Pure Appl. Chem., Vol. 75, Nos. 11–12, pp. 2389–2396, 2003.

Workshop 1.3

QSAR prioritization of chemical inventories for endocrine disruptor testing*

Patricia Schmieder[1,‡], Ovanes Mekenyan[2], Steven Bradbury[3], and Gilman Veith[1]

[1]U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Mid-Continent Ecology Division, 6201 Congdon Blvd., Duluth, MN 55804, USA; [2]Laboratory of Mathematical Chemistry, University "Prof. As. Zlatarov", 8010 Bourgas, Bulgaria; [3]U.S. Environmental Protection Agency, Office of Pesticide Programs, Environmental Fate and Effects Division, Washington, DC 20460, USA

Abstract: Binding affinity between chemicals and the estrogen receptor (ER) serves as an indicator of the potential to cause endocrine disruption through this receptor-mediated endocrine pathway. Estimating ER-binding affinity is, therefore, one strategic approach to reducing the costs of screening chemicals for potential risks of endocrine disruption. While measuring ER binding with in vitro assays may be the first choice in prioritizing chemicals for additional in vitro or in vivo estrogenicity testing, the time and costs associated with screening thousands of chemicals is prohibitive. Recent advances in 3D modeling of the reactivity of flexible structures make in silico methods for estimating ER binding possible. One technique, the common reactivity pattern (COREPA) approach, was applied to development of reactivity patterns for ER relative binding affinity based on global nucleophilicity, interatomic distances between nucleophilic sites, and local electron donor capability of the nucleophilic sites. The reactivity patterns provided descriptor profiles for order-of-magnitude RBA ranges of training set chemicals. An exploratory expert system was subsequently developed to predict RBA and rank chemicals with respect to potential estrogenicity. A strategy is presented for extending initial exploratory 3D QSAR models beyond current training sets to increase applicability to more diverse structures in large chemical inventories.

INTRODUCTION

The concern over potential hazards posed by environmental chemicals that can mimic endogenous estrogens is evidenced by screening and testing programs under development by governments of the United States, European Union, and Japan. Assessing the risk of potential endocrine-disrupting chemicals (EDCs) is particularly challenging due to the diverse nature of chemical structures believed to initiate activity through a common toxic pathway, i.e., direct chemical interaction with the estrogen receptor (ER). When attempting to institute an efficient approach to predict the potential hazard of EDCs, or any chemical hazard, the following question arises. How are efforts best focused on those chemicals most likely to cause adverse effects without empirically testing all chemicals of regulatory concern?

*Report from a SCOPE/IUPAC project: Implication of Endocrine Active Substances for Human and Wildlife (J. Miyamoto and J. Burger, editors). Other reports are published in this issue, *Pure Appl. Chem.* **75**, 1617–2615 (2003).

‡Corresponding author

The use of quantitative structure–activity relationships (QSARs) is a powerful in silico technique useful for prioritizing chemicals for subsequent empirical assessments. Specific challenges faced in developing and applying QSAR to rank and prioritize within large chemical inventories are addressed herein, and include the proper selection of methods that reflect the state-of-the-science, both in regard to what is understood of the underlying biology and chemistry, as well as what computational approaches are possible and applicable.

A critical consideration when applying predictive models is establishing the applicability of any model to a defined universe of chemicals of concern in a particular risk assessment. Thus, the most difficult step may not be QSAR development, per se, but determining when a model is sufficiently improved to satisfy criteria defined for model acceptance. Recently, an international effort has been undertaken to define criteria for acceptance of QSAR in risk assessment [1]. These criteria include the ability to provide predictions for a specified chemical domain of concern within acceptable limits of error. The users (e.g., the regulatory community applying the models) specify the chemical domain of concern and the limits of acceptance. Acceptable error is dependent upon the risk assessment scenario being employed. For instance, predictions used to prioritize chemicals for further testing can have larger uncertainty bounds than those upon which final regulatory determinations are to be made.

A scheme for iterative QSAR model development, as presented in Fig. 1, allows for model development, assessment, and improvement through a series of steps, following rigid scientific criteria, until the model is determined acceptable for use in risk assessment. The scheme is conceptually based upon earlier QSAR approaches used to predict acute toxicity to aquatic organisms for thousands of industrial chemicals [2–7]. Lessons learned from these earlier efforts emphasize the importance of defining the chemical structure space to which models will be applied, and employing approaches which are designed to obtain the maximum amount of new information from any empirical testing conducted for purposes of model improvement.

The scheme in Fig. 1 starts with identification of an adverse effect of concern in a risk assessment. Knowledge, or hypothesis, of the underlying toxicity pathway likely producing the adverse effect is needed to identify a well-defined biological endpoint that is plausibly related to the adverse outcome. An endpoint is also well defined if it can be demonstrated that the chemical concentration and form applied in a study are directly linked to the observed alteration in the biological endpoint. In the case of

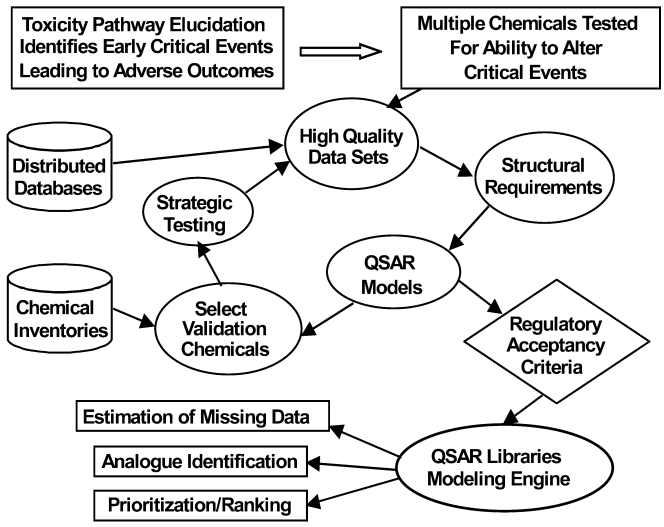

Fig. 1 Iterative QSAR model development scheme: A generalized scheme for development of predictive models to meet regulatory acceptance criteria for applicability to large chemical inventories.

receptor binding, this would require, for instance, assurance that the chemical applied (and not a metabolite of the chemical) is responsible for the apparent affinity, and that displacement of the endogenous ligand is due to competitive binding at the active site and not to some other alteration in protein structure. The next step described in Fig. 1 involves compiling a high-quality data set for the well-defined endpoint, obtained through laboratory determinations or by accessing well-documented databases. It is then possible to identify chemical structural parameters associated with the biological activity, followed by, or simultaneous to, development of a preliminary QSAR model using a training set of high-quality data. The model is evaluated against predetermined criteria (defined in consultation with intended model users). A first iteration through the scheme will likely require proceeding with the selection of validation chemicals. Chemicals used to evaluate and improve the model are selected from the chemical universe of concern, i.e., to which the model will be applied. These strategically selected chemicals are subsequently tested, the results compared to model predictions, and the new information added to expand the training set and improve the model.

Strategic selection of test chemicals intended for model improvement is greatly facilitated by first determining the extent to which a training set covers the larger chemical universe (the domain of applicability) and identifying areas of the chemical universe not well represented in the initial training set. The model is most efficiently improved when strategic chemical selection and testing are focused on the most dissimilar chemicals in the domain of applicability, those chemicals most distinct from each other and most dissimilar from training set chemicals used to develop the model. The results obtained using a strategic testing approach also provide the means to assess the relevance of molecular descriptors from the original model to the newly tested chemicals, as well as allowing an assessment of the model's predictive power. Incorporation of new test data to expand the training set also provides an incremental improvement in predictive power with an increase in the number of chemicals relative to the number of descriptors. The QSAR model is advanced using the expanded training set and reevaluated using criteria established for regulatory acceptance. If criteria are met, the model will be considered for the intended use, e.g., priority setting, hazard identification, or chemical classification. The application of this iterative scheme, as presented in Fig. 1, is essential to the successful development of QSAR for use in risk assessments where such models are applied to large numbers of chemicals showing wide structural variation. The following sections address specific considerations important to the development of robust predictive models for: (1) application in chemical risk assessment (specifically, prioritization of chemicals for further testing); (2) to provide defensible coverage of a large and diverse chemical universe; and (3) based upon a well-defined endpoint for which plausible linkage to an in vivo effect of concern has been established.

RECEPTOR BIOLOGY AND CHEMICAL INTERACTIONS

The endocrine system is a multifaceted and complex set of interconnected pathways that are controlled at multiple points through numerous feedback loops. Determining key points of vulnerability along a pathway, which could be altered by xenobiotic chemical interaction, is necessary to assess, and subsequently predict, the potential for chemical disruption of endocrine systems. Significant advancements have been made in the understanding of actions and interactions of components of the nuclear hormone receptor superfamily, such as, requirements for activation factors, cross-talk with other signal transduction pathways, and critical regulatory enzymes associated with endocrine pathways. Perhaps the most studied means of chemical disruption of endocrine systems, to date, is that initiated by direct chemical–ER interaction. Even in the absence of complete pathway elucidation, enough is known to reliably assert that the binding of xenobiotics to the ER (of both mammalian and nonmammalian species) can potentially alter normal reproduction. It is also now appreciated that a multistep ligand–receptor–effector process is involved in gene activation through this transcription factor, with the induction of protein conformational change (and potentially conformational change of flexible ligands) identified as an important part of this interaction [8].

The role of ligand conformation in the binding interaction has been especially troublesome. With respect to screening chemicals, active chemicals can be missed because the minimum energy conformation did not meet steric requirements for receptor binding or because the electronic descriptors of the minimum energy conformation did not indicate significant binding affinity. Through evaluation of both thermodynamic and kinetic considerations [9] we have shown that multiple stable conformational states can be found within a few kcal/mole of the lowest energy conformer (e.g., ±20 kcal/mol). Activation energies between these states are on the order of 3–4 kcal/mol [9–10], thus establishing the potential for multiple conformers to impart sufficient energy for binding. This suggests that all energetically reasonable conformers of a chemical, not solely a single lowest energy conformer, should be considered. Further, there can be considerable variability in the value of important stereoelectronic parameters among the many conformations [11]. Therefore, failure to assess the possible influence of conformational flexibility of ligands on their potential activity could lead to unreliable screening methods.

COREPA: QSAR evolution for flexible ligands

Until recently, applications of QSAR methods to prediction of adverse ecotoxicological effects for large chemical data sets were limited to 2D depiction of chemicals for classification and model development [2,3,5–7,12]. This early reliance on 2D depiction was driven largely by computing time associated with 3D structures as well as with computing stereoelectronic parameters for a given structure. With greater emphasis on applying QSAR to predict toxic effects initiated by chemical binding to biological receptors, screening methods were required to detect not only those chemicals that fit the structural requirements for a given receptor but also whether each chemical had a plausible conformation and flexibility which permitted binding to the receptor. Recent advances in algorithms for rapidly converting 2D structures to an array of plausible 3D structures [13] now make it possible to develop 3D QSAR screening methods for libraries of thousands to tens of thousands of chemicals. Therefore, the use of 3D approaches is not only warranted, but is feasible. Various receptor-free and receptor-docking 3D QSAR approaches have been recently reviewed [14].

Departing from the minimum-energy presumption for structure gave rise to the significant 3-D QSAR challenges of quantifying the flexibility of chemicals (conformers) when attempting to predict chemical interaction with macromolecules. Software to systematically generate plausible conformations for chemicals and substitute distributions of molecular descriptors for point estimates in developing QSARs is centered around algorithms to identify **CO**mmon **RE**activity **PA**tterns (COREPA) through probabilistic classification [11,15–18]. In this present research on binding affinity to the ER, COREPA was used to solve a number of combinatorial problems. First of all, the experimental data indicate that there are multiple binding mechanisms at the receptor. Secondly, the training set of chemicals contains many flexible chemicals with numerous conformations. Thirdly, each of the projected conformations produce a complete set of stereoelectronic and other molecular descriptors to form a distribution of each descriptor for a given test chemical. COREPA is designed to manage the combinatorial explosion of parameters for each test chemical and identify patterns between molecular descriptors and binding affinity.

The complexity of modeling conformer flexibility at receptors arises from the major computational factors that include rotation around each acyclic single bond, flipping of free corners in saturated rings, and creating mirror images of saturated ring systems. The result is that a single flexible chemical may have hundreds of energetically reasonable conformers and COREPA applies a "genetic" algorithm [22,23] to select representative conformers for stereoelectronic parameter estimations [20,21]. After a 2D structure is automatically converted to an extensive list of plausible 3D conformers, parameters such as E_{HOMO} and E_{LUMO} (energies of the highest occupied, and lowest unoccupied molecular orbitals, respectively), distance between and charges on electronegative atoms, and related stereoelectronic parameters are computed for each conformer and added to the topological indices for the chemical. The COREPA data sets are then used with training databases to explore possible relationships using multi-

variate probabilistic clustering algorithms to identify the structural requirements for different classes of binding affinity and formulation of QSARs.

The COREPA can also be used in a screening mode for large inventories of heterogeneous chemicals by matching the 3D structural requirements for ER binding with each chemical in the inventory. This screening process not only checks if the structure requirements of a minimum energy structure fits the receptor but also whether each chemical is flexible enough to conform to the receptor. Of course, to screen thousands of chemicals in large inventories, the computational time needed can be substantial. While the combinatorial explosion may still be a limiting factor in screening thousands of chemicals, it is feasible to compute permanent conformation files for each chemical in the inventory just once and use these data in screening assessments for many different QSAR models. Advanced statistical approaches for simultaneous multidimensional parameter estimation and comparison to classify biological activity of chemicals are also under development. These approaches, which are currently being developed for classification of large chemical databases are described elsewhere [19].

Prioritizing within large chemical inventories: A European Union case study

The development of a preliminary COREPA model to predict the ER-binding affinity was presented by Bradbury et al. [11], with extension to more diverse structures (from 46 to 115 chemicals) and species (from hER to rat and mouse ER) [24]. The model was experimentally tested within the Endocrine Disruption Activity of Environmental Pollutants project administrated by the European Union (EU). It was used to predict the rank order of chemicals in EU inventories, which might potentially disrupt endocrine systems through the ER signaling pathway. A subset of chemicals predicted to have activity at varying affinity levels was tested in a series of receptor binding, gene activation, and in vivo assays [20].

Consistent with a predicted extremely small occurrence rate of ER binders among all industrial chemicals, the in silico screening did not identify any chemicals that would bind strongly (RBA > 10 %) within EU high production volume chemicals. Of 907 high production volume EU chemicals assessed, seven compounds were predicted to have some, but low, ER relative binding affinity (RBA ~ 1 %). Four of these were randomly selected for empirical evaluation of hER binding and gene expression and found to elicit weak estrogenic activity. An additional 14 chemicals of the 900 predicted to be inactive were randomly sampled and found to have no activity, as predicted. The model was further used to predict potential for ER binding for an additional 63 000 lower production volume chemicals from the EU IUCLID database. Ten chemicals, of 200 predicted to have RBA > 10 %, were randomly selected for testing and all of them were found to be strongly estrogenic in the two assays used previously. These results were confirmed by additional in vitro assays.

The EU exercise represents the progression through the first half of the scheme in Fig. 1. While results were promising, completion of this scheme would require an evaluation of the extent of coverage provided by the initial model training set, in relation to a chemical inventory for which predictions are sought, e.g., EU high-production-volume chemicals. Techniques are available to visualize training sets within the chemical structure space of the universe of applicability (Fig. 2), using stereoelectronic parameters associated with ER binding [11]. As mentioned previously, the next step in application of the iterative model development scheme is to strategically choose chemicals to test that would give maximize the coverage of unknown structure space. The goal is to increase the size of the more certain "interpolation space" of the model, and reduce the size of the less certain "extrapolation space". For instance, the hatched points in Fig. 2 show chemicals in a chemical domain of concern that are all outside of the training set region but "evenly" distributed across the outer edges of the extrapolation region. Algorithms are currently under development to systematically and iteratively select chemicals furthest from any tested chemical. Upon completion of empirical testing, the new data is assessed and, if necessary, added to the training set to improve its applicability. Theoretically the incremental gain in model improvement should diminish with each iteration through the entire scheme shown in Fig. 1. The num-

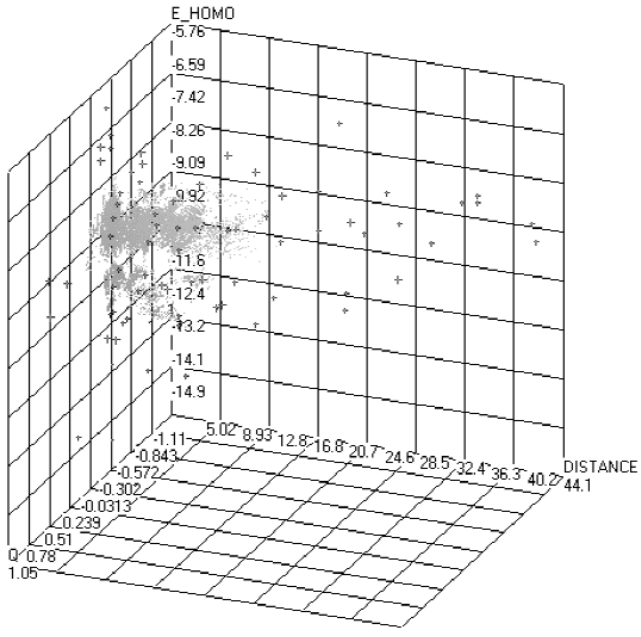

Fig. 2 Chemical selection to cover structure space: An illustration of the chemicals chosen to evenly represent chemical structure space (universe of applicability) of a large industrial chemical inventory, plotted using parameters found associated with ER-binding affinity in the COREPA-QSAR model: E_{HOMO}: energy of the highest occupied molecular orbital; DISTANCE: interatomic distance between nucleophilic sites; and Q: charge on nucleophilic sites. Filled points represent conformationally multiplied training set chemicals used to develop the initial QSAR model. Cross-hatched points represent those chemicals in the universe of applicability that are furthest (most dissimilar) from training set chemicals. Selection of these chemicals for empirical testing serves to expand the structural diversity of the training set to cover that encountered in the universe, thus increasing the applicability of the model.

ber of additional chemicals to test to improve the model will depend on many factors as previously described, and deserves further exploration.

The discussion so far has addressed how to assess and incrementally improve the domain of applicability of a QSAR model. Additional focus should also be given to improving the mechanistic transparency of predictive models. This requires verification of hypothesized chemical/biological interactions linked to the toxicological endpoint, in conjunction with more mechanistic rationale for chemical descriptors used to explain the biological/toxicological action predicted by a QSAR. Attempts to improve model transparency and assess model applicability are considered essential for broader international acceptance of a role for predictive models in human health and environmental risk assessments [1].

SUMMARY

A scheme was presented for development of predictive models for regulatory risk assessment. The prioritization of chemicals, within large inventories, for further testing to identify potential endocrine disruptors was used as an example. Successful application of the scheme requires the integration of empirical testing and further model refinements until regulatory acceptance criteria are met. Current knowledge of the biology of receptor-mediated pathways indicates that consideration of chemical conformational flexibility will improve predictive models for these biological endpoints. Considerable ad-

vances in computational techniques now allow the rapid generation and assessment of energetically reasonable conformers of chemicals. The COREPA approach was used to illustrate how recent advances have been applied to prioritization for further toxicological testing. However, additional advances are needed to achieve efficient application of these techniques to extremely large chemical data sets of regulatory concern. The successful application of the QSAR development scheme, as presented, will facilitate a close collaboration between modelers, data collectors (toxicologists), and users (risk assessors) to iteratively develop and improve models for regulatory acceptance.

DISCLAIMER

This paper has been subjected to review by the USEPA National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. Research associated with this paper was funded in part through an EPA cooperative research agreement (CR828826).

REFERENCES

1. J. Jaworska, M. Comber, C. Auer, C. J. van Leeuwen. International Workshop of the Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints. Proceedings of CEFIC-LRI and ECETOC Workshop, Setabul, Portugal, 4–6 March (2002).
2. G. D. Veith, D. J. Call, L. T. Brooke. *Can. J. Fish. Aquat. Sci*. **40**, 743–48 (1983).
3. G. D. Veith, D. L. DeFoe, M. L. Knuth. *Drug Metab. Rev*. **15**, 1295–1303 (1985).
4. C. M. Auer, J. V. Nabholz, K. P. Baetcke. *Environ. Health Perspect*. **87**, 183–187 (1990).
5. G. D. Veith and S. J. Broderius. *Environ. Health Perspect*. **87**, 207–11 (1990).
6. S. P. Bradbury. *Toxicol. Lett*. **79**, 229–237 (1995).
7. C. L. Russom, S. P. Bradbury, S. J. Broderius, D. E. Hammermeister, R. A. Drummond. *Environ. Toxicol. Chem*. **16**, 948–967 (1997).
8. J. A. Katzenellenbogen, B. W. O'Malley, B. S. Katzenellenbogen. *Mol. Endocrin*. **10**, 119–131 (1996).
9. J. M. Ivanov, O. G. Mekenyan, S. P. Bradbury, G. Schüürmann. *Quant. Struct.–Act. Relat*. **17**, 437–449 (1998).
10. T. E. Wiese and S. C. Brooks. *J. Steroid Biochem. Mol. Biol*. **50**, 61–73 (1994).
11. S. Bradbury, V. Kamenska, P. Schmieder, G. Ankley, O. Mekenyan. *Toxicol. Sci.* **58**, 253–269 (2000).
12. H. J. M. Verhaar, C. J. van Leeuwen, J. L. M. Hermens. *Chemosphere* **25**, 471–491 (1992).
13. J. M. Ivanov, S. H. Karabunarliev, O. G. Mekenyan. *J. Chem. Inf. Comput. Sci*. **34**, 234 (1994).
14. P. Schmieder, G. Ankley, O. Mekenyan, J. Walker, S. Bradbury. *Environ. Toxicol. Chem*. - QSAR Annual Review Issue **22** (8), 1844–1854 (2003).
15. O. Mekenyan, J. Ivanov, G. Veith, S. Bradbury. *Quant. Struct.–Act. Relat*. **13**, 302–307 (1994).
16. O. Mekenyan, J. Ivanov, S. Karabunarliev, S. Bradbury, G. Ankley, W. Karcher. *Environ. Sci. Technol.* **31**, 3702–3711 (1997).
17. O. Mekenyan, N. Nikolova, S. Karabunarliev, S. Bradbury, G. Ankley, B. Hansen. *Quant. Struct.–Act. Relat.* **18**, 139–153 (1999).
18. S. P. Bradbury, O. G. Mekenyan, G. T. Ankley. *Environ. Toxicol. Chem*. **17**, 15–25 (1998).
19. O. Mekenyan, N. Nikolova, P. Schmieder. *J. Mol. Struct. (TheoChem)* **622**, 147–165 (2003).
20. T. Hurst. *J. Chem. Inf. Comput. Sci*. **34**, 90 (1994).
21. O. Mekenyan, V. Kamenska, R. Serafimova, L. Poellinger, A. Brower, J. Walker. *SAR QSAR Environ. Res.* **13**, 579–595 (2002).

22. D. E. Goldeberg. *Genetic Algorithms in Search Optimization and Machine Learning*, Addison-Wesley, Reading, MA (1989).
23. A. W. R. Payne and R. C. Glent. *J. Mol. Graph*. **11**, 74 (1993).
24. O. Mekenyan, V. Kamenska, P. Schmieder, G. Ankley, S. Bradbury. *Toxicol. Sci*. **58**, 270–281 (2000).