IDENTIFICATION OF ENDOCRINE DISRUPTING CHEMICALS BY IN SILICO METHODS

Milen P. Todorov

"Prof. Dr. Assen Zlatarov", University, Department of Inorganic Chemistry

Prof. Yakimov str. #1, Bourgas, Bulgaria

Abstract

Identification of endocrine disrupting chemicals (EDCs) is one of the important goals of environmental chemical hazard screening. It is exhaustive and time consuming to test in vitro all chemicals – potential EDCs – used in industry, agriculture or as food preservatives against their effects on the endocrine system. Computational methods, such as virtual screening, quantitative structure activity relationships and docking, are already well recognized and used in many strategic programs for identification of EDCs. The aim of this study was to evaluate the predictive performances of model for identification of EDCs integrated in non-commercial software tool. The resulting statistics indicated that the binding affinities of the majority of chemicals included in the external validation dataset could be correctly predicted. However, a set of further improvements has been suggested in order to increase the predictive ability of the current profiling scheme for EDCs.

Key words: *estrogen receptor, QSAR, virtual screening, endocrine disruptors, in silico tools*

1. INTRODUCTION

Endocrine disrupting chemicals (EDCs) interfere with the synthesis, secretion, transport, metabolism, binding, or elimination of hormones (Diamanti-Kandarakis et al., 2009). Adverse health effects of EDCs in humans have been demonstrated to involve developmental, reproductive, neurological, cardiovascular, metabolic and immune systems (Schug et al., 2011). A wide range of natural and man-made chemical substances may be causing endocrine disruption and are considered as both human health and environmental hazards (Diamanti-Kandarakis et al., 2009). Costly testing of chemicals for their endocrine disruption potential is required in most industrialized countries (Adler et al., 2011). Because the mechanisms of endocrine disruption are diverse and complex (e.g., interactions with hormone and non-steroid receptors, activation of enzymatic and signaling pathways, etc.), a wide array of *in vitro* and *in vivo* tests is used to identify EDCs (Rotroff et al., 2013).

Structure-activity modeling plays an important role in government programs in support of protecting human populations from exposure to environmental contaminants (Demchuk et al., 2011). Specifically, computational methods to identify chemicals that may pose endocrine disruption hazard for additional *in vitro* or *in vivo* testing are important prioritization approaches (Tsakovska et al., 2011). Because of the diversity and complexity of endocrine disruption mechanisms, as well as the limited data available for *in silico* modeling, most studies have focused on EDCs that act via estrogen or androgen receptors. These modeling approaches include quantitative structure–activity relationship (QSAR) modeling (Salum et al., 2007), molecular dynamics simulations (van Lipzig et al., 2004), docking (Celik et al., 2008) and pharmacophores (Taha et al., 2010). Consequently, many of the models have been implemented as computational tools that are available either publicly or commercially.

In toxicological context, (Q)SAR models represent an effective way of predicting toxicity as a function of molecular structure and they can be used as a computational approach for data-gap filling within the framework of variety national or international research programs (Regulation (EC) No 1907/ 2006). In this context, a module called "*profiler*" for identification of estrogen receptor (ER) binders contained within the freely available Organisation for Economic Development (Q)SAR Toolbox (OECD QSAR Toolbox) offers a fast and efficient way for categorizing chemicals on the basis of their ER binding potential.

The aim of the present work is to evaluate the performance of the ER binding profiling scheme implemented within the QSAR Toolbox by its application for screening of large number of chemicals with known estrogenic effect. The results presented in this article are meant to help a potential user in assessing the uncertainty, which is related to a categorization scheme derived according to the set of rules encoded in the profiler.

2. MATERIALS AND METHODS

2.1. Estrogen binding data

Estrogen binding affinity data were obtained from the implemented in the QSAR Toolbox database related to this endpoint. This database is one of the largest and most heterogeneous datasets and it reports binding affinity data for 1460 chemicals. The experimental values are obtained by standard competitive radiometric binding assay. The values for each chemical are expressed as relative binding affinities in comparison with the estradiol affinity (RBA, %). Due to technical restrictions the analysis was performed by using 1235 out of all 1460 chemicals.

2.2. OECD QSAR Toolbox

This is a software tool especially designated for chemical risk assessment (OECD QSAR Toolbox). A key part of the system is so called categorization of chemicals. The categorization allows grouping of chemical substances into chemical categories. The chemical category is such a group of substances possessing similar physicochemical, toxicological and ecotoxicological properties or their fate in environmental and occupational surrounding or they behave using the common pattern as a result of chemical similarity.

An important advantage of the system is the large number of built-in profilers for different biological/toxic endpoints. Each profile consist a set of rules related to specific or general criteria associated to the respective endpoint.

2.3. Estrogen binding profile

The ER binding profiler requires only chemical structure information describing the two-dimensional (2D) structure of molecules (for instance coded in SMILES format or directly drawn by the user) as an input. According to the classification scheme, cyclic chemical structures weighting less than 500 Daltons (Da) and bearing an OH and/or an NH_2 group are considered as binders. On the other hand, a chemical is considered as a non-binder if it does not satisfy these rules or if its OH or NH_2 groups are impaired by ortho di-substitutions. This set of criteria was derived on the basis of the findings reported in the scientific literature (Schultz et al., 2002). Details for each category are given in Table 1.

Table 1. Structural criteria for the characterization of the binding potency of chemicals according to the ER binding profiler incorporated in the (Q)SAR Toolbox.

Categories of ER binders	**Predictive rule**
Very strong binders	200<MW≤500 Daltons (Da) in combination with two non-impaired OH groups attached to two different five or six carbon-atom rings.
Strong binders	Chemicals with at least one 5-or 6-members carbon ring with an unhindered hydroxyl or amino group and 200<MW≤500 Daltons (Da)
Moderate binders	Chemicals with a single five or six-member carbon ring structure with an unhindered hydroxyl-group (-OH) or amino-group (-NH2) (one in the para- or meta-position on the ring)

Weak binders	Chemicals with at least one 5-or 6-members carbon ring with an unhindered hydroxy or amino group and MW < 170 Da
Non binders	• impaired OH or NH_2 group • no presence of OH or NH_2 group • non-cyclic structure • exceeded upper limit related to molecular weight (MW > 500)

3. RESULTS AND DISCUSSION

The performance of any (Q)SAR model can be assessed by internal and external validation studies. While the internal validation is a part of the modeling process per se the external validation is a generally used measurement for reliability of the model when it is applied for "real" screening purpose. Two main key points should be considered before performing external validation exercise – the number and chemical diversity of compounds used as external set and the quality of experimental data. In the current study, the use of probably the larger public chemical collection covers both criteria which is a strong presumption for the ultimate significance of this validation study.

Since the ER binding profiler is not associated with predefined values of the relative binding affinities (RBA) the analysis was performed in two modes – general and RBA dependent.

In the *general mode* positive binding effect was assigned to all chemicals with non zero RBA values. As a result 881 chemicals from the external set were classified as ER binders and 353 as non binders. All positive classification labels (very strong, strong, moderate and weak) were combined and the final prediction result from any of them is considered as *positive*. The prediction results for this mode of analysis are presented in Table 2.

Table 2. Prediction results for estrogen binding obtained in general mode of the analysis

ER binding effect	Number of chemicals	Predicted ER binding effect
Positive (RBA*>0)	881	88% (775/881) Sensitivity
Negative (RBA=0)	353	63% (222/353) Specificity

*RBA – Relative binding affinity

The performance in terms of sensitivity (correct predicted ER binders) shows that the ER profiler is able to identify 88% of all ER binders in the set. In respect to specificity (correct predicted non ER binders) the performance is 63% only. The rate of false positives (37%) is high which prompts the obvious need for further precise reevaluation of the rules expecting decreased number of wrong predicted non binders.

While the predictions in terms of simple positive/negative results may be useful in prioritization studies most important and informative are those studies which provide qualitative predictions of the ER binding potential of the investigated chemicals. Because the ER profiler is not related to specific RBA ranges for each positive group (very strong, strong, moderate and weak) the next step of the study was focused on definition of activity bins to which the predefined groups can be assigned. Serafimova et al. (2007) reported a QSAR model for identification of estrogen binders in several classes defined by specific RBA ranges. The potency ER binding ranges from this study were used to perform the second - *RBA dependent* mode validation exercise. The specified RBA ranges and respective number of chemicals are presented in Table 3.

Table 3. Distribution of validation chemicals across ER binding potency bins

ER binding category	RBA range (%)	Number of chemicals
Very strong	RBA≥10	257
Strong	0.1≤RBA<10	347
Moderate	0.001≤RBA<0.1	235
Weak	0.0001≤RBA<0.001	42
Non binders	RBA=0	353

*RBA – Relative binding affinity

The obtained prediction results with detailed information for the contribution of each single ER binding category from the ER profiler for all five binding potency bins are presented in Table 4.

Table 4. Prediction results for all external set chemicals segmented by ER binding categories

ER binding category	Number of chemicals	Predictions				
		Very strong RBA≥10	Strong 0.1≤RBA<10	Moderate 0.001≤RBA<0.1	Weak 10^{-4}≤RBA<10^{-3}	Non binders
Very strong	257	153	95	-	-	9
Strong	347	186	99	-	-	62
Moderate	235	87	85	11	8	44
Weak	42	6	7	7	12	10
Non binders	353	11	47	26	47	222

*RBA – Relative binding affinity (%)

The prediction results for most potent (Very strong) ER binders shows sensitivity 60% (153/257). On the other hand it is important to note that significant number of chemicals - 95 was classified as "Strong" binders. Taken together both positive predictions for all chemicals with experimental RBA above 10% reach 96% sensitivity (248/257). This result shows extremely low number of false negatives (9 chemicals) which is an indication that the ER profiler allows correct predictions for chemicals with high ER binding potential.

The analysis for "Strong" ER binding range (0.1≤RBA<10) shows sensitivity 28% (99/347) as a result of correct predictions based on classification "Strong" binders only. Similarly to the previous group significant number of chemicals – 186 were predicted as "Very strong" binders. The resulting combined sensitivity for all positively predicted chemicals is 82% (285/347). It was found also that there are no predictions related to classification rules for moderate and weak ER binders. Regarding false positives which were found to be 18% (62/347) additional structural analysis doesn't help to set the focus on specific class of chemicals which are predominantly misclassified. In general despite of inability of the profiler to make predictions consistent with the RBA values the overall performance in terms of positive predictions is acceptable for this range of activity.

The group of moderate ER binders is represented by 235 chemicals. The correct predictions in terms of coincidence between RBA values and profiler for "Moderate" ER binders showed very limited success – only 5% (11/235). As it can be seen in Table 4 the predominance of the predictions are related to "Very strong" – 37% and "Strong" – 36% ER binders. This result suggests significant insufficiency of the profiler to produce reliable predictions for "Moderate" ER binders. However, this is not surprising result. Taking in mind that the predictive rule contains a general definition of structural features (presence of OH or NH_2 group) it is expectable that inherent characteristics of moderate ER binders are currently unavailable. It appears that improvement of the profiler is necessary in order to avoid overrating of RBA potential of the chemicals in this activity range.

The group of Weak ER binders is represented by smallest number of chemicals in this study – 42 chemicals only. It was found that predictions for weak binding effect are assigned to 12 out of all 42 chemicals (sensitivity 29%). The rest of positive predictions are evenly distributed across all the other higher RBA ranges (see Table 4). Due to the limited number of the investigated chemicals it is not possible to make reliable assessment of this sub level of the ER profiler. On the other hand it was interesting to investigate the list of missed positive chemicals. The responding classification label indicates that the negative predictions are result of absence of hydroxyl or amino groups for eight out of all ten chemicals. Additional analysis of these chemical structures shows that two chemical classes can be distinguished – *phthalates* and *halogenated biphenyls*. It is known that despite the lack of structural features (OH, NH_2) phthalates exert low estrogenic effect. In the current list representative chemicals are Dihexyl phthalate (CAS 84-75-3), Diisooctyl phthalate (CAS 27554-26-3) and Diisodecyl phthalate (CAS 26761-40-0). Serafimova et al. (2007) investigated the effect of these chemicals and they founded that the phthalic structural rule encoded as 2D representative SMILES in combination with range of 4.7-8.5 for log*K*ow (partition-coefficient octanol/water) could be successfully applied for positive predictions for these chemicals. The same predictive rule was applied for missed three chemicals and the obtained result showed positive prediction for all of them. Furthermore the analysis was extended in a way to investigate the predictions for all available phthalates in the whole external set. The total number of 32 structures was found to correspond to the phthalatic structural rule (applied via Query tool in the QSAR Toolbox). For all of them the ER profiler returns negative predictions for ER binding. In order to investigate the role of log*K*ow the values for this parameter were calculated by using built-in calculator in the system. The results presented in Table 5 strongly confirm the predictive ability of the specified log*K*ow range as rule for phthalates.

Table 4. Prediction results for all phthalates found in the validation set of chemicals

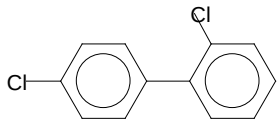
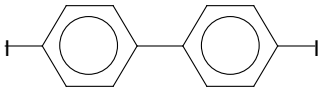
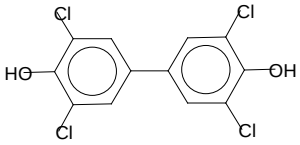
#	CAS	Name	RBA(%)	log*K*ow	ER prediction
1	117-81-7	Di-2-ethylhexylester	0.071	8.39	POS
2	523-31-9	Dibenzyl phthalate	0.0189	5.08	POS
3	84-74-2	Dibutyl ester	0.01	4.61	POS
4	85-68-7	Butylbenzylphthalate	0.00318	4.84	POS
5	131-18-0	Di-n-pentylphthalate	0.00165	5.59	POS
6	4376-20-9	1-(2-ethylhexyl) ester	0.00132	4.73	POS
7	3648-21-3	Diheptyl ester	0.00113	7.56	POS
8	84-61-7	Dicyclohexyl ester	0.00105	4.1	NEG
9	27554-26-3	Diisooctyl phthalate	0.000924	8.39	POS
10	84-75-3	Dihexyl phthalate	0.000918	6.57	POS

11	26761-40-0	Diisodecyl phthalate	0.000343	10.36	NEG
12	131-17-9	Diallylphthalate	0	3.36	NEG
13	84-66-2	Diethyl phthalate	0	2.65	NEG
14	131-11-3	Dimethyl phthalate	0	1.66	NEG
15	84-62-8	Diphenyl phthalate	0	4.1	NEG
16	84-69-5	Isophthalic acid	0	4.46	NEG
17	85-70-1	Butylphthalylbutylglycollate	0	4.15	NEG
18	117-83-9	Bis(2-butoxyethyl) ester	0	4.06	NEG
19	605-45-8	Diisopropyl phthalate	0	3.48	NEG
20	3319-31-1	Tris(2-ethylhexyl) trimellitate	0	11.59	NEG
21	4376-18-5	2-(methoxycarbonyl) benzoic acid	0	1.37	NEG
22	117-84-0	Phthalic acid	0	8.54	NEG
23	131-16-8	Dipropyl phathalate	0	3.63	NEG
24	28553-12-0	Diisononyl phthalate	0	9.37	NEG
25	84-76-4	Dinonyl phthalate	0	9.52	NEG
26	84-77-5	Didecyl phthalate	0	10.5	NEG
27	131-70-4	Butyl hydrogen phthalate	0	2.84	NEG
28	1528-49-0	Trihexyl trimellitate	0	8.87	NEG
29	2432-90-8	Phthalic acid, didodecyl ester	0	12.47	NEG
30	2694-54-4	Triallyl trimellitate	0	4.04	NEG
31	13728-34-2	Dimethyl 2,3-Naphthalenedicarboxylate	0	2.84	NEG
32	53894-23-8	Triisononyl trimellitate	0	13.06	NEG

The log*K*ow values for 82% (9/11) of the positive phthalates fall in the specified range which is highly reliable result. Furthermore, the results show that all non binders were out of the “active” range of log*K*ow. In conclusion it may be stated that manual inclusion the predictive rule related to phthalates will increase the predictive power of the ER profiler when it is used in further screening for identification of chemicals with estrogen binding effect.

The second group of negatively predicted Weak estrogen binders - *halogenated biphenyls* was also analyzed (Table 5). It is well known that hydroxylated metabolites of biphenyls exert estrogenic effect (Arulmozhiraja et al., 2005) and the mechanism is related traditionally to the presence of hydroxyl group in the structure.

Table 5. *Halogenated biphenyls* predicted as non estrogen binders

		
PCB 8 (CAS 34883-43-7)	4,4'-diiodobiphenyl (CAS3001-15-8)	3,3',5,5'-tetrachloro-4,4'-biphenyldiol (CAS 13049-13-3)

The binding mechanism of non phenolic halogenated biphenyls seems to be more complicated. It is expected that the estrogenic effect of these chemicals could be explained as result of non covalent interaction with the ER influenced by specific placement of chlorine atoms among both phenolic rings.

The prediction results for non ER binders shows specificity 63% (222/353). It was found that false predicted chemicals are distributed across all positive ER binding ranges (Table 4). For each activity group additional analysis was performed, however suggestion for improvement of the profiler was proposed for a single chemical class found in the set of wrongly predicted very strong binders. Eleven chemicals were found in this group (Table 4). It was found that five of them shared a structural moiety of flavonoid chemical structures. Taking in mind the biological function of these chemicals as antioxidants along with their clear negative potential for ER binding it is highly suggested inclusion of a specific rule in the profiler in order to provide correct predictions for flavonoids.

4. CONCLUSIONS

During the interaction with ligands the estrogen receptor can be reshaped and reorganized which in some cases may hamper development of highly precise structure–activity rules allowing in certain degree incorrect predictions. Nevertheless, the evaluation of the ER profiler incorporated in the QSAR Toolbox demonstrates generally well-defined rules which yield reasonable performances on large number of external validation set. As far as general predictive aspects are concerned, the prediction of moderate and weak binding affinities is less reliable than the prediction of strong and very strong ER binders. This result suggests that the profiler should be improved especially for lower ER binders by addition of specific rules for this activity bin. Concerning predictive rules for strong and very strong binders it is necessary to be further modified in order to be more distinctive among both groups. A way to increase the predictive performance is to account 3D characteristics of the ligands. For example this can be achieved by using distances between specific active sites (OH groups) as pre-filtering criteria. While the current version of the QSAR Toolbox (3.3) does not allow technical use of distances based on 3D chemical geometry such kind of analysis could be performed by external *in silico* applications and may be applied preliminary before application of the ER-profiler.

In conclusion, this study highlights the fact that the current ER profiler can potentially evolve towards a more refined and expanded set of rules after additional investigation of the training set chemicals. The ultimate benefit of such update is expected to be high degree of reliable predictions for potential endocrine disruptors which is considered to be a primary task in many toxicological research programs worldwide.

5. REFERENCES

Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K.E., Angers-Loustau, A., Aptula, A., Bal-Price, A., Benfenati, E., Bernauer, U., Bessems, J., Bois, F.Y., Boobis, A., Brandon, E., Bremer, S., Broschard, T., Casati, S., Coecke, S., Corvi, R., Cronin, M., Daston, G., Dekant, W., Felter, S., Grignard, E., Gundert-Remy, U., Heinonen, T., Kimber, I., Kleinjans, J., Komulainen, H., Kreiling, R., Kreysa, J., Leite, S.B., Loizou, G., Maxwell, G., Mazzatorta, P., Munn, S., Pfuhler, S., Phrakonkham, P., Piersma, A., Poth, A., Prieto, P., Repetto, G., Rogiers, V., Schoeters, G., Schwarz, M., Serafimova, R., Tahti, H., Testai, E., van Delft, J., van Loveren, H., Vinken, M., Worth, A., Zaldivar, J.M., 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects, *Arch. Toxicol*. vol. 85, pp. 367–485.

Arulmozhiraja, S., Shiraishi, F., Okumura, T., Iida, M., Takigami, H., Edmonds, J.S., Morita, M., 2005. Structural requirements for the interaction of 91 hydroxylated polychlorinated biphenyls with estrogen and thyroid hormone receptors. *Toxicol. Sci.* vol. 84, pp. 49–62.

Celik, L., Davey, J., Lund, D., Schiott, B., 2008. Exploring interactions of endocrine-disrupting compoundswith different conformations of the human estrogen receptor alpha ligand binding domain: a molecular docking study. *Chem. Res. Toxicol.* vol. 21, pp. 2195–2206.

Demchuk, E., Ruiz, P., Chou, S., Fowler, B.A., 2011. SAR/QSAR methods in public health practice. *Toxicol. Appl. Pharmacol*. vol. 254, pp. 192–197.

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.* vol. 30, pp. 293–342.

OECD QSAR Tollbox http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm (last accessed 10.02.2015).

Rotroff, D.M., Dix, D.J., Houck, K.A., Knudsen, T.B., Martin, M.T., McLaurin, K.W., Reif, D.M., Crofton, K.M., Singh, A.V., Xia, M., Huang, R., Judson, R.S., 2013. Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. *Environ. Health Perspect.* vol. 121, pp. 7–14.

Salum, L.B., Polikarpov, I., Andricopulo, A.D., 2007. Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. *J. Mol. Graph. Model.* vol. 26, pp. 434–442.

Schultz T., Sinks G., and Cronin M. 2002. Structure-activity relationships for gene activation oestrogenicity: Evaluation of a diverse set of aromatic compounds, *Environ. Toxicol.* vol. 17, pp. 14–23.

Serafimova R., Todorov M., Nedelcheva D., Pavlov T., Akahori Y., Nakai M. and Mekenyan O. 2007. QSAR and mechanistic interpretation of estrogen receptor binding. *SAR and QSAR in Env. Res.* vol. 18, no. 3-4, pp. 389-421.

Sung, E., Turan, N., Ho, P.W., Ho, S.L., Jarratt, P.D., Waring, R.H., Ramsden, D.B., 2012. Detection of endocrine disruptors — from simple assays to whole genome scanning. *Int. J. Androl.* Vol. 35, pp. 407–414.

Taha, M.O., Tarairah, M., Zalloum, H., Abu-Sheikha, G., 2010. Pharmacophore and QSAR modeling of estrogen receptor beta ligands and subsequent validation and in silico search for new hits. *J. Mol. Graph. Model*. vol. 28, pp. 383–400.

The European Parliament and the Council of the European Union, Regulation (EC) No 1907/ 2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/ 769/EEC and Commission

Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC, Off. J. Eur. Union L306 (30.12.2006), pp. 1–849.

Tsakovska, I., Pajeva, I., Alov, P., Worth, A., 2011. Recent advances in the molecular modeling of estrogen receptor-mediated toxicity. *Adv. Protein Chem. Struct. Biol.* vol. 85, pp. 217–251.

van Lipzig, M.M., ter Laak, A.M., Jongejan, A., Vermeulen, N.P., Wamelink, M., Geerke, D., Meerman, J.H., 2004. Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *J. Med. Chem.* vol. 47, pp. 1018–1030.