Pure Appl. Chem., Vol. 76, No. 10, pp. 1927–1931, 2004.

Workshop 1.6
QSAR and related procedures for prescreening and predicting potential endocrine active compounds*

Toshio Fujita‡

Department of Agricultural Chemistry, Kyoto University, Kyoto, Japan

Abstract: This workshop has been organized to cover various quantitative structure–activity relationship (QSAR) and computer-aided procedures currently carried out for the prediction of the endocrine activity of unknown compounds. Each of the procedures has own scope as well as limitations. It seems inappropriate to consider that a single quantitative prediction model derived from each of these procedures could solve the entire issue. Because the model-building is highly dependent on the data/knowledge about endocrine activity of a large number of existing compounds accumulated to date and the data/knowledge are growing constantly, the model has a destiny to be amended "forever" as the structure–activity data of newly synthesized compounds are accumulated. The skepticism about in silico and QSAR procedures put forward in the past is likely to be cleared at least to some extent if not entirely by participating in this workshop.

INTRODUCTION

The Scientific Committee on Problems of the Environment (SCOPE)/IUPAC international project on endocrine active substances began in the year 2000 to comprehensively synchronize activities organized separately by the World Health Organization (WHO), the Organization for Economic Cooperation and Development (OECD), U.S. Environmental Protection Agency (EPA), and other institutions in preceding years. This project deals with a vast number of natural and synthetic chemicals. Because the financial burden for assessing the potential endocrine activity of these compounds is great and a certain time constraint is mandatory, computer-assisted (in silico) procedures were hoped to be effective in prescreening and prioritizing large numbers of compounds and predicting the endocrine activity and potency of prescreened compounds rapidly and inexpensively. Thus, methodologies relating to QSAR (**q**uantitative **s**tructure–**a**ctivity **r**elationship) analyses and receptor-ligand docking simulations are expected to show promise on this issue. Before the start of the SCOPE/IUPAC project, the classical Hansch-type QSAR [1] and 3D QSAR such as CoMFA (**co**mparative **m**olecular **f**ield **a**nalysis) [2,3] as well as receptor docking [4] models have been applied to analyze/elucidate relationships between structure and activity of a known series of a certain number of estrogenic and endocrine active compounds. After the commencement of the SCOPE project, applications of various (3D) QSAR procedures and docking models have been investigated in various institutions more cooperatively and actively than before.

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

‡Mailing address: 305 Heights Kyogosho, Nakagyoku, Kyoto 604-8057, Japan; E-mail: ped01545@nifty.com

Before the SCOPE project was initiated, however, evaluations for in silico procedures were not necessarily consistent. Whereas the CoMFA approach was expected to have a prime ability to elucidate relationships between structure and activity of a variety of environmental estrogens including nonsteroidal compounds [3], the complexity of the problem was also well recognized [5]. This was especially so, because at least two subtypes of the estrogen receptor, ERα and ERβ, were recognized to exist in 1996–1997 and the structure-binding relationship of nonsteroidal estrogenic compounds with the ERβ-receptor could necessarily require revisions of the preceding points of view [5]. Furthermore, estratriene, in which two hydroxyl groups of estradiol are removed, was found to be estrogenic in vitro and in vivo [6,7]. Among a set of structurally "very similar" hydroxylated polychlorobiphenyls, several analogs exhibit antiestrogenic activity, while others inhibit progesterone receptor binding in the mouse uterus [8]. These complex structure–activity relationships suggested that while "sophisticated" computer analyses might alert one to *obvious* structural analogs, they may not be able to venture into the area of real need such as the prediction of estrogenic activity for structurally remote compounds [6]. Thus, in 1997–1998, before the SCOPE project started, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), which was established by the U.S. EPA, evaluated in silico (SAR/QSAR and docking simulation) procedures [9]. In spite of a certain predictive strength as (virtual) "screening" procedures to identify endocrine active structures, no existing QSAR model was unambiguous and no receptor-binding model for endocrine active compounds was perfect [9]. Recognizing that not every molecular mechanism of receptor binding for various series of endocrine disruptors was established, EDSTAC also pointed out the difficulty of deriving a "single" generalized way of in silico modeling to predict endocrine activity for sets of possible compounds [9]. To overcome this rather difficult situation, EDSTAC requested that in silico SAR/QSAR procedures should be developed using the most complete, accurate, and rigorously corroborated datasets and used only within the range of conditions for which they are validated. At the same time, EDSTAC suggested making the applicable chemical domain as diverse as possible [9]. According to the aims of the SCOPE project and also in response to EDSTAC's requests and suggestions, further elaborative investigations have been performed in subsequent years. In this workshop, five lectures and one poster were organized from among such recent investigations.

WORKSHOP PRESENTATIONS

Among six topics, the first, which is based on the classical Hansch-type QSAR paradigm, was presented by C. Selassie of Pomona College, CA, USA. Its background is the reaction-mechanistic physical organic chemistry in terms of the extended Hammett (linear free-energy) relationship, and the QSAR results have been validated through the CQSAR (comprehensive QSAR) database [10]. The second was by W. Tong and his group from the U.S. Food and Drug Administration (U.S. FDA), the speaker being R. Perkins. This topic deals with an approach consisting of a sequential "four-phase" scheme for priority setting. Each phase uses either one or a certain combination of the different computational and QSAR procedures. In phase 3, the receptor binding affinity is analyzed/predicted by using the CoMFA. The third presentation was by O. Mekenyan from University "Professor As. Zlatarov", Bourgas, Bulgaria, jointly contributed by a U.S. EPA group. Here, a new 3D QSAR technique, the COREPA (**co**mmon **re**activity **pa**ttern) procedure, is applied to classify, rank, and prioritize the endocrine active structures. It does not require the 3D alignment of the minimum-energy conformation for ligand sets, but considers the contribution of every conceivable conformation of ligands in terms of various "physicochemical" descriptors and compares similarities of distribution patterns of descriptors over various conformations between active and inactive ligands. Along with these "QSAR" lectures, two Japanese institutions presented their respective receptor-ligand docking simulation approaches from different force-field concepts. One of them, presented by N. Tomioka of Institute of Medicinal Molecular Design, Tokyo, uses empirical force-field models for electrostatic, steric, and solvation/desolvation interactions of ligands with the receptor. This presentation, titled "3D-SAR Analysis of Endocrine Disruptors based

on Target Receptor Structure", is unfortunately not available in this volume. The other, which was presented via poster by T. Nakano of National Institute of Health Sciences, Tokyo, uses an ab initio MO procedure for the recepter-ligand complex, in which the receptor-protein is dealt with after "fragmentation" according to certain rules. Finally, this workshop incorporated an organic medicinal chemist, J. Katzenellenbogen from the University of Illinois, who is an expert in the designed syntheses of estrogenic compounds and the molecular mechanism of receptor-ligand binding. The topic of the experimental scientist complemented the "computational" presentations and was able to upgrade and enforce the receptor binding models/concepts used in in silico procedures.

OVERALL SUMMARY OF THE WORKSHOP

In fact, the task of "Prescreening and Predicting of Potentially Endocrine Active Compounds" includes incredibly complicated factors. The compounds, the endocrine activity of which should be assessed, include existing chemicals with unknown activity as well as no existing chemicals to be synthesized in future. Their number is enormous and not even estimable, and their structure types are naturally very diverse. In addition, there are still many problems to be clarified in the entire pathway even including the exact modes of receptor binding of active compounds of different structure types as indicated by Katzenellenbogen. In this situation, a theoretically "perfect" approach is not likely to exist. One possible approach to this type of issue would be through empirical procedures using appropriate working models and approximations aided by in silico prediction technology. The skill and data to build the model is highly dependent upon accumulating knowledge/experience. Because knowledge is being accumulated constantly, the model is inevitably inconclusive. It follows then that models have to be updated perhaps with modifications/revisions on the basis of trial and error. As the number and structure type of compounds, of which activity has been measured, are accumulated, the chemical structure space, for which the model has to be updated, expands accordingly. In other words, the in silico prediction procedure has a destiny to never stop and grows to be modified "forever" as the structure–activity data are accumulated.

Each of the presentations proposed different procedures for the same issue, the prediction of the endocrine activity of compounds. Each of them seemed to have its own scope as well as limitations and restrictions. Depending upon the extent of the chemical structure space covered by compounds to be analyzed and predicted, however, each model uses one of two types of the model-building paradigm. For sets of compounds, the structure space of which is limited, e.g., for congeneric sets of compounds, the model could be built on the basis of the principles of physical organic (bio)chemistry to a greater extent than statistical approximations. The classical QSAR approach to the analyses and prediction of endocrine activity/potency for any series of compounds belongs to this category. In these series, skeletal structure is often identical, i.e., the structure space is necessarily not broad. The prediction model can be validated laterally by comparisons with QSAR formulations of other related series of bioactive compounds in terms of the "reaction" mechanism as indicated by Selassie. Approaches using docking simulation models also are in this category as they are based on physical chemistry/quantum chemistry and do not deal with a vast number of compounds statistically. In spite of the difference in the force-field concepts, the calculated binding data agree well with the experimental values although the number of such examples is not very large. In these approaches, the accuracy of the procedure is highly dependent on the accuracy of the molecular model of the receptor protein. In the work of Tomioka et al., a large number of chemicals are first (pre)screened in silico using an automatic docking program named ADAM without detailed optimizations for the geometry of the binding process [11–13]. In the final structure-optimization of receptor-ligand complex, the structure space of ligands is not broad, being limited by the availability of the 3D structure of receptor-ligand complex as exemplified by Nakano.

When a large number of structurally diverse compounds are dealt with, on the other hand, in silico prediction procedures could necessarily be dependent on statistical/probabilistic concepts to a greater extent than physical chemistry principles. In the sequential "four-phase" scheme procedure of

Tong et al. from the U.S. FDA, computer programs for more elaborate as well as more "detailed" quantitative procedures are used as the phase is going up to the higher order. This is accompanied with incrementally reducing the size of the dataset while increasing precision of the prediction during each phase. Thus, as the structure space for the set of compounds becomes reduced, increasing in the extent of "chemistry principle" seems to occur at a cost of decreasing in the extent of "statistical/probabilistic paradigm" in the background of the model building. The COREPA procedure of Mekenyan and the U.S. EPA group is a sort of 3D QSAR, but it differs from such a conventional 3D QSAR as CoMFA in that it does not require the geometrical alignment of molecules, but uses a probabilistic scheme to identify physicochemical and/or quantum-chemical descriptors that best discriminate classes of biologically dissimilar chemicals. In this category, the prediction capability of the models is validated with use of the jack-knife procedure internally or by the experimentally obtained biological data externally. The prediction error is allowed to a certain extent so that the number of the false negatives should be minimized. It should be mentioned here that the CoMFA, which is used in the phase 3 of the four-phase scheme of Tong et al., is not necessarily regarded as an extension of the Hansch-type classic QSAR based on mechanistic physical-organic chemistry as per Selassie. Using principal components of a vast number of molecular (steric and electronic) descriptors, it is rather considered as a variation of the statistical procedure.

It seems impossible and also not appropriate to attempt to solve the issue for "Prescreening and Predicting of Potentially Endocrine Active Compounds" by a single in silico model. Depending upon the nature and size of the set of compounds to be analyzed, different in silico models could be utilized. In this type of very "fuzzy" problem, the rational approach is not necessarily equivalent to an approach that is scientific "(bio)chemically". It should be inevitable to use statistic models to deal with a great number and variety of chemicals at the expense of the chemistry. The personal opinion of the present organizer is that the prescreening procedure to squeeze out the number of "apparently" inactive compounds with combination of various computational (not necessarily QSAR) methods seems to be of prime importance as Tong and his group indicated. After identification of a prescreened set of compounds, and hopefully after classification into structurally congeneric sets with a common skeleton and functional group, more "scientifically precise" quantitative procedures such as classic QSAR and docking simulation could be applied to predict the binding activity and potency "scientifically" and accurately.

Finally, it could be stated that the workshop was successful. The presentations are believed to show that the recent progress of this field of study has been achieved rather quickly. The skepticism about in silico and (Q)SAR procedures put forward previously is likely to be cleared at least to some extent if not entirely. In this workshop, only the receptor binding affinity of estrogenic compounds is discussed. The binding is compulsory for the manifestation of the activity. However, the potency of the exogenous endocrine active compounds observed in vivo is governed by a number of (sub)molecular properties relating to events occurring in the transport process to the receptor sites within the animal body. In addition, the activity of estrogen agonists is controlled not only by the binding affinity to the receptor, but also by the intrinsic potency as mentioned by Katzenellenbogen. The effort to establish the models for various facets of the activity should not be overlooked.

REFERENCES

1. H. Gao, J. A. Katzenellenbogen, R. Garg, C. Hansch. *Chem. Rev.* **99**, 723–744 (1999).
2. T. G. Gantchev, H. Ali, J. E. van Lier. *J. Med. Chem.* **37**, 4164–4176 (1994).
3. C. L. Waller, T. I. Oprea, K. Chae, H.-K. Park, K. S. Korach, S. C. Laws, T. E. Wiese, W. R. Kelce, L. E. Gray, Jr. *Chem. Res. Toxicol.* **9**, 1240–1248 (1996).
4. A. K. Shiau, D. Barstad, P. M. Loria, L. Cheng, P. J. Kushner, D. A. Agard, G. L. Greene. *Cell* **95**, 927–937 (1998).
5. J. A. Dodge. *Pure Appl. Chem.* **70**, 1725–1733 (1998).

6. J. Ashby. *Pure Appl. Chem.* **70**, 1735–1745 (1998).
7. R. Elsby, J. L. Maggs, J. Ashby, D. Paton, J. P. Sampter, B. K. Park. *J. Pharm. Exp. Ther.* **296**, 329–337 (2001).
8. K. Connor, K. Ramamoorthy, M. Moore, M. Mustain, I. Chen, S. Safe, T. Zacharewski, B. Gillesby, A. Joyeux, P. Balaguer. *Toxicol. Appl. Pharm*. **145**, 111–123 (1997).
9. <http://www.epa.gov/oscpmont/oscpendo/history/chap4v14.pdf>, EDSTAC Final Report, Chap. 4, pp. 4–27, August 1998.
10. C. Hansch, D. Hoekman, A. Leo, D. Weininger, C. D. Selassie. *Chem. Rev*. **102**, 783–812 (2002).
11. M. Y. Mizutani, N. Tomioka, A. Itai. *J. Mol. Biol.* **243**, 310–326 (1994).
12. Y. Iwata, M. Arisawa, R. Hamada, Y. Kita, M. Y. Mizutani, N. Tomioka, A. Itai, S. Miyamoto. *J. Med. Chem.* **44**, 1718–1728 (2001).
13. Y. Iwata, S. Naito, A. Itai, S. Miyamoto. *Drug Design Discov.* **17**, 349–359 (2001).