

# Priority Setting of Endocrine Disruptors Using QSARs

**Weida Tong**

Manager of Computational Science Group, Logicon ROW Sciences,  
FDA's National Center for Toxicological Research (NCTR), U.S.A.

Thanks for the nice introduction, Bob. Before I start I would like first to thank Dr. Jun Kanno for inviting me to participate in this important symposium, and also I would like to thank the Ministry of the Environment for giving me this opportunity to talk about our research work involving endocrine disruptors at the U.S. FDA's National Center for Toxicological Research.

It is getting quite clear that endocrine disruptors may have diverse effects on human beings and wildlife. In the United States, two laws have been passed by the U.S. Congress to mandate the U.S. EPA to come up with a strategy for screening and testing a large number of chemicals in the drinking water and food additives for their potential endocrine disruptions.

A two-tier screening and testing approach has been established in the United States, which contains about 20 different *in vitro* and *in vivo* assays. Running through these assays for a chemical would be time consuming, labor intensive and very expensive. In order to more efficiently implement this two-tier screening and testing strategy, a pre-screening technique is required to determine which chemicals need to be tested first.

Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) recommends QSAR models as prescreening tools to provide biological effect information, such as receptor binding activity information for the purpose of priority setting.

About five years ago at the U.S. FDA's National Center for Toxicological Research, we started a project called Endocrine Disruptor Knowledge Based Project. In this project, we did a binding assay for four different receptors: estrogen receptor, androgen receptor, alpha-fetoprotein, and SHBG. These four receptors are associated with the endocrine disruption.

We tested about 200 chemicals for each receptor and standardized the binding assay protocols. The chemicals that we selected for these assays covered a wide range of structural diversity and binding activities. These data have been used to develop QSAR models in our laboratory.

After many years working in this area, we have accumulated a large number of endocrine disrupter data. These data are summarized in our EDKB database, which is a web-based database and software. Users can access our database through <http://edkb.fda.gov> to look up the endocrine disrupter data.

About two years ago, we established an interagency agreement with the U.S. EPA to modify our QSAR models for the purpose of priority setting. The NCTR model will be validated by the EPA to compare the prediction results with the experimental results for about 250 chemicals. The validation process for the ER models expects to be completed by the end of this year. We also developed AR models; the process is nearing completion in our laboratory.

This slide provides basic background of QSARs. The fundamental principle of QSAR is to assume that chemicals with similar structures will have similar biological activities. So, in QSAR we are trying to establish the relationship between chemical structure and biological activity.

The chemical structure can be represented using a set of so-called chemical descriptors. Those descriptors describe 2D and 3D structural information as well as their physical/chemical properties. The biological activity data being modeled in QSARs could be associated with a single biological event, such as receptor binding, or it also can be associated with an endpoint at a higher level of the biological complexity such as acute or chronic toxicity.

In today's presentation, I am going to present the results of our QSAR models to predict the estrogen receptor binding activities. The models have been developed based on the 230 chemicals that are assayed in our laboratory. And then, these models will be used predict the ER binding activity of untested chemicals solely based on the chemical structures. That is, as Dr. Jun Kanno said, the *in silico* predictions.

Using QSARs for priority setting is quite a common practice in the drug discovery community. In drug discovery, QSARs are primarily used to narrow down the number of chemicals for testing. Efforts have been focused on identifying high affinity chemicals, and there have little concern of missing some active chemicals.

The application of the QSAR model for priority setting in regulatory application is different from that used in drug discovery. Even though we still want to narrow down the number of chemicals for testing, we try to identify both the high and low affinity chemicals. False negatives are of greater concern in such applications.

False negatives are those chemicals, which are predicted as inactive, but are actually active in the assay. So, for further experimental evaluations, the false negatives will go to the lower priority group. They will cause more adverse effects in the environment for many years.

There are two criteria we have used to develop the QSAR model. First, the QSAR model should demonstrate efficiency in its capability to narrow down the number of chemicals for testing. The second criteria, I think, is the most important one, the models should not introduce any false negatives because this is a very early stage of identification of endocrine disrupters.

We have evaluated a number of different QSAR approaches for such applications, and have found that using a single QSAR model cannot meet these two criteria. So, at NCTR we combined and integrated several different QSAR models into a so-called "Four-Phase" approach. This is a progressive phase paradigm, and each phase contains several complementary QSAR models to minimize the false negatives. A previous phase is used as a screen to reduce the number of chemicals for prediction by the subsequent phase. Our prediction starts from the qualitative prediction, such as yes or no answers, and then the quantitative predictions will be made in the later phase. The computational time and human expertise will get more heavily involved in the later phase compared to the earlier phase. So far, we have found that this approach can eliminate over 80% of the chemicals for testing, and we did not see any false negatives in such applications.

With respect of the Four-Phase approach for prediction of the estrogen receptor binding activity, in Phase I, we used two simple rejection filters to eliminate those chemicals most unlikely to bind to estrogen receptor. These two rejecting filters are molecular weight range and no-ring structure indicator. That means chemicals whose molecular weight is less than 94 or larger than 1000, or chemicals not containing any ring structure will be eliminated from this phase.

Chemicals that pass Phase I will go to Phase II to be assigned as either active or inactive. The inactive chemicals will be eliminated from this phase, and active chemicals will go to Phase III for the quantitative prediction.

In this final stage of the integrated system, we strongly feel that a set of rules needs to be constructed as a knowledge-based or expert system to make a priority setting decision. The system is useful only after incorporating accumulated human knowledge and expertise (i.e., rules). This system can make decisions on individual chemicals based on the rules in its knowledge base. Computational chemists, toxicologists and regulatory agencies should be included in the definition of the rules. It is believed that without human knowledge intervention any QSAR model cannot reach its optimal performance.

Because of the time limit, today I am just going to briefly describe the QSAR model in the Phase II and Phase III. Then, I will present some results to demonstrate the capability of this system.

As I mentioned earlier, chemicals that pass Phase I will go to Phase II to be assigned either active or inactive using 11 models: three structural alert models, seven pharmacophore models, and one classification tree model. Each model only provides yes or no predictions. Chemicals classified as inactive in this phase will be eliminated. Only chemicals predicted as active will go to Phase III for further evaluation.

In this phase, we try to identify all the active chemicals, but the same time, the models used in this phase also have to demonstrate the capability to reduce the data size. Based on these considerations, each model selected in this phase encodes the unique and key structural requirements for the ER binding activities.

These are three structure alerts in this phase. The structure alert is the key 2D substructure that is important for the binding. A chemical containing this structure alert will be considered as active. For example, estradiol has this steroidal skeleton, highlighted in yellow. Genistein contains two benzene rings separated by two carbons, and nonylphenol contains phenolic rings.

A similar concept also holds true for the pharmacophore models, but there are some slight differences. Pharmacophore models identify active chemicals based on the structure features on the 3D space. In addition to that, we also include the so-called shape restraint that means for an active chemical, the size has to fit into this shape.

The last model we used in this phase is called the classification tree model. The classification tree models classify chemicals into the active or inactive category based on a set of rules, just like Japanese pachinko game, and the rule can be a statement like IF logP is less than 0.5, THEN the chemical is inactive, else it will be active, and so forth.

These eleven models can provide two sets of predictions in this phase. The first is just active or inactive predictions. In order to minimize the false negative rate in this phase, we defined active chemicals as those, which have been predicted as active by any one of the eleven models. Only those chemicals that have been predicted as inactive by all the models are considered as inactive chemicals. Using such approach, we can reduce and minimize the false negatives in this phase. Also, we can use these eleven models to provide semi-quantitative estimation of the ER binding activity. The rationale behind that is assumed that if a chemical has been predicted as active by more models, this chemical should be more active. Based on the number of models predicting a chemical as active, we can estimate the binding activity.

As I mentioned earlier, the chemicals that pass Phase II will go to Phase III for quantitative predictions. In this phase, we have evaluated a number of different QSAR approaches, including the classic QSARs, HQSAR, and CoMFA — that is the one Dr. Kanno mentioned. Each of these models has been fully validated using internal and external validation. In internal validation, we are using the standard statistical measures such as  $r^2$  and  $q^2$  to assess the model's quality, stability and predictivity. In the external validations, we use a test set to challenge the model. The test set contains chemicals not included in the training set for developing the model.

So far, we have found that CoMFA performs the best compared to the classic QSAR and HQSAR. And currently we are just using the CoMFA model in this phase for quantitative prediction. By the way, the CoMFA research is sponsored by American Chemistry Counsel (ACS) through ACS-NCTR CRADA.

This is the result of the 3D QSAR/CoMFA model. The y-axis shows the model-calculated results, and the x-axis is experimental results for ~130 compounds. You can see the model-calculated results are very well in agreement with the experimental results. The correlation coefficient is 0.91. Importantly, this model shows the  $q^2$  value is 0.7. The  $q^2$  value measures the model's predictivity, and as a rule of thumb if a QSAR model has a  $q^2$  value larger than 0.5, the model is considered predictive. Our model has a  $q^2$  value as equal to 0.7.

There are two data sets reports by Kuiper and Waller, respectively. These two data sets contained about 40 compounds not included in our training data sets. Thus, these two data sets can be used as testing data sets to challenge our models.

We did the prediction for these 40 chemicals. The predicted result is shown in this slide. The black triangle is the prediction results for the Kuiper data set and the blue square for the Waller data set. The predicted results are shown in the y-axis and experimental results in the x-axis. You can see the predicted result is very well in agreement with the experimental results.

There are also six chemicals that have been determined as inactive in experiments in these data sets. Five of them are also shown to be inactive by our predictions; for the sixth chemical the experiment show its activity below -2.5, the CoMFA model predicts its logRBA = -2.6, the experimental and predicted results are also consistent.

Now, I am going to show some applications of the Four-Phase approach to several data sets. Actually, we have applied our approach to a number of data sets. I just show five data sets here because these five data sets contain chemicals as they relate to the environment and industrial compounds.

The Nishihara data set and the Soto data set totally have about 600 chemicals. The ER activity data are available for these two data sets. We applied Phase I and Phase II to these two data sets, and were able to reject about 60 to 70% of chemicals for testing. Importantly we did not see any false negatives for these two data sets.

We also applied our Four-Phase approach to these three data sets. These three data sets were provide by U.S. EPA and might need to go through various *in vitro* and *in vivo* assays. As you can see, only using Phase I and Phase II we have been able to eliminate over 85% of the chemicals for testing. If we include Phase III, we can eliminate over 90% of the chemicals for testing.

In the next few slides, I am going to show more details on prediction of the HPV data sets using the Four-Phase approach.

HPV stands for High Production Volume. That means that the data set contains chemicals that are in high-production in the United States. As I mentioned at the beginning of my talk, we have an EDKB database containing a lot of the data relating to endocrine disrupters. The ER activity data has been collected from the binding assay, yeast based reporter gene assay, yeast two-hybrid assay, MCF7 cell proliferation assay, and some *in vivo* data.

We have been able to compare these HPV compounds against our database to see how many chemicals in this HPV list already have the biological data available. We found 68 chemicals that have been tested experimentally before. So now we can look in more detail to see how our prediction results compare to the experimental results for these 68 compounds.

The results are summarized in this table. You can see for the 68 compounds, twenty-seven are predicted as inactive, and the predictions are consistent with experimental observation.

Forty-one chemicals go to Phase II: 24 of them are predicted as inactive. For these 24 chemicals we have four different types of assay data available, and 23 of them are consistent with our predictions. One chemical, dibutyl phthalate is tested by five laboratories; four of them show it as inactive and one shows it as active. Our prediction for this chemical is inactive. For the 17 chemicals that go to Phase III, ten of them have been predicted as inactive. These ten chemicals also have four types of assay data available, and nine of them are consistent with experimental observation. The another chemical, butylbenzyl phthalate, we predict it as inactive. For this chemical, eight laboratories have the test data: six of them show that it is inactive, but other two laboratories show that this chemical as active. Through this process, we identify seven chemicals that go to Phase III.

## Q&A

Kavlock: Can we ask you to hurry to your conclusion?

Ton: Yes. This is our Phase III prediction results compared to the assay results for these seven chemicals. You can see the assay results are consistent with our predictions.

In summary, the data and work I showed in this presentation is a collaboration effort with the U.S. EPA for developing priority setting tools using QSARs, and the system we developed is called the Four-Phase approach. The system has been used to apply a number of data sets, and has been demonstrated to have great efficiency.

The models have been able to reject over 90% of the chemicals for testing, and have a minimal false negative rate. The systems will be validated by the end of this year. The results will be posted on the internet so everyone can see it, and the same system is already developed for the androgen receptor binding activity. Thank you.

Kavlock: Questions?

Q: In Phase III of your model, what defines a negative?

Tong: Less than -3 of the predictions — log RBA less than -3 — we consider as inactive.

Q: So that is less than .001%...

Tong: Minus 3.0. This is, yes, .0001.

Kavlock: No questions? Thank you.