UDC 544.165: 615.917

Studying and Forecasting the Acute Toxicity of Hydrocarbons and Halogenated Derivatives Thereof Using a QSAR Method

O. V. TINKOV[1], P. G. POLISHCHUK[2] and V. E. KUZMIN[1,2]

[1]*Shevchenko State University of Pridnestrovie, Ul. 25 Oktyabrya 128, Tiraspol MD-3300 (Moldova)*

E-mail: tinkov84@mail.ru

[2]*Bogatskiy Physiochemical Institute, NAS of Ukraine, Lyustorfskaya Doroga 86, Odessa 65080 (Ukraine)*

(Received March 27, 2012; revised April 27, 2012)

Abstract

An influence of the molecular structure of hydrocarbons and halogenated derivatives thereof upon the acute toxicity was studied with the use of a 2D simplex representation of the molecular structure and partial least squares technique. Adequate QSAR models were obtained, wherefrom structural fragments have been revealed those could cause the toxicity level to increase. With the help of QSAR models a relative influence of some physicochemical factors upon the acute toxicity was assessed.

Key words: acute toxicity, hydrocarbons and their halogenated derivatives, modelling

INTRODUCTION

Hydrocarbons and halogenated derivatives thereof are widely used in the industry as solvents, refrigerants, intermediates in chemical reactions, fumigants and pesticides. In the course of determining the synthesis biological test strategy, alongside with a target effect it is important to evaluate a toxic effect, in particular, acute toxicity level (LD_{50}). In the world there are known more than 2 million chemical compounds, and for a large part of them there in no information available concerning the most important toxicological characteristics those are required in order to determine their hazard. The costs for the experimental determination of these parameters are not always feasible and available. However, with the help of mathematical modelling and computational methods one could reveal the most efficient and safe products at the early stages of research work. At the same time, the risk of late revealing an adverse effect of these preparations on humans and the environment could be reduced [1].

In connection with the above mentioned, of urgent importance are studies based on the methods of QSAR, wherein the toxicity of organic compounds depending on the structure thereof is modelled. Further, the QSAR models obtained are used in order to predict toxic properties inherent in novel, not yet synthesized chemical compounds.

This work consisted in constructing the mathematical models for the prediction of acute toxicity level inherent in hydrocarbons and their halogenated derivatives and in studying the influence of structural fragments upon the acute toxicity level of hydrocarbons and their halogenated derivatives.

MATERIALS AND METHODS

The information concerning the acute toxicity level and chemical structure of hydrocar-

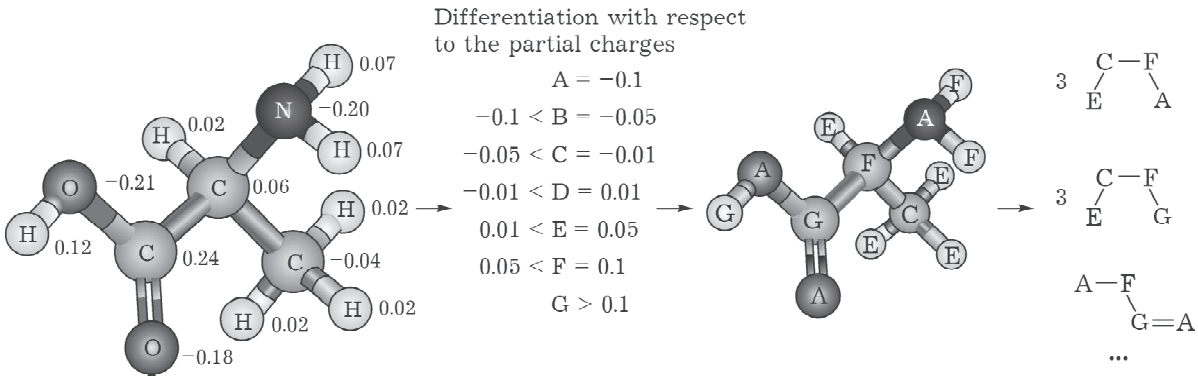


Fig. 1. Example of the 2D generation of simplex descriptors for alanine at 2D level with using the differentiation of atoms with respect to partial atomic charges.

bons and their halogenated derivatives for constructing the mathematical models was obtained by means of processing the electronic database management system Toxic v.1.1.5 [2]. Basing on data available we formed a sampling of 115 compounds with known data concerning the acute toxicity level (LD_{50}) for oral administering to rats.

In order to describe the molecular structure we used 2D simplex descriptors [3], whose calculation is realized in a TheorChem software package. Within the framework of the simplex representation of the molecular structure the molecule is considered a system of different simplexes, *i. e.* four-atomic molecular fragments with a fixed structure (Fig. 1). A descriptor in this case is presented by the number of one type of simplexes. On a 2D level, the atoms (the vertices of the simplex) are differentiated not only with respect to the chemical elements of the atoms, but also taking into account the different physical and chemical properties such as the partial charge at the atom, the atomic lipophilicity, refraction, ability to act as a donor or acceptor of hydrogen in hydrogen bonding.

Adequate QSAR models used in order to predict the activity of such not yet studied molecules, should exhibit high statistical characteristics. At the same time, a good set of statistical characteristics does not mean that the model would have a high predictability. In order to assess the predictability a method of is used the forming an external test sampling (test set, or ts). In the course of this process, a part of the molecules (50 %) is excluded from the process of constructing the model, whereas basing on the remaining molecules, a training sampling (work set, or ws) is formed. After the constructing the model is used in order to predict the activity of compounds in the test sampling.

Owing to a small amount of esters under investigation and their structural diversity, we performed a fivefold external cross-validation. For this purpose, all the compounds of the training sampling were divided into five groups in such a way that the compounds within each group were distributed in a similar manner according to the level of toxicity under investigation. To do this, all the compounds were previously sorted according to the level of their toxicity to select every fifth compound into a separate group. Then the four groups obtained from the five were joined together to construct a QSAR model for the sampling obtained, whereas the remaining fifth group of compounds was used as a test sampling. This procedure was repeated five times so that all the compounds appeared once in the test sample.

For each of the five generated sets of the training and testing samplings we constructed mathematical models, further combined into a consensus model. In this model, the activity values calculated correspond to the average activity values calculated with the use of the models described above.

In order to assess the stability of the models obtained we used a cross-validation procedure wherein each compound is removed from the sampling to predict the value of activity level for this compound. The coefficient of determination Q^2 calculated within the framework of the cross-validation procedure, was calculated according to the following formula

$$Q^2 = \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 - \sum_{i=1}^{m}(y_i - \hat{y}_i^{CV})^2}{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (1)$$

where m is the number of molecules in the sampling; y_i is the preset activity value; $\hat{y}_i$ is the calculated activity value; $\hat{y}_i^{CV}$ is the activity value calculated by means of cross-validation procedure for the i-th molecule.

The quality of predictions based on QSAR models was estimated by means of the coefficient of determination for the external test sampling R^2_{test}:

$$R^2_{\text{test}} = 1 - \frac{\sum_{i}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \hat{y}_{i\,\text{cp}})^2} \quad (2)$$

where y_i is the observed activity value for the i-th compound of the test sampling; $\hat{y}_i$ is the predicted value of the activity for the i-th compound of the test sampling; y_i^{av} is the average activity level for all the compounds of the training sampling.

The standard error of prediction (S) is determined as it follows:

$$S = \sqrt{\frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{m-1}} \quad (3)$$

In order to perform a more rigorous assessment of the statistical significance of the models we have found the coefficients of determination with the use of randomization procedures. The latter consists in the formation of a sampling, wherein the toxicity values are randomly shuffled, whereas the values of descriptors remain constant. After that, the construction of 100 models is performed. The fact that the average statistic indicators for these models are significantly lower as compared to the original one demonstrates the original models to be statistically reliable.

As a statistical method for constructing QSAR models we have chosen a method of partial least squares (PLS) [4].

The calculation of structural fragment contributions to the value of acute toxicity level LD_{50} was performed with the use a Lattice&Simplex Modelling (LSM) program included in the a TheorChem software package.

A detailed description of QSAR methods is presented in [5, 6].

RESULTS AND DISCUSSION

In the course of the studies we obtained a number of adequate QSAR models (Table 1) describing the mathematical relationship between the structure and acute toxicity of hydrocarbons and halogenated derivatives thereof in the case of oral administration the drugs to rats.

In order to reveal molecular fragments those promote enhancing the toxicity of chemical compounds we determined the contributions of

TABLE 1

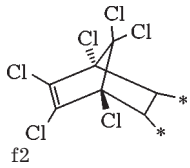
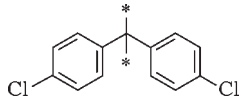
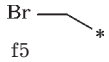
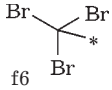
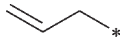
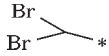
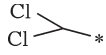
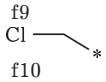
Statistical characteristics of QSAR models for acute toxicity

Models	R^2	Q^2	R^2_{test}	S_{WS}	S_{CV}	S_{ts}	A	R^{2*}
1	0.837	0.722	0.678	0.335	0.441	0.525	3	0.164
2	0.902	0.836	0.673	0.257	0.335	0.482	3	0.210
3	0.907	0.816	0.791	0.253	0.361	0.367	3	0.206
4	0.826	0.775	0.618	0.347	0.396	0.497	3	0.091
5	0.864	0.804	0.733	0.306	0.370	0.421	3	0.129
6 (consensus)	0.885	–	0.700	0.228	–	0.451	–	–

Note. R^2 is coefficient of determination; Q^2 is coefficient of determination, calculated within the framework of the cross-validation procedure; R^2_{test} is coefficient of determination for external test sampling; S_{ws} is standard error of prediction for the training sampling (working set, ws); S_{CV} is standard error calculated within the framework of the cross-validation procedure; S_{ts} is standard prediction error for external test sample, A is the number of latent variables; R^{2*} is the average coefficient of determination calculated within the framework of randomization.

TABLE 2

Contributions of potential toxofors to the acute toxicity of hydrocarbons and halogenated derivatives thereof

Fragments	Number of compounds with the fragment	Contribution		
		Minimum	Maximum	Average
 f1	1	−1.342	−1.342	−1.342
 f2	8	−1.673	−0.816	−1.311
 f3	4	−1.371	−0.915	−1.178
 f4	1	−0.990	−0.990	−0.990
 f5	10	−0.879	−0.435	−0.624
 f6	1	−0.609	−0.609	−0.609
 f7	2	−0.756	−0.323	−0.539
 f8	2	−0.502	−0.502	−0.502
 f9	9	−0.518	−0.233	−0.321
 f10	17	−0.332	−0.027	−0.170

Notes. 1. The lowest contribution value corresponds to the most toxic fragments. 2. "Asterisk" indicates the location of bonding the certain piece with any part of the molecule.

individual fragments to the value of LD_{50} on the basis of the consensus model. Further, we have chosen the fragments therefrom those unambiguously promote increase the toxicity level, *i. e.*, those act as potential toxofors (Table 2). The information presented would allow performing a preliminary selection of potentially hazardous compounds.

According to the results obtained, a significant contribution to hydrocarbon toxicity manifestation is drawn by the fragments containing halogens. Moreover, among the structurally similar halogenated fragments, the greatest contribution to the toxicity level is made by fragments containing iodine. The value of toxicity contribution for hydrocarbon fragments containing fluorine, chlorine, bromine and iodine, increases in the order of F < Cl < Br < I. This is confirmed by data from [7] according to that the mechanism of the action of these toxic compounds consists in biological dehalogenation thereof in an organism. The dehalogenation ability is determined by carbon-halogen bond strength that which decreases in passing from fluorine to iodine. Compounds containing halocarbon fragments can form free radicals (by

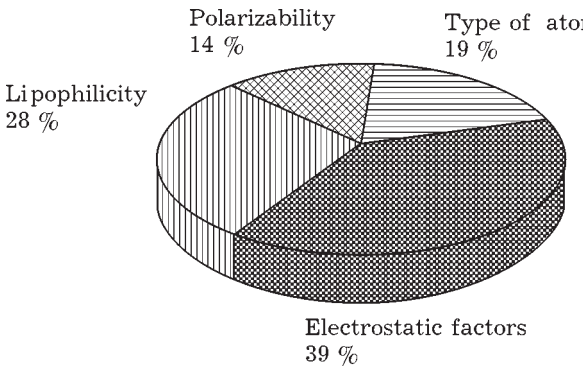


Fig. 2. Relative effect of physicochemical characteristics inherent in the compounds under investigation on the half-lethal dose for oral administration for rats.

enzymatic or non-enzymatic way) those alkylate the reactive groups of proteins and enzymes, giving rise to corresponding pathologies.

From the presented results of the analysis one can see that the fragments containing polychlorinated polycyclic diene radicals (fragment f2) and dichlorodiphenylmethane radicals (fragment f3) are characterized by a high contribution to the toxic action of chemical compounds. The compounds containing fragments f2 and f3 are usually referred to organochlorine ecotoxicants, in particular, to polychlorinated pesticides those exert a stable negative effect on living organisms [8]. These substances exhibit a polytoxic acute effect on an organism, with the disruption of the functioning of different systems and organs in animals and humans [9]. According to the authors of [10], the compounds those contain fragments f2 and f3, exhibit immunotoxicity. A proposed mechanism of the action of polychlorinated pesticides could be connected with the alkylating action thereof on the basic biological targets such as amino acids, peptides, proteins, nucleic acids, lipids.

Basing on the consensus model obtained, an analysis was performed concerning the relative influence of the physicochemical characteristics of hydrocarbons and halogenated derivatives upon acute toxicity thereof (Fig. 2).

In the course of the analysis of the impact of the physicochemical characteristics of atoms forming the hydrocarbons and halogenated derivatives on the acute toxicity level we revealed a dominant role of electrostatic factors in the interaction between toxicants and a target. In addition, the role lipophilicity is quite considerable, which could be connected, to all appearance, with the process of the passive diffusion of toxicants through the cell membrane cell in the course of absorption in tissues and organs.

CONCLUSION

As a result of the studies performed a number of adequate QSAR models were obtained based on 2D simplex descriptors involving a statistical method of partial least squares. It has been found that these models generally exhibit acceptable statistical parameters and predictive ability. Basing on the QSAR models the fragments were determined those cause the toxicity of hydrocarbons and their halogenated derivatives to increase. The resulting information could be useful in the course of performing the molecular design of novel compounds with a low toxicity level. For some molecular fragments those cause increasing the toxicity level, possible mechanisms of toxic pathogenesis were considered.

This work does not pretend to be exhaustively disclosing the issue. It is appropriate to continue this research to extend the sampling of compounds under investigation in order to construct QSAR models for revealing the most important toxofors.

REFERENCES

1 Kirlan A. V., Planirovaniye Napravleniy Sinteza Biologicheski Aktivnykh Oksi- i Amidosoderzhashchikh Geterotsiklicheskikh Soyedineniy s Uchetom Toksichnosti (Candidate's Dissertation in Chemistry), Ufa, 2003.

2 Tinkov O. V., Polishchuk P. G., Artemenko A. G., Ognichenko L. N., Kuzmin V. E., *Vestn. Pridnestr. Gos. Un-ta*, 2 (2011) 112.

3 Kuz'min V. E., Artemenko A. G., Chelombitko V. A., Muratov E. N., Zheltvay A. I., Meshcheryakov A. K., Lyakhovskiy A. V., Nauch. Sem. "Svyaz' Strukrura–Aktivnost' Biologicheski Aktivnykh Veshchestv" (Proceedings), Gurzuf, 2002, pp. 22–26.

4 Rдnnar S., Lindgren F., Geladi P., Wold S., *J. Chemometrics*, 8 (1994) 111.

5 Kuz'min V. E., Artemenko A. G., Muratov E. N., Polischuk P. G., Ognichenko L. N., Liahovsky A. V., Hromov A. I. and Varlamova E. V., in: Recent Advances in QSAR Studies, in T. Puzyn, J. Leszczynski, M. T. D. Cronin (Eds.), Springer, London, 2010, pp. 127–176.

6 Polishchuk P. G., Muratov E. N., Artemenko A. G. , Kolumbin O. G., Muratov N. N. and Kuz'min V. E., *J. Chem. Inf. Model*, 49 (2009) 2481.

7 Lazareva N. V., Levina E. N., Vrednye Veshchestva v Promyshlennosti, Khimiya, Leningrad, 1976, vol. 1, p. 190.
8 Isidorov V. A., Vvedeniye v Khimicheskuyu Ekotoksikologiyu, Khimizdat, St. Petersburg, 1999, pp. 90–97.
9 Ling L. J., Clark R. F., Erickson T. B., Trestrail J. H. (Eds.), Toxicology Secrets, Hanley&Belfus, 2001.
10 Kurlyandskiy B. A., Filov V. A., Obshchaya Toksikologiya, Meditsina, Moscow, 2002, pp. 368–369.