Ivan Juranić*

University of Belgrade, Faculty of Chemistry and Center for Chemistry IChTM, Belgrade, Serbia

Molecular descriptors as proxies for the modeling of the materials and their environmental impact

ABSTRACT

For the prediction of material's properties and of interaction of molecules with the surroundings, one needs to know their properties. Usually, the molecular properties are revealed through experimental measurements. It can be a tedious, time-consuming, and costly work. On the other hand, computational chemistry readily generates a huge number of data which can provide various molecular descriptors. These can be various observable properties (bond lengths and angles, dipole moments, etc...), but also the unobservable properties (partial atomic charges, electronegativity, various latent variables ....).

There is an urgent need to develop accurate and economical screening tools that predict potential toxicity and environmental burden of various chemicals. Equally important is the design of safer alternatives.

Molecular modeling methods offer one of several complementary approaches to evaluate the risk to human health and the environment as a result of exposure to environmental chemicals. These tools can streamline the hazard assessment process by simulating possible modes of action and providing virtual screening tools that can help prioritize bioassay requirements. Tailoring these strategies to the particular challenges presented by environmental chemical interactions make them even more effective.

Advances in the fields of computational chemistry and molecular toxicology in recent decades allow the development of predictive models that inform the design of molecules with reduced potential to be toxic to humans or to the environment.

As an example we present the novel methodology for the computational evaluation of pK_a values of various organic bases, based on calculation of partial atomic charges by a simple semiempirical QM method.

Keywords: *computed molecular properties, molecular descriptors, QSAR/QSPR, partial atomic charges.*

1. INTRODUCTION

With few exceptions [1], chemists are traditionally aware of their society's responsibilities. Modern society is full of potential hazards. For many of those hazards we need public policies that protect us.

Almost everything we do to improve our living has a downside. Industries that make goods cause pollution, whether it is a copper smelter releasing toxic gasses or runoff from a cultivated field contaminating ground water with pesticides and/or fertilizer. It is easy to recognize the innumerable issues we face in our society [2]:

*Corresponding author: Ivan Juranić
E-mail: ijuranic@chem.bg.ac.rs

Food additives and processed food
Medical practices and drugs
Agricultural practices: pesticides, fertilizers, genetically modified organisms
Pollution: chemical, radioactivity
Environmental hazards of consumer products (polyhalogenated hydrocarbons, electromagnetic fields, lead in gasoline …)
Product safety (autos, electrical appliances, children toys …)
Transportation/traffic safety

Chemical professionals and researchers study these issues through various subjects – to mention some of them:

New materials: Comprise new molecular and supramolecular structures having interesting functional properties. Nowadays this is intensively researched area, both theoretically and for practical applications. Nevertheless, it won't be in focus of this article.

Prediction of Vapor Pressures [3,4]. Vapor pressure plays an important role in the transport, distribution, and fate of environmental pollutants in the atmosphere and hence, the environmental acceptability/availability of chemical products and processes.

Atmospherically Relevant Molecular Clusters [5]. Atmospheric aerosols affect our life in a number of ways, both in terms of their effect on health and due to their effect on climate patterns.

Pharmacokinetics [6-8]. Human beings live in changing environments in which some factors change very rapidly and others more slowly. Organisms must adapt themselves to these changes if they are to survive and reproduce. A specific aspect of the adaptation is the mechanism by which organisms try to maintain homeostasis in respect to the chemical environment. Organisms are endowed with the enzyme machinery which disposes of chemicals that have entered the body. These enzymes are called drug- or xenobiotic-metabolizing enzymes and they have been dealt with in numerous monographs, review articles, and symposia during the past 30 years or so.

Drug toxicity [9,10]. Physico-chemical indicators have been increasingly used during the early stages of drug discovery to provide a more comprehensive understanding of the key properties that affect the biological functions(i.e. ADME—absorption, distribution, metabolism, and excretion). The most commonly measured physicochemical properties are permeability and solubility (due to their importance in the gastrointestinal absorption of orally administered drugs), and also lipophilicity, integrity, and stability (since these properties generally define the pharmaceutical potential of a compound). Often, the ADME concept is expanded by toxicity. Apparently, this potential hazardous quality of new molecular entry (NME) should be taken into account with no less care than possible therapeutic effects.

There are several end points that relate to potential hazardous effects (such as the liver effects of drug candidates; cardiac toxicity, and blood–brain barrier of compounds that has influence upon both drug efficiency and drug toxicity) which should be estimated during the early stages of drug discovery. Finally carcinogenicity and mutagenicity of a drug also should be estimated during the initial stages of drug discovery.

Due to huge cost and time necessary for research and development related to drug design, many *in silico* methods have been developed to provide accurate prediction of pharmacokinetic properties, such as Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET), in very early stage of drug discovery.

Prioritizing Bioassay Requirements [11]. Computational molecular modeling methods aid the risk assessment process by providing a rational approach for some extrapolations in the evaluation of chemical hazard. For instance, when elements of a data set required for evaluating the potential hazard of a chemical are unavailable and inferences can be made based on interactions with putative targets, molecular modeling can be used to simulate the relevant missing information. Both ligand and structure-based molecular modeling methods used in pharmaceutical discovery can be adapted to provide this type of simulated data. However, because of the greater diversity of chemical space and binding affinity domains being considered and the differences in the strategic application of the results (the need to minimize false negatives), these molecular modeling strategies require additional considerations when assessing chemical hazards. Molecular docking of potential environmental chemicals to putative macromolecular targets for toxicity provides a measure of their capacity to interact and hence is an aid in the (pre)screening process for specific modes of toxicity. These results provide a rationale for developing further, more complete testing strategies.

Acute aquatic toxicity [12]. The approach differs from QSAR in that rather than predicting a toxicity value or a threshold, it elucidates the probability that a compound with particular properties will exhibit a certain toxicity profile. Thus, the approach does not yield deterministic values, but rather probabilistic information about whether a chemical is likely to exhibit toxicity above or below a certain threshold value. The output of this approach can help reduce *in vitro* and *in vivo* testing by using key properties to define a high-probability safe chemical space of low to no toxicity. Further, by using property-based criteria as opposed to QSAR descriptors, informed design decisions can more readily be pursued for modifications to reduce potential toxicity.

Screening Environmental Chemicals [13]. A complete set of regulatory tests for a single chemical (including those for carcinogenicity and for chronic, reproductive, and development toxicity) uses thousands of animals and costs millions of dollars. In addition, traditional animal tests often yield limited information on mechanism of action, and hence on the cellular pathways that could lead to toxicity in humans. Such mechanistic information is crucial to move beyond default approaches for extrapolating from high-dose animal toxicity tests to

estimation of human risk at realistic exposure levels. There is a pressing need to screen the large backlog of chemicals for their potential toxicity and, ultimately, their contribution to human diseases. The ability to use molecular and computational sciences holds the potential to usher in a new era of prioritizing, assessing, and managing chemicals.

The bioconcentration factor (BCF). The bioconcentration factor (BCF) is an important bioaccumulation hazard assessment metric in many regulatory contexts. Its assessment is required by the REACH regulation (Registration, Evaluation, Authorization and Restriction of Chemicals) and by CLP (Classification, Labeling and Packaging) [14].

Oral Bioavailability Prediction. Drug bioavailability is a key aspect often evaluated with *in silico* models [15]. It comprises the assessment of common physico-chemical properties such as solubility [16] or pK$_a$, which describe the substance regardless of the biological environment. Moreover, interactions with the biological system are also evaluated, such as intestinal permeability**Error! Bookmark not defined.** (with models derived on the basis of *in vivo* [17], or *in vitro* data [8]). Interest is now growing also in the role of active transporters [19] together with passive diffusion [20].

Biological Clogging [21]. The biological clogging of natural porous media, often in conjunction with physical or chemical clogging, is encountered under a wide range of conditions. Wastewater disposal, artificial groundwater recharge, *in situ* bioremediation of contaminated aquifers, construction of water reservoirs, or secondary oil recovery, are all affected by this process.

This is closely related to the study of **membrane permeation**, the important issue in process industry [22], biology [23], environment [24], etc…

All these (and many other) studies require sophisticated procedures and considerable human and material resources. The innumerable variety of chemical compounds renders the task of checking every compound, impossible. Therefore, a number of models were developed to enable relatively quick and (as much as possible) credible assessment of environmental impact, toxicity, and stability of compounds. The most successful are the models using the descriptors derived from the structure of the molecule. A single, or combination of several molecular descriptors, proved to be dependable proxies for molecular and material properties.

2. QUANTITATIVE STRUCTURE – ACTIVITY RELATIONSHIPS (QSAR)

Molecular descriptors are numerical indexes encoding some information related to the molecular structure. They can be both experimental physico-chemical properties of molecules and theoretical indexes calculated by mathematical formulas or computational algorithms. The main tool for modeling behavior of molecules and materials is quantitative correlation of structure and (re)activity (QSAR). Because the observable properties of molecules usually require tedious measurements, it is convenient to relate structure of molecule with macroscopic measurable properties (QSPR).

(re)activity of compound = f(molecular structure) (1)

A first model of this kind was formulated by Hammett [25]. It was originally formulated for the prediction of dissociation constants of benzoic acids

$$\sigma_X = log\, K_X - log\, K_H \qquad (2)$$

where X denotes the substituent on benzene ring.

Although the Hammett methodology has been criticized by theoreticians because of its empirical foundation, it is astonishing that constants, obtained simply from the ionization of organic acids in solution, can frequently predict successfully equilibrium and rate constants for a variety of families of reactions in solution. Almost every kind of organic reaction has been treated *via* the Hammett equation, or its extended form. The literature is so voluminous and extensive that there is no complete review of all that has been accomplished [26].

Analogous approach was formulated by Free and Wilson for the assessment of biological activity of compounds [27].

$$log(1/C) = \Sigma\, a_j\, X_{ij} + \mu \qquad (3)$$

where a_j is group contribution of structural element X in position *j* in molecule *i*. Constant μ is theoretical biological activity of referent compound.

Another, widely used model is proposed by Hansch [28].

$$log(1/C) = A{\times}logP + B{\times}\sigma + C{\times}E_s + D \, (4)$$

where Log*P* describes the lipophilicity of whole molecule, but could be replaced with lipophilic parameter of substituent, π if a congener series of compounds is considered; σ are Hammett parameters of substituent electronic effects; E$_s$ are Taft steric parameters of substituents. All (and many other) parameters could be derived from computational results on molecules (computed molecular descriptors).

Therefore, the quantitative correlation between molecular structure and some measurable property

of molecule or material is usually obtained through (multiple) regression analysis involving several/various molecular descriptors.

The health and environmental concerns are most frequently related to toxicity of compounds.

When dealing with *in silico* models for toxicity [20] we can distinguish methods which address the overall toxic phenomenon, for instance carcinogenicity, and methods which address factors of the process leading to toxicity manifestations. Thus, some models address systemic toxicity while others focus on organ-specific toxicity. There are *in silico* models for a large number of toxicity phenomena [29]. Among these genotoxicity and carcinogenicity, although very complex, are among the most extensively studied. However, the availability of a model is different from the reliability of the model prediction. Moreover the possibility to satisfactorily apply a model to a pharmaceutical of interest may depend on toxicological data availability for molecules chemically related to this entity. There are papers in which model performance is assessed for general chemicals [30], or for pharmaceuticals, with respect to carcinogenicity and mutagenicity [31]. In contrast, there is a paucity of models for developmental or reproductive toxicity [32]. In comparison with the models presented above, *in silico* models for organ-specific effects are generally focused on pharmaceuticals, since data availability is most abundant for drug-like compounds. Among them hepatotoxicity has been frequently investigated [33], and nowadays increasing interest is also placed on cardiotoxicity [34] and nephrotoxicity [35]. Limited models are available for neurotoxicity or other effects such as phospholipidosis [36]. Some investigations involved the use of adverse effect databases [37] or combinations of effects against several receptor targets [38], using integrated risk indices.

2.1. Regulatory context for in silico models

The use of *in silico* methods to estimate toxicity is solicited in different legislation in the EU such as those concerning chemicals or cosmetics. Their use in the pharmaceutical field is more related to the R&D of new drugs avoiding potential adverse effects. Furthermore, regulatory criteria have been defined for the use of *in silico* predictions in drug safety for the genotoxicity evaluation of drug impurities where the presence or absence of specific structural alerts triggers the subsequent management of potential risk posed by the presence of this impurity [39,40]. In this context, new models have been recently proposed to estimate genotoxic or carcinogenic potency in

order to estimate safety [40, 41] Very good results have been obtained in the evaluation of genotoxic impurities in the industrial sector when *in silico* data were coupled with expert evaluation [40] achieving a negative predictive value of 99%.

There are numerous computational methods for the assessment of the toxicity of compound, and many of them comprise the software packages and databases which are publicly accessible [42].

3. MOLECULAR DESCRIPTORS

Molecular descriptors are formal mathematical representations of a molecule, obtained by a well-specified algorithm, and applied to a defined molecular representation or a well-specified experimental procedure: *the molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment* [43].

An ideal molecular descriptor should, in numerical form, encode all the features of structure which determine chemical and physical properties, conventionally the constituent atoms, their bonding and spatial relationships, molecular shape and size, and electronics allowing for modifications due to interaction of the molecule with its environment [44]. In terms of quantum mechanics, the most complete descriptor of molecular structure and properties is the molecular wave function, which characterizes nuclear and electronic properties fully. In principle, the interaction of any molecule with any electromagnetic field can be predicted precisely by solving the appropriate Schrödinger equation. Therefore, all molecular properties (including interactions with biological systems) can be quantified if we know the molecular wave function.

Thus, molecular descriptors, which are closely connected to the concept of molecular structure, play a fundamental role in scientific research, being the theoretical core of a complex network of knowledge, as is shown [45] in Figure 1.

In chemistry we deal with various molecular descriptors, which proved to be very useful in prediction of various molecular properties. Traditionally are used melting and boiling points, dipole moments, polarizability, partition coefficients, R_f values, and other properties that can be measured. Moreover, various non-measurable properties, like electronegativity, partial atomic charges, and other, were widely used. There is a variety of ways for classification of descriptors; and one classification (far from being exhaustive) is given in Table 1.
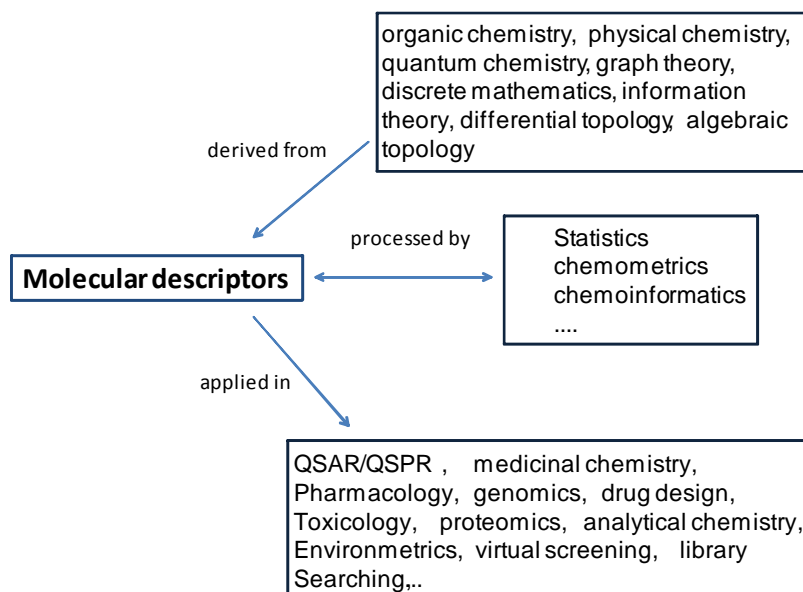
Figure 1 - General scheme of the relationships among molecular structure, molecular descriptors, cheminformatics and QSAR/QSPR modelling

Table 1 - Molecular descriptors [46] and their classes [47]

Descriptor Class	Molecular Descriptor
simple molecular properties, constitutional descriptors	molecular weight, number of ring structures, number of rotatable bonds, number of H-bond donors, number of H-bond acceptors, element counts, functional group counts (amine group, carbonyl group, carboxylic acid group, hydroxyl group, nitrile group, nitro group)
molecular connectivity and shape	Kier and Hall connectivity indices (zero order, first order, and second order) and valence connectivity indices (zero order, first order, and second order), molecular shape kappa indices, kappa alpha indices, flexibility index
electrostatic descriptors, electrotopological state	partial charges, charged surface areas, electrotopological state indices and atom type electrotopological state indices
quantum chemical properties	atomic charge on the most positively charged H atom, largest negative charge on an non-H atom, polarizability index, hydrogen bond acceptor basicity (covalent HBAB), hydrogen bond donor acidity (covalent HBDA), molecular dipole moment, absolute hardness, softness, ionization potential, electron affinity, chemical potential, electronegativity index, electrophilicity index, quadrupole moment, HOMO and LUMO energies, dielectric energy
geometrical properties	molecular size vectors (distance of the longest separated atom pairs, combined distance of the longest separated three atoms, combined distance of the longest separated four atoms), molecular van der Waals volume, solvent accessible surface area, molecular surface area, van der Waals surface area, polar molecular surface area, sum of solvent accessible surface areas of positively charged atoms, sum of solvent accessible surface areas of negatively charged atoms, sum of charge weighted solvent accessible surface areas of positively charged atoms, sum of charge weighted solvent accessible surface areas of negatively charged atoms, sum of van der Waals surface areas of positively charged atoms, sum of van der Waals surface areas of negatively charged atoms, sum of charge weighted van der Waals surface areas of positively charged atoms, sum of charge weighted van der Waals surface areas of negatively charged atoms, isosurface volume, molar volume

The measurable descriptors are more valuable because they reflect a genuine property of molecule. Values of non-observable descriptors heavily depend on the definition of them and on the method of their calculation. Determination of measurable descriptors is frequently very tedious

and time consuming (*e.g.* reaction rate constants, equilibrium constants, partition coefficients, etc.), and there is a continuous desire to estimate them using other, much easily reachable descriptors.

Since the computers become widely accessible, computational modeling of molecules becomes a standard procedure in chemical research. Some of the most frequently used, computationally obtained descriptors, are results of quantum-chemical calculations, and are presented

- **Computationally derived MD**

    Descriptors derived by quantum mechanics
    - Partial Atomic Charges
    - HOMO and LUMO Energies
    - Orbital Electron Densities
    - Superdelocalizabilities
    - Atom-Atom Polarizabilities
    - Molecular Polarizabilities
    - Dipole Moments and Polarity Indices
    - Energies
    - Hydrogen bond strength [49]
    - Lipophilicity and molar refractivity [50]
    - Electrotopological state indexes [51]
    - Molecular surface properties
        - Generation of 3D structures and conformational analyses [52]
        - Surface area calculations [53]
- Topological descriptors [54]
- Molecular mechanic computation
    - Molecular interaction fields
    - Potential energy surface
    - Latent Variables

To demonstrate how the computed data can be used for the evaluation of molecular properties we shall present one (previously not published) study. The pK_a values of amines are defined by reaction:
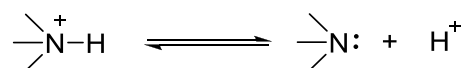
And experimentally can be determined by measuring the concentrations of all reactants and products, what can be done by various physical-chemical methods. It may be costly, and usually require a lot of time. On the other hand we can use some quick quantum chemical method to calculate the versatile descriptors. As a descriptor of interest we choose the partial atomic charges (PAC), because of their recognized importance for the evaluation of the reactivity of compounds. Electronegativity, Hammett rationalization of reactivity, and many other ideas in chemistry strongly rely on idea of PAC.

on the left column in following list. In the right-hand column are listed some of molecular properties that can be computed on the basis of these descriptors.

As is previously noted, biological activity, or toxicity of compound have been correlated with various measurable properties like pK_a. To extend these studies to new compounds is much easier (and much less expensive) if pK_a values can be estimated from data supplied by computational modeling [48].

QSAR Results
- Biological Activity
- Chemical Reactivity
- Partition Coefficients
- Chromatographic Retention Indexes and Response Factors
- Other Physicochemical Properties
    - boiling point (vapor pressure) [55] (PCA)
    - critical temperature
    - molar critical volume
    - melting points
    - flash points
- Substituent Constants
- Solvational Characteristics
- CoMFA [56]

3.1. Method of calculation of PAC

Because partial atomic charges are not observable, it's hard to argue which method of their calculation is the "right" one. Definitions of PAC: Coulson's, Mulliken's, Gasteiger's, Bader's, Hirshfeld's, and other, generally provide similar results, but can sometimes markedly differ. The same is when the calculations are done by different quantum chemical methods. The single rule must be applied - that all calculations have to be done by the same procedure. And only results obtained in the same quantum-chemical frame can be compared.

For this study a semiempirical MNDO-PM6 method (with simulation of solvent) was applied for calculation of electronic properties of the base's conjugated acids (protonated, aminonium compounds).

As a basis for the evaluation of pK_a values the partial atomic charges on hydrogen (q_H) and

nitrogen (q_N) of protonated amino group were used. The correlation was achieved by equation:

$$pK_a = A \cdot q_H + B \cdot q_N + C \cdot I + D \qquad (5)$$

This equation contains the index variable I, which has to take in account the effects on pK_a

values, other than the polarization of NH bond. In previous work [57] on aliphatic amines was found that such variable can correspond to the number of hydrogen atoms bound to nitrogen in aminonium ion.

Table 1 - pK$_a$ values and partial atomic charges calculated with semiempirical PM6/COSMO model

	pK$_{a\,(exp)}$	q$_H$	q$_N$	I	pK$_{a\,(calc)}$
Piperazine	9.73	0.2910	0.0310	1	9.768
Quinuclidine	10.58	0.2958	0.0762	1	8.990
Benzoquinuclidine	7.79	0.3023	0.0843	1	8.406
N,N'-Dimethyl hydrazine	7.52	0.2933	0.1354	1	8.665
Benzimidazole	6.5	0.3437	-0.1276	1	7.011
2-Methylimidazole	7.75	0.3403	-0.1213	1	7.224
Imidazole	7.05	0.3381	-0.0570	1	6.830
Hexamethylenetetramine	4.9	0.3286	-0.0949	1.56	4.873
Histamine	7.3	0.2810	-0.0379	1.70	7.314
2,3-Diaminobutane, *meso*	6.92	0.2812	-0.0811	1.85	6.869
Phenazine	1.23	0.3427	-0.0095	1.85	1.386
Pyrimidine	1.37	0.3494	-0.0930	1.85	1.594
Pyrazine	0.92	0.3396	0.1171	1.85	0.513
Quinazoline	1.55	0.3499	-0.1288	1.85	1.871
Quinoxaline	0.8	0.3415	0.0727	1.85	0.755
Benzothiazole	1.84	0.3413	-0.0301	1.85	1.678
Aniline	4.58	0.2882	-0.0345	2.07	4.689
1-Phenanthrylamine	3.23	0.2913	-0.0309	2.33	2.994
9-Phenanthrylamine	3.19	0.2897	-0.0334	2.33	3.142
1-Naphthylamine	3.4	0.2885	-0.0311	2.33	3.217
3-Aminopyrene	2.91	0.28847	-0.0251	2.33	3.166
1-Anthrylamine	3.22	0.2921	-0.0297	2.33	2.920
9-Anthrylamine	2.7	0.2886	-0.0192	2.33	3.104

The regression analysis was done for 23 organic bases, involving aliphatic and aromatic amines and nitrogen heterocyclic bases, and is presented in Table1. Much effort was invested to verify the accuracy of many experimental values. In diacidic base was frequently hard to identify is the published value the result for mono-, or diprotonated base. In such case we used the already published method [49] for evaluation of dissociation constant of acids – in this case conjugated acids, to distinguish between the dissociation of monoprotonated and diprotonated bases. The experimental results for which was not clear how the measurement was done (solvent, *etc.*) were discarded.

It turned out that the index variable that was valid for aliphatic amines is not applicable here.

The biggest difference is among acyclic and heterocyclic amines without adjacent N atoms (to them can be assigned value $I = 1$), and aniline analogs having NH$_2$ group on aromatic ring (to which can be assigned value $I = 2.33$).

To heterocycles having adjacent N atoms in the ring could be assigned index variable $I = 1.85$.

Tree data points with markers '2' and '4' are outliers of this rule, and were given slightly different I values. For them could be speculated that taking into account the specificity of molecular structure the scattering can be explained. For example, hexamethylene tetramine may have value for $I = 1 + \log(4) = 1.6$ due to four equivalent protonation sites. It is closest to $I = 1.56$, what is inside of the experimental error interval for the determination of pK_a.

The following regression parameters (5) were obtained:

$$A = -78.878; \qquad B = -8.828;$$
$$C = -5.473; \qquad D = 34.468$$

Regression coefficient is $r = 0.985$

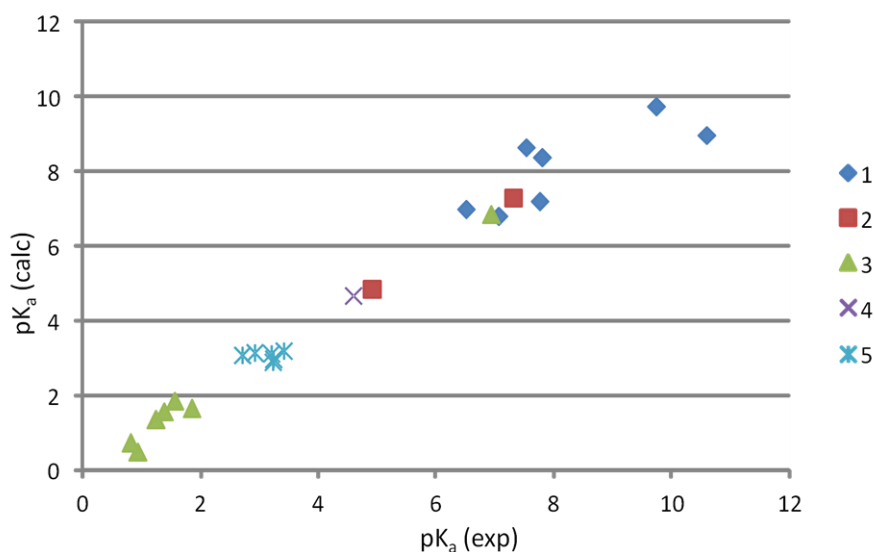Comparison of calculated and experimentally obtained pK$_a$ values is presented on Figure 2.

Figure 2 - Correlation of calculated and measured pK$_a$ values. Markers correspond to various values of parameter I = 1 (1); 1.7 (2); 1.85 (3); 2.07 (4); 2.33 (5).

This example illustrates how the use of results about the electronic structure of molecule, *i.e.*, distribution of electron density in it, may provide accurate prediction of various physicochemical properties of molecules. The acidity constants are correlated with many chemical and biological features of molecules, and their measurement is highly enhanced by knowledge of approximate value of them. Various other molecular descriptors can be used, but for efficient correlation one has to choose those which have intrinsic relation to the property in study.

Many molecular properties is hard to measure and, on the other hand, they are highly useful for the prediction of the behavior of compounds in the environment. Computationally generated molecular descriptors are easily available, and can be exploited to calculate (at least approximately) every molecular observable property.

Because of their straightforward relation to the distribution of electron density in molecule, the partial atomic charges (regardless of the fact that they are not observable) were proven to be particularly useful proxies to evaluate molecular properties. It enables easy interpretation of results, because the so called 'chemical intuition' is based on the presumed distribution of charges in molecule.

4. CONCLUSION

In silico models for toxicity prediction are in use in pharmaceutical industry and other research areas. This methodology can be efficiently used for assessing the environmental issues, like aquatic toxicity, bioavailability, screening environmental chemicals, bioconcentration factor(s), and others.

With the advance in computational efficacy, the calculation of hundreds of molecular descriptors is now a routine job. The computationally obtained descriptors are easy to use as proxies for observable properties of molecules. The rational selection of calculated descriptors provide basis for the modeling of molecular property of interest.

A wide selection of different modeling tools exist, addressing general and specific effects on environment. Computational models to predict ADME are common and assist in understanding toxicity phenomena.

The existing and future models will benefit from new experimental data, allowing finer tuning of model(s). It must be stressed that the combination of multiple approaches will increase the performance of the predictions, compared to results related to one model.

5. REFERENCES

[1] For example, Fritz Haber, who designed, produced, and supervised the use of war gasses in the IWW.

[2] J.Bull: Scientific decission making. http://www.utexas.edu/courses/bio301d/SDM.pdf.

[3] C.Liang, D.A.Gallagher (1998) QSPR Prediction of Vapor Pressure from Solely Theoretically-Derived Descriptors, J. Chem. Inf. Comput. Sci., 38 (2), 321–324.

[4] T.G.Öberg (2002) Prediction of Vapour Pressures for Halogenated Diphenyl Ether Congeners from Molecular Descriptors, ESPR - Environ Sci & Pollut Res., 9 (6), 405 - 411.

[5] H.Henschel, J.C.Acosta Navarro, T.Yli-Juuti, O. Kupiainen-Määttä, T.Olenius, I.K.Ortega, S.L. Clegg, Th. Kurtén, I. Riipinen, H. Vehkamäki (2014) Hydration of Atmospherically Relevant Molecular Clusters: Computational Chemistry and Classical Thermodynamics, J. Phys. Chem. A, 118, 2599−2611.

[6] O.Pelkonen, D.D.Breimer (1994) Role of Environmental Factors in the Pharmacokinetics of Drugs: Considerations with Respect to Animal Models, P-450 Enzymes, and Probe Drugs. Chapter 10. In P.G. Welling et al (eds.) Pharmacokinetics of Drugs. Springer-Verlag Berlin, Heidelberg 1994. Volume 110 of the series Handbook of Experimental Pharmacology, p. 289-332.

[7] H.Peng, B.Cheung (2009) A review on pharmacokinetic modeling and the effects of environmental stressors on pharmacokinetics for operational medicine: Operational pharmacokinetics. Defence R&D Canada – Toronto Technical Report DRDC Toronto TR 2009-037.

[8] G.Ginsberg, D.Hattis, A.Russ, B.Sonawane (2005) Pharmacokinetic and Pharmacodynamic Factors That Can Affect Sensitivity to Neurotoxic Sequelae in Elderly Individuals, Environmental Health Perspectives, 113(9), 1243-1249.

[9] G.H.I.Wolfgang, D.E.Johnson (2002) Web resources for drug toxicity, Toxicology, 173, 67–74.

[10] A. A. Toropov, A.P.Toropova, Jr.Ivan Raska, D. Leszczynska, J.Leszczynski (2014) Comprehension of drug toxicity: Software and databases, Computers in Biology and Medicine, 45, 20–25.

[11] J.R.Rabinowitz, Michael-Rock Goldsmith, S.B. Little, M.A.Pasquinelli (2008) Computational Molecular Modeling for Evaluating the Toxicity of Environmental Chemicals: Prioritizing Bioassay Requirements, Environ Health Perspect, 116(5), 573–577. Published online 2008 February 1. doi: 10.1289/ehp.11077

[12] J.Kostal, A.Voutchkova-Kostal, P.T.Anastas, J. Beth Zimmerman (2004) Identifying and designing chemicals with minimal acute aquatic toxicity, Procedings Natl. Acad. Sci, 112(20) 6289-6294.

[13] R.S.Judson, K.A.Houck, R.J.Kavlock, Th.B. Knudsen, M.T.Martin, H.M.Mortensen, D.M.Reif, D. M.Rotroff, I.Shah, A.M.Richard, D.J.Dix (2010) In Vitro Screening of Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project, Environmental Health Perspectives, 118(4), 485-492.

[14] A.Gissi, A.Lombardo, A.Roncaglioni, D.Gadaleta, G. Felice Mangiatordi, O.Nicolotti, E.Benfenati (2015) Evaluation and comparison of benchmark QSAR models to predict Relevant REACH endpoint: The bioconcentration factor (BCF), Environmental Research, 137, 398–409.

[15] T.Hou, Y.Li, W.Zhang, J.Wang (2009) Recent developments of in silico predictions of intestinal absorption and oral bioavailability, Comb Chem High Throughput Screen, 12, 497-506.

[16] J.Wang, T.Hou (2011) Recent advances on aqueous solubility prediction, Comb Chem High Throughput Screen , 14, 328-338.

[17] A.Talevi, M.Goodarzi, E.V.Ortiz, P.R.Duchowicz, C.L.Bellera, G.Pesce, E.A.Castro, L.E.Bruno-Blanch (2011) Prediction of drug intestinal absorption by new linear and non-linear QSPR, Eur J Med Chem, 46, 218-228.

[18] C.A.Larregieu, L.Z.Benet (2013) Drug discovery and regulatory considerations for improving in silico and in vitro predictions that use Caco-2 as a surrogate for human intestinal permeability measurements, AAPSJ, 15, 483-497.

[19] P.M.Bahadduri, J.E.Polli, P.W.Swaan, S.Ekins (2010) Targeting drug transporters - combining in silico and in vitro approaches to predict in vivo, Methods Mol Biol, 637, 65-103.

[20] A.Roncaglioni, A.A.Toropov, A.P.Toropova, E. Benfenati (2013) In silico methods to predict drug toxicity, Current Opinion in Pharmacology, 13, 802–806.

[21] Ph.Baveye, Ph.Vandevivere, B.L.Hoyle, P.C. De Leo, D.S. de Lozada (1998) Environmental Impact and Mechanisms of the Biological Clogging of Saturated Soils and Aquifer Materials, Critical Reviews in Environmental Science and Technology, 28(2), 123–191.

[22] H.Choi, E.Stathatos, D.D.Dionysiou (2006) Sol–gel preparation of mesoporous photocatalytic TiO_2 films and TiO_2/Al_2O_3 composite membranes for environmental applications, Applied Catalysis B: Environmental, 63, 60–67.

[23] Y.Shai (1999) Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by K-helical antimicrobial and cell non-selective membrane-lytic peptides, Biochimica et Biophysica Acta, 1462, 55-70.

[24] W.Kujawski (2000) Application of Pervaporation and Vapor Permeation in Environmental Protection, Polish Journal of Environmental Studies, 9(1), 13-26.

[25] N.B.Chapman, J.Shorter (1978) Correlation Analysis in Chemistry; Plenum Press, New York.

[26] C. Hansch, A. Leo, R. W. Taft (1991) A Survey of Hammett Substituent Constants and Resonance and Field Parameters, Chem. Rev, 97, 165-195.

[27] S.M.Free, J.W.Wilson (1964) A Mathematical Contribution To Structure-Activity Studies, J. Med. Chem., 7, 395-399.

[28] C.Hansch, T.Fujita (1964) p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure, J. Am. Chem. Soc, 86(8), 1616-1626.

[29] S.Kar, R.K.Kunal (2010) Predictive toxicology using QSAR: a perspective. J Indian Chem Soc, 87,1455-1515.; S. Kar , K.Roy (2011) Development and validation of a robust model for prediction of carcinogenicity of drugs, Indian J Biochem Biophys, 48, 111-122.

[30] N.G.Bakhtyari, G.Raitano, E.Benfenati, T.Martin, D.Young (2013) Comparison of in silico models for prediction of mutagenicity, J Environ Sci Health C

Environ Carcinog Ecotoxicol Rev, 31, 45-66; C.Milan, O.Schifanella, A.Roncaglioni, E.Benfenati (2011) Comparison and possible use of *in silico* tools for carcinogenicity within REACH legislation, J Environ Sci Health C Environ Carcinog Ecotoxicol Rev, 29, 300-323.

[31] A.Hillebrecht, W.Muster, A.Brigo, M.Kansy, T.Weiser, T.Singer (2011) Comparative evaluation of *in silico* systems for ames test mutagenicity prediction: scope and limitations, Chem Res Toxicol, 24, 843-854.

[32] A.Worth, M.Fuart-Gatnik, S.Lapenna, E.Lo Piparo, A.Mostrag-Szlichtyng, R.Serafimova (2011) The Use of Computational Methods in the Toxicological Assessment of Chemicals in Food: Current Status and Future Prospects, Scientific and Technical Research Series EUR 24748 EN. Luxembourg: Publications Office of the European Union;

[33] J.G.Diaz Ochoa, J.Bucher, A.R.R.Pe´ry, J.M.Zaldivar Comenges, J.Niklas, K.Mauch (2013) A multi-scale modeling framework for individualized, spatiotemporal prediction of drug effects and toxicological risk, Front Pharmacol, 3, 204.

[34] A.A.Frid, E.J.Matthews (2010) Prediction of drug-related cardiac adverse effects in humans-B: use of QSAR programs for early detection of drug-induced cardiac toxicities, Regul Toxicol Pharm, 56, 276-289. L.C.Huang, X.Wu, J.Y.Chen (2011) Predicting adverse side effects of drugs, BMC Genomics , 12(5), 11.

[35] E.Myshkin, R.Brennan, T.Khasanova, T.Sitnik, T.Serebriyskaya, E.Litvinova, A.Guryanov, Y.Nikolsky, T.Nikolskaya, S.Bureeva (2012) Prediction of organ toxicity endpoints by QSAR modeling based on precise chemical-histopathology annotations, Chem Biol Drug Des, 80, 406-416.

[36] S.S,Choi, J.S.Kim, L.G.Valerio, N.Sadrieh (20130 *In silico* modeling to predict drug-induced phospho-lipidosis, Toxicol Appl Pharmacol, 269, 195-204.

[37] F.Cheng, W.Li, X.Wang, Y.Zhou, Z.Wu , J.Shen, Y.Tang (2013) Adverse drug events: database construction and *in silico* prediction, J Chem Inf Model, 53 (4), 744–752.

[38] A.Vedani, M.Dobler, M.Smiesˇko (2012) Virtual ToxLab — a platform for estimating the toxic potential of drugs, chemicals and natural products, Toxicol Appl Pharmacol, 261, 142-153.

[39] K.L.Dobo, N.Greene, C.Fred, S.Glowienke, J.S. Harvey, C.Hasselgren, R.Jolly, M.O.Kenyon, J.B. Munzner, W.Muster (2012) *In silico* methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: An industry survey, Regul Toxicol Pharmacol, 62, 449-455.

[40] J.F.Contrera (2012) Improved *in silico* prediction of carcinogenic potency (TD50) and the risk specific dose (RSD) adjusted threshold of toxicological concern (TTC) for genotoxic chemicals and pharmaceutical impurities, Regul Toxicol Pharmacol, 59, 133-141; Jr.L.G.Valerio, K.P.Cross (2012) Characterization and validation of an *in silico* toxicology model to predict the mutagenic potential

of drug impurities, Toxicol Appl Pharmacol, 260, 209-221.

[41] J.P.Bercu, S.M.Morton, J.T.Deahl, V.K.Gombar, C.M.Callis, R.B.van Lier (2010) *In silico* approaches to predicting cancer potency for risk assessment of genotoxic impurities in drug substances, Regul Toxicol Pharmacol, 57, 300-306.

[42] A.A.Toropov, A.P.Toropova, I.Jr.Raska, D. Leszczynska, J.Leszczynski (2014) Comprehension of drug toxicity: Software and databases, Computers in Biology and Medicine, 45, 20–25. http://dx.doi.org/10.1016/j.compbiomed.2013.11.013 ;H.I.G. Wolfgang, D. E. Johnson (2002) Web resources for drug toxicity, Toxicology, 173, 67–74. PII: S0300-483X(02)00022-7

[43] R.Todeschini, V.Consonni (2000) Handbook of molecular descriptors. Wiley-VCH, Weinheim p. XI

[44] A.M.Ferguson, T.Heritage, P.Jonathon, S.E.Pack, L.Phillips, J.Rogan, P.J.Snaith (1997) A new theoretically based molecular descriptor for use in QSAR/QSPR analysis, Journal of Computer-Aided Molecular Design, 11, 143–152.

[45] V.Consonni, R.Todeschini (2010) Molecular Descriptors in T. Puzyn et al. (eds.), *Recent Advances in QSAR Studies*, 29–102. Springer, Dordrecht, Heidelberg, London, New York.

[46] C.Liang, D.A.Gallagher (1998) An integrated software system for modeling organic molecules, J. Chem. Inf. Comput. Sci., 38 (2), 321–324 (1998)

[47] Y.Xue, Z.R..Li, C.W.Yap, L.Z.Sun, X.Chen, Y.Z.Chen (2004) Effect of Molecular Descriptor Feature Selection in Support Vector Machine Classification of Pharmacokinetic and Toxicological Properties of Chemical Agents, J. Chem. Inf. Comput. Sci., 44, 1630-1638

[48] L.Pfendt, G.Popović, M.Kostić, Lj.Stevović, B. Drakulić, I.Juranić (2000) Determination and structural correlation of the pK_a values of *p*-substituted trans-2,3-epoxy-4-oxo-4-phenylbutanoic acids", J. Serb. Chem. Soc., 65(11), 781-788.

[49] MolSurf v3.11 (Paper III) and MolSurf v98 (Paper IV), Qemist AB, Hertig Carls allé 29, SE-691 41 Karlskoga, Sweden. http://members.nbci.com/ Qemist. HYBOT 98, pION Inc,5 Constitution Way, Woburn, MA 01801, USA.

[50] J.A.Padron, R.Carasco, R.F.Pellon (2002) Molecular descriptor based on a molar refractivitz partition using Randic-type graph/theoretical invariant, J. Pharm. Pharmaceut. Sci., 5(3), 258-266; BioByte v4.0, BioByte Corp. 201 W. Fourth Street, Suite #204, Claremont, CA 91711- 4707, USA; Sybyl v6.5 Update 3, Tripos Inc., 1699 S Hanley Road, Suite 303, St. Louis, MO 63144-2913, USA.

[51] L.H.Hall, L.B.Kier (1995) Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological and Valence State Information, J. Chem. Inf. Comput. Sci., 35, 1039-1045; Molconn-Z v3.15S, Hall Associates Consulting, Quincy, Mass., USA.

[52] G.Chang, W.C.Guida, W.C.Still (1989) An internal coordinate Monte Carlo method for searching conformational space, J. Am. Chem. Soc., 111,

4379-4386; F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C.Still (1990) Macro Model—An integrated software system for modeling organic and bioorganic molecules using molecular mechanics, J. Comp. Chem., 11,440-467; S.J. Weiner , P.A. Kollman, D.A. Case, C. Ghio, G. Alagona, S. Profeta, P.K. Weiner (1984) A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins, J. Am. Chem. Soc., 106,765-784; Concord v3.2.4, Tripos Inc., 1699 S Hanley Road, Suite 303, St. Louis, MO 63144-2913, USA.

[53] MAREA v1.4. The program MAREA is available upon request. The program is provided free of charge for academic users. Contact Johan Gråsjö (email: johan.grasjo@galenik.uu.se).; K.B. Lipkowitz, B. Baker, R. Larter (1989) Dynamic molecular surface areas, J. Am. Chem. Soc., 111, 7750-7753.

[54] I.Gutman (2005) Topology and stability of conjugated hydrocarbons. The dependence of total π-electron energy on molecular topology (Review), J. Serb. Chem. Soc., 70(4), 441-456.

[55] T.G.Öberg (2002) Prediction of Vapour Pressures for Halogenated Diphenyl Ether Congeners from Molecular Descriptors, ESPR - Environ Sci & Pollut Res, 9 (6) 405 - 411.

[56] a) R.D.Cramer, D.Paterson, J.D.Bunce (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc., 110, 5959-5967; b) R.D.Cramer (2003) Topomer CoMFA: A Design Methodology for Rapid Lead Optimization , J. Med. Chem., 46, 374-388.

[57] I.Juranić (2014) Simple Method for the Estimation of pKa of Amines, Croat. Chem. Acta, 87(4), 343–347.

IZVOD

RAČUNARSKI GENERISANI MOLEKULSKI DESKRIPTORI KAO *PROXI*-JI ZA MODELOVANJE MATERIJALA I NJIHOVOG UTICAJA NA OKOLINU

Za predviđanje osobina materijala i njihove interakcije sa okolinom, treba poznavati njihove osobine. Obično se osobine molekula otkrivaju eksperimentalnim merenjima. To može biti mukotrpan dugotrajan i skup posao. Sa druge strane, računarska hemija lako daje veliki broj podataka koji mogu da obezbede različite deskriptore molekula. To mogu biti razne merljive veličine (dužine i uglovi veza, dipolni momenti, i sl...), ali i nemerljive osobine (parcijalna atomska naelektrisanja, elektronegativnost, razne latentne varijable ....).

Postoji velika potreba za razvijanjem pouzdanih i ekonomičnih načina za skringe kojima se predskazuje potencijalna otrovnost i opterećenje životne okoline raznim hemikalijama. Jednako važan je i dizajn bezbednijih alternativa.

Metode molekulskog modelovanja nude jedan od nekoliko komplementarnih pristupa za procenu rizika za zdravlje ljudi i životne sredine kao posledicu izlaganja hemikalijama u okolišu. Ovim postupcima se može neprekidno vršiti procena opasnosti simuliranjem mogućih načina delovanja, a obezbeđivanje virtualnog skrininga može pomoći u određivanju prioriteta kod bio-eseja. Ukrajanjem ovih strategija u određene izazove interakcija hemikalija i životne sredine može iste učiniti efikasnijima.

Napredak u računarskoj hemiji i molekulskoj toksikologiji postignut poslednjih decenija dozvoljava razvoj prediktivnih modela za racionalni dizajn molekula sa umanjenim potencijalom otrovnosti za ljude ili za životnu sredinu.

Kao primer predstavljamo novu metodologiju za računarsko procenjivanje pK_a vrednosti različitih organskih baza na osnovu izračunavanja parcijalnih atomskih naelektrisanja prostim semiempirijskim QM metodom.

Ključne reči: *izračunavanje osobina molekula, molekulski deskriptori, QSAR/QSPR, parcijalna atomska naelektrisanja.*