QUANTITATIVE STRUCTURE- ACTIVITY RELATIONSHIPS (QSARs) IN RISK ASSESSMENT OF CHEMICALS

Mats Tysklind and Patrik Andersson

Department of Chemistry, Environmental Chemistry, Umeå University, SE-901 87 Umeå, Sweden

Introduction

The potential use of quantitative structure-activity relationship (QSAR) models in risk assessment processes (RA) has been stressed in recent years. QSARs provide a non-testing method that could be an aid in priority settings and classifications, to fill existing data gaps, reduce the number of animal tests and save costs. QSAR modelling is based on the assumption that the chemical properties of a molecule are correlated to its biological activity. Chemicals with similar physicochemical properties will thus have common biological features and act via comparable mechanisms of action. The QSAR model includes calculated or experimental values making up a matrix of descriptors related to the physical and chemical properties of the chemicals, and a response matrix, of biological activity or a physical or chemical property. The QSAR model is the mathematical expression relating the two matrices. A QSAR model produces quantitative measures of the studied response whereas a qualitative relationship between the chemical information and the response is named structure-activity relationship (SAR). A SAR could be regarded as a structural alert approach where certain fragments or substructures are correlated to the studied response.

QSARs have been used for many years in the field of environmental chemistry and toxicology, although the acceptance and number of applications are much greater in the pharmaceutical science. The regulatory use of QSARs is today relatively low with the exception of a few countries (Cronin *et al.* 2003a,b). In Europe, the Danish Environmental Protection Agency (EPA) has created a QSAR database of more than 166 000 substances including various environmental and human health endpoints. Canada is extensively using QSARs for categorising the 23 000 chemicals on their Domestic Substance List. In the United States, the US EPA, the US Food and Drug Administration (FDA), and other governmental organizations apply models for predicting various properties. In the EU, the White Paper on the Strategy for a future Chemicals Policy published in 2001 indicated the potential use of QSARs as a means to improve the risk assessment of the large number of chemicals in use (European Commission, 2001). In the proposal for a new chemicals legislation in Europe entitled Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), which followed the White Paper, both (Q)SARs and read across/grouping of chemicals are described as methods to develop necessary information (European Commission, 2003). In order to internationally harmonize the development and validation of QSARs, the Organisation for Economic Co-operation and Development (OECD) initiated in 2003 a work programme on QSARs (OECD, 2004).

The development of QSAR models

The process of developing a QSAR model is initiated by the definition of the studied chemical domain and chemical descriptors calculated, should capture the intrinsic structural characteristics of the compounds. Subsequently, responses are measured or captured from databases or the literature and a QSAR model is calculated. The final step is the process of validation and then the model could be used for predictions within its applicability domain. Classical QSARs models were most often built on homologues series of chemicals, where for example the chain length of hydrocarbons was changed. A large number of models exist constructed for well defined chemical domains, e.g. models for halogenated aliphatic hydrocarbons (Eriksson *et al.* 1996), di- and tri-hydroxybenzenes (Aptula *et al.* 2005), polychlorinated biphenyls (Andersson *et al.* 2000), and brominated diphenylethers (Harju *et al.* 2002).

Mathematical methods are needed to bridge the information describing the compounds chemical and structural variation with their measured response. Common regression based methods used for continuous response data are multiple linear regression (MLR), principal component regression (PCR), partial least squares regression to latent structures (PLS), and various neural network methods. These methods are generally used to tailor models for a certain set of chemicals and endpoint. Expert systems, in contrast, incorporate a multitude of models or expert knowledge in combination with defined structure-based rules to reach a prediction for a wider range of chemicals. More details on the multivariate projection methods can be found in e.g. Eriksson *et al.* 2001 and on neural networks in e.g. Zupan and Gasteiger, 1999.

The definition of class of chemicals considered is the crucial first step in the development of a QSAR model since the chemical domain included in the development of a model defines its limits and applicability. If not response data exists in the domain of interest it is important to perform testing representative for the whole domain. It is important to realize that outside these boundaries the predictions will be extrapolations as compared to the interpolations of chemicals defined to belong to the assessed chemical domain. For more details on the development of QSARs we refer to the recent reviews by Cronin *et al.* 2003a,b, Schultz *et al.* 2003a,b, Walker *et al.* 2003, Eriksson *et al.* 2003, and Cronin and Livingstone, 2004.

Chemical descriptors

Today thousands of descriptors are available and most of them can be calculated *in silico*, i.e. directly in the computer without experiments. This is an advantage as predictions of compounds can be made before they are available on the market or even synthesized. In general, chemical descriptors are considered to describe the compounds steric, hydrophobic, and electronic nature. It is commonly accepted that these properties are related to the compounds biological activity. Chemical descriptors range from simple counts of atoms and functional groups to electronic characteristics as calculated using quantum chemistry. The most commonly used chemical descriptor in QSAR modelling is however the partition coefficient between octanol and water (Kow). Recent reviews of chemical descriptors include Todeschini and Consonni, 2000 and Karelson, 2000.

Validation and applicability domain

A QSAR model should be validated by internal and external means. The most demanding validation procedure is to use an external set of compounds (validation set or test set) that were not use as the model was calculated. These compounds should be structurally representative of the studied chemical domain. However, a true external set of compounds may not be available or the resources limited to measure the activity of additional compounds. One crucial measure of a QSAR model, which is included in the OECD principles, but which definitions has not reached general consensus is the applicability domain. The establishment of this domain sets the use of a model and predictions of compounds defined as within that domain can be interpreted as interpolations. Accordingly, the response of compounds outside the domain are extrapolations and thus less valid. The applicability domain is rarely defined and methods differ depending on the type of model. Recently, the applicability domain was defined as "the response and chemical structure space in which the model makes predictions with a given reliability" (Netzeva *et al.* 2005). The domain can be defined by using ranges of descriptors, structural rules, or by statistical means to define a chemical variation. One starting point, independent on applied QSAR method, to define the applicability domain is the chemical variation as covered by the training set. The strategy behind the selection of the training set is hence of crucial importance for the validity of the model.

Conclusions

Non-testing methods are urgently needed in order reach better understanding of the fate and effects of the huge number of chemicals that are in commercial use. QSARs provide one attractive alternative that have shown great potential for filling existing data gaps for several properties. However, it is of crucial importance that the users of QSARs are well aware of their limitations and intended use of the model. The data set used to train the model may be erroneous, the applied descriptors have limitations, the response that is modelled has an error, and the regression method may yield an overfitted model. The applicability domain of the model needs also to be understood so that responses of only members of the domain are predicted. One major and basic problem that hampers the development of new QSAR models is the lack of response data with high quality. A systematic QSAR strategy strives to develop such data sets for structurally representative chemicals from defined chemical domains (Tysklind and Andersson 1998). If correctly used QSARs have a great potential in future RA processes and generally accepted validation procedures could be one way to trigger QSAR models future employment.

Acknowledgments

Financial support from the Swedish Foundation for Strategic Environmental Research (Mistra) is gratefully acknowledged.

References

Andersson PL. (2000). Physico-chemical characteristics and quantitative structure-activity relationships of PCBs. Thesis. Umeå University. Umeå. Sweden. (Available at http://www.diva-portal.org/umu/theses/).

Aptula AO, Roberts DW, Cronin MTD, Schultz TW. (2005). Chemistry-toxicity relationships for the effects of di-and trihydroxybenzenes to Tetrahymena pyriformis. *Chemical Research in Toxicology* 18:844-854.

Cronin MTD, Walker JD, Jaworska JS, Comber MHI, Watts CD, Worth AP. (2003a). Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. *Environmental Health Perspectives* 111:1376-1390.

Cronin MTD, Jaworska JS, Walker JD, Comber MHI, Watts CD, Worth AP. (2003b). Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environmental Health Perspectives* 111:1391-1401.

Cronin MTD, D.J. Livingstone. (Eds). (2004). Predicting Chemical Toxicity and Fate. CRC Press, London, UK.

Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. (2001). Multi and Megavariate Data Analysis—Principles and Applications. Umetrics AB, Umeå, Sweden.

Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives* 111:1361-1375.

European Commission. (1996). Technical guidance document in support of commission directive 93/67/EEC on risk assessment for new notified substances and commission regulation (EC) No 1488/94 on risk assessment for existing substances. Office for Official Publications of the European Communities. Luxembourg. (Available at http://ecb.jrc.it).

European Commission. (2001). White Paper Strategy for a future Chemicals Policy. 27.2.2001 COM(2001) 88 final. Brussels. Belgium. (Available at http://europa.eu.int).

European Commission. (2003). Proposal for a regulation of the European parliament and of the council concerning the Registration, Authorisation and Restriction of Chemicals (REACH). 29.10.2003. COM(2003) 644 final. Brussels. Belgium. (Available at http://europa.eu.int).

Harju M, Andersson PL, Haglund P, Tysklind M. (2002). Multivariate physicochemical characterisation and quantitative structure-property relationship modelling of polybrominated diphenyl ethers. *Chemosphere* 47:375-384.

Karelson M. (2000). Molecular Descriptors in QSPR/QSAR. John Wiley and Sons, London, UK.

Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, van de Sandt JJM, Tong WD, Veith G, Yang CH. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships - The report and recommendations of ECVAM Workshop 52. *Atla-Alternatives to Laboratory Animals* 33:155-173.

Organisation for Economic Co-operation and Development (OECD). (2004). The report from the expert group on (Quantitative) structure-activity relationships [(Q)SARs] on the principles for the validation of (Q)SARs. OECD Environment Health and Safety Publications Series on Testing and Assessment No. 49. ENV/JM/MONO(2004)24. Paris. France.

Schultz TW, Cronin MTD, Walker JD, Aptula AO. (2003a). Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular Structure-Theochem* 622:1-22.

Schultz TW, Cronin MTD, Netzeva TI. (2003b). The present status of QSAR in toxicology. *Journal of Molecular Structure-Theochem* 622:23-38.

Todeschini R, Consonni V. (2000). Handbook of Molecular Descriptors. Wiley-VCH, Weinheim, Germany.

Tysklind M, Andersson PL. (1998). The use of quantitative structure-activity relationships in fate and risk assessment of persistent organic pollutants (POPs). *Organohalogen Compounds*. 36:455-458.

Walker JD, Jaworska J, Comber MHI, Schultz TW, Dearden JC. (2003). Guidelines for developing and using quantitative structure-activity relationships. *Environmental Toxicology and Chemistry* 22:1653-1665.

Zupan J, Gasteiger J. (1999). Neural Networks in Chemistry and Drug Design. Wiley, Weinheim, Germany.