JRC Scientific and Technical Reports

Review of QSAR Models and Software Tools for predicting Developmental and Reproductive Toxicity

Elena Lo Piparo and Andrew Worth

EUR 24522 EN - 2010

JRC
EUROPEAN COMMISSION

ihcp
Institute for Health and Consumer Protection

The mission of the JRC-IHCP is to protect the interests and health of the consumer in the framework of EU legislation on chemicals, food, and consumer products by providing scientific and technical support including risk-benefit assessment and analysis of traceability.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information
Address: Via E. Fermi 2749, 21027 Ispra (VA), Italy
E-mail: andrew.worth@ec.europa.eu
Tel.: +39 0332 789566
Fax: +39 0332 786717

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 59820

EUR 24522 EN
ISBN 978-92-79-16702-7
ISSN 1018-5593
doi:10.2788/9628

Luxembourg: Publications Office of the European Union

Printed in Italy

ABSTRACT

This report provides a state-of-the-art review of available computational models for developmental and reproductive toxicity, including Quantitative Structure-Activity Relationship (QSARs) and related estimation methods such as decision tree approaches and expert systems. At present, there are relatively few models for developmental and reproductive toxicity endpoints, and those available have limited applicability domains. This situation is partly due to the biological complexity of the endpoint, which covers many incompletely understood mechanisms of action, and partly due to the paucity and heterogeneity of high quality data suitable for model development. In contrast, there is an extensive and growing range of software and literature models for predicting endocrine-related activities, in particular models for oestrogen and androgen activity. There is a considerable need to further develop and characterise *in silico* models for developmental and reproductive toxicity, and to explore their applicability in a regulatory setting.

LIST OF ABBREVIATIONS

AD	Applicability Domain
AR	Androgen Receptor
CASE	Computer Automated Structure Evaluation
CoMFA	Comparative Molecular Field Analysis
EC	European Commission
ECHA	European Chemicals Agency
EAS	Endocrine Active Substances
ED	Endocrine Disruptor
EINECS	European Inventory of Existing Chemical Substances
ER	Oestrogen Receptor
GRIND	GRid INdependent Descriptors
HTS	High-throughput screening
IC50	Half Maximal Inhibitory Concentration
ILSI	International Life Sciences Institute
JRC	Joint Research Centre
LBD	Ligand Binding Domain
LD50	Median Lethal Dose
NMR	Nuclear Magnetic Resonance
NR	Nuclear receptor
MCASE	Multi Computer Automated Structure Evaluation
MC4PC	Multi CASE for Personal Computer
OECD	Organisation for Economic Cooperation and Development
OPS	Optimal Predictive Space
PASS	Prediction of Activity Spectra for Substances
PDB	Protein Data Bank
PPAR	Peroxisome Proliferator Activated Receptor
PR	Progesterone Receptor
(Q)SAR	(Quantitative) Structure-Activity Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
SDF	Structure Data file
SERM	Selective Estrogen Receptor Modulator
TEST	Toxicity Estimation Software Tool
TIMES	Tissue MEtabolism Simulator
US EPA	United States Environmental Protection Agency
WoE	Weight of Evidence

CONTENTS

1. Introduction **1**
2. Software **1**
3. Databases **7**
4. Literature models for reprotoxicity endpoints **10**
5. Endocrine-related effects **12**
5.1 Endocrine Active Substances and potential Endocrine Disruptors 12
5.2 *In silico* modelling of endocrine-related effects 12
6. Regulatory use of *in silico* predictions **18**
7. Conclusions **19**
8. References **20**

1. Introduction

Reproductive and developmental toxicity studies (referred to collectively here as reprotoxicity) are used to identify the adverse effects a chemical may have on sexual function and fertility in adult males and females, developmental toxicity in the offspring, as well as effects on, or mediated via, lactation. Thus, reproductive toxicity refers to a range of endpoints relating to the impairment of male and female reproductive capacity (fertility) and the induction of non-heritable harmful effects on the progeny (developmental toxicity). The variety of observable effects are brought about by a plethora of mechanisms of action, many of which are unknown or only partially understood at the molecular and cellular level. Along with carcinogenicity studies, reprotoxicity studies are among the most costly and time-consuming experimental procedures. Furthermore, reprotoxicity testing requires the highest number of test animals. For all these reasons, the development of alternative (non-animal) methods for reprotoxicity assessment is a high political priority.

In this report, a state-of-the-art review of available models for developmental and reproductive toxicity is provided. It consists of three sections: the first one summarizes the software (both freely available and commercial) that can be used to predict reproductive and developmental toxicity; the second one the database that can be used to build the models and the last one literature available models.

2. Software

A number of computer programs generate structure-based predictions of reprotoxicity endpoints, as summarised in Table 1, and reviewed briefly below. Some of these models are classification models, making categorical predictions, whereas others make quantitative predictions.

ADMET Predictor: This commercial program is designed to estimate certain ADMET (Absorption, Distribution, Metabolism, Elimination, and Toxicity) properties of a drug-like chemical from its molecular structure. It includes a qualitative assessment of oestrogen receptor toxicity in rats (TOX_ER_filter), together with a quantitative measure of oestrogen receptor toxicity in rats (TOX_ER (IC50(estrogen)) that is applied only for compounds classified as 'Toxic' by the previous model.

ACD/Tox Suite: The ACD/Tox Suite (formerly called ToxBoxes), provided by ACD/Labs and Pharma Algorithms, provides predictions of various toxicity endpoints including ER binding affinity (http://www.acdlabs.com/products/admet/tox/). The predictions are associated with confidence intervals and probabilities, thereby providing a numerical expression of prediction reliability. The software incorporates the ability to identify and visualize specific structural toxicophores, giving insight as to which parts of the molecule are responsible for the toxic effect. It also identifies analogues from its training set, which can also increase confidence in the prediction. The algorithms and datasets not disclosed. A web version of the software is freely accessible at http://www.pharma-algorithms.com/webboxes/

CAESAR: A series of statistically-based models, developed within EU-funded CAESAR project (http://www.caesar-project.eu), have been implemented into open-source software and made available for online use via the web. The freely accessible CAESAR model for developmental toxicity was built using 292 compounds. Two models were developed, one using WEKA (Waikato Environment for Knowledge Analysis) and Random Forest, and the other using the Adaptive Fuzzy Partition (AFP) classification model.

Derek: Derek for Windows (DfW) is a rule-based system developed by Lhasa Ltd, a non-profit company and educational charity (https://www.lhasalimited.org/). All the rules are based either on hypotheses relating to mechanisms of action of a chemical class or on observed empirical relationships. This system includes structural alerts for three specific endpoints: developmental toxicity (3 alerts), teratogenicity (5 alerts), testicular toxicity (1 alert) and oestrogenicity (4 alerts) associated with nine levels of confidence: certain, probable, plausible, equivocal, doubted, improbably,

impossible, open, contradicted. An alert consists of a toxicophore (a substructure known or thought to be responsible for the toxicity) and is associated with literature references, comments and examples. A key feature of DfW is the transparent reporting of the reasoning underlying each prediction.

Pearl *et al*. (2001) conducted a small validation study with 34 chemicals, and reported 100% specificity (equivalent to 0% false positives) and 72% sensitivity (28% false negatives). However, due to the small size of the dataset, it is difficult to draw general conclusions from these results.

The Dutch National Institute for Public Health and the Environment (RIVM) published a study (Hulzebos and Posthumus, 2003) where Derek predictions for the reproductive toxicity effects of 60 substances were compared with experimental data. The authors concluded that reprotoxicity is poorly predicted by this software. A further study by the RIVM (Maslankiewicz, 2005) reached the same conclusion. The study examined the ability to correctly predict the developmental toxicities of 108 industrial chemicals by using Derek and by applying the chemical categories developed by the US EPA to support the implementation of the Toxic Substances Control Act (TSCA; http://www.epa.gov/compliance/civil/tsca/tscaenfstatreq.html). The conclusion was based on the observation that Derek only recognised 10% of substances which may cause impaired fertility, and only 19% of chemicals which may harm the foetus (on the basis of the harmonised EU classifications of chemicals in Annex I of the Dangerous Substances Directive). However, this conclusion is unfair to the extent that it ignores the fact that Derek is only designed to identify positives and does not make negative predictions – the absence of a prediction simply means there are no rules identifying chemical features of toxicological concern, and does not necessarily reflect the absence of toxicity. For the same reason, use of the ten TSCA categories also revealed low sensitivities (percentage of correctly predicted positive substances) – 19% and 18% for fertility and teratogenicity effects, respectively. The authors also noted that Derek and TSCA had one structural alert in common for the studied chemicals and thus the applicability domain is different for the two predictive approaches. For this reason, it would be worthwhile to build on the RIVM study by investigating the combined use of prediction based on the use of TSCA categories and Derek.

Endocrine Disruptor Knowledge Base (EDKB): This online database, developed and made publicly available by the US FDA's National Center for Toxicological Research (NCTR), contains computer-based predictive models to predict the binding affinity of compounds to the oestrogen and androgen nuclear receptor proteins.

Leadscope: The Leadscope software has a module containing QSAR models for predicting the developmental toxicity of the rodent foetus, including dysmorphogenesis (structural and visceral birth defects), developmental toxicity (foetal growth retardation and weight decrease), and foetal survival (foetal death, post-implantation loss, and preimplantation loss). The Leadscope QSAR models for reproductive toxicity include rodent male reproductive, rodent male sperm, female reproductive.

Molcode Toolbox: This is a commercial tool developed and marketed by Molcode Ltd (http://molcode.com/). It has a range of modules for predicting toxicological endpoints and ADME properties between them endocrine activity. The models are well documented and the underlying experimental data is made available with references and structure files (MDL molfile).

MultiCASE: This software, developed by MultiCASE Inc. (http://multicase.com/), implements the so-called CASE (Computer Automated Structure Evaluation) approach, and is referred to in different ways (MCASE or MC4PC), depending on the software version and computer platform and its successor. The program automatically generates predictive models from datasets provided by the user. It is based on a fragment-based technology sometimes referred to as the CASE approach. The program performs a hierarchical statistical analysis of a database to discover substructures that appear mostly in active molecules thus being with high probability responsible for the observed activity. Initially, it identifies the statistically most significant substructure within the training set. This fragment, labelled the top biophore, is considered responsible for the activity of the largest possible number of active molecules. The active molecules containing this biophore are then removed from the database, and the remaining ones are submitted to a new analysis for identification of the next biophore. The procedure

is repeated until either the activity of all the molecules in the training set has been accounted for or no additional statistically significant substructure can be found. Then for each set of molecules containing a specific biophore, the program identifies additional parameters called modulators, which can be used to derive QSAR within the reduced set of congeneric molecules. The modulators consist of certain substructures or physicochemical parameters that significantly enhance or diminish the activity attributable to the biophore. QSARs are then derived by incorporating the biophores and the modulators into the model. For the endpoints prediction, the software uses it own toxicity scale, from 0 to 100 CASE units, to cover the range from inactive, marginally active and active. In many cases, it is difficult to relate these CASE units to traditional measures of toxicity.

The US FDA have applied MultiCASE methodology (the MC4PC software) to the below-mentioned FDA database to develop a battery of QSAR models for reproductive and developmental toxicity hazard identification (Matthews *et al*., 2007a, 2007b). Their models were designed to predict seven general reprotoxicity classes: male and female reproductive toxicity, foetal dysmorphogenesis, functional toxicity, mortality, growth, and newborn behavioural toxicity. These are different to the models included in the marketed version of the software. The QSARs were derived from weighted reproductive toxicity findings, in order to incorporate a WoE paradigm based on data from as many as three mammalian species (rats, mice, and rabbits) and to identify trans-species reprotoxicants with a high probability of being reprotoxic in humans. The authors reported a good predictive performance for the majority of the QSARs in this battery: high specificity (>80%), low false positive rate (<20%), and high database coverage (>80%). Because of the large size of the training sets (containing 627 to 2023 chemicals) and the diversity of molecular structures they represent, the authors argue that the QSARs to have a wide applicability domain. However, the models are not documented in sufficient detail to be reproduced and they are not readily transferable.

OECD QSAR Application Toolbox: The OECD QSAR Application Toolbox is a standalone software application for data gaps for assessing the hazards of chemicals. Data gaps are filled by following a flexible workflow in which chemical categories are built and missing data are estimated by read-across or by applying local QSARs (trends within the category). The Toolbox also includes a range of profilers to quickly evaluate chemicals for common mechanisms or modes of action. In order to support read-across and trend analysis, the Toolbox contains numerous databases with results from experimental studies. It is freely available from the OECD website: http://www.oecd.org/env/existingchemicals/qsar and it will soon include a profiler for predicting ER binding potential, based on a decision tree developed by the US EPA described below (OECD, 2009).

PASS: The PASS (Prediction of Activity Spectra for Substances) is developed and marketed by the Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences. Chemicals structures are presented in mol format and used to generate Multilevel Neighbourhood of Atoms (MNA) descriptors (Filimonov *et al*., 1999). The system predicts the probability (Pa) of a biological activity for a new compound, by estimating the similarity/dissimilarity of the new substance to substances with well known biological activities present in the training set. The tool also gives a cross reference between biological activities on the basis of the knowledgebase of mechanism-effect relationships. A Bayesian algorithm is used to predict various biological activities in terms of the probabilities of presence (Pa) and absence (Pi) of each particular activity (Filimonov & Poroikov, 2008; Poroikov *et al*., 2007). It predicts several specific toxicities among them teratogenicity and embryotoxicity. An online version of PASS is available at: http://195.178.207.233/PASS/index.html.

T.E.S.T.: The Toxicity Estimation Software Tool is an open-source application developed by the US EPA. It estimates the toxicity of a compound by applying several QSAR methodologies thus allowing the user to have greater confidence in predicted toxicities. Among other toxicities it predicts developmental toxicity. The tool is freely downloadable from the EPA website (http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST). The models are well documented and the training set is made available as structure files (SDF file).

TIMES: TIssue MEtabolism Simulator is a heuristic algorithm to generate metabolic maps from a library of biotransformations and abiotic reactions. It allows prioritization of chemicals according to toxicity of their metabolites. The TIMES platform is also used to predict different endpoints including receptor mediated endpoints for oestrogen, androgen and aryl hydrocarbon binding affinity. They are based on the Common Reactivity Pattern (COREPA) approach developed by the Laboratory of Mathematical Chemistry at the Bourgas University Bulgaria. The COREPA approach is a probabilistic classification method which assesses the impact of molecular flexibility on stereo electronic properties of chemicals. Similarity between chemicals is analysed by comparing their conformational distributions, and the system automatically identifies the parameter that best discriminate chemicals in groups. A Bayesian decision tree is then developed for classifying untested chemicals. The use of COREPA to predict oestrogenicity has been well described elsewhere (Mekenyan *et al*., 2003a, b; Schmieder *et al*., 2003).

TOPKAT: This QSAR-based system, developed by Accelrys Inc. (http://accelrys.com/), makes predictions of a range of toxicological endpoints, including developmental toxicity. The Developmental Toxicity Potential (DTP) module of the TOPKAT software was developed from experimental studies selected after review of literature citations on rat oral data. TOPKAT comprises three QSAR models, each applicable to a specific class of chemicals. The output is the probability of a submitted chemical structure being a developmental toxicant in the rat. A probability below 0.3 indicates no potential for developmental toxicity (NEG), whereas a probability above 0.7 signifies developmental toxicity potential (POS). The probability range between 0.3 and 0.7 refers to the "indeterminate" zone (IND). The TOPKAT model automatically determines whether the submitted structure belongs to the Optimum Prediction Space (OPS) of the model in order to evaluate the reliability of prediction. The OPS is TOPKAT's formulation of the model applicability domain - a unique multivariate descriptor space in which a given model is considered to applicable. Any prediction generated for a query structure outside of the OPS space is considered unreliable. The original models were published by Enslein *et al*. (1983) and by Gombar *et al*. (1995), although it is not clear whether the models now implemented in the software are the same as, or refinements of, the original models.

Toxmatch: This freely available software does not in itself generate predictions of reprotoxicity endpoints, but it can be used to develop categories and support read-across assessments. This has been demonstrated in a study by Enoch *et al*. (2009). This study illustrates the use of 2D similarity indices within Toxmatch to form categories for 57 query chemicals. The underlying hypothesis is that chemicals selected as being similar should act via a single mechanism of action, even if that mechanism is unknown. Read-across predictions were performed for the 17 query chemicals for which a category could be formed. The authors concluded that 2D similarity methods offer a useful method for building chemical categories for reproductive toxicity in which a priori mechanistic knowledge is limited. Although the categories proposed are limited in terms of their applicability (40 query chemicals were not allocated to categories), the results form a good basis for further investigations.

VirtualToxLab: This is a commercial tool for predicting endocrine disrupting potential by simulating and quantifying the interactions with aryl hydrocarbon, oestrogen alpha/beta, androgen, thyroid alpha/beta, glucocorticoid, liver X, mineralocorticoid and peroxisome proliferator-activated receptor gamma (Vedani *et al*., 2009; Vedani & Smiesko 2009). It also includes metabolic considerations by simulating interactions with the enzymes CYP450 3A4 and 2A13. The tool is based on the combined use of automated flexible docking with multi-dimensional QSAR (mQSAR).

Table 1 Software for developmental and reproductive toxicity

Software	Availability	Applicability
ADMET Predictor http://www.simulations-plus.com/	Commercial	Qualitative and quantitative prediction of oestrogen receptor toxicity in rats. Based on two models: a qualitative model and, if toxic, the quantitative ratio of IC50 estradiol/IC50 compound).
ACD ToxSuite (ToxBoxes); http://www.acdlabs.com/products/admet/tox/	Commercial Free web application: http://www.pharma-algorithms.com/webboxes/	ER binding affinity prediction. Identify and visualize specific structural toxicophores. Identify analogues from its training set. Algorithms and datasets not disclosed. Predictions associated with confidence intervals and probabilities, providing prediction reliability.
CAESAR http://www.caesar-project.eu/	Freely available	Two classification models for developmental toxicity based on the dataset of Arena *et al*. (2004) including 292 compounds.
Derek http://www.lhasalimited.org/	Commercial	Classification models (different levels of likelihood) based on 23 alerts for developmental toxicity; 4 alerts for oestrogenicity.
Endocrine Disruptor Knowledge Base (EDKB) database (US FDA) http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm	Freely available	Quantitative models to predict the binding affinity of compounds to the estrogen and androgen nuclear receptor proteins.
Leadscope http://www.leadscope.com/	Commercial	Classification models for developmental toxicity in the rodent fetus: dysmorphogenesis (structural and visceral birth defects), developmental toxicity (fetal growth retardation and weight decrease), and fetal survival (fetal death, post-implantation loss, and preimplantation loss). Models of reproductive toxicity: rodent male reproductive, rodent male sperm, female reproductive.
MolCode Toolbox http://molcode.com/	Commercial	Quantitative prediction of rat ER binding affinity and AhR binding affinity
MultiCASE (MC4PC) http://www.multicase.com/	Commercial	Classifcation models for developmental toxicity associated with a variety of datasets, mainly drugs. The marketed software includes modules for predicting mammal sperm toxicity, developmental toxicity, developmental fetal growth retardation, development fetal weight decrease and survival fetal death.

Software	Availability	Applicability
OSIRIS property explorer http://www.organic-chemistry.org/prog/peo/	Freely available	Classification model which predicts "undesirable" effects (mutagenicity, tumorigenicity, irritating effects and reproductive effects), mainly based on the RTECS database of >3500 compounds.
PASS Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Moscow http://ibmc.p450.ru/PASS//	Commercial	Classification models giving probability of reprotoxic effects. The embryotoxicity model predicts the probability that a substance crosses the placental membrane and causes any toxic effect (e.g. fetal bradycardia, low birth weight) or death of an embryo. The teratogenicity model predicts the probability that a substance crosses the placental membrane and causes abnormal development of one or more body systems in the embryo.
T.E.S.T.: The Toxicity Estimation Software Tool http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST)	Freely available	Developmental toxicity estimation. The prediction is done by applying several QSAR methodologies resulting in a greater confidence of the results.
TIMES (COREPA) Laboratory of Mathematical Chemistry, Bourgas University http://oasis-lmc.org/	Commercial	Classification models for the prediction of estrogen, androgen and aryl hydrocarbon binding.The chemical is predicted to fall in one of several activity bins (ranges of binding affinity).
TOPKAT (Accelrys) http://www.accelrys.com	Commercial	Classification model for developmental toxicity of pesticides, industrial chemicals.
Toxboxes Pharma Algorithms http://pharma-algorithms.com/tox_boxes.htm	Commercial	Classification model for the prediction of ER binding.
VirtualToxLab http://www.biograf.ch	Commercial	Classification model for endocrine-disruptiong potential based on simulations of the interactions towards aryl hydrocarbon, estrogen •/•, androgen, thyroid •/•, glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated receptor •, as well as the enzymes CYP450 3A4 and 2A13.

3. Databases

To improve the availability of (Q)SARs and other *in silico* methods for reprotoxicity endpoints, there is a need to develop reprotoxicity databases of high quality and high resolution, in terms of capturing the wide variety of adverse effects and underlying mechanisms of action. Currently available databases are summarised in Table 2.

This need has been acknowledged by the International Life Sciences Institute Risk Science Institute (ILSI RSI), who convened a working group to review methodology used to construct statistically based SAR systems for developmental toxicity (Julien *et al*., 2004). It was concluded that an improved process is needed for utilizing developmental toxicity data in the construction of statistically based SAR models. As result of the ILSI RSI report (Julien *et al*., 2004), ILSI is developing a QSAR-ready and peer-reviewed database with the assistance of Leadscope Inc. (Columbus, Ohio, USA), and with data contributions coming from a range of governmental and academic organisations, as well as contract research laboratories and major pharmaceutical companies. At the time of writing, this database is not yet available.

The US FDA has developed and made publicly available the ICSAS Reprotox database (named after the developer research unit, the Informatics and Computational Safety Analysis Staff [ICSAS]), as reported by Matthews *et al*. (2007a, 2007b). The majority of the data were taken from five publicly available sources: Reproductive Toxicology Center System (REPROTOX), Shepard's Catalog of Teratogenic Agents, Teratogen Information System (TERIS), The Registry of Toxic Effects of Chemical Substances (RTECS), and The Physicians' Desk Reference (PDR). In addition, a small portion of internal FDA reprotoxicity data was included. A review of the many duplicate records provided an opportunity to investigate the consistency of information that was reported in the different public databases but extracted from the same original source. This investigation revealed a consistent interpretation of the data from the original sources with the exception of RTECS, indicating in a lesser reliability of this database. The reprotoxicity data were classified into seven general classes (male reproductive toxicity, female reproductive toxicity, fetal dysmorphogenesis, functional toxicity, mortality, growth, and newborn behavioural toxicity), and 90 specific categories. Each specific category contained over 500 chemicals, but the percentage of active chemicals is low, generally only 0.1–10%. In total, the database contains 51,724 study records from over 10,000 individual reprotoxicity studies in which each record is linked to the test chemical structure. The majority of reprotoxicity studies were conducted in rats, mice and rabbits. The majority of test substances were pharmaceuticals, with a relatively limited number of industrial chemicals. The chemical structures are represented as "mol" files and as SMILES (Simplified Molecular Input Line Entry System) codes. The database contains 2134 organic chemicals that are suitable for QSAR modelling. In the QSAR-ready database, built for QSAR analysis, the inorganics, organometallics, high molecular weight polymers, and mixtures of organic chemicals, were excluded.

In support of the ToxCast predictive toxicology effort (Dix *et al*., 2007) the US EPA has developed and made publicly available the Toxicity Reference Database (ToxRefDB) for capturing information from publicly available *in vivo* toxicity studies. This database contains standard toxicity test results for pesticides and other environmental chemicals. It includes the Developmental Toxicity Endpoints dataset (Knudsen *et al*., 2009) resulting from 383 rat and 368 rabbit prenatal studies on 387 chemicals, mostly pesticides; and the Reproductive Toxicity Endpoints dataset (Martin *et al*., 2009) results from multigeneration reproductive toxicity studies on 316 chemicals. The multigeneration reproductive toxicity data set includes assessment of gonadal function, the oestrous cycle, mating behaviour, conception, gestation, parturition, lactation, weaning, and on the growth and development of the offspring. The information in the ToxRefDB is well structured, searchable and downloadable, which makes it a potentially useful resource for QSAR modelling and other developments in predictive toxicology. In order to develop models capable of supporting risk assessment, dose-response data will need to be added.

Table 2 Databases for reproductive toxicity (including receptor binding)

Database	Availability	Information
Toxicology Data Network (Toxnet) Developmental and Reproductive Toxicology Database (DART) http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?DARTETIC.	Freely available	Bibliographic database containing over 200,000 references to literature published since 1965. It covers teratology and other aspects of developmental and reproductive toxicology. Users can search by subject terms, title words, chemical name, Chemical Abstracts Service Registry Number (RN), and author.
Endocrine Disruptor Knowledge Base (EDKB) database (US FDA) http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/default.htm	Freely available	Biological activity database including in vitro and *in vivo* experimental data for more than 3,000 chemicals and chemical-structure search capabilities. It includes two datasets: Estrogen Receptor (ER) binding dataset (containing 131 ER binders and 101 non-ER binders), and Androgen Receptor (AR) bataset (containing 146 AR binders and 56 non-AR binders). Searchable by assay type and by structure; provides a search ranking based on a structure similarity index.
Endocrine Active Chemicals Database (JRC)	Under development. In-house prototype with web version planned.	Searchable database giving information on chemical identity (e.g. CAS number), chemical structure, toxicity (both to humans and wildlife), physicochemical properties, mode and mechanism of action, for about 520 chemicals, including those on the EU priority list of substances (http://ec.europa.eu/environment/endocrine/strategy/substances_en.htm)
ICSAS Reprotox Database (US FDA) http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm092217.htm	Freely available	WOE values for 2134 organic chemicals (most of them pharmaceuticals; plus limited numbers of industrial chemicals). SMILES and mol files available.
NureXbase http://nurexbase.prabi.fr	Freely available	Information on EACs linked to their receptor targets. Sequence, expression and 3D structures data are linked.
NURSA (Nuclear Receptor Signaling Atlas) http://www.nursa.org/	Freely available	Information on chemical structure, crystal structure, SMILES, physical descriptors, nuclear receptors and mechanism of endocrine action.
OECD (Q)SAR Toolbox	Freely available	Altough primarily a tool for chemical categories and read-acros, it also includes several databases, including reprotoxicity data: 166,072 ER binding data from Danish EPA (pre-generated predictions, not experimental values) as well as 1606 experimental ER binding affinity values from the OASIS commercial database.

Database	Availability	Information
REDIPED (Relational Database of Information on Potential Endocrine Disrupters) developed by the Institute for Environment & Health, University of Leicester, Leicester, UK. http://www.cranfield.ac.uk/health/researchareas/environmenthealth/	Commercial	Includes references and data on chemical identity, physical properties, production volumes, uses, regulations, sources of exposure, exposure assessment, environmental fate & transport (i.e. accumulation, degradation, fate), and biological activity (*in vitro* and *in vivo* activity, binding abilities, relative activity, and general toxic effects).
US EPA ToxRefDB http://www.epa.gov/NCCT/toxrefdb/	Freely available	Standard toxicity test results for pesticides and other environmental chemicals including developmental toxicity (387 chemicals) and multigeneration reproductive toxicity (316 chemicals).

4. Literature models for reprotoxicity endpoints

Compared with other toxicological endpoints, there are relatively few (Q)SARs for reprotoxicity, which is due to the diversity and biological complexity of this family of endpoints as well as the paucity of data suitable for model development. The published (Q)SARs and other structure-based methods (e.g. decision tree approaches) can be grouped into the following categories: a) local models for the reprotoxic effects of individual series of compounds; b) global models for the reprotoxic effects of heterogeneous groups of compounds; c) models for ADME properties; d) models relating to endocrine activity and endocrine disruption potential; and e) chemical categories and read-across assessments. The status of these methods up to 2008 has been reviewed elsewhere (Cronin & Worth, 2008).

In this section, selected and representative examples of published (Q)SAR models and decision tree approaches are described. Most of these models are based on specific chemical classes and consequently have limited applicability domains. A summary is given in Table 3. Models for endocrine-related effects are covered separately below, since they are not necessarily related with reprotoxicity.

Several studies have developed decision trees for classifying reprotoxic effects, which are potentially useful for regulatory applications, due to their transparency, reproducibility and transferability. For example, Sussman *et al*. (Sussman *et al*., 2003) used toxicity data from the Teratogen Information System (TERIS) as well as FDA data. They also explored the development of model batteries (termed ensembles). They found the most important physicochemical variables for predicting developmental toxicity were logP, which expresses a chemical's ability to distribute itself between aqueous and lipid biophases, electronic variables such as the energies of the HOMO (highest occupied molecular orbital), LUMO (lowest unoccupied molecular orbital), and measures of hydrogen bonding.

Examples of read-across within categories are provided by Enoch *et al*. (2009), described above, and by Fabjan *et al*. (2006), who illustrated the ability to make read-across predictions for developmental toxicity within a series of phthalate esters. The category comprised 10 ortho-phthalate esters with different side chain lengths. Phthalates with side-chain lengths C4 to C6, which are known to cause severe reproductive effects, were included. This study showed that by careful category formation, qualitative read-across predictions could be made for several chemicals not used to develop the initial phthalate category. The authors suggest that all the chemicals within the category act via the same mechanism of action, thus adding confidence to the read-across predictions.

A limited number of QSAR models were developed within the ReproTect project (1/07/04-30/06/09), funded under the 6th EU Framework Programme. ReproTect aimed to develop and optimise an array of *in vitro* tests and testing batteries, in order to provide information on the hazard of compounds to the mammalian reproductive cycle (http://www.reprotect.eu/). ReProTect has covered the reproductive cycle by addressing three major research areas: i) fertility; ii) implantation; (iii) prenatal development.

Table 3. Summary of QSAR models for reproductive toxicity

Reference	Endpoint	Dataset size and applicability
Matthews *et al*. (2007a, 2007b)	Male reproductive toxicity, female reproductive toxicity, fetal dysmorphogenesis, fetal functional toxicity, fetal and newborn mammal mortality, fetal growth toxicity, newborn behavioural toxicity	2134 marketed drugs, pesticides and industrial chemicals. Salts, metals, high MW substances, organometallics, gases and complex mixtures were excluded.
Hewitt *et al*. (2007)	Placental transfer (clearance index and a transfer index)	Clearance index dataset (n=86) and a transfer index dataset (n=58).
Arena *et al*. (2004)	Developmental toxicity potential from the evaluation of human and animal data on potentially teratogenic chemicals (FDA/TERIS database)	293 chemicals (117 active and 176 with no evidence of developmental toxicity)
Sussman *et al*. (2003)	Developmental toxicity potential from the evaluation of human and animal data on potentially teratogenic chemicals (FDA/TERIS database)	293 chemicals (117 active and 176 with no evidence of developmental toxicity)
Devillers *et al*. (2002a)	Developmental toxicities to *Hydra attenuata*	17 glycols, glycol ethers, and xylenes
Devillers *et al*. (2002b)	Developmental toxicities to *Hydra attenuata*	30 glycols, glycol ethers, xylenes and phenols
Mekenyan *et al*. (1996)	Acute lethality evaluated using the Frog Embryo Teratogenesis Assay: *Xenopus* (FETAX)	36 semicarbazides and thiosemicarbazides
Richard and Hunter (1996)	Developmental toxicity in mouse whole embryo culture assay.	acetic acid and a series of ten haloacetic acids
Dawson *et al*. (1996)	Developmental toxicity to *Xenopus* embryos including 96-h lethality, malformation and developmental hazard index	45 carboxylic acids including aliphatics and aromatics
Ridings *et al*. (1992)	Developmental toxicity potential in rats	12 dopamine mimetics
Dawson (1991)	A modified FETAX (Frog Embryo Teratogenesis Assay: *Xenopus*)	10 aliphatic carboxylic acids
Dawson *et al*. (1991)	Toxicity and teratogenicity using early embryos of the frog *Xenopus laevis*	Nine benzoic acid hydrazides and carbazates
Schultz and Dawson (1990)	Developmental toxicity to *Hydra attentuata*	14 glycols and glycol ethers
Schultz and Dawson (1990)	96 hour embryo malformation (EC50) using FETAX	9 short-chain carboxylic acids:
Schultz and Dawson (1990)	96 hour embryo malformation (EC50) using FETAX	9 acid hydrazides
Kavlock (1990)	maternal toxicity in the Chernoff/ Kavlock assay	22 substituted phenols
Brown *et al*. (1989)	teratogenicity in an *in vitro* assay to inhibit rat embryo mid-brain cells.	20 phenylhydantoins

5. Endocrine-related effects

5.1 Endocrine Active Substances and potential Endocrine Disruptors

Endocrine Active Substances (EAS) are chemicals having the potential to interfere with the endocrine systems, as judged from *in vitro* or *in vivo* tests. Such chemicals may be regarded as endocrine disruptors (EDs) if there is evidence that the substance causes adverse health effects in an intact organism, or its progeny, secondary to changes in endocrine function. In particular, EDs have been associated with reprotoxicity, as well as cancer, diabetes and obesity. Numerous mechanisms of action have been associated with endocrine disruption, and a wide variety of *in vitro* tests have been developed to identify chemicals acting via these mechanisms. The status of these *in vitro* tests has been reviewed by Jacobs *et al*. (2008), who also emphasise the need to incorporate metabolic considerations into the assessment of EAS. Chemicals with ED potential are of particular concern for human health and the environment, especially if their potential adverse effects are not detected by other endpoint assays. In REACH, EDs are considered to be Substances of Equivalent Concern as other Substances of Very High Concern.

EAS act via a range of mechanisms with the result of enhancing or suppressing normal hormone responses, including homeostatic and feedback mechanisms. In many cases, EAS act by binding to nuclear hormone receptors (NRs), which are ligand-inducible transcription factors involved in the regulation of specific target genes and of critical cellular processes such as cell growth, differentiation and metabolic processes. Members of the NR superfamily include receptors for various steroid hormones oestrogen (ER), androgen (AR), progesterone (PR), several corticosteroids, retinoic acid, thyroid hormones, vitamin D, and dietary lipids (the peroxisome proliferator activated receptor; PPAR).

The largest and best studied group of NRs is the Oestrogen Receptor (ER) family. The ER is a ligand-dependent transcription factor – when a hormone binds to the ligand binding domain (LBD), it induces a conformational change in the receptor that initiates a series of events that culminate in the activation or repression of responsive genes (Anstead *et al*., 1997). The crystallographic structures available for the ER have provided insights into mechanisms of action and have given an input to the development of highly specific *in silico* models. The mobility and plasticity of the ER ligand-binding cavity have been identified as important factors allowing the binding of compounds of different structural types to the receptor site (Pike *et al*., 1999). In absence of the ligand, ERs are in an inactive conformation in the target cell nuclei. The binding of an agonist switches the ER into an active conformation, while the binding of an antagonist blocks agonist access. A third category of ligands, termed selective ER modulators (SERMS), have the ability to act as both agonists and antagonists, depending on the cellular and promoter context.

5.2 *In silico* modelling of endocrine-related effects

There is an extensive literature on the modelling of NR binding and endocrine activity, including studies based on traditional QSAR, molecular modelling, and decision tree approaches. This section reviews, with illustrative examples, shows the main types of *in silico* methods that have been developed to support the identification of EDs. Strictly, these should not be regarded as *in silico* models for endocrine disruption, since they do not in themselves provide sufficient information to determine whether adverse effects are produced secondary to changes in endocrine function. However, they could be regarded as models for the identification of EAS.

An extensive (but not exhaustive) list of available models is provided in Table 4. Some of the studies included in this table were reviewed in the context of a JRC-funded study entitled the "Validation of non-commercial (Q)SAR models for ER and AR binding", which was performed by Mario Negri Institute (Benfenati *et al*., 2005). In this study, non-proprietary models for ER and AR binding activity

were reviewed in order to identify interesting publications related to ER and AR endpoints. A scheme for scoring each model/publication was based on the availability of key information (experimental biological data, structures, descriptors, chemical domain and models). A total of 158 models (published until 2005) were scored. Some of the most promising (highest-scoring) ones are included in the Table 4, along with some additional ones (those published after 2005).

Several studies have reported decision trees for categorising chemicals based on the NR binding potential. These are potentially useful for regulatory applications, due to their simplicity, transparency, reproducibility and transferability. For example, Netzeva *et al*. (2006) developed a binary classification model (Figure 1) for predicting oestrogen-active versus inactive chemicals on the basis of two descriptors with a clear physicochemical meaning (logP and the number of hydrogen bond donors). The data used were from an *in vitro* reporter gene assay. Subsequently Gallegos Saliner *et al*. (2006) assessed the predictivity of the model using an external test set and by taking into account the limitations associated with the applicability domain (AD) of the model. This validation study indicated a concordance of 71%, a sensitivity of 84 and a specificity of 69%.

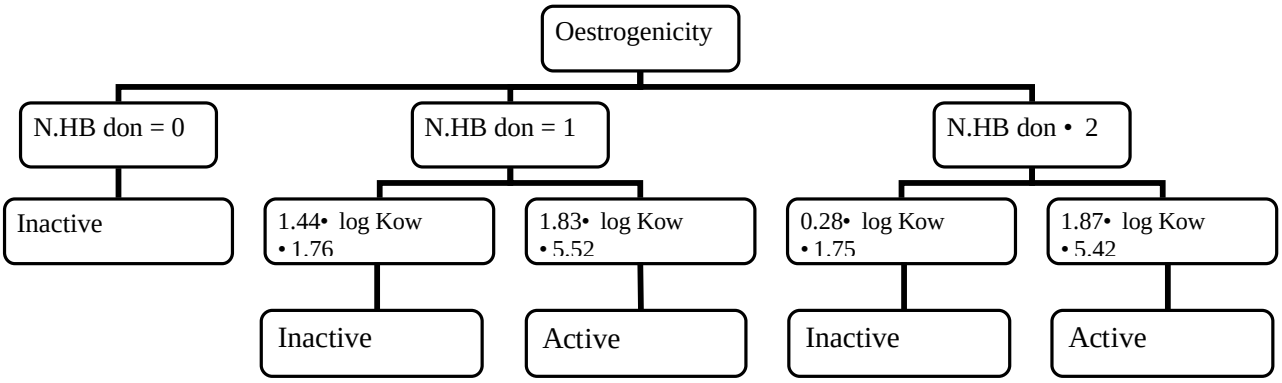

Figure 1 Decision tree for the classification of oestrogen-active chemicals (modified from Netzeva *et al*., 2006)

A decision tree for predicting ER binding has been developed by the US EPA and has been extensively peer-reviewed, both by the OECD QSAR Management Group (OECD, 2009) and by an independent peer review panel convened by the EPA. The decision tree is based on the hypothesis that the structural domain of chemicals that can bind to the ER is determined by the energy and steric constraints of the ER itself. Based on experimental data available in literature, the nature of the chemical interactions in the various “subpockets” within the ER-binding domain(s) was hypothesised. Three primary ER binding subpockets were identified, having different requirements for hydrogen binding. The decision tree described uses basic structural features and simple properties to match chemicals with “similar” chemical groups. The system examines each chemical and places them into groups of inactive chemicals, “drug-like” chemicals (which have the potential for strong ER binding affinity), or groups of chemicals which may have weak-to-moderate binding affinity, depending on specific properties or structural features. The QSAR Management Group has decided to incorporate this decision tree as a “profiler” in the OECD (Q)SAR Application Toolbox, thereby making it freely available and readily applicable.

QSARs have been developed by the University of Bourgas (Serafimova *et al*., 2007) to predict the binding affinity of chemicals with the human oestrogen receptor alpha (hER•). The hER• binding affinities were modelled using a training set of 645 chemicals. The ultimate model was organised as a battery of QSARs associated with interaction mechanism and potency categories. The QSAR model was combined with a metabolic simulator in the TIMES modelling platform. When a new chemical is submitted to the system, it undergoes simulated metabolism; part of the generated metabolites are detoxified by phase II reactions thereby interrupting the metabolism; other metabolites are filtered by the QSAR model to predict their binding potency to hER•. According to the authors, this model can be used as supporting information in a weight of evidence approach helping in the replacement or minimization of animal testing.

When the 3D structure of the protein receptor is known, *in silico* approaches such as molecular docking can be applied. Docking is used to find the best match between a biological macromolecule and a ligand. The ligand is placed inside the receptor pocket and the free energy of binding of the molecular complex is estimated computationally. The receptor structure needs to available from experimental studies, usually X-ray crystallography or NMR. In the case of ERs, several crystal structures of the receptor with different ligands (both agonists and antagonists) are available from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do).

Another *in silico* approach often used for ER affinity prediction is 3D-QSAR based on so-called field-based descriptors that describe the micro-environment surrounding the (ligand) molecules (molecular electrostatic and steric potential and Van der Waals volume). For example, Comparative Molecular Field Analysis (CoMFA) is a modelling method that examines molecules in three-dimensional detail, describing the magnitude and directions of electronic and steric interactions (Cramer *et al*., 2002). CoMFA produces an imaginary 3D box around the ligand, consisting of steric and electrostatic interaction energies at each grid point. These values become the descriptors for QSAR analysis. The main advantages of CoMFA methods are: a) the crystal structure of the protein target is not needed, since the analysis is derived entirely from the ligand; and b) by describing properties in terms of 3D fields, it is possible to visualise areas within the 3D space around the ligand that are positively or negatively related to the activity. The main disadvantage of CoMFA is the need to align (superimpose) numerous 3D structures which makes it difficult to study heterogeneous datasets. Some examples of CoMFA investigations of ER and AR binding are given in Table 4. CoMFA is a research tool that requires considerable expertise to implement. It is useful for investigations into mechanisms of binding and in the development of QSARs, but is not suited for the routine assessment of chemicals by non-specialists.

An alternative to CoMFA, which avoids alignment difficulties, is to use VolSurf (Cruciani *et al*., 2000a,b) and ALMOND (Pastor *et al*., 2000), which are commercially available 3D-QSAR methods developed by Molecular Discovery (http://www.moldiscovery.com/index.php). These are sophisticated yet easy-to-handle computational procedures that can be used to explore the physicochemical property space of a molecule, using a simple molecular input such as SMILES. There is no need to use and manipulate 3D structures since these operations are automatically performed by the software. VolSurf automatically generates 3D maps and compresses the information into numerical descriptors. ALMOND generates and handles alignment-independent descriptors called GRIND (GRid INdependent Descriptors). These are a new generation of 3D-molecular descriptors - being alignment independent, they are quickly and automatically computed. These methodologies are promising research tools for future QSAR development.

A more recent development is VirtualToxLab, developed by Vedani and colleagues (Vedani *et al*., 2009; Vedani & Smiesko 2009). This is an *in silico* tool for predicting the endocrine-disrupting potential of compounds by simulating their interactions towards a series of proteins known to trigger adverse effects. It is based on a fully automated protocol, calculating the binding affinity of a molecule towards a series of proteins and estimating the resulting toxic potential. Currently, 12 protein targets are included: the androgen, aryl hydrocarbon, oestrogen alpha/beta, glucocorticoid, mineralocorticoid, thyroid alpha/beta liver X and the peroxisome proliferator-activated receptor gamma (PPAR-γ), as well

as the enzymes cytochrome P450 3A4 and 2A13. Toxic potential is estimated automatically by simulating the interactions with the macromolecular targets, by quantifying these interactions in terms individual binding affinities and combining the flexible docking routine with multidimensional QSAR. The technology is accessible over the Internet (http://www.biograf.ch/).

Table 4. Some examples *of in silico* models for the prediction of ER and AR binding

Reference	Endpoint	Method / type of model	Dataset size and applicability
Taha *et al.* (2010)	ER• binding	Pharmacophore modeling by CATALYST	Training set:119 compounds; Test set: 23 compounds
Salum *et al.* (2008)	binding affinity values for both ERα and ER•	3D QSAR: CoMFA and GRID	81 hER modulators
Soderholm *et al.* (2008)	AR binding	3D QSAR and docking	219,680 compounds from Asinex commercial library (http://www.asinex.com).
Vinggaard *et al.* (2008)	Human AR binding	MultiCASE analysis to identify the most representative chemical fragments responsible for the AR antagonism	Training consisting of 523 chemicals covering a wide range of chemical structures (e.g., organochlorines and polycyclic aromatic hydrocarbons) and various functions (e.g., natural hormones, pesticides, plastizicers, plastic additives, brominated flame retardants, and roast mutagens)
Islam *et al.* (2008)	ER binding	Pharmacophore by Catalyst	35 compounds in the training set plus 102 compounds in the test set.
Salum *et al.* (2007)	ERα modulators	3D QSAR (CoMFA) and 2D Hologram QSAR	Two training sets containing either 127 or 69 compounds
Serafimova *et al.* (2007)	Human ER binding	COmmon REactivity PAttern (COREPA) modeling approach	645 chemicals included 497 steroid and environmental chemicals
Netzeva *et al.* (2006)	Oestrogen-responsive gene expression in vitro reporter gene assay.	Classification tree	117 aromatic compounds published including bisphenols, benzophenones, flavonoids, biphenyls, phenols, and other aromatic chemicals
Tamura *et al.* (2006)	AR binding	3D QSAR (CoMFA)	35 chemicals for antagonists model and 13 chemicals for agonist and antagonist activity models
Saliner *et al.* (2006)	Human ERα binding	Models developed using quantum similarity methods	117 aromatic chemicals
Ghafourian & Cronin (2005)	Rat ER binding	TSAR 3D and 2D descriptors, PLS analysis by SIMCA-P, cluster analysis in MINITAB	131 chemicals from NCTR dataset
Vedani *et al.* (2005)	Rat ER binding	Protein Modeling and 6D-QSAR	106 compounds

Reference	Endpoint	Method / type of model	Dataset size and applicability
Lill *et al*. (2005)	Aryl hydrocarbon, ER, AR binding affinity.	Multi-dimensional QSAR: Quasar and Raptor	Database containing 121 Aryl hydrocarbon compounds (91 training and 30 external test), 116 ER (93/23) and 72 AR (56/16)
Demyttenaere-Kovatcheva *et al*. (2005)	ER • and •	CoMFA	Diphenolic Azoles: 72 in training and 32 in test set
Marini *et al*. (2005)	ER binding	Various multivariate methods e.g. a back-propagation neural network	132 heterogeneous compounds
Hong *et al*. (2005)	ER binding	Decision forest	232 structurally diverse compounds, validated using a test set of 463 compounds
Kurunczi *et al*. (2005)	Rat ER binding	Partial least-squares (PLS) model	45
Akahori *et al*. (2005)	Human ERα binding	A two-step QSAR using discriminant and multilinear regression (MLR) analyses.	alkylphenols, phthalates, diphenylethanes and benzophenones
Mukherjee *et al*. (2005)	ER binding	QSAR based on multiple linear regression	25 triphenylacrylonitriles
Zhao *et al*. (2005)	AR binding	QSARs based on multiple linear regression, radical basis function neural network and support vector machine (SVM)	146 structurally diverse natural, synthetic and environmental chemicals
Lill *et al*. (2004)	ER binding	Multidimensional QSAR (Raptor)	116 chemicals from NCTR dataset
Asikainen *et al*. (2004)	ERα and ER• binding	Consensus kNN QSAR	calf (53), mouse (68), rat (130), human ER• (61), human ER• (61)
Tong *et al*. (2004)	ER binding	Decision Forest classifier	Dataset 1 : 232 chemicals tested in-house (131 active, 101 inactive) Dataset 2:, literature compilation of 1,092 chemicals (350 active, 736 inactive) Both datasets are structurally diverse
Hong *et al*. (2003)	Rat AR binding	3D QSAR (CoMFA)	146
Fang *et al*. (2001)	Rat ER binding	Pharmacophore by CATALYST	232 chemicals from NCTR dataset
Kramer & Giesy (1999)	Bovine calf uterine ER binding	Quantitative structure-binding relationship (QSBR)	25 hydroxy PCBs

AR = androgen receptor; ER = oestrogen receptor; ERα = oestrogen receptor alpha; ER• = oestrogen receptor beta

6. Regulatory use of *in silico* predictions

In silico models for reprotoxicity endpoints and NR binding have mainly been used for setting priorities for testing, rather than to fill data gaps for hazard and risk assessment.

A stepwise "four-phase" scheme for identifying oestrogenic substances was developed by the US FDA (Tong *et al*., 2004). These four steps (phases) work in a hierarchical manner to reduce the size of a dataset incrementally while increasing the accuracy of prediction. Within each step (phase), different models were selected to work in a complementary fashion in order to minimise the rate of false negatives. The four phases for identification of ER ligands are described below:

- Phase I: Filtering – gross structural filters for low and high molecular weight compounds (i.e. <94 or >1000) or the absence of a ring structure eliminate chemicals very unlikely to bind to ER.
- Phase II: Active/Inactive Assignment –chemicals passing from the previous phase are classified into active and inactive categories. Three structural alerts, seven pharmacophore queries, and a Decision Tree classification model are used in parallel to discriminate active from inactive chemicals.
- Phase III: Quantitative Predictions – a CoMFA model is used to make a more accurate quantitative activity prediction for chemicals passing from Phase II. Chemicals with higher predicted binding affinity are given higher priority for further evaluation in Phase IV. The CoMFA model demonstrated good statistical reliability in both cross-validation and external validation (Fang *et al*., 2001).
- Phase IV: Rule-Based Decision-Making System - the final stage of the integrated system relies on the use of a knowledge-base (expert system) to make priority setting decisions (Fang *et al*., 2001).

This stepwise approach shows how models of increasing complexity and accuracy can be used as successive and more accurate filters. However, the application of this approach requires specialised expertise to apply the CoMFA model in Phase III.

An example of how (Q)SARs can be used in classification and labelling has been reported by the Danish National Food Institute in Denmark (Jensen *et al*., 2008). They performed a screening exercise of 57, 014 European Inventory of Existing Chemical Substances (EINECS) chemicals by using in-house and commercial QSAR models (mainly MultiCASE) in order to identify possible reprotoxicants. Three QSAR models were used for reproductive toxicity for the endpoints teratogenic risk to humans, dominant lethal effect in rodents and *Drosophila melanogaster* sex-linked recessive lethal effect. In addition, the chemicals were also screened by using three models for endocrine activity. Chemicals were considered predicted positive for reproductive toxicity if a positive prediction was obtained in any of the models within the applicability domain. On this basis, 5240 EINECS chemicals (9.2% of the chemicals screened) were predicted as reprotoxicants by one or more of the models. The authors also interpreted the model outputs in terms of EU classifications for reproductive toxicity - category Xn (Harmful) and R63 (Possible risk of harm to the unborn child). The list of chemicals with EU classifications suggested on the basis of QSAR, have been submitted to the Danish EPA to support a future update of the advisory classification list (which industry can use to support the self-classification of chemicals).

7. Conclusions

At present, the availability of (Q)SARs for reprotoxicity endpoints (excluding models related to endocrine activity) is limited as a result of the diversity and biological complexity of the endpoints, and the paucity of data suitable for modelling. Available models are potentially useful as a means of supporting hazard identification and priority setting, but not yet for the establishment of toxic potencies for use in risk assessment.

Given the nature of the reprotoxicity endpoints, it is unlikely that an entirely structure-based approach will be capable of fully describing and predicting the *in vivo* effects. Thus, available models should not be used in isolation but to contribute to WoE assessments, and to guide experimental testing, where necessary. Batteries of models and *in vitro* tests will need to be developed, and this has been the aim of an EU-funded Reprotect project (http://www.reprotect.eu/). It is also unlikely that the combined use of *in vitro* methods and QSAR analysis will be able to completely replace vertebrate animal testing, but it should increasingly be possible to reduce it.

At the current state of development, it is not possible to give clear recommendations on how to use the results of models for reprotoxicity endpoints. For short-term progress (next 3 years), it is recommended that further research on the regulatory applicability of current models is performed, for example along the lines of the Danish EPA study (Jensen et al., 2008). Further work will also need to be aimed at the development and assessment of integrated strategies including *in vitro* data as well as *in silico* models.

The future development of (Q)SAR models and databases will also depend on the development of a standardised vocabulary for describing the plethora of reprotoxic effects at different levels of biological organisation. ILSI and Leadscope have already started such an initiative. In relation to databases, an important achievement has been the construction, from publicly available information sources, of the US FDA's weight-of-evidence (WOE) Reprotox database suitable for QSAR modelling (Matthews et al., 2007a, 2007b).

In the longer term (5 years and more), the development of systems biology approaches incorporating "omic" and HTS data is likely to become increasingly important. Preliminary investigations have started, for example in connection with the US ToxCast initiative (Martin et al., 2009; Knudsen et al., 2009). It is too early to judge whether this approach, which reflects a shift from modelling apical endpoints to toxicity pathways, will ultimately be useful in the routine regulatory assessment of chemicals.

In contrast to reprotoxicity, there is an extensive and growing range of software and literature models for predicting endocrine-related activities, and especially binding to the ER and AR receptors. In many cases, these models are at the research stage and require specialised expertise to recreate them in molecular modelling software. However, there are a number of potentially useful models, including simple decision tree approaches (e.g. OECD, 2009) as well as commercial models (e.g. the VirtualToxLab approach; Vedani et al., 2009). One of the main challenges here is to develop agreed approaches for interpreting model results for regulatory applications other than priority setting.

8. References

Akahori Y, Nakai M, Yakabe Y, Takatsuki M, Mizutani M, Matsuo M & Shimohigashi Y (2005). Two-step models to predict binding affinity of chemicals to the human estrogen receptor • by three-dimensional quantitative structure–activity relationships (3D-QSARs) using receptor-ligand docking simulation. SAR and QSAR in Environmental Research 16(4), 323-337.

Anstead GM, Carlson KE & Katzenellenbogen JA (1997). The estradiol pharmacophore: Ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. Steroids 62(3), 268-303.

Arena VC, Sussman NB, Mazumdar S, Yu S & Macina OT (2004). The Utility of Structure–Activity Relationship (SAR) Models for Prediction and Covariate Selection in Developmental Toxicity: Comparative Analysis of Logistic Regression and Decision Tree Models. SAR and QSAR in Environmental Research 15(1), 1 - 18.

Asikainen AH, Ruuskanen J & Tuppurainen KA (2004). Consensus kNN QSAR: a versatile method for predicting the estrogenic activity of organic compounds in silico. A comparative study with five estrogen receptors and a large, diverse set of ligands. Environmental Science & Technology 38(24), 6724-6729.

Benfenati E, Roncaglioni A, Boriani E, Porcelli C, Spreafico M & Lo Piparo E (2005). Validation of selected, non-commercial (Q)SAR models for ER and AR binding. Final report of JRC contract CCR.IHCP.C430414.X0.

Brown LP, Lewis DFV, Flint OP, Orton TC & Gibson GG (1989). Teratogenicity of phenylhydantoins in an in vitro system: Molecular orbital-generated quantitative structure-toxicity relationships. Xenobiotica 19(12), 1471-1481.

Cramer RD, Patterson DE & Bunce JD (2002). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society 110(18), 5959-5967.

Cronin MTD & Worth AP (2008). (Q)SARs for Predicting Effects Relating to Reproductive Toxicity. QSAR & Combinatorial Science 27(1), 91-100.

Cruciani G, Crivori P, Carrupt PA & Testa B (2000). Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. Journal of Molecular Structure: THEOCHEM 503(1-2), 17-30.

Dawson DA (1991). Additive incidence of developmental malformation for *Xenopus embryos* exposed to a mixture of ten aliphatic carboxylic acids. Teratology 44(5), 531-546.

Dawson DA & Wilke TS (1991). Initial evaluation of developmental malformation as an end point in mixture toxicity hazard assessment for aquatic vertebrates. Ecotoxicology and Environmental Safety 21(2), 215-226.

Dawson DA, Schultz TW & Hunter RS (1996). Developmental toxicity of carboxylic acids to *Xenopus embryos*: A quantitative structure-activity relationship and computer-automated structure evaluation. Teratogenesis, Carcinogenesis, and Mutagenesis 16(2), 109-124.

Demyttenaere-Kovatcheva DA, Cronin MTD, Benfenati E, Roncaglioni A & Lo Piparo E (2005). Identification of the structural requirements of the receptor-binding affinity of diphenolic azoles to estrogen receptors alpha and beta by three-dimensional quantitative structure-activity relationship and structure-activity relationship analysis. Journal of Medicinal Chemistry 48(24), 7628-7636.

Devillers J, Chezeau A & Thybaud E (2002). PLS-QSAR of the adult and developmental toxicity of chemicals to Hydra attenuata. SAR and QSAR in Environmental Research 13(7), 705 - 712.

Devillers J, Chezeau A, Thybaud E & Rahmani R (2002). QSAR modeling of the adult and developmental toxicity of glycols, glycol ethers and xylenes to hydra attenuata. SAR and QSAR in Environmental Research 13(5), 555 - 566.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW & Kavlock RJ (2007). The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. Toxicological Sciences 95(1), 5-12.

Enoch SJ, Cronin MTD, Madden JC & Hewitt M (2009). Formation of Structural Categories to Allow for Read-Across for Teratogenicity. QSAR & Combinatorial Science 28(6-7), 696-708.

Enslein K, Lander TR & Strange JR (1983). Teratogenesis: a statistical structure-activity model. Teratogenesis, Carcinogenesis, and Mutagenesis 3(3), 289-309.

Fabjan E, Hulzebos E, Mennes W & Piersma AH (2006). A Category Approach for Reproductive Effects of Phthalates. Critical Reviews in Toxicology 36(9), 695-726.

Fang H, Tong W, Shi LM, Blair R, Perkins R, Branham W, Hass BS, Xie Q, Dial SL, Moland CL & Sheehan DM (2001). Structure• Activity Relationships for a Large Diverse Set of Natural, Synthetic, and Environmental Estrogens. Chemical Research in Toxicology 14(3), 280-294.

Filimonov D & Poroikov V (2008). Probabilistic Approaches in Activity Prediction. In: Chemoinformatics Approaches to Virtual Screening (Alexandre Varnek and Alexander Tropsha eds.). Cambridge, RSC Publishing, 182-216.

Filimonov D, Poroikov V, Borodina Yu & Gloriozova T (1999). Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. Journal of Chemical Information and Computer Sciences 39(4), 666-670.

Gallegos Saliner A, Netzeva TI & Worth AP (2006). Prediction of Estrogenicity: Validation of a Classification Model. SAR and QSAR in Environmental Research 17, 195-223.

Ghafourian T & Cronin MT (2005). The impact of variable selection on the modelling of oestrogenicity. SAR and QSAR in Environmental Research 16(1-2), 171-190.

Gombar VK, Enslein K & Blake BW (1995). Assessment of developmental toxicity potential of chemicals by quantitative structure-toxicity relationship models. Chemosphere 31(1), 2499-2510.

Hewitt M, Madden JC, Rowe PH & Cronin MTD (2007). Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier. SAR and QSAR in Environmental Research 18(1), 57 - 76.

Hewitt M, Ellison CM, Enoch SJ, Madden JC & Cronin MTD (2010). Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. Reproductive Toxicology 30(1), 147-160.

Hong H, Fang H, Xie Q, Perkins R, Sheehan DM & Tong W (2003). Comparative molecular field analysis (CoMFA) model using a large diverse set of natural, synthetic and environmental chemicals for binding to the androgen receptor. SAR and QSAR in Environmental Research 14(5-6), 373-388.

Hong H, Tong W, Xie Q, Fang H & Perkins R (2005). An in silico ensemble method for lead discovery: decision forest. SAR and QSAR in Environmental Research 16(4), 339 - 347.

Hulzebos EM & Posthumus R (2003). (Q)SARs: gatekeepers against risk on chemicals? SAR and QSAR in Environmental Research 14(4), 285-316.

Islam MA, Nagar S, Das S, Mukherjee A & Saha A (2008). Molecular design based on receptor-independent pharmacophore: application to estrogen receptor ligands. Biological & Pharmaceutical Bulletin 31(7), 1453-1460.

Jacobs MN, Janssens W, Bernauer U, Brandon E, Coecke S, Combes R, Edwards P, Freidig A, Freyberger A, Kolanczyk R, Mc Ardle C, Mekenyan O, Schmieder P, Schrader T, Takeyoshi M & van der Burg B (2008). The use of metabolising systems for in vitro testing of endocrine disruptors. Current Drug Metabolism 9(8), 796-826.

Jensen GE, Niemelä JR, Wedebye EB & Nikolov NG (2008). QSAR models for reproductive toxicity and endocrine disruption in regulatory use – a preliminary investigation. SAR and QSAR in Environmental Research 19(7), 631 - 641.

Julien E, Willhite CC, Richard AM, DeSesso JM & Group IRW (2004). Challenges in constructing statistically based structure-activity relationship models for developmental toxicity. Birth Defects Research Part A: Clinical and Molecular Teratology 70(12), 902-911.

Kavlock RJ (1990). Structure-activity relationships in the developmental toxicity of substituted phenols: in vivo effects. Teratology 41(1), 43-59.

Knudsen TB, Martin MT, Kavlock RJ, Judson RS, Dix DJ & Singh AV (2009). Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB. Reproductive Toxicology 28(2), 209-219.

Kramer VJ & Giesy JP (1999). Specific binding of hydroxylated polychlorinated biphenyl metabolites and other substances to bovine calf uterine estrogen receptor: structure-binding relationships. Science of the Total Environment 233(1-3), 141-161.

Kurunczi L, Seclaman E, Oprea TI, Crisan L & Simon Z (2005). MTD• PLS: A PLS Variant of the Minimal Topologic Difference Method. III. Mapping Interactions between Estradiol Derivatives and the Alpha Estrogenic Receptor. Journal of Chemical Information and Modeling 45(5), 1275-1281.

Lill MA, Dobler M & Vedani A (2005). In silico prediction of receptor-mediated environmental toxic phenomena-application to endocrine disruption. SAR and QSAR in Environmental Research 16(1-2), 149-169.

Lill MA, Vedani A & Dobler M (2004). Raptor: Combining Dual-Shell Representation, Induced-Fit Simulation, and Hydrophobicity Scoring in Receptor Modeling: Application toward the Simulation of Structurally Diverse Ligand Sets. Journal of Medicinal Chemistry 47(25), 6174-6186.

Marini F, Roncaglioni A & Novic• M (2005). Variable Selection and Interpretation in Structure• Affinity Correlation Modeling of Estrogen Receptor Binders. Journal of Chemical Information and Modeling 45(6), 1507-1519.

Martin MT, Mendez E, Corum DG, Judson RS, Kavlock RJ, Rotroff DM & Dix DJ (2009). Profiling the Reproductive Toxicity of Chemicals from Multigeneration Studies in the Toxicity Reference Database. Toxicological Sciences 110(1), 181-190.

Maslankiewicz L HE, Vermeire TG, Muller JJA, Piersma AH (2005). Can chemical structure predict reproductive toxicity? Available at: http://hdl.handle.net/10029/7374

Matthews EJ, Kruhlak NL, Daniel Benz R & Contrera JF (2007a). A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. Regulatory Toxicology and Pharmacology 47(2), 115-135.

Matthews EJ, Kruhlak NL, Daniel Benz R, Ivanov J, Klopman G & Contrera JF (2007b). A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. Regulatory Toxicology and Pharmacology 47(2), 136-155.

Mekenyan O, Dimitrov S, Schmieder P & Veith G (2003a). In silico modelling of hazard endpoints: current problems and perspectives. SAR and QSAR in Environmental Research 14, 361-371.

Mekenyan O, Nikolova N & Schmieder P (2003b). Dynamic 3D QSAR techniques: applications in toxicology Journal of Molecular Structure: THEOCHEM 622, 147-165.

Mekenyan OG, Schultz TW, Veith GD & Kamenska V (1996). 'Dynamic' QSAR for semicarbazide-induced mortality in frog embryos. Journal of Applied Toxicology 16(4), 355-363.

Mukherjee S, Saha A & Roy K (2005). QSAR of estrogen receptor modulators: exploring selectivity requirements for ER[alpha] versus ER[beta] binding of tetrahydroisoquinoline derivatives using E-state and physicochemical parameters. Bioorganic & Medicinal Chemistry Letters 15(4), 957-961.

Netzeva TI, Gallegos Saliner A & Worth AP (2006). Comparison of the Applicability Domain of a QSAR for Estrogenicity with a Large Chemical Inventory. Environmental Toxicology & Chemistry 25, 1223-1230.

OECD (2009). (Quantitative) Structure Activity Relationships [(Q)SARs] of Estrogen binding affinity to support prioritization of heterogenous chemicals within defined inventories for screening and testing. Schmieder P., R. Kolanczyk , H. Aladjov , M. Hornung , M. Tapper , J. Denny , B. Sheedy. Organisation for Economic Cooperation and Development. Available at: http://www.olis.oecd.org/olis/2009doc.nsf/LinkTo/NT00004CD6/$FILE/JT03268747.PDF

Pastor M, Cruciani G, McLay I, Pickett S & Clementi S (2000). GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. Journal of Medicinal Chemistry 43(17), 3233-3243.

Pearl GM, Livingston-Carr S, Durham S (2001). Integration of computational analysis as a sentinel tool in toxicological assessments. Current Topics in Medicinal Chemistry 1, 247-255.

Poroikov V., Filimonov D., Lagunin A., Gloriozova T & Zakharov A (2007). PASS: identification of probable targets and mechanisms of toxicity. SAR and QSAR in Environmental Research 18(1), 101-110.

Pike ACW, Brzozowski AM, Hubbard RE, Bonn T, Thorsell A-G, Engstrom O, Ljunggren J, Gustafsson J-A & Carlquist M (1999). Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. EMBO Journal 18(17), 4608-4618.

Richard AM & Hunter ES III (1996). Quantitative structure-activity relationships for the developmental toxicity of haloacetic acids in mammalian whole embryo culture. Teratology 53(6), 352-360.

Ridings JE & Baldwin JA (1992). A qualitative assessment of developmental toxicity within a series of structurally related dopamine mimetics. Toxicology 76(3), 197-207.

Saliner AG, Netzeva TI & Worth AP (2006). Prediction of estrogenicity: validation of a classification model. SAR and QSAR in Environmental Research 17(2), 195 - 223.

Salum LdB, Polikarpov I & Andricopulo AD (2007). Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. Journal of Molecular Graphics and Modelling 26(2), 434-442.

Salum LvB, Polikarpov I & Andricopulo AD (2008). Structure-Based Approach for the Study of Estrogen Receptor Binding Affinity and Subtype Selectivity. Journal of Chemical Information and Modeling 48(11), 2243-2253.

Scialli AR (2008). The challenge of reproductive and developmental toxicology under REACH. Regulatory Toxicology and Pharmacology 51(2), 244-250.

Schmieder P, Mekenyan O, Bradbury S & Veith G (2003). QSAR prioritization of chemical inventories for endocrine disruptor testing. Pure and Applied Chemistry 75, 2389-2396.

Schultz TW & Dawson DA (1990). Structure-Activity Relationships for Teratogenicity and Developmental Toxicity, in: W. Karcher, J. Devillers (Eds.), Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicity, European Commission, Brussels, pp. 389–409.

Serafimova R, Todorov M, Nedelcheva D, Pavlov T, Akahori Y, Nakai M & Mekenyan O (2007). QSAR and mechanistic interpretation of estrogen receptor binding. SAR and QSAR in Environmental Research 18(3), 389 - 421.

Soderholm AA, Viiliainen J, Lehtovuori PT, Eskelinen H, Roell D, Baniahmad A & Nyronen TH (2008). Computationally Identified Novel Diphenyl- and Phenylpyridine Androgen Receptor Antagonist Structures. Journal of Chemical Information and Modeling 48(9), 1882-1890.

Sussman NB, Arena VC, Yu S, Mazumdar S & Thampatty BP (2003). Decision tree SAR models for developmental toxicity based on an FDA/TERIS database. SAR and QSAR in Environmental Research 14(2), 83-96.

Taha MO, Tarairah M, Zalloum H & Abu-Sheikha G (2010). Pharmacophore and QSAR modeling of estrogen receptor [beta] ligands and subsequent validation and in silico search for new hits. Journal of Molecular Graphics and Modelling 28(5), 383-400.

Tamura H, Ishimoto Y, Fujikawa T, Aoyama H, Yoshikawa H & Akamatsu M (2006). Structural basis for androgen receptor agonists and antagonists: Interaction of SPEED 98-listed chemicals and related compounds with the androgen receptor based on an in vitro reporter gene assay and 3D-QSAR. Bioorganic & Medicinal Chemistry 14(21), 7160-7174.

Tong W, Xie Q, Hong H, Shi L, Fang H & Perkins R (2004). Assessment of Prediction Confidence and Domain Extrapolation of Two Structure–Activity Relationship Models for Predicting Estrogen Receptor Binding Activity. Environmental Health Perspectives 112, 1249-1254.

Vedani A, Dobler M & Lill MA (2005). Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor†. Journal of Medicinal Chemistry 48(11), 3700-3703.

Vedani A & Smiesko M (2009). In silico toxicology in drug discovery - concepts based on three-dimensional models. Alternatives to Laboratory Animals 37(5), 477-496.

Vedani A, Smiesko M, Spreafico M, Peristera O & Dobler M (2009). VirtualToxLab - in silico prediction of the toxic (endocrine-disrupting) potential of drugs, chemicals and natural products. Two years and 2,000 compounds of experience: a progress report. ALTEX 26(3), 167-176.

Vinggaard AM, Niemelä J, Wedebye EB & Jensen GE (2008). Screening of 397 Chemicals and Development of a Quantitative Structure• Activity Relationship Model for Androgen Receptor Antagonism. Chemical Research in Toxicology 21(4), 813-823.

Zhao CY, Zhang RS, Zhang HX, Xue CX, Liu HX, Liu MC, Hu ZD & Fan BT (2005). QSAR study of natural, synthetic and environmental endocrine disrupting compounds for binding to the androgen receptor. SAR and QSAR in Environmental Research 16(4), 349 - 367.

European Commission

EUR 24522 EN – Joint Research Centre – Institute for Health and Consumer Protection
Title: Review of QSAR Models and Software Tools for Predicting Developmental and Reproductive Toxicity
Author(s): Elena Lo Piparo and Andrew Worth
Luxembourg: Publications Office of the European Union
2010 – 24 pp. – 21 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-16702-7
doi:10.2788/9628

Abstract

This report provides a state-of-the-art review of available computational models for developmental and reproductive toxicity, including Quantitative Structure-Activity Relationship (QSARs) and related estimation methods such as decision tree approaches and expert systems. At present, there are relatively few models for developmental and reproductive toxicity endpoints, and those available have limited applicability domains. This situation is partly due to the biological complexity of the endpoint, which covers many incompletely understood mechanisms of action, and partly due to the paucity and heterogeneity of high quality data suitable for model development. In contrast, there is an extensive and growing range of software and literature models for predicting endocrine-related activities, in particular models for oestrogen and androgen activity. There is a considerable need to further develop and characterise *in silico* models for developmental and reproductive toxicity, and to explore their applicability in a regulatory setting.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-24-522-EN-C

Publications Office

ISBN 978-92-79-16702-7