COMPARATIVE EVALUATION OF THE APPLICABILITY OF IN SILICO TOOLS FOR IDENTIFICATION OF ENDOCRINE DISRUPTING CHEMICALS

Milen Todorov
Assen Zlatarov University, Department of Inorganic Chemistry, 8010 Burgas, Bulgaria,
mtodorov@btu.bg

ABSTRACT

Rapidly and correctly identifying endocrine-disrupting chemicals (EDCs) is an important issue in environmental risk assessment. Because of the high cost and time-consuming nature of experimental tests, *in silico* methods are valuable alternative tools for the identification of EDCs. In silico tools like knowledge-based expert systems and (quantitative) structure-activity relationship models have been created or upgraded on the yearly basis and also widely advertised to be used as primary screening technique in studies related to receptor mediated effects. The aim of the present work is to evaluate the performance of the ER binding profiling schemes implemented within the QSAR Toolbox. The results presented in this article are meant to help a potential user in assessing the uncertainty, which is related to the categorization rules encoded in the profilers.

Keywords*: endocrine disruption, computational toxicology, QSAR*

INTRODUCTION

Humans and other species are exposes every day to tens of thousands of man-made chemicals. Some of these chemicals are able to mimic natural hormones and disrupt normal functions of the endocrine system [1]. Such chemicals, called endocrine disruptors (ED), can pose a serious threat to the reproductive ability of humans and wildlife, and are thought to have led to a decline in some wildlife populations through a variety of mechanisms, mainly estrogen-receptor (ER)-mediated mechanisms of toxicity. Among the chemicals capable of causing endocrine disruption, environmental estrogens, which are derived from a variety of sources such as pesticides, plastics, combustion byproducts, and plants, that is, phytoestrogens and agricultural products form a large group [2, 3].

Testing the actions of all used chemicals – possible EDCs – against all the potential targets related to endocrine disruption is an important but also expensive and difficult, if not impossible, task, also due to the limited availability of suitable bioassays. Especially biological testing is time and cost intensive; therefore, more rational approaches to help to identify potentially harmful chemicals in a fast way are urgently needed. In this context, methods established in drug discovery and development, where the task is to identify bioactive compounds from millions of available substances, can be applied to EDCs research.

On being faced with the challenging task of screening large molecule libraries for biological activity, the benefits of virtual screening with quantitative structure-activity relationship (QSAR) techniques become immediately obvious when endeavoring to identify possible endocrine disruptors. Instead of expensive laboratory work, biological activity will be predicted solely on the basis of the molecular structure of the compound and will, thus, decrease the number of tests, in accordance with the EU recommendations in the new REACH system for chemical regulation [4].

Recent findings have proven QSARs valuable assistance in predicting the estrogenic activity of organic molecules. Models based on different computational methods [5] provide different perspectives on the interactions between the estrogen receptor and its ligands. However, despite the success of QSAR techniques, their limitations must be identified for the regulatory communities to accept them. Another issue that deserves a special attention is the high cost of the major computational instruments used for predictions. Considering that the combinatorial use of such tools will bring reliable results it is obvious that significant financial support is needed. On the other

hand, proprietary rules or models have been developed in many private companies based on in house-generated experimental data for new chemicals. Most often they are used to follow up on negative results obtained with public or commercial tools, in order to increase the sensitivity of predictions. Priority should be given to public or commercial tools in order to promote transparency, standardization and knowledge sharing. Focusing on tools that are widely available will also encourage sharing of data (e.g. estrogen binding *in vitro*) to further improve the systems on a community basis.

As an excellent demonstration of development of standardized non-commercial tools for chemical risk assessment one should pointed out the development of the QSAR Toolbox platform [6]. Currently it is accepted and used in many companies, organizations and national authorities for *in silico* predictions of different biological endpoints.

In a recent study [7] the predictive ability of the estrogen binding predictive module (ER binding profiler) incorporated in the Toolbox has been assessed by predictions for larger number of chemicals taken from public source. As far as general predictive aspects are concerned, the predictions for high binding affinity show good performance, whereas the predictions for lower affinities need to be further improved. On the other hand the presence of another module in the platform related to the ER binding effect could improve the overall assessment for receptor binding potential for new chemicals. The aim of the present study is evaluation of the predictive ability of both profiling schemes. This combined approach will give an opportunity to researchers to obtain more reliable predictions in terms of mechanistic interpretation and clear interpretation of the results.

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 2437 chemicals. The experimental values are obtained by standard competitive radiometric binding assay. The values for each chemical are expressed as relative binding affinities (RBA, %) in comparison with the estradiol affinity.

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. General 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 [8].

2.4. Expert system for estrogen binding prediction

The expert system consist a set of molecular definitions that mimic structural criteria identified in potential estrogen receptor. It is rule-based decision tree that encodes the expert's mechanistic understanding with respect to both the chemical and biological aspects of the estrogen binding. The ultimate prediction is supported by information of the identified chemical category and additional details of the interaction mechanism.

RESULTS AND DISCUSSION

The performance of both modules for estrogen binding prediction has been assessed by their application by making use a database of 1497 out of the total 2437 chemical compounds. This reduction was necessary because the experimental data for relative binding is available for different species. In order to avoid mixing data the analysis was performed by using only experimental results obtained by using recombinant human estrogen receptor. Initially, the positive binding effect was assigned to all chemicals with non-zero relative binding affinity (RBA) values. The obtained results are shown in Table 1.

Table1. Prediction results for estrogen binding obtained by both modules - General estrogen binding profile and the Expert system for estrogen binding

ER binding effect	Number of chemicals	Predicted ER binding effect	
		by GEBP	by ESEBP
Positive (RBA>0)	348	62% (214/348)	21%(73/348)
Negative (RBA=0)	1085	78% (849/1085)	89%(965/1085)

RBA – Relative binding affinity; GEBP - General estrogen binding profile; ESBP - Expert system for estrogen binding prediction

As it can be seen the general estrogen binding profile performs better especially in terms of sensitivity. In addition it was found that overlapping positive predictions by both schemes are obtained for eight chemicals only. This clearly demonstrates that the predictive rules are quite different in both schemes. An important advantage of the general profiler is its ability to predict chemicals in predefined binding affinity ranges. Currently, the profiler could classify a chemical in one of the following categories: very strong, strong, moderate or weak estrogen binder. Since there is no indication of the qualitative thresholds for each category additional study was performed in order to assess the predictive ability of the profiler. Based on the experimental data for relative binding affinity an analysis was performed for all 214 positive predicted chemicals. They were segmented in three activity ranges presented in Table 2.

Table 2. Prediction results obtained by general ER profile 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}
Very strong	18	10	8	-	-
Strong	45	24	21	-	-
Moderate	135	40	58	21	16

*RBA – Relative binding affinity (%)

Results show that the most consistent predictions are obtained for very strong binders – 10 out of total 18 chemicals (56%). For lower binding affinities the percentage of consistent predictions drops down from 47% (21 out of total 45 chemicals) for strong binders to 16% for moderate binders (21 out of 135 chemicals). As it can be seen there is a need of further improvement of the profile especially for lowest range of estrogen binding affinity. A basis for such analysis could be inspection of the chemical structures. It was found that all forty chemical contains two phenolic rings. This structural requirement is known and associated with binding effect with high potency. In addition it should be emphasized that along with the presence of both hydroxyl groups another requirement is a specific distance between oxygen atoms. 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. More generally, despite the distance between atoms more precise definitions of the rules for moderate estrogen binders (e.g. strict definition of substituents) is expected to increase the consistency with the experimental data.

Another important aspect of the analysis is assessment of the performance in respect to false positive predictions. The prediction results for non-estrogen binders shows specificity of 78% (correct predicted 849 out of all 1085 non binders). It was found that false predicted chemicals are distributed across all positive binding ranges. For each activity group additional analysis needed in order to find a way for improvement of the profiler. Similarly to the group of false negatives it is expected that definition of structural masks will reduce the number of positive predictions for non-estrogen binders.

The expert system for estrogen binding affinity was analyzed in a same manner. Concerning low sensitivity (21%) there is obvious need of improvement. One may assume that the list of structural rule definitions should be expanded in order to cover the diversity of the chemicals with positive binding effect. Moreover, due to the lack of qualitative information associated with current list of prediction rules the expansion of the available rules could be done by further evaluation of all estrogen binders with negative predictions.

Concerning false positive predictions it was found that major part of them are classified as Alkylphenols (17 chemicals), DDT-like compounds (21 chemicals) and multicyclic hydrocarbons (26 chemicals).

The estrogenic effect of alkylphenols has been investigated by Serafimova et al. [9]. In order

to cause binding effect to the estrogen receptor a specific structural rule is defined in combination with range of 0.63-1.53 [a.u./eV] for molecular descriptor volume polarizability (defined as a sum of atomic self-polarizabilities, and describe the averaged ability of a compound to change electron density as its atoms during chemical interactions.) The same predictive rule was applied for all 17 false positives chemicals and the obtained result showed reduced number of wrong predictions (Table 3).

Table 3. Prediction results for alkylphenols

#	CAS	Name	RBA(%)	Volume Polar. [a.u./eV]	ER prediction
1	88-18-6	2-ter-butylphenol	0	0.70	POS
2	88-60-8	2-tert -butyl-5-methylphenol	0	0.77	POS
3	89-72-5	O-Sec-butylphenol	0	0.69	POS
4	89-83-8	Thymol	0	0.69	POS
5	90-00-6	2-ethylphenol	0	0.55	POS
6	95-48-7	o-Cresol	0	0.47	POS
7	106-44-5	p-Cresol	0	0.47	POS
8	108-95-2	Phenol	0	0.40	POS
9	123-07-9	p-ethyl-phenol	0	0.55	POS
10	499-75-2	Carvacrol	0	0.69	POS
11	585-34-2	3-tert-butylphenol	0	0.69	POS
12	620-17-7	m-Ethylphenol	0	0.55	POS
13	645-56-7	4-n-Propylpehnol	0	0.62	POS
14	2409-55-4	2-tert-butyl-4-methyl-phenol	0	0.76	POS
15	527-54-8	3,4,5-trimethyl-phenol	0	0.62	POS
16	1138-52-9	3,5-di-tert-butyl-phenol	0	0.99	POS
17	501-24-6	3-n-Pentadecylphenol	0	1.50	POS

As it can be seen almost half of the chemicals (# 5, 6, 7, 8, 9, 12, 13, and 15) could be predicted correctly. Additionally, the range for volume polarizability could be further analyzed and modified in order to improve the predictive ability of the profile for this chemical class. All necessary calculations could be done by application of module for 2D parameters which is implemented in the Toolbox.

The group of DDT-like compounds was analyzed in the same manner. The predictive rule for this category was found to be defined very general. In addition, there are no other requirements (molecular descriptors) that could be associated to those compounds with estrogen binding effect. In such a case it is possible to investigate externally selected chemicals (from available datasets) for identification of specific molecular descriptors which may explain their positive effect. For example, by making use of the technical capability for defining 2D structures in the Toolbox, twenty six chemicals were found with available experimental data for estrogenic effect. By contrasting positives and negatives in respect to different appropriate molecular descriptors a robust predictive rule could be defined.

The last group of multicyclic hydrocarbons was analyzed in a same manner as DDT-like compounds. Again, more specific structural definitions are expected to improve predictions and addition of ranges for specific molecular descriptors as well. Since the aim of this study is not development of new models it may serve to point out some insufficiencies that should be taken into

account when this profile is used.

CONCLUSIONS

In this study existing profiling schemes for identification of chemicals able to bind the estrogen receptor have been evaluated and compared. In terms of sensitivity (correct predicted estrogen binders) the general estrogen binding profile performs better. However, possible further improvements are discussed. In terms of specificity (correct predicted non-estrogen binders) both profiles provide satisfactory results. Concerning predictive rules which bring larger part of incorrect predictions the focus was set on Alkylphenols, DDT-like compounds and multicyclic hydrocarbons. For each class further modifications are needed in order to be more distinctive among binders non-binders.

In conclusion, this study highlights the fact that current definitions of investigated profiles can potentially evolve towards a more refined and expanded set of rules after additional investigation of new 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.

REFERENCES

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

2. Saliner, A. G., Amat, L., Carbo´-Dorca, R., Schultz, T. W., and Cronin, M. T. D. (2003) Molecular quantum similarity analysis of estrogenic activity. J. Chem. Inf. Comput. Sci. 43, 1166-1176.

3. Petkova S., and Ramon S. Nanotoxicology and regulatory issues of emerging technologies. Organisation for the Prohibition of Chemical Weapons International Cooperation and Assistance / ICB, 2013, AP2013/RP/06

4. https://echa.europa.eu/bg/regulations/reach (last accessed on 15 January 2017)

5. 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., 85, pp. 217–251.

6. The OECD QSAR Toollbox http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm (last accessed 29.01.2017).

7. Todorov M. Evaluation of in silico approach for identification of estrogen receptor ligands from chemical inventories. 2016, Industrial Technologies, 2016, Vol III (1), pp. 87-93.

8. 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.

9. 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.