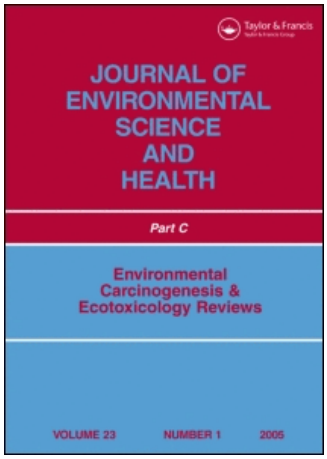

Journal of Environmental Science and Health, Part C
Environmental Carcinogenesis and Ecotoxicology Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597270

Directions in QSAR Modeling for Regulatory Uses in OECD Member Countries, EU and in Russia

NATALJA FJODOROVA [a]; MARJANA NOVICH [a]; MARJAN VRACHKO [a]; VJACHESLAV SMIRNOV [b]; NINA KHARCHEVNIKOVA [c]; ZOYA ZHOLDAKOVA [c]; SERGEI NOVIKOV [c]; NATALJA SKVORTSOVA [c]; DMITRII FILIMONOV [d]; VLADIMIR POROIKOV [d]; EMILIO BENFENATI [e]
[a] National Institute of Chemistry, Ljubljana, Slovenia
[b] Regional Hygienic and Toxicology Information Center "TOXI", Institute of Toxicology, (Ministry of Health Protection of Russian Federation), S-Petersburg, Russia
[c] A.N.Sysin Research Institute of Human Ecology and Environmental Health RAMS (Russian Academy of Medical Science), Moscow, Russia
[d] Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Moscow
[e] Istituto di Ricerche Farmacologiche, Mario Negri, Milan, Italy

Online Publication Date: 01 April 2008

To cite this Article: FJODOROVA, NATALJA, NOVICH, MARJANA, VRACHKO, MARJAN, SMIRNOV, VJACHESLAV, KHARCHEVNIKOVA, NINA, ZHOLDAKOVA, ZOYA, NOVIKOV, SERGEI, SKVORTSOVA, NATALJA, FILIMONOV, DMITRII, POROIKOV, VLADIMIR and BENFENATI, EMILIO (2008) 'Directions in QSAR Modeling for Regulatory Uses in OECD Member Countries, EU and in Russia', Journal of Environmental Science and Health, Part C, 26:2, 201 — 236

To link to this article: DOI: 10.1080/10590500802135578
URL: http://dx.doi.org/10.1080/10590500802135578

Journal of Environmental Science and Health Part C, 26:201–236, 2008

ISSN: 1059-0501 (Print); 1532-4095 (Online)
DOI: 10.1080/10590500802135578

Taylor & Francis
Taylor & Francis Group

Directions in QSAR Modeling for Regulatory Uses in OECD Member Countries, EU and in Russia

Natalja Fjodorova,[1] Marjana Novich,[1] Marjan Vrachko,[1] Vjacheslav Smirnov,[2] Nina Kharchevnikova,[3] Zoya Zholdakova,[3] Sergei Novikov,[3] Natalja Skvortsova,[3] Dmitrii Filimonov,[4] Vladimir Poroikov,[4] and Emilio Benfenati[5]

[1]National Institute of Chemistry, Ljubljana, Slovenia
[2]Regional Hygienic and Toxicology Information Center "TOXI", Institute of Toxicology, (Ministry of Health Protection of Russian Federation), S-Petersburg, Russia
[3]A.N.Sysin Research Institute of Human Ecology and Environmental Health RAMS (Russian Academy of Medical Science), Moscow, Russia
[4]Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Moscow
[5]Istituto di Ricerche Farmacologiche, Mario Negri, Milan, Italy

The aim of this article is to show the main aspects of quantitative structure activity relationship (QSAR) modeling for regulatory purposes. We try to answer the question; what makes QSAR models suitable for regulatory uses. The article focuses on directions in QSAR modeling in European Union (EU) and Russia. Difficulties in validation models have been discussed.

Key Words: biokinetic modeling; computational toxicology; expert systems; human health effects; metabolism; prediction; quantitative structure-activity relationships (QSARs); registration, evaluation and authorization of chemicals (REACH); regulatory agencies; risk assessment; structure-activity relationships (SARs); toxicity

INTRODUCTION

It is estimated that over 30,000 industrial chemicals used in Europe require additional safety testing to meet requirements of the new chemical regulation REACH (registration, evaluation and authorization of chemicals). If conducted

Received December 12, 2007; accepted April 15, 2008.
Address correspondence to Dr. Natalja Fjodorova, National Institute of Chemistry, Ljubljana, Slovenia. E-mail: natalja.fjodorova@ki.si

on animals, this testing would require the use of an extra 10–20 million animal experiments. Quantitative structure activity relationships (QSAR) is one major prospect between alternative testing methods to be used in a regulatory context.

Development of QSARs is solving several problems:

- Ethical saving of animal lives;
- Economical cost reduction on testing;
- Political implementation of new chemical regulation REACH.

In the context of REACH and the Cosmetics Directive (Council Directive 2003/15/EC), it is anticipated that (Q)SARs will be used more extensively, in the interests of time- and cost-effectiveness and animal welfare.

Many different (Q)SAR models for prediction of properties relevant for chemical management exist and have been published in the literature. However, most of them are poorly described in terms of the five principles for validation of (Q)SAR models, which have been adopted by Organization for Economical Cooperation and Development (OECD).

The aim of this article is to show the main aspects of QSAR modeling for regulatory purposes. We try to answer the question: what makes QSAR models suitable for regulatory uses? The article focuses on directions in QSAR modeling in European Union (EU) and Russia. Difficulties in validation models have been discussed.

1. QSAR MODELING FOR REGULATORY USES IN OECD MEMBER COUNTRIES AND IN EU

1.1 Definition of (Q)SAR

Structure-activity relationships (SARs) and quantitative structure-activity relationships (QSARs) are theoretical models that can be used to predict the physicochemical and biological (e.g., toxicological) properties of molecules from knowledge of chemical structure. A SAR usually represents an association between a chemical substructure and the potential of a chemical containing the substructure to exhibit a certain biological effect. A QSAR quantitatively relates the properties of a chemical (encoded in its chemical structure) to a physical property or to a biological effect (e.g., a toxicological endpoint).

QSARs are all quantitative models yielding a continuous or categorical result. The most common techniques for developing QSARs are regression analysis, neural nets, and classification methods. SARs are qualitative relationships in the form of structural alerts that incorporate molecular substructures or fragments related to the presence or absence of activity.

(Q)SARs for human health endpoints and certain eco-toxicological endpoints can be regarded as alternative methods to animal experiments since they could be used to replace or reduce animal testing (1).

The development of predictive models is intended for application in chemical management, including priority setting, risk assessment and classification and labeling (2).

1.2 Scientific and Regulatory Uses of (Q)SARs

From the scientific perspective, (Q)SARs can be developed for prediction of the following types of property or effects:

1. Physicochemical properties;
2. Toxic potential and potency;
3. Environmental distribution and fate;
4. Biokinetic processes (absorption, distribution, metabolism and excretion).

Regulatory uses of QSARs include:

1. Supporting priority setting of chemicals;
2. Guiding experimental design of regulatory tests or testing strategies;
3. Providing mechanistic information;
4. Grouping chemicals into categories based on similarity;
5. Filling in a data gap needed for classification and labeling;
6. Filling in a data gap needed for risk assessment.

It must be emphasized that principles and procedures for scientific validation of (Q)SARs are separate from the considerations and procedures necessary for regulatory acceptance (3).

1.3 Organizations Involved with Validation of QSARs

It was considered necessary to develop a framework for the independent development, validation and dissemination of QSARs. The European Commission's Joint Research Centre (JRC) is a suitable organization to coordinate such a framework, due to its recognized independence from national and sectoral interests, and its established role in the provision of scientific and technical support for the development and implementation of EU legislation on chemicals.

The JRC established the *QSAR Action* as a project within the JRC Work Programme to coordinate and expedite activities in the area of chemical-grouping, SARs, and QSARs approaches with potential regulatory use.

The European Chemicals Bureau (ECB) is coordinating the JRC Action on Computational Toxicology (including QSARs), which aims to promote the development, validation and implementation of computational models that are useful for regulatory purposes (4). The work involves collaboration with internal partners, such as European Center for the Validation of Alternative Methods (ECVAM), and external partners, such as the OECD.

1.4 Guidelines and Documents for Developing and Application QSARs for Regulatory Uses

Preliminary guidance "The Characterization of (Quantitative) Structure-Activity Relationships" has been published in (5).

The following *endpoints* associated with EU Test Methods and OECD test guidelines have been proposed:

- *Physicochemical properties* such as melting point, boiling point, vapor pressure, K octanol/water partition coefficient, Koc organic carbon/water partition coefficient, water solubility;
- *Ecological effects* such as acute fish, long-term toxicity, acute Daphnid, algal, terrestrial toxicity;
- *Environmental fate* such as biodegradation, hydrolysis in water, atmospheric oxidation. bioaccumulation;
- *Human health effects* such as acute oral, acute inhalation, acute dermal, skin irritation, eye irritation, skin sensitization, repeated dose toxicity, genotoxicity (*in vitro*, bacterial cells), genotoxicity (*in vitro*, mammalian cells), genotoxicity (*in vivo*), reproductive toxicity, developmental toxicity, carcinogenicity.

Guidelines for developing and using QSARs with examples of models for prediction toxicity was published in (6). Regulatory uses and applications of (Q)SAR models in the assessment of new and existing chemicals in OECD member countries was reported in (7). Principles of validation of (Q)SARs was published in (8).

The general acceptability criteria or validation principles of (Q)SARs for Human Health and Environmental Endpoints was developed at the workshop "Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints," hosted by the European Centre for Ecotoxicology and Toxicology of Chemicals and organized by the International Council of Chemical Associations (ICCA) and the European Chemical Industry Council (CEFIC) held 4–6 March 2002 in Setubal. In November 2004, at the 37th Joint Meeting of Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology, the OECD Member Countries and the European Commission adopted *five*

principles for the validation of (Q)SARs intended for use in the regulatory assessment of chemicals.

Accordant with these principles a (Q)SAR model for regulatory use should be associated with the following information:

1. A defined endpoint;
2. An unambiguous and easily applicable algorithm;
3. A defined domain of applicability;
4. Appropriate measures of goodness-of–fit, robustness and predictivity;
5. A mechanistic interpretation, if possible.

The international agreement on a set of validation principles was important, not only to provide regulatory bodies with a scientific basis for making decisions on the acceptability of data generated by (Q)SARs, but also to promote the mutual acceptance of (Q)SAR models by improving the transparency and consistency of (Q)SAR reporting (9).

Recently, the EC funded project DEMETRA addressed the specific case of the QSAR models for the European legislation for pesticides (10).

Several major points have been defined. The QSAR model should clearly mention the specific legislation involved. This point should be more extensively considered in all QSAR models for regulatory purposes because different regulations have different ways to address and express the phenomenon. Related to the legislation, in many cases there are specific guidelines that have to be considered. In some legislations, these guidelines are strict, while in others, a certain degree of freedom is given. The developer should be aware of this fact.

DEMETRA addressed some other issues, not directly defined within the OECD principles, which can be used to evaluate QSAR models for regulatory purposes. For instance, DEMETRA dedicated efforts on the definition of the model utility for regulatory purposes, in order to identify the QSAR models which should be more useful for legislation. DEMETRA defined tools to address and reduce false negatives. False negatives are very important in case of QSAR models for regulatory purposes because regulators want to avoid predictions that predict safety chemicals which are toxic; the reverse case is not so critical. DEMETRA addressed the situation of model uncertainty in a detailed way. For regulatory purposes, it is not enough to have a predicted value: its uncertainty has to be characterized, and relation to its use and the uncertainty of the input values. All these points have been thoroughly addressed and discussed within DEMETRA (10).

1.5 QSARs for Human Health and Environmental Endpoint

In the mini-monograph, Cronin et al. (11) reported the use of QSARs in international decision-making frameworks to predict ecological effects and environmental fate. QSARs for prediction health effects of chemical substances are presented in another mini-monograph (12).

Most expert systems, SARs, and QSARs are based on chemical classes or on mode of action. More details on in silico methodologies with regard to their usage in REACH is provided in report on the "Review of the Status of the Development of Alternatives to using Animals in Chemical Safety Testing and Identification of New Areas for Development or Research in the Context of the Proposed REACH Regulation" (13).

The comprehensive investigation of quantitative methods of hazard characterization used in food safety assessment and used for regulatory decision-making in Europe was reported in monograph (14).

1.6 Difficulties to Validation of QSARs

It should be noted that there are many practical difficulties to the validation of (Q)SARs, in particular obtaining data for a meaningful external validation, as well as obtaining transparent models for some methodologies (e.g., commercial expert systems, neural networks, etc.).

There are three main reasons why (Q)SARs and expert systems have not been used to their full potential:

1. None have yet been formally validated;
2. They need to be improved to cover a wider spectrum of toxic mechanisms of action, especially for endocrine disruption and non-genotoxic carcinogenesis (that are both based on receptor-binding);
3. Their coordinated and combined use has not been explored sufficiently (15).

1.7 OECD's Database on Chemical Risk Assessment Models

Models (computerized or capable of being computerized) that are used by OECD countries to predict health or environmental effects, exposure potential, and possible risks were organized into searchable database. But it should be taken into account that this database is created for developmental use and the methods described there have not been evaluated or validated by OECD; no endorsement of the methods by OECD should be inferred by the inclusion of certain methods in this database. This database is intended as an information resource only. The models are listed by countries and by property or effect included.

Screening level methods described there are useful, when chemical-specific data are lacking, for establishing priorities for chemical evaluation and for identifying issues of potential concern (16). Table 1 presents information included in models.

Table 1: Information included in models of OECD's database

Exposure/risk models for predicting human health or environmental exposure potential and potential environmental, worker or consumer risk	
Areas of assessment	**Human health, environment**
Human health	
Exposure covered	• Indirect human exposure via the environment • Consumer product exposure • Worker exposure
Routes of exposure covered	• Inhalation • Ingestion • Dermal • Multi-media
Environment	
Organisms covered	• Freshwater organisms • Marine organisms • Sediment organisms • Terrestrial organisms • Micro-organisms in sewage treatment plant • Fish-and-worm eating predators
Pathways of exposure covered	Air, water, sediment, soil, biota, sewage treatment plant, multi-media
Type of information provided	Daily intake, potential dose, margin of safety, predicted environmental concentration, risk quotient (predicted environmental concentration/predicted no-effect concentration)
Health or environmental effects models for predicting physical/chemical properties, chemical and fate properties, and human and aquatic hazard effects	
Category of information provided	
Physical/chemical properties • Melting point, boiling point, vapor pressure • Octanol-water partition coefficient (KOW) • Water solubility • Organic carbon adsorption coefficient (KOC)	
Environmental fate properties • BCF (bioconcentration factor) • AOP (atmospheric oxidation potential) • Biodegradation • Hydrolysis • Percent removal in wastewater treatment	
Hazard-human health • Mutagenicity • Neurotoxicity • Reproductive toxicity • Developmental toxicity • Systemic toxicity • Skin/eye irritation • Oncogenicity	Hazard-environmental • Aquatic biota • Terrestrial biota

(Continued on next page)

Table 1: Information included in models of OECD's database *(Continued)*

Type of Information Provided • Qualitative • Quantitative • Range • Point estimate • Categorical information	Endpoint(s) • Reproduction • Growth Mortality
Species/compartment addressed by model • Air • Water • Sediment • Soil • Multi-media • Aquatic biota • Terrestrial biota	Model approach • Deterministic or probabilistic • QSARs • SARs

1.8 QSARs Based on Metabolism and *In Vitro* Data

Several *in silico* systems for predicting metabolism are available, including (Q)SAR models and expert systems, but none of these have been compared extensively for their relative performances, and none have been formally accepted for regulatory use, although some models can be used to provide supporting information in chemical risk assessments. There is currently no consensus on how *in silico* models for predicting biotransformation should be validated. Also, a variety of systems are in different stages of development, assessment and validation. If they are to be of more practical use outside the pharmaceutical industry for regulatory testing, then further research needs to be undertaken to make them more amenable for a wider range of chemicals. Problems with regard to the availability of good quality data for benchmarking purposes, apply to techniques for using *in silico* prediction systems and biokinetic models to assess the metabolic fate of chemicals after uptake by different routes of exposure in different species. Before these systems can be validated, more chemicals with good quality data need to be found, for use as test sets. Nevertheless, it was agreed that at least one biokinetic modeling system is, in principle, ready for more formal consideration for validation. Clearly, *in silico* systems for predicting toxicity should take account of the possibility that biotransformation could modulate toxicity. This could be achieved by modifying these systems, so that they can model the toxicity of the principal metabolites of chemicals, or by linking them with systems specifically designed to predict metabolite formation (17).

An overall scheme for predictive toxicity testing has been discussed in the manuscript (18). It has been emphasized that (Q)SAR analyses are used in conjunction with expert system and biokinetic modeling, and information on metabolism and identification of the principal metabolites in humans. Several recommendations are made, the most important of which is that the European Union (EU) should actively promote the improvement and validation of

(Q)SAR models and expert systems, and computer-based methods for biokinetic modeling.

In the monograph (15), it was highlighted that if mechanistically-based toxicokinetic and toxicodynamic data are obtainable, risk characterization can be improved considerably. This is illustrated by physiologically-based toxicokinetic modelling (pBTK models), which can be used at various stages of risk assessment.

1.9 Perspectives in QSAR Modeling for Regulatory Use

Thousands of predictive models have been published in recent years, but typically they are not suitable for regulatory purposes because they have not taken into account essential factors for validation or quality assurance and specific requirements for regulation.

The project CAESAR (computer assisted evaluation of industrial chemical substances according to regulations) ongoing in the scope of FR6 Six Framework Programme will develop (Q)SAR models as non-animal alternative tools for assessment of chemical toxicity under REACH. CAESAR will include the high quality factors that are needed to make the use of (Q)SARs acceptable for regulatory purposes (such as the implementation of the REACH proposal) for the prediction of the toxicity of chemical substances in a transparent manner by applying new and unique modeling and validation methods. Five endpoints will be addressed within CAESAR, chosen on the basis of the animal use that is expected for the REACH legislation. In order to have high quality data sets, data have been selected from high quality sources, and structures checked independently by at least two groups in the consortium. Preliminary results on a model for the bioconcentration factor are superior to those previously published. The predictions of properties together with all modeling details can be easily used in chemical regulation. The CAESAR project goal is to design and develop a web site incorporating the models developed. This site will be freely accessible and (Q)SAR models and protocols will be available for non-commercial use (18).

2. PREDICTION METHODS FOR REGULATORY USES AND DIRECTIONS IN QSAR MODELING IN RUSSIA

2.1 Computational Methods in Toxicology to Assess Health Effects from Exposure to Hazardous Substances

Computational Toxicology Methods Overview

The development of reliable computational methods for determination of toxicity of chemical compounds is an intricate process that utilizes knowledge

from many different scientific disciplines, including toxicology, chemistry, mathematics and combination of listed sciences like chemometrics.

An outstanding Russian scientist toxicologist, professor Nikolay Vasilievich Lazarev, introduced the computational methods in toxicology in Russia (19).

Prediction of toxicological properties of substances can be performed using different physicochemical parameters. The description of computational methods for regulatory uses is reported in monograph (20).

Methods for substantiation, determination and calculation of maximum allowable concentrations and tentative safe levels for different media (air, water, soil) are given in guidelines called Methodical Instructions. Environmental and occupational exposure levels (standards, norms) for chemicals in Russia are approved by the Chief State Sanitary Doctor and are published as legislative rules in official documents named Hygienic Norms (HN). Maximum allowable concentrations (MAC), tentative safety exposure level (TSEL), tentative safety levels (TSL) and tentative permissible levels (TPL) of chemicals in various environmental media are accepted as a permanent or temporary safe exposure levels. Temporary norms are usually established for certain period of time (for example, 2 years).

The determination of temporary norms or safety limits in different media (air, water, soil) is based on calculation methods using regressions equations (21–24).

In the monograph, Smirnov et al. (20) computational methods for sanitary NORMs are divided into 3 groups:

1. On the basis of physico-chemical properties of substances;
2. Establishment of safety levels by toxicological parameters from short terms experiments;
3. On the basis of safety levels found out in different media. In this case a data transformation obtained in one media is done, applying the data to another one.

The first step in risk assessment is estimation of toxicity parameters of acute toxicity such as $\mathbf{LD_{50}}$, $\mathbf{LC_{50}}$ (lethal dose and concentration) values, **Lim-** (threshold of hazardous effect of substances) and others. Safety levels (concentrations or doses) of substances used for regulatory purposes are calculated on the bases of $\mathbf{LD_{50}}$, $\mathbf{LC_{50}}$ or **Lim**.

The calculation of $\mathbf{LD_{50}}$ or $\mathbf{LC_{50}}$ is divided into several sections:

1. For volatile organic compounds with boiling point below 200°C ($t \leq 200$°C).
2. For low-volatile and non-volatile organic compounds with boiling point above 200°C ($t \geq 200$°C)
3. For inorganic compounds of metals (oxides and salts).

The value of biological endpoint such as toxicity, strongly depends on the aggregation state of substances (solid, liquid, gas) and route of administration (oral, dermal, inhalation).

Thus, equations of regressions have been composed with $\mathbf{LD_{50}}$ or $\mathbf{LC_{50}}$ as a response. As dependent variables the following physical and chemical properties of chemicals have been used:

M—molecular mass;
d—density (g/cm^3);
RD—mole refraction;
$\mathbf{t^{\circ}_{boiling.}}$—boiling point (°—);
$\mathbf{t^{\circ}_{melting.}}$—melting point (°—);
$\mathbf{C_{20}}$—maximum saturated concentration of substances in the air at 20°;
P—pressure of vapor at 20°— (mm, millimeter of mercury);
S—dissolubility in water (g/l);
K—coefficient of distribution oil/water;
M.o.—molecular volume (M/d);
mM—millimole;
nD—refraction coefficient;
$\mathbf{t_{flash}}$—flashing point (°—);
μ—dipole moment (debye);
Σ_α—sum of increments of nuclear quadrupole resonance (NQR);
Σ_σ—sum σ constant of Hammet.

In some cases, transformation of data from one species to another one or from animal to human has been done.

Safety levels or NORMs can be found on the basis of the acute toxicity values. Numerous regression equations have been composed.

Prediction of TPLs in Air

Parameters such as molecular weight (MW), boiling point and melting point are used for prediction of safety levels in ambient air. Equations obtained for non-congeneric sets of substances have been proposed. For organic substances it was suggested to use the next equation:

$$\log \mathrm{TSL} = -8.0 \log \mathrm{M} + 14.75 + \mathrm{K},$$

where K is the correction factor dependent on MW, M is the molecular mass, and TSL is the tentative safety level.

The values of correction factor depending on MW are presented in Table 2 (20).

TSLs in Air for Polybromo- and Chlorobenzenes using Electronic Parameters

The selection of substances with certain structure types suitable for creating predictive models for safety levels mainly determined accidently for

Table 2: Correction factor values depending on MW

K—correction factor	MW—molecular weight
$K = -3$	$MW < 45$
$K = -1$	$45 \leq MW \leq 70$
$K = 0$	$70 < MW \leq 146.9$
$K = 1$	$147 \leq MW \leq 199.9$
$K = 2$	$200 \leq MW \leq 265$
$K = 3$	$MW \geq 265$

example by scientific interest of researchers to a selected group of substances. Thus, guidelines for development and establishment of TSL in ambient air contain equations for prediction of safety level for polybromo- and polychlorbenzene using electronic parameters is calculated by Hukkel method.

$$\log \mathrm{TSL} = -6.33 + 17.04|\mathrm{Q_{max}}| - 16.20|\Delta \mathrm{Q}| + 12.24\mathrm{N_{max}},$$

where $\mathrm{Q_{max}}$ is the maximum charge of carbon atom of benzene ring not connected with substitute, $\Delta \mathrm{Q}$ is the difference of sums of charge for atoms of benzene ring of chlorine or bromine benzenes and their substituted derivatives, and $\mathrm{N_{max}}$ is the maximum index of free valency of benzene ring atoms.

TSLs in Air for Polybromo- and Polychlorobenzenes using Acute Toxicity Value

The prediction of safety levels can be done using acute toxicity value. For polybromo- and polychlorobenzenes the following equation has been obtained:

$$\log \mathrm{TSL} = -8.0 \log \mathrm{LD_{50}} - 4.72$$

The multiple linear regression (MLR) equations have been suggested for prediction of $\mathrm{LD_{50}}$ of new substances or poorly known chemicals of the polybromo- and polychlorobenzenes series:

$$\log \mathrm{LD_{50}} = 3.34 - 0.25 - 0.25\mu - 0.33|\Sigma\sigma|$$

$$\log \mathrm{LD_{50}} = 3.34 - 0.25 - 0.22\mu - 0.52|\Sigma\sigma| - 0.0021|\Sigma\alpha|$$

$$\log \mathrm{LD_{50}} = 3.69 + 0.0003\mathrm{M} + 0.22\mu - 0.0003\mathrm{t_{Melting}} - 0.22|\mu - 0.53|\Sigma\sigma|$$

Where μ is the dipole moment, $\Sigma\sigma$ is the sum of Hammet constants, and $\Sigma\alpha$ is the sum of increments of nuclear quadrupole resonance of ^{35}Cl and ^{79}Br nucleus.

TSLs for analogs of homologous series substances

For analogs within homologous series substances, it was suggested to use the equation based on assumption of additivity of contribution of chemical bonds to biological activity

$\mathrm{TSL} = 1000/\Sigma \mathrm{l}_i$, where $\Sigma \mathrm{l}_i$ is the additivity of chemical bonds.

In the guidelines for establishment of tentative safety levels for hazard substances in the workplace air, the following equations are recommended for calculation of TSL for halogenated toluenes containing halogen atoms in methyl group or benzene ring.

In the case of *irritating substances* the following equation has been proposed:

$$\log(1/\mathrm{TSL}) = -2.4 + 0.903\mathrm{E_{DA}}$$

while for toxic substances the equation below is applied:

$$\log(1/\mathrm{TSL}) = -0.2 + 0.027\mathrm{E_{DA}},$$

where E_{DA} is the parameter characterizing strength of bond with receptor and can be calculated with the equation (hal = number of halogen atoms in the molecule; OCC = number of occupied MO in the molecule):

$$\mathrm{E_{DA}} = \sum_{i}^{\mathrm{hal}} \sum_{j}^{\mathrm{occ}} c_{ij}^2/(\varepsilon_r - \varepsilon_i),$$

where c_{ij} = coefficients for p-AO atom orbital of halogens in occupied MO, ε_r energy of (p-type) acceptor level of receptor, and ε_i is the energy of occupied MO of substance.

Equations were obtained on small samples, and coefficients of correlation are not cited. However the coefficients of correlation for analogues equations for prediction of $LD_{50}$for the same group of substances are equal to 0.95–0.97 (25).

Prediction of TPLs in Water

Computational methods for prediction of TPLs in water are based on regression equations with one parameter such as physicochemical descriptors (electronic, hydrophobic or steric) or empirical parameters (solubility, melting point, boiling point, spectroscopic descriptors, etc.). Table 3 contains parameters used in regression equations for different groups of substances listed in (26).

Table 3: Parameters used in regression equation for different groups of substances

Group of substances	Parameters use in regression equation
Substitute benzenes	μ-dipole moment
Amine and amide substances	MW-molecular weight
Oxygen-containing substances	Solubility in water
Solid hydrocarbon	Melting point
Liquid organic compounds and gases (with exception of acids and monohydric alcohols)	Boiling point

It must be taken into account that guidelines in Russia didn't contain statistical characteristics of the reported equations (sample size, coefficient of correlation, standard error, F-value). For this reason, the new models for acute toxicity prediction and for calculation of safety levels have been developed, which include statistical characteristics like: **n** = number of chemicals, **r** = correlation coefficient, and **s** = standard error, etc. QSAR techniques are used to estimate the toxicity of poorly characterized substances based on comparisons to well-studied substances having similar chemical structures.

In this article, we focus on some aspects of QSAR modeling in Russia. Many QSAR models have been developed for regulatory use, and establishment of safety levels in different media as described below.

2.2 QSAR Methods used in Russia

Hansch and Free Wilson Methods

Between methods used in Russia we mention Hansch analysis (investigation of the quantitative relationship between the biological activity of a series of compounds and their physicochemical substituent or global parameters representing hydrophobic, electronic, steric and other effects using multiple regression correlation methodology) and Free-Wilson (FW) analysis (a regression technique using the presence or absence of substituents or groups as the only molecular descriptors in correlations with biological activity).

For structural series of phenols, a regression equation was reported for acute toxicity in rats in case of oral administration (n = 52, r = 0,887). For MAC in air of work zone regression equations have been reported (n = 15, r = 0,907) using the sum of Hammett electronic substituent constant, reflecting the electron-donating or -accepting properties of a substituent (27). For the same group of substances, 23 phenols were selected and data collected from chronic toxicity experiment in rats in case of oral administration. A correlation was found between MNED (maximum not effective dose) and logarithm of octanol-water partition coefficient P, sum of Hammett substituent constants and *Free-Wilson indices* for the fragments NO_2, CH_3 and Cl. The statistical characteristics in this example were n= 23, r = 0.909, s = 0.577 (28).

Later topological descriptors were used for prediction models (29). The molecular topology considers that biological activity is related to the molecular topological characteristics, numerically represented using the distance and connectivity indices.

Pattern Recognition Technique

Pattern recognition is the identification of patterns in large data sets, using appropriate mathematical methodology. Examples are principal component

analysis (PCA), SIMCA, partial least squares (PLS) and artificial neural networks (ANN).

A computer-based system was developed for calculation of LD_{50} for drugs on the basis of structural descriptors. Accuracy of prediction shows that in 91% cases calculated values don't differ from experimental by more then 3 times (30).

SAR Study

A model predicting toxicity level of organic substances by analyzing their structural fragments was developed (31). Such a model determines hazardous classes of substances through the MAC value. Substance with known priori probability can be ranked in one of the four possible classes in relation to the presence of some sets of fragments in the structure.

Hydrogen Bond Thermodynamics (HYBOT) Descriptors

Many program packages have been elaborated by Raevsky et al. They suggested to use hydrogen bond thermodynamics (HYBOT) program for the estimation of hydrogen bonds strength (32, 33) and developed the program package molecular transform analysis (MOLTRA) (by Raevsky, Sapegin, Zefirov), the QSAR discriminant-regression model conformational analysis (CONFAN), dissociation constants (DISCO), and others. They also proposed a program for calculating solubility, lipophilicity, and liposome penetrability (34). Most of these programs are applied for pharmaceutical needs, but HYPOT descriptors have a wide spectrum of application. Predictive models of aquatic toxicity of environmental pollutants with different mechanisms of action were developed on the basis of molecular similarity and HYBOT descriptors.

The molecular polarisability and hydrogen bond descriptors for the chemicals of interest and related compounds have been used to calculate any additional contribution in toxicity by means of linear regression relationships. Final comparison of calculated and experimental toxicity values gave good results, with standard deviation close to the experimental error (35).

The software program SLIPPER-2001 for prediction of the lipophilicity (log P), solubility (log Sw), and oral absorption of drugs in humans (FA) has been developed. It is based on structural and physicochemical similarity. Reliable results were obtained for simple compounds, for complex chemicals, and for drugs. Thus, the principle of "similar compounds display similar properties" together with estimating incremental changes in properties by using differences in physicochemical parameters results in "structure-property" predictive models, even in the absence of a precise understanding of the mechanisms involved (36).

Prediction by Infrared Spectrum

A computer-based system SPECTR for prediction of acute toxicity of pesticides and intermediate products in the process of their manufacturing was

developed, using infrared spectrum (37). The prediction is performed using the method of closest neighbors in the space of spectrum features.

Another method was developed for acute toxicity prediction on the basis of geometrical, topological, and quantum-chemical characteristics by means of modified method of potential function and principle components analysis. The accuracy of recognition was high enough, between 77 to 96% depending on the structural series. The research was made for derivatives of thiazol, thiazolidine and 1,3,4-thiazolidine (38). Logical structural approach for prediction of acute toxicity hazardous class was also implemented (39, 40).

Neural Networks

Multi-level neural networks were used for prediction of toxicity, biological activity, mutagenicity and carcinogenicity and were described by authors from Laboratory of Organic Chemistry in Chemistry Department of M. V. Lomonosov Moscow State University (MSU) (41).

2.3 QSARs Based on ADME and *In Vitro* Data

Pathogenetic Model of Intoxication

Pathogenetic model of intoxication was introduced by Zoldakova. For prediction of toxicity parameters, a regression equation was used based on spline approximation. Different stages of mechanism of toxicological action were systematized, and physico-chemical indices characterized each stage (such as absorbtion, distribution, overcoming the biological barriers, interaction with ferments, included into cell's membrane and others active centers, transformation, cumulation, excretion) were applied (42).

The quantitative correlation for pathogenetical model can be expressed by equation:

$$\lg DE = A_0 + \sum_{i=1}^{m} \sum_{j=1}^{m} A_{i,j}\, x_i\, x_j + \sum_{i=1}^{m} A_i, x_i$$

where DE = effective dose of substance (LD_{50}, MAC, etc.), and x_i = index.

Reliable correlations between acute toxicity and chronic toxicity (threshold and safety levels) and physico-chemical parameters for 13 groups of congeneric chemicals have been obtained.

The use of bioactivation processes in prediction of toxicity parameters was also described in the following articles (43–45).

The hazardous class of substance in case of carcinogenicity of polycyclic aromatic hydrocarbons and haloginated aliphatic hydrocarbons was predicted using a logical combinatorial method and analysis of electronic parameters. Structural and numerical descriptors have been used (46).

QSARs Based on Bioactivation Processes of Metabolism

Structures and properties of single chemicals usually apply in QSAR models. However, the toxic effects are quite often determined by formation of metabolites in the processes of bioactivation of chemicals by various enzymes. Mechanism-based approach SARs for several toxic effects in various structural series have been developed (43, 44).

Due to complexity of processes of biotransformation of chemicals in biological systems, a single reaction of bioactivation cannot account for overall toxicity. Therefore, a logical-combinatorial method of automatic hypothesis generation was developed (47) based on the John Stuart Mills (JSM) logic. The JSM method enables one to predict some property and provide the explanation of this prediction. The prediction is based on the learning using the sets of positive and negative examples. The method does not require big training sets. The standard JSM method does not operate with numerical parameters, but only with chemical structure described by means of special descriptors named as functional code of substructures superposition (FCSS). A combined approach has been developed by Kharchevnikova, D'yachkov, Maksin, et al. They analyze not only the similarity of structures, but also the closeness of numerical parameters characterizing bioactivation process and/or stable metabolites (48–50).

A new approach has been developed for prediction of the most probable metabolic sites on the basis of statistical analysis of various metabolic transformations. It is related to the prediction of aromatic hydroxylation sites for diverse sets of substrates. Training is performed using the aromatic hydroxylation reactions from the metabolism database (Accelrys). Validation was carried out on heterogeneous sets of aromatic compounds reported in the metabolite database (MDL). The average accuracy of prediction of experimentally observed hydroxylation sites estimated for 1552 substrates from metabolite is 84.5%. The proposed approach is compared with two electronic models for P450 mediated aromatic hydroxylation: the oxenoid model using the atomic oxygen and the model using the methoxy radical as a model for the heme active oxygen species. For benzene derivatives, the proposed method is inferior to the oxenoid model and as accurate as the methoxy-radical model. It was shown that for hetero- and polycyclic compounds, the oxenoid model was not applicable, and the statistical method was the most accurate (51).

An approach based on the oxenoid model of monooxygenase action and semiempirical quantum chemical calculations was applied to the prediction of aromatic hydroxylation sites of cytochrome P450 substrates. The results were compared with experimental data on the metabolism in mammals and human from metabolite database (52).

Fundamental review about application of quantum chemistry for toxicology is given in the Russian magazine Toxicology (53). Knowledge of metabolic pathways of chemical can substantially enhance the accuracy of structure activity analysis.

Carcinogenicity Prediction Models

A version of logical-combinatorial JSM type intelligent system has been developed to predict the carcinogenicity class of untested chemicals. This version was based on the combined description of chemical substances including both structural and numeric parameters. The new version relies on the fact that the toxicity and danger caused by chemical substances often depends on their biological activation in the organism. The classification of chemicals was made according to their carcinogenic activity, for polycyclic aromatic hydrocarbons using a model of bioactivation via the formation of diolepoxides, and for the halogenated alkanes using a model of bioactivation via oxidative dehalogenation. The system is able to define the limit (boundary level) of an energetic parameter. The exceeding of this limit related to the inhibition of halogenated alkanes's metabolism and the absence of carcinogenic activity. Hazardous class can be predicted by the value of carcinogenic potential, number of affected organs (tissues) and the number of tumor bearing species (54).

Genotoxicity Prediction Based on Ames Test and Micronuclear Test Data

In 1974, Bruce Ames developed a bioassay performed on bacteria to assess the capability of environmental chemicals to induce genetic mutation. In genotoxicity prediction models as independent variable the results of battery of tests with strain of Salmonella typhimurium and several possible tests, for instance with activation of microsomal fraction of mammalian liver cells are used. An original ensemble of structural descriptors was used, indicating simultaneously presence in molecule activating and deactivating groups. Some chemicals that are known to cause cancer do not test positive in the Ames test, and some chemicals that test positive do not cause cancer. Nonetheless, the test is still considered an important part of assessing the safety of new chemicals. The test is useful as a screening tool for setting priorities because it is an inexpensive and quick way to help single out chemicals that should be targeted for further testing. It is also used in industry as a primary preventive approach to eliminate potential carcinogens early in the process of developing new commercial chemicals (55).

In Russia, the use of the Ames test is recommended in the course of stepwise substantiation of safety levels. These results must help the investigator to decide if it is necessary to carry out long-term experiments on carcinogenicity using laboratory animals. A disadvantage of the Ames test is the difference of bacterial and mammalian cells. The micronuclear test on genotoxicity is widely used in vitro toxicological tests and is carried out using cells of mammals. Usually the bone marrow cells are used. A new approach, the polyorgan micronuclear test, is used in Russia for the evaluation of mutagenic effects of chemicals (56). This test can be also used during multi-stage substantiation of safety levels of chemicals.

2.4 Computer Program PASS

Input and Output Content of PASS Program

Computer program PASS (prediction of activity spectra for substances) predicts simultaneously more than three thousand biological activities (main and side pharmacological effects, mechanisms of action, specific toxicities, biotransformations) (57–62).

PASS is based on the concept of biological activity spectrum of the compound, which must reflect all kinds of its biological activity resulting from the compound's interaction with biological entities. Since not one compound has been tested experimentally against all known kinds of biological activity, for any real compound known biological activity spectrum contains only part of such information. Biological activity spectrum for the compound under study predicted *in silico* with PASS can identify some additional kinds of biological activity, based on the structural similarity to the sub-sets of compounds, for which the appropriate activities were determined experimentally. Biological activities are described in PASS in qualitative mode ("active" and "inactive"), which provides the possibility of combing the heterogenous information collected from literature in the PASS training set.

Therefore, PASS predictions are based on the results of structure-activity relationships analysis accumulated in the SARBase, which is generated during the training procedure. Currently (PASS 2007 version), PASS training set includes the information on ~120000 biologically active compounds with ~5000 kinds of biological activity. These molecules are presented by the completely determined simply connected 2D structural formulae of uncharged molecules. The user can explore the existing SARBase, provided with PASS, or create his own SARBase using in house developed training set(s).

Since new information about biologically active compounds emerges constantly, continual updating of the existing PASS training set is performed. The first version of PASS (1995) was based on the data for ~10000 biologically active compounds with ~100 kinds of biological activity; in 1998 these figures came to ~30000 and ~500, respectively; in 2004 these figures came to ~57000 and ~1000, respectively; etc. (see Table 4).

Table 4: Computer program PASS versions

PASS versions presented by years	Amount of biologically active compounds	Number of biological activity types
1995	~10,000	~100
1998	~30,000	~500
2004	~57,000	~1000
2007	~120,000	~5000

Information about biologically active compounds is collected from papers and electronic sources and, after the experts' evaluation, is regularly added to the PASS training set.

Statistical Performance of PASS Algorithm

In parallel with the extending of PASS training set, PASS algorithm is also modified to provide more accurate results of prediction. The average accuracy of prediction estimated on the basis of leave one out cross-validation (LOO CV) for the whole training set and all predictable kinds of biological activity was ~78% in 1995, ~85% in 1998, and ~94% in the current version of PASS.

PASS 2007 version predicts ~3300 kinds of biological activity, while biologically active compounds from the PASS training set are described by ~5000 kinds of biological activity. However, some of these biological activities are represented by one or two compounds in the PASS training set, which is not enough to provide an accurate estimation of biological activity (three is the minimum number of compounds currently specified in PASS); also, for some kinds of biological activity accuracy of prediction in LOO CV procedure is less than 70%. Such kinds of biological activity are not included into the default list of PASS predictable activities.

Due to the unavoidable incompleteness of any training set, which can be used for biological activity spectra prediction, a robustness of the used algorithm is particularly important. By special computational experiments made with a set of about 20000 principal compounds from the MDDR database it was shown that, despite the random removal of up to 60% of structural or biological information, PASS algorithm still provides a reasonable accuracy of prediction (63).

Descriptors Used in PASS

Chemical descriptors used in PASS analysis, called Multilevel Neighborhoods of Atoms (MNA) are described in detail elsewhere (64). They are automatically generated on the basis of MOL-file of a molecule. The list of MNA descriptors currently consists of more than 52,000 different items. The new descriptors added to this list being found in a novel compound refreshing the training set.

During the prediction of biological activity spectra the number of new (in relation to the existing SARBase) descriptors is calculated for the compound under study, which provides the possibility of broad definition of PASS applicability domain. If the compound under study contains three or more new descriptors, the results of prediction give a rather crude estimaton of potential biological activity spectrum for this compound.

MNA descriptors are effectively utilized in SAR, QSAR and similarity analysis for drug-like compounds (64–71). Recently, new local integrative

descriptors (QNA) were proposed (72), which may provide some advantages in (Q)SAR/(Q)SPR analysis (73). On the basis of QNA descriptors and self-consistent regression computer program GUSAR (general unrestricted structure-activity relationships) is developed (76). For nine data sets, which contain the information about biological activity, toxicity and metaboloism for non-congeneric compounds, it was shown that GUSAR provides accuracy of prediction comparable to that of CoMFA, CoMSIA, GRID, HQSAR, EVA and 2D QSAR methods (74).

Prediction of Biotransformations

On the basis of PASS, the computer program for prediction of drug-like compounds biotransformations by enzymatic systems in the organism was developed. It was shown that PASS Biotrafo provides reasonable accuracy of prediction for about 1000 kinds of biotransformation (75). Also, it was demonstrated (77) that use of PASS-like algorithms allow to predict the aromatic hydroxylation sites with accuracy close or better to those of quantum chemical methods (78).

The Latest Results of PASS

One of the latest results concerning PASS was represented at the 12th International Workshop on Quantitative Structure-Activity Relationships in Environmental Toxicology (QSAR2006), 8–12 May 2006 in Lyon, France. It was shown that toxicity of chemical compound is a complex phenomenon that may be caused by its interaction with different targets in the organism. Two distinct types of toxicity can be broadly specified: the first one is caused by the strong compound's interaction with a single target (e.g., AChE inhibition); while the second one is caused by the moderate compound's interaction with many various targets. Since PASS predicts with reasonable accuracy several thousand kinds of biological activity based on the structural formula of chemical compounds, PASS provides an estimated profile of compound's action in biological space. Such profiles can be used to recognize the most probable targets, interaction with which might be a reason of compound's toxicity (79). Predicted biological activity spectra were used recently as a secondary variable for modeling of endocrine disruption profiles of xenobiotics (80).

PASS Availability in Internet

Since 2000, PASS predictions can be performed via Internet (62, 81, 82). One may obtain the results of PASS INet prediction by submitting the MOL file as an input data or drawing the molecule directly on the display using the MARVIN applet. For about 3000 registered users, this service is provided free of charge, and in 2007 alone, more than 70,000 molecules were submitted for prediction. A dozen papers were published by the independent researchers, in

which PASS predictions were later confirmed by the experiments (for the review see (83)).

2.5 Expert System SARET-TERA

SARET

Expert system SARET (structure-activity relationships for environmental toxicology) has been developed for quantitative analysis of structure-property (QSPR), structure-activity (QSAR) and property-property (QPPR) relationships and prediction of toxicity and environmental effects of chemical compounds. It was introduced by Prof. Sergey Novikov, MRC "MEDTOXECO", Department of General Hygiene, Moscow, Russia and by Prof. Vladimir Poroikov, IBMC RAMS, Moscow, Russia, http://www.ibmh.msk.ru (84–85).

The expert system SARET consists of

1. SARETbase—data bank that includes toxicological parameters of chemicals,
2. SARETmodel—special computer system for modeling and calculations,
3. Computer programs for calculation of descriptors (sub-structural, electronic, topological, etc.),
4. The integrated risk assessment program for determination of health hazardous of chemicals.

Input Content of SARET

SARETbase includes the information on more than 190 characteristics for 8500 substances: chemical structure, physico-chemical properties (density, boiling and melting points, partition coefficients of octanol/water, etc.), adverse effect doses and concentrations for acute and chronic exposure, odor thresholds in water and air, character of odor, some of threshold limit values for occupational and environmental exposure (air, water), etc. (86).

Output Content of SARET

SARETmodel is designed for statistical analysis of data and calculation of unknown parameters of substances on the basis of (Q)SARs. The application of SARET provides the information necessary to evaluate the hazard of chemicals and to estimate their unknown characteristics. Mathematical models for prediction of toxicological properties of chemicals have been developed. Maximum allowable concentrations for hazard substances in different environmental compartments (air, water, etc.) for different classes of chemical compounds have been calculated. The relationship between physicochemical properties and safe exposure limits have been studied. The new methods for prediction of

maximum allowable concentrations for air pollutants have been introduced. The distinguishing characteristics of biological activity of chemicals was taken into account.

SARET programs was written in DOS. Application of operation system Windows stimulated renovation of prediction programs and development of expert system TERA (tools for environmental risk assessment).

TERA

TERA is aimed at risk assessment of different pollutants. TERAbase is a part of expert system created by prof. S.M.Novikov and coauthors from the A.N.Sysin Research Institute of Human Ecology and Environmental Health of Russian Academy of Medical Sciences. TERA contains information useful for human, environmental and ecological risk assessment and management.

Input Content of TERA

TERA includes the information on approximately 200 characteristics for more than 13,000 chemical substances. The information collected in SARET and TERA is verified and specified on the basis of both Russian and foreign literature data including official documents, open publications, and "grey" literature. TERA contains information for 194 mixtures, 182 polymers, 346 dyes, 1080 non-organic compounds, 1407 remedies, 1260 agrochemicals (including pesticides). More than 1000 compounds contained in TERA are not presented in the Registry of Toxic Effects of Chemical Substances (RTECS).

TERA contains the following characteristics:

- Chemical structures and their codes (SMILES), the CAS and RTECs numbers;
- Physicochemical properties;
- Human health toxicity values (adverse effect doses and concentrations for acute and chronic exposure);
- Odor thresholds in water and air;
- Skin, eye irritating properties of substances;
- Threshold limit values for occupational and environmental exposure in different media such as maximum allowable concentration used in Russia, safe limits set by American Conference of Governmental Industrial Hygienists (ACGIH), Occupational Safety and Health Administration (OSHA), National Institute for Occupational Safety and Health (NIOSH) and risk assessment values such as Immediately Dangerous to Life and Health (IDLH);
- Target organs and systems;

- Characteristics of specific effects such as carcinogenicity, mutagenicity, teratogenicity, embryotoxicity, etc. Evaluation of carcinogenic potency is given in accordance with Russian classifications as well as those set by the following agencies and bodies: International Agency for Research on Cancer (IARC); National Institute for Occupational Safety and Health (NIOSH); Office of Environmental Health and Hazard Assessment (OEHHA); Occupational Safety & Health Administration (OSHA); American Conference of Governmental Industrial Hygienists (ACGIH); National Toxicology Program (NTP);
- Hazardous classes of chemicals according to international classifications;
- Epidemiological data and human health risk assessment;
- Toxicological properties of substances for different kinds of biosystems, etc.;
- Ecological effects such as acute fish, long-term toxicity, acute Daphnid, Alga, terrestrial toxicity.

Besides TERA contains more than 50 special databases, i.e., on cancer slope factors, the regional USA safety levels, reference doses (RfDs), reference concentrations (RfCs) from integrated risk information system (IRIS), the EPA superfund health effects assessment summary tables (HEAST), California Environmental Protection Agency (Cal EPA), etc. Exposure standards as defined by World Health Organization (WHO) and agencies and regulatory bodies of EU, Canada, Sweden and United States (US) are presented in TERA.

Output Content of TERA

TERA is integrated system which incorporates:

- Calculation of physical and chemical properties;
- Assessment of multi-domain risk;
- Assessment of carcinogenic potency risk;
- Prediction of lead concentrations in blood of fetus, children, adults (system LRISK);
- Health risk connected with lead exposure;
- Prediction of emission of chemical substances and there distribution in different media;
- Parameters used for setting priority of chemical substances in risk assessment;
- Risk assessment using epidemiological data;

- Health risk assessment of air pollutants;
- Health risk assessment of chemicals in case of emergency;
- Evaluation of industrial chemicals emission, etc.

TERA includes additionally biokinetic models taken from US Environmental Protection Agency (EPA) and simple risk assessment model CalTOX (CalEPA) that calculates the emissions of a chemical, the concentration of a chemical in soil, and the risk of an adverse health effect due to a chemical.

Continuous Development of TERA

TERA is continuously updating. The new substances structures and properties are inserting into database.

The main ongoing activities of TERA are listed below:

1. Development of new models for air pollutants emissions;
2. Improvement of predictive models on behavior and fate of chemicals in environment;
3. Improvement of predictive models on physicochemical toxicological properties of chemicals in relation to human exposure;
4. Health care costs calculation in case of exposure to harmful chemicals.

For regulatory use introduction of models for calculation of tentative safe exposure levels of unknown chemicals will be done (87, 88).

Chemical Substances Information Resources in Russia

Registers of Chemical Substanses in Russia

Registers of chemical substances in Russia are databases where chemicals are collected and registered.

The biggest registers are listed below:

- National register of pesticides and agrochemicals;
- National register of potentially hazardous chemical and biological substances;
- National register of human and veterinary medicinal products;
- National register on hazardous industrial sites;
- National register on waste disposal facilities and federal classification catalogue of wastes;
- Regional Toxicological Information Center “TOXI”, St-Petersburg.

Russian register of potentially hazardous chemical and biological substances in Moscow, Russia «Российский Регистр Потенциально Опасных Химических и Биологических Веществ" includes on-line "АРИПС" data base which contains information about 2977 substances (05.06.2007) and is available at http://www.rpohv.ru/arips/online/ (89).

Harmful Chemical Substances Encyclopedic Editions

The outstanding Russian scientist toxicologist, professor Nikolay Vasilievich Lazarev published in 1933 the first reference book *Harmful Chemical Substances in Industry* (" Химические вредные вещества в промышленности"- in Russian), which included encyclopedic information about chemical substances and pesticides and their properties. His book has been published 7 times from 1933 to 1977 and content has been updated. Later the publishing house Chemistry published a few reference books edited by V.A. Filov (90–97).

The common name of all books is Harmful Substances.

The content of books is listed below:

- Oxygen-containing organic substances;
- Nonorganic substances of elements by group I–IV;
- Nonorganic substances of elements by group V–VIII;
- Radioactive substances;
- Carbohydrates. Halogen containing carbohydrates;
- Nitrogen-containing organic substances;
- Halogen- and oxygencontaining organic substances;
- Natural organic substances.

This is the biggest collection of chemicals substances. It contains IUPAC names of substances, trade names and synonyms. Articles in the books contain information about chemicals pollutants, physico-chemical properties, synthesis, application, information about toxic properties of substances in environment and concerned with effect to human health. The reference book includes also hygienic norms, i.e., safety levels of substances in different media. It includes references to test methods of substances from 1970 (98).

Publishing house "Professional" (" НПО " Профессионал") prepared for publishing 11 volumes of Filov V.A. reference book *Harmful Chemical Substances in Environment* (" Вредные химические вещества в окружающей среде") and a book Hygienic normatives of chemical substances in environment (" Гигиенические нормативы" химических веществ в окружающей среде").

Volumes 1–6 of Filov's reference book have been published.

Volume 7 and 8 will be published soon. Information about publishing editions (98–101) is avalable at web http://www.naukaspb.ru/Podpiski/VHV.htm

The biggest collection of hygienic norms approved by Ministry of Health until 2007 is represented in the book of Rahmanina et.al *Hygienic Norms of Chemical Substances in Environment* (102).

In the book, the data for more than 4000 substances have been collected including MACs, TSLs, TPLs for different media (air, water, soil), safety levels for food, data for pesticides, etc. Each chemical has chemical abstracts service (CAS) numbers and an IUPAC name.

Authors are experts in chemistry, biology, toxicology, biochemistry, ecology, etc.

Hazard Database

The database HAZARD includes data on teratogenicity, carcinogenicity and mutagenicity of chemicals and is available at web http://www.iephb.nw.ru/~spirov/hazard/ (103).

HAZARD Project was supported by Russian Foundation for Basic Researches (Grants No 98-07-90373, 01-07-90373). The database contains data on carcinogenicity for about 400 chemicals, data for teratogenicity (results of tests on embryotoxicity, palatal cleft, neural system abnormalities, skeletal abnormalities) for about 1000 chemicals and data on mutagenic activity for about 700 chemicals.

Conclusion

Development of reliable QSARs models in Europe for the regulatory needs is actual in the light of the REACH and at the international level in the scope of OECD chemical assessment programs and Globally Harmonized System of Classification and Labeling of Chemicals (GHS). QSAR models for most endpoints will undoubtedly be used to provide us with test expectations for thousands of untested chemicals. In so doing, QSAR will complement the 3Rs (replacement, refinement and reduction of animals in research) with a powerful new tool to minimize animal testing. The integration of QSAR models with *in vitro* methods holds great promise in the prudent use and interpretation of our testing and assessment resources. Improved QSAR models will follow quickly. Sharing knowledge about QSAR modeling in different countries helps to choose the optimal decision in future implementation of QSARs for regulatory uses.

Acknowledgements

The financial support of the European Union through CAESAR project (SSPI-022674) as well as of the Slovenian Ministry of Higher Education, Science and Technology (grant P1-017) is gratefully acknowledged.

Terms

Artificial Neural Networks–Artificial neural networks (ANN) are algorithms simulating the functioning of human neurons and may be used for pattern recognition problems, e.g., to establish a quantitative structure-activity relationship.

Chemometrics–Chemometrics is the application of statistics to the analysis of chemical data (from organic, analytical or medicinal chemistry) and design of chemical experiments and simulation.

Computational Chemistry–Computational chemistry is a discipline using mathematical methods for the calculation of molecular properties or for the simulation of molecular behavior. It also includes, synthesis planning, database searching, and combinatorial library manipulation.

Free-Wilson (FW) Analysis–Free-Wilson analysis is a regression technique using the presence or absence of substituents or groups as the only molecule descriptors in correlations with biological activity.

Hammett Constant–The Hammett constant is an electronic substituent descriptor reflecting the electron-donating or -accepting properties of a substituent.

Hansch Analysis–Hansch analysis is the investigation of the quantitative relationship between the biological activity of a series of compounds and their physicochemical substituent or global parameters representing hydrophobic, electronic, steric and other effects using multiple regression correlation methodology.

Hansch-Fujita Constant–The Hansch-Fujita constant describes the contribution of a substituent to the lipophilicity of a compound.

Hydrophilicity–Hydrophilicity is the tendency of a molecule to be solvated by water.

Hydrophobicity–Hydrophobicity is the association of non-polar groups or molecules in an aqueous environment that arises from the tendency of water to exclude non-polar molecules.

LC_{50}Lethal Concentration–The concentration of chemical that causes the death of 50% of test animals.

LD_{50} Lethal Dose–(50% of population is expected to die) The quantity of material what will result in death of 50% of the test animals.

Molecular Descriptors–Molecular descriptors are terms that characterize a specific aspect of a molecule.

Multivariate Statistics–Multivariate statistics is a set of statistical tools to analyze data (e.g., chemical and biological) matrices using regression and/or pattern recognition techniques.

Pattern Recognition–Pattern recognition is the identification of patterns in large data sets, using appropriate mathematical methodology. Examples are

principal component analysis (PCA), SIMCA, partial least squares (PLS) and artificial neural networks (ANN).

QSAR–Quantitative structure–activity relationship (QSAR). A quantitative functional relation between an activity (e.g., toxicity) and one or more other structural descriptors. The generic term is quantitative structure–property relationship (QSPR), and specific sub-terms are used, such as quantitative structure–toxicity relationship (QSTR) and quantitative structure–biodegradability.

3Rs–Refinement: minimize suffering and distress; Reduction: minimize number of animals used; Replacement: avoid the use of living animals. The principle of the 3Rs was first enunciated by William Russell and Rex Burch in 1959.

Regression Analysis–Regression analysis is the use of statistical methods for modeling a set of dependent variables, Y, in terms of combinations of predictors, X. It includes methods such as multiple linear regression (MLR) and partial least squares (PLS).

Risk Assessment (in the context of human health)-The evaluation of scientific information on the hazardous properties of environmental agents (hazard characterization), the dose-response relationship (dose-response assessment), and the extent of human exposure to those agents (exposure assessment). The product of the risk assessment is a statement regarding the probability that populations or individuals so exposed will be harmed and to what degree (risk characterization).

SAR–Structure–activity relationship (SAR); a qualitative relationship between molecular structure and biological activity.

SMILES–Simplified molecular input line entry system (SMILES) is a string notation used to describe the nature and topology of molecular structures.

Topological Index–A topological index is a numerical value associated with chemical constitution for correlation of chemical structure with various physical properties, chemical reactivity or biological activity. The numerical basis for topological indices is provided (depending on how a molecular graph is converted into a numerical value) by either the adjacency matrix or the topological distance matrix. In the latter the topological distance between two vertices is the number of edges in the shortest path between these.

Toxicity–Deleterious or adverse biological effects elicited by a chemical, physical, or biological agent.

Toxicodynamics–The determination and quantification of the sequence of events at the cellular and molecular levels leading to a toxic response to an environmental agent (sometimes referred to as pharmacodynamics).

Toxicokinetics–The determination and quantification of the time course of absorption, distribution, biotransformation, and excretion of chemicals (sometimes referred to as pharmacokinetics).

Toxicometry–Term sometimes used to indicate a combination of investigative methods and techniques for making a quantitative assessment of toxicity and the hazards of potentially toxic substances.

REFERENCES

1. Database service on alternative methods to animal experimentation. European Center for Validation of Alternative Methods (ECVAM). http://ecvam.jrc.it/index.htm

2. Jaworska J, Comber M, Auer C, Van Leeuwen CJ. Summary of workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. Environmental Health Perspectives 2003;10(111):1358–60.

3. Worth A, Van Leeuwen CJ, Hartung T. The prospects for using (Q)SARs in a changing political environment- high expectations and key role for the European Commission's Joint Research Center. SAR and QSAR in Environmental Research 2004;15(5–6):331–43.

4. Computational Toxicology and Environmental Risk Modeling. The European Chemicals Bureau home page. http://ecb.jrc.it/qsar/.

5. Worth A, Bassan A, Gallegos A, Netzeva T, Patlewicz G, Pavan M, Tsakovska I, Vracko M. The Characterization of (Quantitative) Structure-Activity Relationships: Preliminary Guidance. EUR 21866 EN European Communities. http://ecb.jrc.it/DOCUMENTS/QSAR/QSAR_characterisation_EUR_21866_EN.pdf.; 2005.

6. Walker J, Jaworska J, Comber M, Schultz T, Dearden, J. Guidelines for developing and using quantitative structure–activity relationships. Environmental Toxicology and Chemistry 2003;22(8):1653–65.

7. Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationchip ((Q)SAR) Models in the Assessment of New and Existing Chemicals. (2006). *OECD Environment Health and Safety Publications. Series on Testing and Assessment No. 58*. Environment Directorate, OECD, Paris.

8. Report from the Expert group on (Quantitative) Structure-Activity Relationships ((Q)SARs) on the Principles for the Validation of (Q)SARs. (2004). *OECD Environment Health and Safety Publications, Series on Testing and Assessment No. 49*. Environment Directorate, OECD, Paris.

9. Jaworska J, Comber M, Auer C, Van Leeuwen CJ. Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. Environmental Health Perspectives Mini Monograph 2003;10 (111);1358–60.

10. Benfenati E. (Editor). *Quantitative Structure-Activity Relationship (QSAR) of Pesticide for Regulatory Purposes*. 2007; Amsterdam: Elsevier.

11. Cronin M, Walker J, Jaworska J, Comber M, Watts C, Worth A. Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. Environmental Health Perspectives 2003;10(111):1376–90.

12. Cronin M, Jaworska J, Walker J, Comber M, Watts C, Worth A. Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. Environmental Health Perspectives 2003;10(111):1391–1401.

13. Cronin M. Report on the "Review of the Status of the Development of Alternatives to using Animals in Chemical Safety Testing and Identification of New Areas for Development or Research in the Context of the Proposed REACH Regulation" , DEFRA contract number: CPEC 42, May 2005. http://www.defra.gov.uk/science/project_data/DocumentLibrary/CB01067/CB01067_2795_FRA.pdf.

14. Edler L, Poirier K, Dourson M, Kleiner J, Mileson B, Nordmann H, Renwick A, Slob W, Walton K, Würtzen G. Mathematical modeling and quantitative methods. Food and Chemical Toxicology 2002;40:283–326.

15. Barratt R, Balls, M. An overall strategy for the testing of chemicals for human hazard and risk assessment under the EU REACH system. ATLA 2003;31(7):19.

16. OECD's Database on Chemical Risk Assessment Models. http://webdomino1.oecd.org/comnet/env/models.nsf.

17. Coecke S, Ahr H, Blaauboer BJ. Metabolism: a bottleneck in in vitro toxicological test development. The Report and Recommendations of ECVAM Workshop 54. ATLA 2006;34:49–84.

18. CAESAR project home page. http://www.caesar-project.eu/.

19. Filov VA. Biography of Nikolay Vasilievich Lazarev. International Journal of Toxicology, 2002;21(4):235–6.

20. Smirnov V, Maimulov V, Nechiporenko S, Loit A, Loianich A, Kolbasov S. Computational methods for estimation of hazardous of harmful substances and establishment of hygienic norms in defferent media, 2002; pp. 130c (Russian).

21. Krasovskii G, Pinigin M, Tepikina L, Zholdakova Z, Egorova N, Ulanova I, Sidorov K, Scherbakov B. Computational methods for prediction of safety levels of substances in different environmental media. Gigien. Aspekti Ohrani Okr. Sredi, 1979;7:43–45. Moscow. (Russian).

22. Kasparov A, Sanotski, I. Toxicometry of environmental pollutants. CMP GKNP, 1986; p. 428. Moscow (Russian).

23. Sanotski I, Kagan Y, Krasovskii, G. Thresholds of harmful effects of chemicals in human and animal organizm. Prediction of safety levels exposure. In: *Toxicology of Chemicals as Environmental Pollutants*, 1986; pp.334–76. Moscow. (Russian).

24. Loit A, Arhipova M. Computational methods for establishment of hygienic norms in different environmental media. SPb. MAPO, 1997; p. 40. (Russian).

25. Methodical instructions for establishment of tentative safety levels of hazardous substances in workplace area. 1985; Minzrav SSSR, 04.11.85, No. 4000–85. Moscow. (Russian).

26. Sanotskii I, Ulanova I, Sidorov K, Ivanov N, Khalepo I, Avilova G, Kagan Yu, Sasinovich L, Miheev M, Trahtenberg I, Speranckii S, Rotenberg Yu, Novikov S, Maksimov G, Kovyazin V, Zeltzer P, Zaeva G. Methodical instruction for application of computational and express-experimantal methods for establishment of hygienic norm of chemicals used in water bodies. 1979; Minzdrav SSSR, 08À2.78, No 1943–78. Moscow. (Russian).

27. Novikov S. Modern appraches in prediction of toxicity of hazardous substances using relationship between chemical structure and biological activity. Gig Sanit 1980;10:16–19 (Russian).

28. Jegorov N. Prediction of toxicological parameters of individual groups of chemicals on the basis of SAR. Gig Sanit 1980;11:23–25 (Russian).

29. Zholdakova Z, Koh, R. (1986). Molecular connectivity and acute toxicity of chemicals. Gig Sanit 1986;2; p. 77–79 (Russian).

30. Tsaritschenko G, Malochnik V, Liberman M et al. The prediction of lethal doses (LD50) of drugs and products of there synthesis using a computer. Gigiena truda i profzabolevania, 1983;9:64–66 (Russian).

31. Koviazin VG. Predicting the toxicity level of organic substances by analyzing their structural fragments. Gig Tr Prof Zabol 1992;1:27–30.

32. Raevsky O, Skvortsov V. Quantifying hydrogen bonding in QSAR and molecular modeling, SAR QSAR Environ. Res 2005;3:287–300.

33. Raevsky O, Grigor'ev Ju, Kireev D, Zefirov N. Complete thermodynamic description of H-bonding in the framework of multiplicative approach. Quant Struct-Act Relat 1992;11:49–64.

34. Raevskii O, Trepalina E, Trepalin S. A new program for calculating solubility, lipophilicity, and liposome penetrability. Pharmaceutical Chemistry Journal 2000;1:33–36.

35. Raevsky O, Dearden J. Creation of predictive models of aquatic toxicity of environmental pollutants with different mechanisms of action on the basis of molecular similarity and HYBOT descriptors. SAR QSAR Environ Res 2005;15(5–6):433–48

36. Raevsky O, Trepalin V, Trepalina S, Gerasimenko V, Raevskaja O. SLIPPER-2001-software for predicting molecular properties on the basis of physicochemical descriptors and structural similarity. J Chem Inf Comput Sci 2001;42(3):540.

37. Solominova T, Maximov G, Semenov V. Prediction of acute toxicity of organic compounds on the basis of pattern recognition methods. Khim.-pharm journal 1984;2:181–188.

38. Belik A, Novikova L, Petrova Y, Sidorova L, Novikova A, Chupakhin O. Prediction o the class of acute toxicity of 1,3,4-thiadiazine derivatives. Khim.-pharm journal 1992;3:62–64.

39. Zul'karnaev T, Tiurina L, Solominova T, Novikov S, Kosheleva O, Kirlan S. An approach to predicting the acute toxicity of chemical substances. Gig Sanit 1999;3:54–61.

40. Zul'karnaev T, Solominova T, Tiurina L, Novikov S. The prediction of the class of hazard of organic compounds by their structure. Gig Sanit 1999;1:56–9.

41. Baskin I, Palyulin V, Zefirov N. A neural device for searching direct correlations between structures and properties of chemical compounds. Journal of Chemical Information and Computer Science 1997;37(4):715–21.

42. Zholdakova Z, Gusev E. Interval- polynomial approximation method as a way to increase efficiency of accelerated estimations of safety levels of hazardous substances in water. Gig Sanit 1987;7:9–13 (Russian).

43. Kharchevnikova N, Zholdakova Z, Zhurkov V. Theoretical interpretation of structure-genotoxicity relationships for mammals and bacteria in case of aliphatic halogen compounds with short chain. Vestn. Ross Akad Med Nauk 1997;7:8–13 (Russian).

44. Kharchevnikova N, Zholdakova Z, Zhurkov V. Theoretical approaches to predicting the metabolism and toxicity of aromatic amines. Gig Sanit 1998;4:62–65 (Russian).

45. Khalepo A, Ulanova I, D'iachkov P, Ibatullina R. Use of quantitative chemical characteristics of phenol ethers in the prediction of toxicity and maximum permissible exposure levels (MPEL) in the air of work areas. Gig Tr Prof Zabol 1990;3:10–15.

46. Kharchevnikova N. A system for predicting the toxicity and hazard of chemical substances, based on the joint use of logistic and numerical methods. Gig Sanit 2005;6:21–24.

47. Blinova V, Dobrynin D, Finn V, Kuznetsov S, Pankratova E. Toxicology analysis by means of the JSM-method. Biometrics 2003;19:1201–7.

48. Maksin M, Kharchevnikova N. Quantum mechanical module realizing a numerical approach to the problem of prediction of chemical's properties. J. Nauchno-tehnicheskoi informacii, 2002;ser. 2, 6:25–31.

49. Kharchevnikova N, Maksin M, Blinova V, Dobrynin D, Zholdakova Z. Prediction of carcinogenicity of polycyclic aromatic hydrocarbons using a quantum-chemical module of metabolites generation, a part of the intellectual JSM system. Ibid.,2003;11:15–17.

50. Kharchevnikova N, Blinova V, Maksin M, Dobrynin D, Zholdakova Z. Prediction of chronic toxicity and carcinogenicity of halogenated aliphatic compounds using JSM l method of authomatic hypothesis generation and quantum chemical calculations Ibid. 2004;12:15–19.

51. BorodinaYu, Rudik A, Filimonov D, Kharchevnikova N, Dmitriev A, Blinova V, Poroikov V. A new statistical approach to predicting aromatic hydroxylation sites. Comparison with model-based approaches. J Chem Inf Comput Sci 2004;44:1998–2009.

52. Kharchevnikova N, Dmitriev A, Borodina Iu, D'iachkov P. Quantum chemical model for prediction of the site of hydroxylation of aromatic substances mediated by cytochrome P450. Biomed Khim 2005;51(3):341–55.

53. Dyachkov P. Quantum-chemical calculations in the study of action mechanism and toxicity of xenobiotics. Itogi nauki I techniki VINITI. Ser Toksikologya 1990;16:1–280 (Russian).

54. Zholdakova Z, Kharchevnikova N. Application of quantum-chemical methods to prediction of the carcinogenicity of chemical substances. Vestn Ross Akad Med Nauk 2006;4:46–54.

55. Tarasov V, Mustafaev O, Abilev S, Mel'nik V. Use of ensemble structural descriptors for increasing the efficiency of QSAR study. Genetika 2005;41(7):997–1005 (Russian).

56. Sycheva L. Evaluation of the mutagenic effects of environmental factors using polyorgan micronuclear test. Vestn. Ross Akad Med Nauk 2006;7:27–32.

57. Filimonov DA, Poroikov VV, Karaicheva EI et al. Computer-aided prediction of biological activity spectra of chemical substances on the basis of their structural formulae: computerized system PASS. Experimental and Clinical Pharmacology 1995;58(2):56–62 (Russian).

58. Gloriozova TA, Filimonov DA, Lagunin AA, Poroikov VV. Evaluation of computer system for prediction of biological activity PASS on the set of new chemical compounds. Chim.-Pharm. J. 1998;32(12):32–39 (English translation by Consultants Bureau, New York: *Pharmaceutical Chemistry Journal*) (Rus).

59. Stepanchikova AV, Lagunin AA, Filimonov DA, Poroikov VV. Prediction of biological activity spectra for substances: Evaluation on the diverse set of drugs-like structures. Curr Med Chem 2003;10(3):225–33.

60. Poroikov V, & Filimonov D. PASS: prediction of biological activity spectra for substances. In: H. Christoph, (Ed), *Predictive Toxicology* 2005; 459–478. Boca Raton: Taylor & Francis Group, LLC.

61. Filimonov DA, Poroikov VV. Prediction of biological activity spectra for organic compounds. Russian Chemical Journal 2006;50(2):66–75 (Russian).

62. PASS INet project home page (2000). http://www.ibmc.msk.ru/PASS.

63. Poroikov VV, Filimonov DA, Borodina Yu V, Lagunin AA & Kos A. Robustness of biological activity spectra predicting by computer program pass for non-congeneric sets of chemical compounds. J Chem Inf Comput Sci 2000;40(6):1349–55.

64. Filimonov D, Poroikov V, Borodina Yu, Gloriozova T. Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors. J Chem Comput Sci 1999;39(4):666–70.

65. Poroikov VV & Filimonov DA. How to acquire new biological activities in old compounds by computer prediction. J Comput Aid Mol Des 2003;16(11):819–24.

66. Anzali S, Barnickel G, Cezanne B, Krug M, Filimonov D, Poroikov V. Discriminating between drugs and nondrugs by prediction of activity spectra for substances (PASS). J Med Chem 2001;44(15):2432–7.

67. Poroikov VV, Filimonov DA, Ihlenfeldt WD, Gloriozova TA, Lagunin AA, Borodina YuV, Stepanchikova AV, Nicklaus MC. PASS biological activity spectrum predictions in the enhanced open NCI database browser. J Chem Inform Comput Sci 2003;43(1):228–36.

68. Lagunin AA, Gomazkov OA, Filimonov DA, Gureeva TA, Dilakyan EA, Kugaevskaya EV, Elisseeva YuE, Solovyeva NI, Poroikov VV. Computer-aided selection of potential antihypertensive compounds with dual mechanisms of action. J Med Chem 2003;46(15):3326–32.

69. Geronikaki A, Babaev E, Dearden J, Dehaen W, Filimonov D, Galaeva I, Krajneva V, Lagunin A, Macaev F, Molodavkin G, Poroikov V, Saloutin V, Stepanchikova A, Voronina T. Design of new anxiolytics: from computer prediction to synthesis and biological evaluation. Bioorg Med Chem 2004;12(24):6559–68.

70. Lagunin AA, Dearden J, Filimonov DA, Poroikov VV. Computer-aided rodent carcinogenicity prediction. Mutation Research, 2005;586(2):138–46.

71. Zakharov AV, Lagunin AA, Filimonov DA, Poroikov VV. Quantitative Structure-Activity Relationships of Cyclin-Dependent Kinase I Inhibitors. Biochemistry (Moscow), Supplement, Series B, Biomedical Chemistry, 2007;1(1):17–28.

72. Filimonov DA, Lagunin AA, Poroikov VV. Prediction of activity spectra for substances using new local integrative descriptors. QSAR and Molecular Modelling in Rational Design of Bioactive Molecules. Esin Aki Sener, Ismail Yalcin Eds., Ankara (Turkey); CADD & D Society, 2005; p. 98–99.

73. Lagunin AA, Zakharov AV, Filimonov DA, Poroikov VV. New approach for QSAR modeling of the acute toxicity. SAR & QSAR in Environmental Research 2007;18(3–4):285–298.

74. Borodina Yu, Sadym A, Filimonov D, Blinova V, Dmitriev A, Poroikov, V. Predicting Biotransformation Potential from Molecular Structure. J Chem Inf Comput Sci 2003;43:1636–46.

75. Zakharov AV, Filimonov DA, Lagunin AA, Poroikov VV. GUSAR (general inrestricted structure-activity relationships) program package. 2004; Russian State Patent Agency. N 2006613591 of 16.10.2006.

76. Zakharov AV, Filimonov DA, Lagunin AA, Poroikov VV. GUSAR: new approach for multiple QSAR. 2007; *Quant. Struct. Activ. Relations*. (In press).

77. Borodina Yu, Rudik A, Filimonov D, Kharchevnikova N, Dmitriev A, Blinova V, Poroikov V. A new statistical approach to predicting aromatic hydroxylation sites. Comparison with model-based approaches. J Chem Inform Comput Sci 2004;44(6):1998–2009.

78. D'yachkov PN, Kharchevnikova NV, Dmitriev AV, Kuznetsov AV, Poroikov VV. Quantum chemical simulation of cytochrome P450 catalyzed aromatic oxidation: Metabolism, toxicity, and biodegradation of benzene derivatives. International Journal of Quantum Chemistry, 2007;107(13):2454–78.

79. Poroikov V, Filimonov D, Lagunin A, Gloriozova T, Zakharov A. PASS: Identification of probable targets and mechanisms of toxicity. SAR & QSAR in Environmental Research. 2007;18(1–2):101–10.

80. Devillers J, Dore JC, Marchand-Geneste N, Porcher JM, Poroikov V. Modelling the endocrine disruption profile of xenobiotics. Lecture Series on Computer and Computational Sciences 2006;6:1–3.

81. Lagunin A, Stepanchikova A, Filimonov D, Poroikov V. PASS: prediction of activity spectra for biologically active substances. Bioinformatics 2000;16(8):747–8.

82. Sadym A, Lagunin A, Filimonov D, Poroikov V. Prediction of biological activity spectra via Internet. SAR and QSAR in Environmental Research 2003;14(5–6):339–47.

83. Geronikaki A, Druzhilovsky D, Zakharov A, Poroikov V. Computer-aided predictions for medicinal chemistry via Internet. SAR and QSAR in Environmental Research 2008;19(1&2):27–38.

84. SARET: Computerized System for QSAR, QSPR and Property-Property Analysis in Environmental Toxicology. http://www.chem.ac.ru/Chemistry/Databases/SARET.en.html

85. Computerized System SARET for QSAR, QSPR and Property-Property Analysis in Environmental Toxicology. http://erh.ru/dok/metod02.php. (Russian).

86. Novikov SM, Poroikov VV, Tertichnikov SN, Filimonov DA, Zholdakova ZI, Kharchevnikova NV, Tepikina LA. Computer Database SARET base. State Registry Number 0229601501. Reg. No 1114 of 16 December 1996, Moscow: Informregistr.

87. Novikov SM, Poroikov VV, Semenovyh LN. Contemporary Problems of computer systems application for estimation of toxicological and ecotoxicological risks for chemical compounds. *Hygiene and Sanitary* 1994;5:4–8 (Russian).

88. Filimonov DA, Novikov SM, Tertichnikov SN, Poroikov VV. Prediction of mutagenicity and carcinogenicity for pharmaceuticals and other chemical substances using computerized system SARET *Abstr. Workshop "Quantitative Modelling Approaches for Understanding and Predicting Mutagenicity and Carcinogenicity"*, 1997; p.22, Rome (Italy).

89. Russian National register of potentially hazardous chemical and biological substances, Moscow, http://www.rpohv.ru/.

90. Filov VA, Ivina BA, Musijchuk YI. Harmful substances. Oxygen-containing organic substances. S-Petersburg (SPb): NPO Professional, 2004; part 1: p. 402; part 2: p. 342. (Russian).

91. Filov VA et al. Harmful substances. Nonorganic substances of elements by group I-IV. SPb: Khimia, Leningrad, 1988; p. 512 (Russian).

92. Filov VA et al. Harmful substances. Nonorganic substances of elements by group V-VIII. SPb: Khimia, Leningrad, 1989; p. 592 (Russian).

93. Filov VA et al. Harmful substances. Radioactive substances. SPb: Khimia, Leningrad, 1990; p.464 (Russian).

94. Filov VA et al. Harmful substances. Carbohydrates. Halogen containing carbohydrates. SPb: Khimia, Leningrad, 1999; p. 732 (Russian).

95. Filov VA et al. Harmful substances. Nitrogen-containing organic substances. SPb: Khimia, Leningrad, 1992; p. 432 (Russian).

96. Filov VA et al. Harmful substances. Halogen- and oxygencontaining organic substances. SPb: Khimia, Leningrad, 1994; p. 687 (Russian).

97. Filov VA et al. Harmful substances. Natural organic substances. SPb: Khimia, Leningrad, 1994; p. 504Ò (Russian).

98. Filov VA, Ivin BA. Environmental chemical pollutants, toxicology and information. Ros.Khim.J.,2004, 2 (Russian).

99. Ljublina E, Minkina M, Miheev M, Musijchuk Y, Rusin V, Tiunov V, Filov V, Frolova T, Hrustaliova T. Oxygen-containing organic substances. Part1, 2007; p. 404 (Russian).

100. Alkilbaeva L, et al. Nonorganic substances of elements by group I-IV, 207; p. 462. (Russian).

101. Buldakov L, Vasilenko I, Vasilenko O, Kalistratova V, Moskaljov Y, Moskvin A, Nisimov P, Osipov V, Romanova I, Stepanov S, Filov V, Shishkin V. Radioactive substances. S-Petersburg (SPb): NPO Professional, Russia, 2007: p. 334. (Russian).

102. Rahmanin Y et al. Hygienic norms of chemical substances in the environment. SPb.: NPO Professional, 2006; p. 736. (Russian).

103. Database HAZARD, http://www.iephb.nw.ru/~spirov/hazard/