were considered, 2) Two lists of chemicals within the model domain representing positive and negative predictions, respectively, were generated, 3) The chemicals in each list were randomized, and chemicals in the top were selected for testing. If a chemical was not commercially available, the next chemical on the list was taken. The distribution of selected chemicals for external validation was approximately 10% predicted positive and 90% predicted negative. This approach was taken to reflect the prevalence of chemicals with positive and negative activity as predicted by the QSAR model. Chemicals were blinded until test and data treatments were completed. (Vinggaard *et al.* 2008)

7.6 Experimental design of test set

7.7 Predictivity – Statistics obtained by external validation

- Sensitivity (true positives / (true positives + false negatives)): 57.1%
- Specificity (true negatives / (true negatives + false positives)): 97.6%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 91.7%

7.8 Predictivity – Assessment of the external validation set

7.9 Comments on the external validation of the model

The sensitivity for the external validation was based on a rather small set of 14 chemicals and a measure of 57% (8/14) does in that context not seem far from the sensitivity measure of 64% (corresponding to 9/14) in the cross validation. The specificity of the external validation was based on a bigger set of 82 chemicals and gave a clearly better result of 98% compared to the specificity result of 84% in the cross-validation. This brought the total concordance up from 76% in the cross-validation to 92% in the external validation.

7.8 Predictivity – Assessment of the external validation set

7.9 Comments on the external validation of the model

The sensitivity for the external validation was based on a rather small set of 14 chemicals and a measure of 57% (8/14) does in that context not seem far from the sensitivity measure of 64% (corresponding to 9/14) in the cross validation. The specificity of the external validation was based on a bigger set of 82 chemicals and gave a clearly better result of 98% compared to the specificity result of 84% in the cross-validation. This brought the total concordance up from 76% in the cross-validation to 92% in the external validation.

8. Mechanistic interpretation

8.1 Mechanistic basis of the model

The model identifies statistically relevant substructures (i.e. alerts) and for each set of molecules containing a specific alert it further identifies additional parameters found to modulate the alert (e.g. logP and molecular orbital energies, etc.). Many predictions may indicate modes of action that are obvious for persons with expert knowledge about 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 if a chemical has an antagonistic effect on the human androgen receptor *in vitro*.

9.2 Bibliography

Andersen, H.R., Vinggaard, A.M., Rasmussen, T.H., Gjermandsen, I.M., and Bonefeld-Jørgensen, E.C. (2002) Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity *in vitro. Toxicology and Applied Pharmacology*, 179, 1-12.

Araki, N., Ohno, K., Takeyoshi, M., and Iida, M. (2005a) Evaluation of a rapid in vitro androgen receptor transcriptional activation assay using AR-EcoScreen™ cells. *Toxicology In Vitro*, 19, 335-352.

Araki, N., Ohno, K., Nakai, M., Takeyoshi, M., and Iida, M. (2005b) Screening for androgen receptor activities in 253 industrial chemicals by in vitro reporter gene assays using AR-EcoScreen™ cells. *Toxicology In Vitro*, 19, 831-842.

Jensen, G.E., Nikolov, N.G., Wedebye, E.B., Ringsted, T., and Niemela, J.R. (2011) QSAR models for anti-androgenic effect--a preliminary study. *SAR and QSAR in Environmental Research*, 22:1-2, 35-49.

Jensen, G.E., Nikolov, N.G., Dreisig, K., Vinggaard, A.M., and Niemelä, J.R. (2012) QSAR Model for Androgen Receptor Antagonism - Data from CHO Cell Reporter Gene Assays. *J Steroids and Hormonal Science*, S2:006, doi:10.4172/2157-7536.S2-006.

Kawamura, Y., Mutsuga, M., Kato, T., Iida, M., and Tanamoto, K. (2005) Estrogenic and Anti-Androgenic Activities of Benzophenones in Human Estroge and Androgen Receptor Mediated Mammalian Reporter Gene Assays. *J Health Sci,* 51:1, 48-54.

Kojima, H., Iida, M., Katsura, E., Kanetoshi, A., Hori, Y., and Kobayashi, K. (2003) Effects of a diphenyl ether-type herbicide, chlornitrofen and its amino derivative on androgen and estrogen receptor activities. *Environ. Health Perspect*, 11:4, 497–502.

Kojima, H., Takeuchi, S., Uramaru, N., Sugihara, K., Yoshida, T., and Kitamura, S. (2009) Nuclear hormone receptor activity of polybrominated diphenyl ethers and their hydroxylated and methoxylated metabolites in transactivation assays using Chinese hamster ovary cells. *Environ Health Perspect*, 117:8, 1210-1218.

Körner, W., Vinggaard, A.M., Térouanne, B., Ma, R., Wieloch, C., Schlumpf, M., Sultan, C., and Soto, A.M. (2004) Interlaboratory comparison of four in vitro assays for assessing androgenic and antiandrogenic activity of environmental chemicals. *Environ Health Perspect*, 112:6, 695-702.

Satoh, K., Ohyama, K., Aoki, N., Iida, M., and Nagai, F. (2004) Study on anti-androgenic effects of bisphenol a diglycidyl ether (BADGE), bisphenol F diglycidyl ether (BFDGE) and their derivatives using cells stably transfected with human androgen receptor, AR-EcoScreen. *Food Chem Toxicol,* 42, 983-993.

Satoh, K., Nonaka, R., Ohyama, K., Nagai, F., Ogata, A., and Iida, M. (2008) Endocrine disruptive effects of chemicals eluted from nitrile-butadiene rubber gloves using reporter gene assay systems. *Biol Pharm Bull,* 31:3, 375-379.

Takeuchi, S., Iida, M., Kobayashi, S., Jin, K., Matsuda, T., and Kojima, H. (2005) Differential effects of phthalate esters on transcriptional activities via human estrogen receptors alpha and beta, and androgen receptor. *Toxicology*, 210, 223-233.

Takeuchi, S., Takahashi, T., Sawada, Y., Iida, M., Matsuda, T., and Kojima, H. (2009) Comparative study on the nuclear hormone receptor activity of various phytochemicals and their metabolites by reporter gene assays using Chinese hamster ovary cells. *Biol Pharm Bull*, 32:2, 195-202.

Vinggaard, A.M., Bonefeld-Joergensen, E.C., and Larsen, J.C. (1999) Rapid and sensitive reporter gene assays for detection of antiandrogenic and estrogenic effects of environmental chemicals. *Toxicol Appl Pharmacol*, 155, 150-160.

Vinggaard, A.M., Nellemann, C., Dalgaard, M., Jørgensen, E.B., and Andersen, H.R. (2002) Antiandrogenic effects in vitro and in vivo of the fungicide prochloraz. *Toxicol Sci*, 69, 344-353.

Vinggaard, A.M., Niemelä, J.R., Wedebye, E.B., and Jensen, G.E. (2008) Screening of 397 Chemicals and Development of a Quantitative Structure - Activity Relationship Model for Androgen Receptor Antagonism. *Chem. Res. Toxicol*., 21, 813-823.

9.3 Supporting information