Toxicology and Applied Pharmacology 378 (2019) 114630

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/taap

Predicting estrogen receptor binding of chemicals using a suite of *in silico* methods – Complementary approaches of (Q)SAR, molecular docking and molecular dynamics

Cotterill J.V.[a,1], Palazzolo L.[b,1], Ridgway C.[a], Price N.[a], Rorije E.[c], Moretto A.[d], Peijnenburg A.[e], Eberini I.[f,*]

[a] Fera Science Limited, Sand Hutton, York YO41 1LZ, UK
[b] Università degli Studi di Milano, Dipartimento di Scienze Farmacologiche e Biomolecolari, Via Balzaretti 9, 20133 Milano, Italy
[c] Centre for Safety of Substances and Products, National Institute for Public Health and Environment (RIVM), P.O. Box 1, 3720 BA, Bilthoven, The Netherlands
[d] Università degli Studi di Milano, Dipartimento di Scienze Biomediche e Cliniche, Ospedale L. Sacco, Padiglione 17, Via G.B. Grassi 74, 20157 Milano, Italy
[e] Wageningen University & Research, Wageningen, The Netherlands
[f] Università degli Studi di Milano, Dipartimento di Scienze Farmacologiche e Biomolecolari & DSRC, Via Balzaretti 9, 20133 Milano, Italy

ARTICLE INFO

Keywords:
Estrogen receptor
In Silico
QSAR
Molecular docking
Low-mode molecular dynamics simulation

ABSTRACT

With the aim of obtaining reliable estimates of Estrogen Receptor (ER) binding for diverse classes of compounds, a weight of evidence approach using estimates from a suite of *in silico* models was assessed. The predictivity of a simple Majority Consensus of (Q)SAR models was assessed using a test set of compounds with experimental Relative Binding Affinity (RBA) data. Molecular docking was also carried out and the binding energies of these compounds to the ERα receptor were determined. For a few selected compounds, including a known full agonist and antagonist, the intrinsic activity was determined using low-mode molecular dynamics methods. Individual (Q)SAR model predictivity varied, as expected, with some models showing high sensitivity, others higher specificity. However, the Majority Consensus (Q)SAR prediction showed a high accuracy and reasonably balanced sensitivity and specificity. Molecular docking provided quantitative information on strength of binding to the ERα receptor. For the 50 highest binding affinity compounds with positive RBA experimental values, just 5 of them were predicted to be non-binders by the Majority QSAR Consensus. Furthermore, agonist-specific assay experimental values for these 5 compounds were negative, which indicates that they may be ER antagonists. We also showed different scenarios of combining (Q)SAR results with Molecular docking classification of ER binding based on cut-off values of binding energies, providing a rational combined strategy to maximize terms of toxicological interest.

1. Introduction

Exposure to endocrine disrupting chemicals (EDCs) has been linked to an increase in reproductive problems, hormone-dependent cancers, diabetes and obesity (Diamanti-Kandarakis et al., 2009; Piparo and Worth, 2010; Schug et al., 2011, Vuorinen et al., 2013). There are diverse and complex mechanisms of endocrine disruption, including direct activation or inactivation of key endocrine target receptors such as estrogen, androgen, progesterone and several corticosteroid receptors, as well as disruption of hormone synthesis and inhibition or activation of hormone metabolizing enzymes such as hydroxysteroid dehydrogenases.

The majority of research on EDCs has been based on interactions of compounds with nuclear hormone receptors (NR), especially estrogen receptors (ER*α* and ER*β*) and the androgen receptor (AR). Potential EDCs are often identified by *in vitro* and *in vivo* screening tests (Borgert et al., 2011), however this can be time consuming and expensive. *In silico* screening is far quicker and has lower cost implications and so can be a valuable tool for prioritising potential EDCs for further biological evaluation. Additionally, *in silico* screening can be applied to substances that are not synthesized (yet) or which would have physico-chemical properties that makes *in vitro* testing difficult and/or unreliable.

There is a range of *in silico* methods available to predict potential EDCs including (Q)SAR ((Quantitative) Structure Activity

* Corresponding author.
E-mail address: ivano.eberini@unimi.it (I. Eberini).
[1] These Authors equally contributed.

https://doi.org/10.1016/j.taap.2019.114630
Received 8 March 2019; Received in revised form 17 May 2019; Accepted 17 June 2019
Available online 18 June 2019

Relationships), Read across, molecular docking, pharmacophore modelling and virtual screening (Diaza et al., 2012; Porta et al., 2016). Although most methods will classify compounds as either binders or non-binders of a particular receptor, methods such as Molecular docking and some methods in the Endocrine Disruptor Knowledge Base, EDKB (Ding et al., 2010) can provide a quantitative estimate of the binding (Galli et al., 2014; Trisciuzzi et al., 2017). On the other hand, molecular dynamics simulation allows to evaluate the intrinsic activity of a chemical bound to a nuclear receptor, estimating the alpha helix 12 conformations. There are pros and cons to the different approaches. (Q)SARs are quick and easy to run but individual models have a limited chemical space *i.e.* the types of compounds which fall within the applicability domain of the models. Molecular docking is applicable to almost all compounds and is quantitative, however the more accurate methods required for toxicology, rather than the preliminary low accuracy pharmacology approach for large numbers of compounds, are more computationally intensive and time consuming (Trisciuzzi et al., 2005).

Recently a lot of effort is being put into the estimation of EDCs, for example the estimation of ER activity in a large-scale modelling project called CERAPP (Collaborative Estrogen Receptor Activity Prediction Project), Mansouri et al., 2016. In this extensive project 48 QSAR models to predict ER activity developed using a common training set of 1677 compounds, were combined and evaluated using a validation set of 7522 compounds. There is also a large literature on using docking against ERα applied to toxicology in order to reduce animal tests. For example Trisciuzzi et al. (2005) present a study on estrogen receptors by deriving *ad hoc* docking-based classification models to discern potential estrogenic from non-estrogenic activity. On the other hand, many authors used molecular docking simulations to evaluate both affinity and molecular recognition mechanism of chemical::ERα complexes in order to develop drugs (Maruthanila et al., 2018), test xenobiotics effect (Conroy-Ben et al., 2018; Pang et al., 2018; Ye and Shaw, 2019) or study the molecular recognition mechanism of endogenous ligand at an atomistic level (Li et al., 2019).

Estrogen Receptor binding is one of the endpoints being considered in the EU-funded project EuroMix (https://www.euromixproject.eu/), where *in silico* predictions are being used as input in the risk calculation of combined exposure to multiple chemicals. Exposure can occur to a diverse range of compounds which may be present in mixtures in food and feed, for which experimental data may not be available. These compounds include plant protection products, biocides, environmental pollutants, mycotoxins, alkaloids, non-intentionally added substances (NIAS), food contract materials and food additives. In the component-based approach to mixture toxicity assessment proposed by the EuroMix project, QSAR predictions are used as (lower tier) information to determine which substances are likely to contribute to similar toxicological effects, and therefore should be assessed together in Cumulative Assessment Groups (CAG). This CAG approach to mixture toxicity assessment is explained in draft guidance on mixture toxicity risk assessment from the European Food and Safety Authority (EFSA, 2019) and information on this approach can be found at the EFSA website (http://www.efsa.europa.eu/en/topics/topic/chemical-mixtures). Also required in the proposed project risk assessment process are the relative potency factors on the compounds in the cumulative assessment group. As no single (Q)SAR method is likely to be capable of providing reliable predictions for such a wide range of compounds, an approach using a weight of evidence of estimates from a suite of *in silico* models was proposed. We identified a broad selection of (Q)SAR models which were fundamentally different to each other, *i.e.* built using different chemicals, types of data and using different approaches and algorithms, in order to enable a diverse range of compounds to be estimated with more confidence.

In the initial phase of the EuroMix project a simple Majority Consensus approach for the interpretation of multiple QSAR results was used. To test the validity of this approach a test set of compounds with experimental Relative Binding Affinity (RBA) data was evaluated and the predictivity of both the individual models and the Majority Consensus prediction was assessed. Experimental values from reporter gene (RA) assays (*i.e.* agonist specific activation assays) were also examined to investigate whether false negatives were likely to be antagonists. In addition to application of the (Q)SAR models, molecular docking was carried out and the binding energies of the test set compounds to the ERα receptor were determined and we investigated whether the Majority Consensus of QSAR models correctly predicted the strongest binding compounds. As well as using Molecular docking data to provide an assessment of the strength of binding, we investigated using different binding energies as a cut off to determine whether a compound is a binder or non-binder and also how QSAR model and molecular docking results can be best combined according to a particular requirement, for example to minimise false negatives, or to obtain the highest accuracy. A few test compounds were also evaluated using low-mode molecular dynamics simulations to determine their intrinsic activity and to investigate whether some of the negatives from the QSAR Majority Consensus were actually ERα antagonists.

2. Methods

2.1. (Q)SAR models

Details of the models used are shown below and a summary of the type of ER activity estimated by the models is shown in Table 1. The input for the models was an .sdf file of the test set compounds (see 2.4).

2.1.1. COSMOS nuclear receptor model

The COSMOS project (www.cosmostox.eu) Nuclear Receptor model implemented into KNIME workflows are available in the COSMOS KNIME WebPortal. Although primarily developed to identify potential binding to NRs important in hepatosteatosis, ER receptors are included in the model, which was developed using structural and physico-chemical features of NR ligands using data from ChEMBL and the Protein Data Bank (PDB). A total of 1489 ER agonists were identified and used in the workflow. Further details of the methodology of the workflow are available (Mellor et al., 2015).

2.1.2. DEREK Nexus endocrine alerts

DEREK Nexus is a rule-based expert system where Structural alerts

Table 1
ER activity estimated by the (Q)SAR models.

(Q)SAR Model	ER-related Endpoints considered	Access to models
COSMOS Nuclear Receptor model	ER-agonists	Freely available
DEREK Nexus	Various endpoints related to ER activity	License fee
OCHEM estrogen receptor alpha agonists	ER- α agonists	Freely available
OECD QSAR Toolbox DART scheme (ER binding)	Various endpoints related to ER binding	Freely available
OECD QSAR Toolbox alerts (ER binding alert)	Various endpoints related to ER binding	Freely available
OECD QSAR Toolbox alerts (rtER alert)	Various endpoints related to ER binding,	Freely available
VEGA – RBA	Relative Binding Affinity, hER- α	Freely available
VEGA - CERAPP	Various endpoints related to ER signalling	Freely available

for a particular endpoint identify important structural fragments within molecules that are associated with a specific toxicological effect (http://www.lhasalimited.org/products/derek-nexus.htm). If a compound contains a structural alert then the likelihood that the compound will cause toxicity is provided (based on the species and other rules such as bioavailability). If no Structural Alert is fired then DEREK returns "Nothing to Report". This does not necessarily mean a negative result, just that that the compound contains none of the structural fragments built into the rule-based system. Derek Nexus 5.01 contains 9 alerts for Estrogen receptor modulation and 2 alerts for Oestrogenicity. Presence of these alerts is for our purposes interpreted as a substance being an ER binder, and absence of any of the alerts is interpreted as being a non-binder.

2.1.3. OCHEM estrogen receptor alpha agonists qualitative model

QSAR modelling efforts in the Tox21 Data Challenge 2014 ("TOX 21," 2014) resulted in a number of models for ER receptor binding, which were implemented in the online chemical modelling environment (OCHEM, http://ochem.eu). These include two ER-α agonist models using two different cell lines and a third model developed using log RBA data (Abdelaziz et al., 2016). For the purposes of this study only one model was used (Consensus Estrogen receptor α agonists qualitative), as the other Estrogen receptor agonist and RBA model estimates for the test set used in this study were found to be highly correlated. Other OCHEM models for ER agonists were available, but again as these are similar to the VEGA CERAPP model (2.1.7) they were not included in the study model selection.

2.1.4. OECD QSAR toolbox DART scheme

DART (Developmental and reproductive toxicity) is a decision tree developed on the basis of the combination of known modes of action (MoA) and associated structural features, as well as an empirical association of structural fragments within molecules of reproductive or developmental toxic chemicals when MoA information was lacking. The decision tree is based on a detailed review of 716 chemicals (664 positive, 16 negative, 36 with insufficient data) that have DART end-point data and are grouped into defined receptor binding and chemical domains. When tested against a group of chemicals not included in the training set, the decision tree is shown to identify a high percentage of chemicals with known DART effects (Wu et al., 2013). The DART scheme is incorporated into the OECD (Q)SAR Toolbox (http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm). For the purposes of this study a positive score was assigned if a DART alert was present and the alert specifically mentioned ER binding.

2.1.5. OECD QSAR toolbox ER profilers

There are two profilers related to ER binding freely available in the OECD (Q)SAR Application Toolbox software.

2.1.5.1. ER binding alert. The incorporated Toolbox ER binding profiling scheme is based on structural and parametric rules extracted from literature sources and supported by experimental data (Hamblen et al., 2003; Saliner et al., 2006; Schultz et al., 2002). The ER-binding profiler classifies chemicals as non - binders or weak, moderate, strong or very strong binders depending on molecular weight (MW) and structural characteristics of the chemicals. The performance of this profiler was evaluated by Mombelli (2012) using large human and rat binding datasets and the majority of compounds were correctly predicted. For the purposes of the present study chemicals were classed as positive if they had any alert for ER-binding (weak, moderate, strong or very strong).

2.1.5.2. rtER alert. The rtER Expert System v1, USEPA Estrogen Receptor Expert System (ERES) Profiler is an effects-based automated system used to predict estrogen receptor binding affinity, based on rainbow trout ER (rtER) (Hornung et al., 2014; Schmieder et al., 2014). It was specially designed to prioritise pesticides (inert and antimicrobial) that do not include any steroidal structures and thus are not capable of higher affinity ER interactions. The ERES is a logic rule-based decision tree that encodes the experts' mechanistic understanding with respect to both the chemical and biological aspects of the well-defined endpoint, or the ER bioassay domain.

2.1.6. VEGA Estrogen RBA model (IRFMN) – v.1.0.1

This classification QSAR model for binding to human estrogen receptor alpha (hER-α) was developed using experimental values for relative binding affinity (RBA), with 17β-estradiol as reference (Roncaglioni et al., 2008). This model is incorporated into the VEGA *in silico* platform, which is freely available online at http://www.vega-qsar.eu/ (the version used for this study was 1.1.3).

2.1.7. VEGA Estrogen receptor-mediated effect (IRFMN/CERAPP) - v.1.0.0

This Structural alert rules-based model was built using Sarpy software using a large dataset of high- quality ER signalling data (1529 chemicals screened across 18 high-throughput screening assays integrated into a single score), from the ToxCast program (Judson et al., 2014). The model was developed within the framework of the Collaborative Estrogen Receptor Activity Prediction Project (CERAPP), Mansouri et al., 2016.

2.2. Applicability domain

The (Q)SAR models considered in this study are in general applicable only for small organic molecules; inorganic compounds, organometallic and polymeric structures are outside of the domain of the models. The QSAR models available in the VEGA platform have a built-in tool to measure the reliability of the prediction through the applicability domain index, based on similarity to molecules in the training set, accuracy of prediction of similar molecules, concordance for similar molecules, errors of prediction among similar molecules, model's descriptor range check and atom centered fragments similarity checks. For the purposes of this study, estimates of low reliability were not considered in the majority consensus. For the other models considered in the study based on Structural Alerts, training sets are often not readily available and the domain thus difficult to define. Although the lack of an alert thus does not necessarily mean a negative estimation, the presence of these alerts is for our purposes interpreted as a substance being an ER binder, and absence of any of the alerts is interpreted as being a non-binder. Some compounds, for example Organophosphorus compounds are not covered by the version of the DART scheme used in the study and so are outside the applicability domain of this model.

2.3. Majority consensus

The model outputs were collated and a score of 1 assigned to positively predicted compounds and a score of 0 assigned to negatively predicted compounds. Where no estimate was obtained from a model, if it was outside the applicability domain of the model, or the compound was in the training set of the model and so its estimate excluded from the analysis, N/A was assigned. The total score of each compound (*i.e.* the number of models with a positive prediction) was divided by the total number of models producing an estimate. Where this was greater than or equal to 0.5 a positive estimate was assigned and where less than 0.5 a negative estimate was assigned.

2.4. Selection of the validation set

As the various models to predict ER-binding cover a range of activities, for example some of the (Q)SAR models and the standard molecular docking procedure do not distinguish between agonists and antagonists, a validation set which isn't specific for agonists or antagonists was required. The validation set selected contained Relative

Binding Affinity (RBA) data for the ERα receptor and is the external validation set used to test the VEGA RBA model (Roncaglioni et al., 2008). It is a diverse dataset of compounds, including natural and synthetic steroids, drugs and chemical contaminants such as pesticides, PCBs and phthalates, originally obtained from the Japanese METI database (METI 2002). The validation set was selected on the basis of it containing RBA data for a heterogeneous group of compounds, including chemical contaminant groups important in the Euromix project, it's previous use as a validation set (Roncaglioni et al., 2008), data for all of the compounds also being available from reporter gene assay and it containing not too large a number of compounds to enable molecular docking to be also carried out for all of the compounds within the time constraints of the project. The validation set was downloaded as a text file from VEGA version 1.1.3 (from the dataset of 806 compounds, the 150 compounds labelled with TEST status were selected) and then converted to an .sdf file for input to the various models. Details of how the chemical structures were originally obtained and modified for use in QSAR modelling are available in Roncaglioni et al., 2008. Two compounds containing tin were removed for this study as these metallo-organic compounds are not predicted in several models, leaving 148 compounds of which 52 were active and 96 were inactive for ERα receptor binding.

Evaluation of predictive performance where the validation set compounds were actually used to build the suite of (Q)SAR models and therefore would lead to overestimation of the accuracy of the models is to be avoided. Therefore, where the training sets for models are known, any compounds in the validation set which were used to build models were removed. For the VEGA CERAPP model, 50 of the 148 compounds were used to build the model, so these were removed leaving 98 validation compounds (38 active, 60 inactive) for this model. For the OCHEM ER agonist model 68 of the 148 compounds were used to build the model, so these were removed leaving 82 validation compounds (35 Active, 47 Inactive) for this model. The training sets for the other models considered in the study were not available.

Reporter gene (RA) assay experimental data for the validation set were also obtained (Roncaglioni et al., 2008) in order to use alongside RBA results with a view to identifying if compounds may be antagonists.

2.5. Cooper statistics

In order to compare the individual (Q)SAR models, the molecular docking prioritization (Trisciuzzi et al., 2017) and majority consensus predictions, the standard Cooper statistics (Cooper et al., 1979) and Matthews correlation coefficient (Matthews, 1975) were used to assess the quality of the predictions. Sensitivity is defined as the percentage of correctly classified positive predictions among the total number of positive instances. Specificity is the percentage of correct negative predictions compared to the total number of negatives. Accuracy is defined as the total number both positive and negatives correctly predicted among the total number of compounds. MCC (Matthews correlation coefficient) is a weighted value that overcomes any imbalance in the data classes which might lead to over optimistic values of accuracy. The so-called Negative Predictive Value (NPV) was also computed for the (Q)SAR and molecular docking results to evaluate the goodness of the classification and, in particular, to represent the probability that a chemical predicted as a non-binder (under-threshold) is actually a non-binder (Trisciuzzi et al., 2017; Trisciuzzi et al., 2005).

2.6. Methods for molecular docking

Among all the solved structures of estrogen receptor alpha LBD in complex with estradiol, the one with both good resolution and the lowest number of crystallographic non-solved amino acids was retrieved from the RCSB Protein Data Bank [PDB entry: 3UUD.A] (Delfosse et al., 2012). 3D structures were then verified and structurally-prepared using MOE Structure Preparation Module, in order to correct crystallographic-related errors, adding hydrogens and/or to fill up any unresolved residues. The 3D structure was then submitted to an energy minimization step with the Amber10:EHT force field and the reaction field solvation model. Refinement was carried out down to a Root Mean Square (RMS) gradient of 0.05 kcal/mol/$Å^2$.

For each chemical, stereochemistry was carefully checked according to those reported in PubChem, the dominant protomer/tautomer and protonation state was computed for at physiological pH for each chemical. 20,000 rotamers were also generated for each chemical.

In silico molecular docking was carried out with the MOE Dock Program. 'Triangle Matcher' was selected as placement methodology, in which the substance poses are generated by superposing triplets of ligand atoms on triplets of receptor site points, which are alpha spheres centres representing locations of tight packing.

30 complexes were generated for each tested ligand, removing the duplicate poses if the same set of ligand-receptor atom pairs is involved in both hydrogen bond and hydrophobic interactions. Then, putative poses were scored according to the London dG scoring empirical function, to estimate the binding free energy of the ligand from a given pose.

A refinement step was then applied to all the kept poses, basing on molecular mechanics in which all receptor atoms were held fixed during this step and the solvation effects were calculated using the reaction field functional form for the electrostatic energy term. Then, the GBVI/WSA dG scoring function with the Generalized Born solvation model (GBVI) (Wojciechowski and Lesyng, 2004) was used to evaluate the final energy (docking score) of ligand::protein complexes.

To verify the robustness of the molecular docking approach on ERα, the binding pose of the 3UUD co-crystallized estradiol was computed, obtaining a perfect overlapping (RMSD lower than 0.3 Å) (Galli et al., 2014). Moreover, with the aim of detecting the best cut-off energy values for a toxicological evaluation, *i.e.* whether a compound could be classed as an ER-binder or ER-non-binder, Cooper statistics were applied. To better visualize the docking behaviour, the Receiver Operating Characteristic (ROC) curves were used to graphically compare docking performances, for a range of different cut-off values.

2.7. Low-mode molecular dynamics simulations (LM-MD)

To study the flexibility of α-helix 12 of NR LBD due to ligand activity, LM-MD simulation is a very efficient way to reproduce the low-mode vibrations with respect to classical molecular dynamics for *minima* troughs on the potential energy surface. To run these computations, MOE Conformational Search program was used, estimating the low-frequency modes through an efficient implicit method, based on the attenuation of high-range velocities as described in detail in Labute, 2010. The human ERα LBD bound to

i) a well-known full-agonist (17β-estradiol),
ii) a well-known antagonist (4-hydroxytamoxifen),
iii) selected chemicals (listed in Table S2) and
iv) in its *apo*- form were simulated after the MOE QuickPrep preparation.

(i) ERα::17β-estradiol complex was obtained from the above mentioned structure preparation procedure; (iv) the *apo*- form, the protein moiety of a molecular complex, was obtained from the same PDB, by removing the endogenous hormone *in silico*; (ii) the complex with antagonist was obtained superimposing RCSB PDB 3ERT (Shiau et al., 1998) and RCSB PDB 3UUD crystal structures and then importing the coordinates of co-crystalized antagonist 4-hydroxytamoxifen from 3ERT to 3UUD in *apo*- form.

Complexes with selected chemicals were obtained with the following procedure: ERα::Ligand complexes resulting from docking procedures were superimposed to 3UUD bound to both full-agonist (i) and

antagonist (ii); proteins were removed and two Flex-Alignments for putative agonist/antagonist were performed keeping hold of the endogenous hormone or antagonist, respectively; complexes with selected ligands were rebuilt using the coordinates of the *apo*-3UUD 3D structure.

Both helix 12 (set as a rigid body) and the loop joining helix 12 to the preceding helix were left free to move during the low-mode molecular dynamics, whereas the residues more than 4.5 Å away were fixed (not free to move but used for the energy calculations); the other residues were defined as inert (fixed and not used for energy calculations). The simulation was carried out with default parameters, except for strain energy cut-off, which was set at 200 kcal/mol. 100 conformations were generated and analysed. The Amber10:EHT force field was used for all the computational procedures.

In order to classify the tested chemicals as agonist, partial agonist or antagonist, the Root Mean Square Deviation (RMSD) values of helix 12 alpha carbons was computed between 3UUD crystal structure and simulated complex.

3. Results

3.1. Predictivity of individual (Q)SAR models and the majority consensus

The predictivity of the individual (Q)SAR models was variable, as expected; some models with a high sensitivity and others with high specificity. The Majority Consensus gave very good results with an accuracy of 0.8 and a reasonably balanced sensitivity and specificity and a high NPV value (Table 2). The VEGA-RBA model gave slightly better results (higher Accuracy, Specificity and MCC, the same Sensitivity and a similar NPV value) than the Majority Consensus model. Although it could be argued that the VEGA-RBA model alone could thus be used instead of the Majority Consensus model, the consensus of a number of different models is likely to be suitable for a wider range of compounds, *i.e.* will have a broader applicability domain, and if a compound is out of the domain of the VEGA-RBA model, it may be predicted by other models. It is also possible that the prediction statistics for the VEGA-RBA model may be inflated, as although the validation set was not used to build the model, the final model selection would have been based on giving good results for the validation set.

Of the 52 compounds in the test set with active experimental RBA values, there were 12 compounds which were predicted to be non-binders by the majority QSAR Consensus. These compounds covered a range of chemical classes, including phthalates, benzaldehydes, organophosphate, organochlorine, dicarboximide, organosulfur and polycyclic aromatic hydrocarbons. These compounds were investigated further and Reporter Gene assay (RA) results values were obtained (Roncaglioni et al., 2008), the transcriptional activity values of which are positive for agonists only. From the RA data it was found that 10 of these 12 compounds were not able to activate the ER, which indicates that they may be antagonists. The remaining two compounds, which were not structurally similar to each other (2-hydroxy fluorene and 2,2-Bis (4-aminophenyl) hexafluoropropane), are not indicated to be antagonists and so potential false negatives.

3.2. Molecular docking

3.2.1. Binding energies

From molecular docking to ERα, both the molecular poses and the free binding energies of the 148 validation set compounds were obtained. These energies ranged from −8.9 to 9.6 kcal/mol, with the ten strongest binders shown in Table 3. Eight of these ten compounds were classified as active from the RBA experimental values. Moreover, a free binding energy of −8.1 kcal/mol was calculated for 17β-estriadiol, the endogenous hormone for ERα. Based on this value, it was possible to classify within the database how many compounds have a lower value of ΔG. As a result, only 7 compounds have a ΔG lower than −8.1 kcal/mol, of which 5 are classified as active, while 2 are classified as inactive, on the basis of VEGA RBA data.

3.2.2. Cut-off values of binding energies used to class as binders or non-binders

In order to investigate the binding free energy value to be used as a cut-off, regardless of the endogenous substrate value, an R-script was written to compute the sensitivity, the specificity and the accuracy of the docking procedure by changing cut-off value. Table 4 shows the values of the Cooper statistics, calculated for four different cut-offs. Based on accuracy, the optimal cut-off value is −6.5 kcal/mol, which corresponds to a 10*E-06 M for dissociation constant (K_i). Using this cut-off value, although the accuracy of prediction was close to 0.7, the Cooper statistics were not as good as for the QSAR Majority Consensus.

The NPV values ranged from 0.69 to 0.82 using the cut-off values of −7 to −5.5 kcal/mol (Table 4) and so a value of −5.5 kcal/mol would minimise the false negative prediction. Considering also this term for molecular docking, a good compromise between accuracy and NPV, would be a cut-off value of −6 kcal/mol, for which there is also a good sensitivity (0.75). For these reasons, in section 3.4 we also consider the cut-off of −6 kcal/mol for the majority consensus between (Q)SAR and molecular docking.

3.3. Majority consensus (Q)SAR prediction of the strongest binding compounds

Experimental values and Majority Consensus (Q)SAR predictions for all compounds with binding energies below −6.5 kcal/mol (50 compounds in total) were examined. Results (Supplementary data Table S1) show that out of these 50 strongest binding compounds, which also had positive RBA experimental values, only 5 compounds were predicted by the Majority Consensus QSAR to be negative. In addition to the RBA experimental values, the data for the same compounds for the RA test (agonist only) were also considered. The RA experimental results for these 5 compounds were all negative. As the RA test is agonist specific and RBA is general (could be agonists or antagonists) it indicates that these 5 compounds are in fact ER antagonists. This suggests that the majority consensus QSAR approach has not missed any of the highest

Table 2
Cooper statistics and NPV values for individual (Q)SAR models and Majority Consensus predicting experimental Relative Binding Affinity.

(Q)SAR Model	Sensitivity	Specificity	Accuracy	MCC	NPV
COSMOS Nuclear Receptor model	0.85	0.40	0.55	0.25	0.83
DEREK Nexus	0.33	0.98	0.75	0.44	0.73
OCHEM estrogen receptor alpha agonists [a]	0.88	0.51	0.66	0.40	0.86
OECD QSAR Toolbox DART scheme (ER binding)	0.29	0.83	0.64	0.14	0.68
OECD QSAR Toolbox ER binding OR rtER alert	0.75	0.64	0.68	0.37	0.82
VEGA – RBA	0.77	0.88	0.84	0.64	0.88
VEGA – CERAPP [a]	0.73	0.68	0.70	0.40	0.82
Majority Consensus	0.77	0.82	0.80	0.58	0.87

[a] – test set compounds used to build the model were not used in the evaluation and assigned a not predicted (N/A) score.

Table 3
Binding Energies for ten strongest binding compounds and experimental values.

CAS number	Chemical name	Binding energy DG [kcal/mol]	RBA Experimental value
17606-31-4	Bensultap	−8.9	Active
1816-85-9	11-Hydroxytestosterone	−8.5	Inactive
566-76-7	16alpha-Hydroxyestrone	−8.3	Active
2772-45-4	2,4-Bis(alpha,alpha-dimethylbenzyl)phenol	−8.3	Active
67747-09-5	Prochloraz	−8.2	Inactive
71030-11-0	beta-Zearalenol	−8.2	Active
571-20-0	5alpha-Androstane-3beta,17beta-diol	−8.2	Active
104-43-8	4-Dodecylphenol	−8.0	Active
1476-34-2	6-Keto estrone	−8.0	Active
5447-02-9	3,4-Bis(benzyloxy)benzaldehyde	−7.7	Active

Table 4
Cooper statistics and NPV value for Molecular docking binding energy cut-off values to assign whether compounds are binders or non-binders.

Cut-off (kcal/mol)	Sensitivity	Specificity	Accuracy	MCC	NPV
−5.5	0.87	0.38	0.55	0.25	0.82
−6	0.75	0.58	0.64	0.32	0.81
−6.5	0.54	0.77	0.69	0.31	0.76
−7	0.31	0.83	0.65	0.16	0.69

binding ER agonists in the validation set.

3.4. Majority consensus between methodologies

In order to highlight the weight of each model and to provide different scenarios for interpreting the results, a majority consensus between methodologies were evaluated. As a first step, the same weight as a single (Q)SAR model was associated to molecular docking. Using −6 kcal/mol as docking cut-off, two different scenarios were obtained considering a chemical "positive" if it was classified as "positive" in three or half of the models, respectively (Table 5). In the first case, we obtained a high sensitivity value of 0.87, maximizing the true positive rate, although the accuracy was reduced compared to the majority consensus of QSAR models. In the second case there were more balanced Cooper statistics.

Using the same approach but with a − 6.5 kcal/mol as binding free energy cut-off value, again where three or models were positive, we obtained a high sensitivity value of 0.83, minimizing the number of false negatives. Again there were more balanced the Cooper parameters when we considering a chemical "positive" a chemical if at least half of the models were positive.

Changing perspective and using the logical operator "OR", we considered a chemical "positive" if it was positive in least half of the Q)SAR models OR positive for molecular docking (binding free energy below the cut-off). Using −6 kcal/mol as docking cut-off an extremely high sensitivity value of 0.94 was obtained, but with a low specificity. Slightly more balanced Cooper statistics were obtained using the −6.5 kcal/mol as docking cut-off with the sensitivity still high (0.87). With these analyses we have shown that depending on the requirement *e.g.* highest accuracy, or highest sensitivity to decrease the chances of false negatives, it is possible to combine the (Q)SAR and molecular docking results accordingly, providing a rational combined strategy to maximize terms of toxicological interest.

3.5. Low-mode molecular dynamics simulations to determine intrinsic activity of ER binders

To evaluate the procedure of LM-MD for identifying the intrinsic activity of some strongly binding compounds, *i.e.* whether agonists or antagonists, ten compounds were selected (Table S2 Supplementary materials). Five likely agonists selected from the test set were the strongest binding compounds with positive RBA and RA experimental data and five possible antagonists selected were the negative compounds from the Majority Consensus QSAR, which had positive RBA data but negative results in the RA assay. In addition to these compounds, a full-agonist (17-beta-estradiol), an antagonist (4-hydroxytamoxifen) and the *apo*-form were tested, in order to have a solid background to work on considerations related to intrinsic activity.

As a first step, the *ab initio* flexible alignment with MOE Conformational Search program was verified, on the basis of the molecular structures of reference compounds (17β-estradiol and 4-hydroxytamoxifen). Subsequently, three reference conformations were computed (Fig. 1) and verified, starting from the lowest energetic conformation of 17β-estradiol, 4-hydroxytamoxifen and *apo*- form, respectively. Closed conformation of full-agonist corresponds to starting 3UUD conformation, while both partially open and completely open conformations, respectively for *apo*- and antagonist, have similar shape with respect to AR open conformation reported in Galli et al., (2014).

Through an R script, the RMSD of the first 100 conformations was evaluated for each generated complex, using as reference the 3UUD crystallographic structure in closed conformation. For the five selected putative agonists (Fig. 2, left), we found that 2 compounds (V2 and V3) have low RMSD values, so they can be classified as full agonist, another 2 compounds (V4 and V5) have low RMSD values for approximately 70% of the generated poses, so they can be classified as partial agonist, while the last one (v6) has a RMSD value similar as the reference

Table 5
Cooper statistics combining the (Q)SAR model and molecular docking results under different scenarios.

Methods	Sensitivity	Specificity	Accuracy	MCC
Majority Consensus using 7 QSAR models	0.77	0.82	0.80	0.58
Molecular docking cut off −6	0.75	0.58	0.64	0.32
Molecular docking cut off −6.5	0.54	0.77	0.69	0.31
Consensus including docking (−6 cut off) as one of the models (positive if half or more models positive	0.77	0.79	0.78	0.55
Consensus including docking (−6 cut off) as one of the models (positive if 3 or more positive)	0.87	0.63	0.71	0.47
Consensus including docking (−6.5 cut off) as one of the models (positive if half or more models positive)	0.75	0.81	0.79	0.55
Consensus including docking (−6.5 cut off) as one of models (positive if 3 or more positive)	0.83	0.66	0.72	0.46
Consensus half or more QSAR models OR docking positive (−6 cut off)	0.94	0.49	0.65	0.44
Consensus half or more QSAR models OR docking positive (−6.5 cut off)	0.87	0.63	0.71	0.47

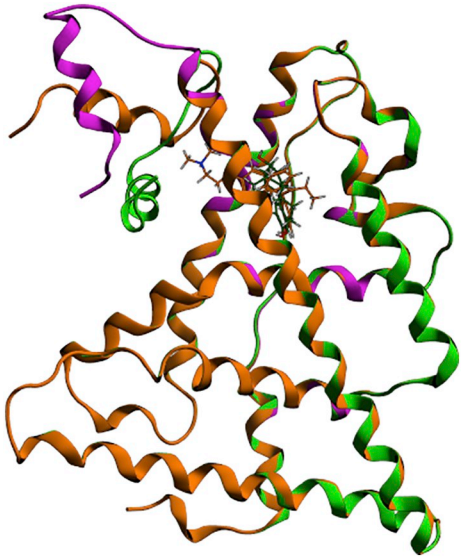

Fig. 1. Superimposition of lower energetic configuration for agonist (green), antagonist (orange) and *apo*- form (light violet). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

antagonist, so it could be classified as an antagonist. On the other hand, for the five selected putative antagonists (Fig. 2, right), 4 compounds (V9, V10, V11 and V12) have RMSD values very high or comparable with the antagonist reference value, while 1 compound (V8) showed for approx. 50% of generated conformations a low RMSD value, comparable with the agonist value. In this case, we can assess that 4 compounds are antagonists, while the fifth is a weak partial agonist.

The 17-β estradiol, the reference full agonist, has a median RMSD value of 0.67 Å, as does compound V2, whereas compounds V3-V5 have a median value of around 0.71 Å. Compound V6 has a median value of 7 Å, an order of magnitude higher than the reference agonist. On the other hand, the 14-hydroxytamoxifen, the reference antagonist, has a median value of about 8 Å, while the compounds V8-V12 have a median value ranging from 7 Å to 18 Å. *Apo*-ERα has a median value of 13 Å. Compound V8, defined as very weak partial agonist, therefore presents a higher interquartile range, because of the dispersion of the generated configurations, which are in part those of an agonist, in part those of an antagonist. A box plot of the RMSD of the generated complexes is also shown, in Fig. 3.

4. Discussion/conclusions

The approach of using the results from a suite of *in silico* models which account for different ER binding related endpoints, are built from different compounds and using different methodologies, has the advantage of increasing the chemical space covered and thus the probability that any active compounds from diverse classes of chemicals such as those considered in the EuroMix project will be correctly identified. Weight of evidence from different (Q)SAR models has been successfully used for a number of toxicological endpoints. For example, Price and Chaudhry 2014 showed that this approach using different *in silico* models can provide a rapid and reliable means of rapid screening for mutagenicity and carcinogenicity for compounds that may migrate from food packaging. Hewitt et al., 2010 and Marzo et al., 2016 integrated *in silico* models to enhance predictivity for developmental toxicity. Benfenati et al., 2015 integrated QSAR and Read-across results for the assessment of bioconcentration factors of chemicals.

In attempting to identify substitute compounds for known phthalate, bisphenol and parabens EDCs, Porta et al., 2016 applied a battery of different models, along with EC priority lists and other rule sets derived from authority's opinions. Similar to Porta et al., 2016, we selected models which are fundamentally different to each other, *i.e.* they were developed using different chemicals, using experimental results from a range of different assays and thus different ER binding endpoints and using different methodology (*e.g.* QSAR models generated using molecular descriptors by a range of algorithms, SARs using molecular fragments *etc.*), in order to enable a diverse range of compounds to be estimated with more confidence. The (Q)SAR models used in the EuroMix project were however also selected on the basis of being readily available and implemented into software programs, easy to use

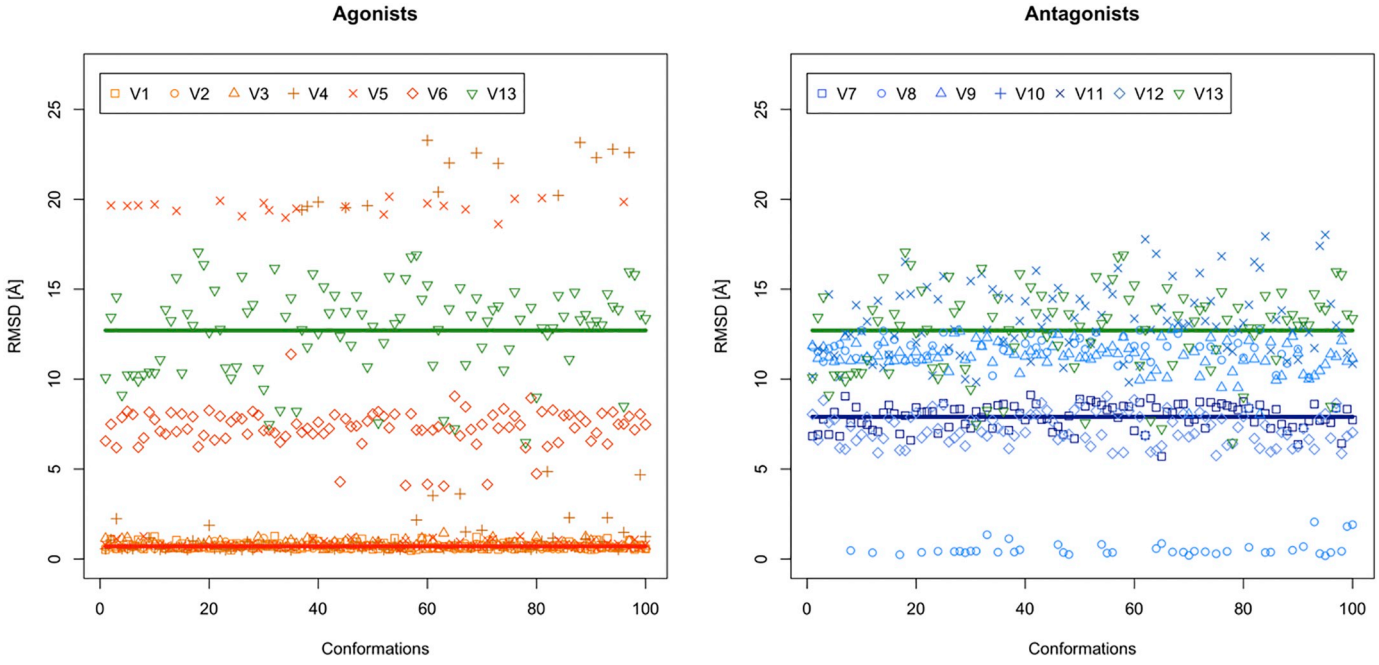

Fig. 2. RMSD values for both putative ER-alpha agonists (left) and putative ER-alpha antagonists (right). Lines represent the reference conformations for agonists (red), antagonists (blue) and *apo*-ERα (dark green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

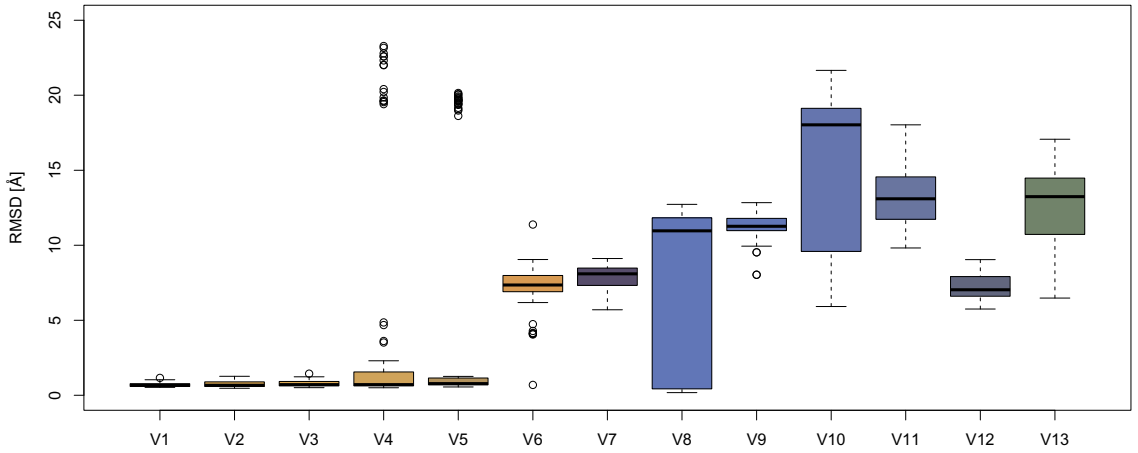

Fig. 3. Box plot of the RMSD of the generated complexes. Reference agonist (V1) and putative agonists (V2-V6) are coloured in orange, while reference antagonist (V7) and putative antagonists (V8-V12) are coloured in blue. Apo-ERα is coloured in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and with the benefit of being able to run in batch mode and thus be able to screen large numbers of compounds.

The results showed that individual (Q)SAR model predictivity varied, as expected, with accuracies ranging from 0.55 to 0.84. Some models such as the COSMOS Nuclear Receptor and OCHEM models showed high sensitivity, whilst others such as DEREK Nexus and the OECD Toolbox DART scheme alert showed very high specificity. The Majority Consensus prediction shows a high accuracy (0.8) as well as well-balanced sensitivity and specificity.

To further investigate the false negative predictions from the QSAR Majority Consensus, experimental values from reporter gene (RA) assays (an agonist specific assay) were obtained for these compounds. The vast majority of these compounds had negative RA values, which indicates that these compounds may be ER receptor antagonists. As the (Q)SAR models covered a range of ER-compound interactions, including relative binding to, and activation of the ER, then although the Majority consensus used in the study predicts ER interaction in general well, it is perhaps not surprising that it appears to be less predictive for antagonists, as no specific ER-antagonist QSAR models were used in the study.

Molecular docking was also used to provide quantitative information on the strength of binding to the ERα receptor, thus allowing to derive first-tier estrogenic potencies in the EuroMix project. Using a range of cut-off values of binding energies to predict whether a compound is a binder or non-binder, Cooper statistics showed that a threshold of −6.5 kcal/mol produced the highest accuracy. Using the molecular docking energies with the threshold value for predicting ER binding *vs.* non binding had a lower accuracy than the QSAR Majority Consensus approach, but it provides invaluable (quantitative) information on the strength of receptor binding. Finally we demonstrated that using Molecular docking cut-off values to assign ER binding can be combined with (Q)SAR results either as an additional *in silico* model in an overall consensus, or to assign a compound as an ER-binder if either the (Q)SAR Majority consensus was positive OR the Molecular docking classified it as a binder, for example if it is desired to optimise the sensitivity of the model (at the cost of overall accuracy) to reduce the chances of false negative predictions.

Further investigations on the 50 highest binding affinity compounds showed that the QSAR Majority Consensus correctly predicted these compounds to be binders in 90% of the cases. Of all compounds with positive RBA experimental values, only 5 were predicted as non-binders by the QSARs. Furthermore, the experimental values from RA assays for these 5 compounds were all negative, indicating that the negatives from the consensus of QSARs may be ER antagonists. Low Mode Molecular dynamics simulation was used to determine intrinsic activity of these negative compounds, together with 5 likely agonists and the results were mostly consistent with expectation. Four of the five proposed agonists were confirmed as such (16alpha-Hydroxyestrone, beta-Zearalenol, 5alpha-Androstane-3beta,17beta-diol and 4-Dodecylphenol) and four of the five proposed antagonists were confirmed as such (3,4-Bis(benzyloxy)benzaldehyde, Chlorpyriphos, 2-(4-Chlorophenyl)-1,1-diphenylethanol and Captafol).

QSAR models are available which can provide quantitative estimation of ER binding (*e.g.* those developed in the CERAPP project, Mansouri et al., 2016), which could also be used to provide strength of binding estimates, in addition to, or instead of the Molecular docking results. For other endpoints considered in the Euromix project, such as steatosis, quantitative QSAR models are not available at present for all of the NR's associated with steatosis and so the approach of using Molecular docking data was adopted in the project. Similarly, QSAR models have been developed to predict ER agonists or ER antagonists, rather than binding in general (*e.g.* Mansouri et al., 2016), which could be used in place of the Molecular Dynamics simulations to identify if a compound is an agonist or antagonist. Again, such models are not available for all endpoints considered in the Euromix project and so the Molecular Dynamics simulation approach was investigated here.

Overall the results show that the Majority Consensus of the (Q)SAR models is a good method to predict whether a compound is an ER-receptor binder or non-binder. It predicts ER binding well for the majority of the highest binding compounds and the majority of the relatively few false negatives may be antagonists. This method has the benefit of being quick to provide results, being simple to use and is based on readily available (Q)SAR models. Compounds predicted positive by QSARs could then be screened by molecular docking to assess whether they are weak or strong binders. We also showed different scenarios of combining (Q)SAR results with Molecular docking classification of ER binding based on cut-off values of binding energies, providing a rational combined strategy to maximize terms of toxicological interest, for example to minimise false negatives. As complementary approach, low-mode MD can be applied to distinguish between agonists and antagonists, improving both the (Q)SAR- and molecular docking-related information. A logical improvement over a simple Majority Consensus approach of interpreting multiple (Q)SAR predictions would be to take into account the individual predictive performance (sensitivity, specificity) of the (Q)SAR models and apply Bayesian statistical theory. Examples of application of this approach are *e.g.* in Rorije et al., 2012; Buist et al., 2012. In this case the predictive results from the Majority Consensus approach are such that not much improvement was expected and hence, Bayesian statistics were not applied. All the prediction data (QSAR predictions and Molecular

Docking energies, but not the low mode MD results) are available for a set of ~1600 food and feed relevant substances – the EuroMix Inventory. This data can be accessed and subsequently used in risk assessment calculations for combined exposure to multiple chemicals as a part of the EuroMix Software tool / MCRA 9.0 Beta – https://mcra-test.rivm.nl/Select.

Acknowledgments

This work has been funded by the EU project EuroMix (Grant Agreement 633172; www.euromixproject.eu) and has been supported by Grants from Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR)- "Progetto Eccellenza" (IE). IE gratefully acknowledges departmental "Linea 2-Azione A 2017" funding.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.taap.2019.114630.

References

Abdelaziz, A., Spahn-Langguth, H., Schramm, K.-W., Tetko, I.V., 2016. Consensus modeling for HTS assays using in silico descriptors calculates the best balanced accuracy in Tox21 challenge. Front. Environ. Sci. https://doi.org/10.3389/fenvs.2016.00002.

Benfenati, E., Roncaglioni, A., Petoumenou, M.I., Cappelli, C.I., Gini, G., 2015. Integrating QSAR and read-across for environmental assessment. SAR QSAR Environ Res. 26 (7-9), 605–618. https://doi.org/10.1080/1062936X.2015.1078408. Epub 2015 Sep 11.

Borgert, Christopher J., Mihaich, Ellen M., Ortego, Lisa S., Bentley, Karin S., Holmes, Catherine M., Levine, Steven L., Becker, Richard A., 2011. Hypothesis-driven weight of evidence framework for evaluating data within the US EPA's Endocrine Disruptor Screening Program. Regulatory Toxicology and Pharmacology 61 (2), 185–191. https://doi.org/10.1016/j.yrtph.2011.07.007.

Buist, H.E., Ld, Wit-Bos, Bouwman, T., Vaes, W.H., 2012. Predicting blood:air partition coefficients using basic physicochemical properties. Regul Toxicol Pharmacol. 62 (1), 23–28. https://doi.org/10.1016/j.yrtph.2011.11.019. Epub 2011 Dec 9.

Conroy-Ben, O., Garcia, I., Teske, S.S., 2018. In silico binding of 4,4′-bisphenols predicts in vitro estrogenic and antiandrogenic activity. Environ. Toxicol. 33, 569–578. https://doi.org/10.1002/tox.22539.

Cooper, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening tests. Br. J. Cancer. https://doi.org/10.1038/bjc.1979.10.

Delfosse, V., Grimaldi, M., Pons, J.-L., Boulahtouf, A., le Maire, A., Cavailles, V., Labesse, G., Bourguet, W., Balaguer, P., 2012. Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol a substitutes. Proc. Natl. Acad. Sci. U. S. A. 109, 14930–14935. https://doi.org/10.1073/pnas.1203574109.

Diamanti-Kandarakis, E., Bourguignon, J.P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr. Rev. https://doi.org/10.1210/er.2009-0002.

Diaza, R.G., Roncaglioni, A., Benfenati, E., 2012. In silico approaches to screening dietary endocrine disruptors. Issues Toxicol. 170–183.

Ding, D., Xu, L., Fang, H., Hong, H., Perkins, R., Harris, S., Bearden, E.D., Shi, L., Tong, W., 2010. The EDKB: an established knowledge base for endocrine disrupting chemicals. BMC Bioinforma. https://doi.org/10.1186/1471-2105-11-5.

EFSA, 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA Journal 17 (3), 5634. https://doi.org/10.2903/j.efsa.2019.5634.

Galli, C.L., Sensi, C., Fumagalli, A., Parravicini, C., Marinovich, M., Eberini, I., 2014. A computational approach to evaluate the androgenic affinity of iprodione, procymidone, vinclozolin and their metabolites. PLoS One 9. https://doi.org/10.1371/journal.pone.0104822.

Hamblen, E.L., Cronin, M.T.D., Schultz, T.W., 2003. Estrogenicity and acute toxicity of selected anilines using a recombinant yeast assay. Chemosphere. https://doi.org/10.1016/S0045-6535(03)00333-3.

Hewitt, M., Ellison, C.M., Enoch, S.J., Madden, J.C., Cronin, M.T., 2010. Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. Reprod Toxicol. 30 (1), 147–160. https://doi.org/10.1016/j.reprotox.2009.12.003. Epub 2009 Dec 16.

Hornung, M.W., Tapper, M.A., Denny, J.S., Kolanczyk, R.C., Sheedy, B.R., Hartig, P.C., Aladjov, H., Henry, T.R., Schmieder, P.K., 2014. Effects-based chemical category approach for prioritization of low affinity estrogenic chemicals. SAR QSAR Environ. Res. 25, 289–323. https://doi.org/10.1080/1062936X.2014.898692.

Judson, R., Houck, K., Martin, M., Knudsen, T., Thomas, R.S., Sipes, N., Shah, I., Wambaugh, J., Crofton, K., 2014. In vitro and modelling approaches to risk assessment from the U.S. environmental protection agency ToxCast programme. Basic Clin. Pharmacol. Toxicol. https://doi.org/10.1111/bcpt.12239.

Labute, P., 2010. LowModeMD - Implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. Journal of Chemical Information and Modeling. https://doi.org/10.1021/ci900508k.

Li, D., Cai, Y., Teng, D., Li, W., Tang, Y., Liu, G., 2019. Computational insights into the interaction mechanisms of estrogen-related receptor alpha with endogenous ligand cholesterol. Chem. Biol. Drug Des. https://doi.org/10.1111/cbdd.13506.

Mansouri, K., Abdelaziz, A., Rybacka, A., Roncaglioni, A., Tropsha, A., Varnek, A., Zakharov, A., Worth, A., Richard, A.M., Grulke, C.M., Trisciuzzi, D., Fourches, D., Horvath, D., Benfenati, E., Muratov, E., Wedebye, E.B., Grisoni, F., Mangiatordi, G.F., Incisivo, G.M., Hong, H., Ng, H.W., Tetko, I.V., Balabin, I., Kancherla, J., Shen, J., Burton, J., Nicklaus, M., Cassotti, M., Nikolov, N.G., Nicolotti, O., Andersson, P.L., Zang, Q., Politi, R., Beger, R.D., Todeschini, R., Huang, R., Farag, S., Rosenberg, S.A., Slavov, S., Hu, X., Judson, R.S., 2016. CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. Environ. Health Perspect. 24 (7), 1023–1033. https://doi.org/10.1289/ehp.1510267. Epub 2016 Feb 23.

Maruthanila, V.L., Elancheran, R., Roy, N.K., Bhattacharya, A., Kunnumakkara, A.B., Kabilan, S., Kotoky, J., 2018. In silico molecular modelling of selected natural ligands and their binding features with estrogen receptor alpha. Curr. Comput. Aided. Drug Des. https://doi.org/10.2174/1573409914666181008165356.

Marzo, M., Kulkarni, S., Manganaro, A., Roncaglioni, A., Wu, S., Barton-Maclaren, T.S., Lester, C., Benfenati, E., 2016. Integrating in silico models to enhance predictivity for developmental toxicity. Toxicology 370, 127–137. https://doi.org/10.1016/j.tox.2016.09.015. Epub 2016 Sep 28.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta. https://doi.org/10.1016/0005-2795(75)90109-9.

Mellor, C.L., Steinmetz, F.P., Cronin, M.T.D., 2015. The identification of nuclear receptors associated with hepatic steatosis to develop and extend adverse outcome pathways. Crit. Rev. Toxicol. 8444, 1–15. https://doi.org/10.3109/10408444.2015.1089471.

Mombelli, E., 2012. Evaluation of the OECD (Q)SAR application toolbox for the profiling of estrogen receptor binding affinities. SAR QSAR Environ. Res. 23, 37–57. https://doi.org/10.1080/1062936X.2011.623325.

Pang, X., Fu, W., Wang, J., Kang, D., Xu, L., Zhao, Y., Liu, A.-L., Du, G.-H., 2018. Identification of estrogen receptor alpha antagonists from natural products via in vitro and in silico approaches. Oxidative Med. Cell. Longev. 2018https://doi.org/10.1155/2018/6040149. 6040149.

Piparo, E. Lo, Worth, A.P., 2010. Review of QSAR Models and Software Tools for Predicting Developmental and Reproductive Toxicity. https://doi.org/10.2788/9628.

Porta, N., ra Roncaglioni, A., Marzo, M., Benfenati, E., 2016. QSAR methods to screen endocrine disruptors. Nucl. Recept. Res. https://doi.org/10.11131/2016/101203.

Roncaglioni, A., Piclin, N., Pintore, M., Benfenati, E., 2008. Binary classification models for endocrine disrupter effects mediated through the estrogen receptor. SAR QSAR Environ. Res. 697–733. https://doi.org/10.1080/10629360802550606.

Rorije, E., van Hienen, F.J., Dang, Z.C., Hakkert, B.H., Vermeire, T., Piersma, A.H., 2012. Relative parameter sensitivity in prenatal toxicity studies with substances classified as developmental toxicants. Reprod Toxicol. 34 (2), 284–290. https://doi.org/10.1016/j.reprotox.2012.05.092. Epub 2012 May 28.

Saliner, A.G., Netzeva, T.I., Worth, A.P., 2006. Prediction of estrogenicity: validation of a classification model. SAR QSAR Environ Res. 17 (2), 195–223.

Schmieder, P.K., Kolanczyk, R.C., Hornung, M.W., Tapper, M.A., Denny, J.S., Sheedy, B.R., Aladjov, H., 2014. A rule-based expert system for chemical prioritization using effects-based chemical categories. SAR QSAR Environ. Res. 25, 253–287. https://doi.org/10.1080/1062936X.2014.898691.

Schug, T.T., Janesick, A., Blumberg, B., Heindel, J.J., 2011. Endocrine disrupting chemicals and disease susceptibility. J. Steroid Biochem. Mol. Biol. 127, 204–215. https://doi.org/10.1016/j.jsbmb.2011.08.007.

Schultz, T.W., Sinks, G.D., Cronin, M.T.D., 2002. Structure–activity relationships for gene activation oestrogenicity: Evaluation of a diverse set of aromatic compounds. Environ. Toxicol. 17 (2002), 14–23.

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., Greene, G.L., 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell 95, 927–937. https://doi.org/10.1016/S0092-8674(00)81717-1.

TOX 21 [WWW Document], 2014. Tox 21 Data Chall.

Trisciuzzi, D., Domenico, Alberga, Mansouri, K., Judson, R., Saverio, Cellamare, Catto, M., Carotti, A., Benfenati, Emilio, Novellino, E., Mangiatordi, G.F., Nicolotti, O., 2005. Docking-based classification models for exploratory toxicology studies on high-quality estrogenic experimental data. J. Med. Chem. 6, 829. https://doi.org/10.1021/jm00342a066.

Trisciuzzi, D., Alberga, D., Mansouri, K., Judson, R., Novellino, E., Mangiatordi, G.F., Nicolotti, O., 2017. Predictive structure-based toxicology approaches to assess the androgenic potential of chemicals. J. Chem. Inf. Model. 57, 2874–2884. https://doi.org/10.1021/acs.jcim.7b00420.

Vuorinen, A., Odermatt, A., Schuster, D.J., 2013. In silico methods in the discovery of endocrine disrupting chemicals. J. Steroid Biochem. Mol. Biol. 137, 18–26. https://doi.org/10.1016/j.jsbmb.2013.04.009.

Wojciechowski, M., Lesyng, B., 2004. Generalized born model: analysis, refinement, and applications to proteins. J. Phys. Chem. B 108, 18368–18376. https://doi.org/10.1021/jp046748b.

Wu, S., Fisher, J., Naciff, J., Laufersweiler, M., Lester, C., Daston, G., Blackburn, K., 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chemical Research in Toxicology. https://doi.org/10.1021/tx400226u.

Ye, H., Shaw, I.C., 2019. Food flavonoid ligand structure/estrogen receptor-alpha affinity relationships - toxicity or food functionality? Food Chem. Toxicol. https://doi.org/10.1016/j.fct.2019.04.008.