Binding organophosphate pesticides to acetylcholinesterase: Risk assessment using the Monte Carlo method

Alla P. Toropova (alla.toropova@marionegri.it)
Istituto di Ricerche Farmacologiche Mario Negri IRCCS https://orcid.org/0000-0002-4194-9963
Andrey A. Toropov
Istituto di Ricerche Farmacologiche Mario Negri
Alessandra Roncaglioni
Istituto di Ricerche Farmacologiche Mario Negri
Emilio Benfenati
Istituto di Ricerche Farmacologiche Mario Negri

Research Article

Keywords: Organophosphate, Pesticide Toxicity, QSAR, Risk Assessment, SMILES, Monte Carlo method

Posted Date: July 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1744436/v1

Abstract

Binding to acetylcholinesterase may cause toxic effects in humans. Organophosphates simultaneously are both dangerous and useful substances: dangerous since they are employed in chemical warfare; and useful when they are applied as pesticides. Here we suggest the models for organophosphates binding to acetylcholinesterase developed via representing the molecular structure by a simplified molecular input-line entry system (SMILES) using so-called optimal SMILES-based descriptors calculated with the Monte Carlo technique using the CORAL software available on the Internet (http://www.insilico.eu/coral). Quantitative structure-activity relationships (QSARs) serve to develop predictive models for organophosphates. The predictive potential of these models is quite good: the determination coefficient for the validation set ranged from 0.87 to 0.90. These models were built up according to the principle "QSAR is a random event", i.e. predictive potential of an approach should be checked up with several splits of available data into the training and test sets. The special scheme of mechanistic interpretation definition is represented. The mechanistic interpretation is based on probabilities of molecular features to be in the sub-group of promoters of increase for endpoint or in sub-group of promoters of its' decrease.

Introduction

Pesticides are widely used in agriculture for plant protection and to increase the yields and quality of agricultural products, but also in domestic applications (Naik et al. 2009). Organophosphate compounds (OPC) are the widely used choice as insecticides and are widely used across the world. For this reason, these substances are frequently encountered in surface water and groundwater due to their agricultural use. OPC are also used for ornamentals, in floriculture, and for residential treatments. Unfortunately, despite these advantages, these pesticides are highly toxic and also bio accumulate (Torres-Palma and Serna-Galvis 2018; Bora et al. 2020). Exposure to OPCs can affect the immune response and carbohydrate metabolism, and cause cardiovascular toxicity (Marimuthu et al. 2019).

Residues of these pesticides usually build up on the surface of the soil and ultimately find their way into food products. They may have adverse impacts on various groups of soil microbes, wild animals, and also humans, through environmental contamination, or residues in food and feedstuff (Torres-Palma and Serna-Galvis 2018; Bora et al. 2020; Toropov et al. 2017, 2018). Experimental data on these chemicals can be used to improve the design of new pesticides, moving towards safer, greener substances, considering their adverse effects on humans and ecosystems (Mudasir et al. 2013; Wibowo et al. 2017). Furthermore, organophosphates are chemical warfare tools (Mendonca et al. 2020).

Quantitative structure-activity relationships (QSARs) serve to develop predictive models for the toxicity of compounds in general and organophosphates in particular (Niraj et al. 2015; Marimuthu et al. 2019; Bora et al. 2020; Maxwell et al. 1992; Ruark et al. 2011; Zhao and Yu 2013; Morrill et al. 2015; Lee and Barron 2016). The CORAL software (http://www.insilico.eu/coral) has been used up as a tool to develop QSAR models for different endpoints (Toropov et al. 2017, 2018; Veselinović et al. 2015; Kumar and Chauhan

2017; Kumar et al. 2019; Islam and Pillay 2016). Here we studied the ability of this software to build up predictive models for toxicity OPC binding to acetylcholinesterase.

Method

Data

A database of 278 molecular structures together with numerical data on toxicity represented by rate constants for human binding to acetylcholinesterase (AChE k1) were taken from the literature (Ruark et al. 2013). These substances have randomly distributed into an active training set (≈ 25%) which has used to construct the model, a passive training set (≈ 25%) that independent of those used for building the model, a calibration set (≈ 25%) is to identify situations under which the model is over-training, and a validation set (≈ 25%) is to provide the final assessment of the statistical quality of the model (Toropov et al. 2020). The molecular structure was represented with simplified molecular input-line entry systems (SMILES) (Weininger 1988).

Optimal SMILES-based descriptors

The optimal SMILES-based descriptor of the correlation weights (*DCW*) applied here is calculated as follows:

$$DCW_{SMILES}\left(T^*, N^*\right) = CW(APP) + \sum CW(S_k) + \sum CW(SS_k)$$

1

Correlation weights (*CW*) are calculated with the Monte Carlo method using the CORAL software (http://www.insilico.eu/coral). *T* is an integer to define rare fragments and active ones. *N* is the number of epochs of the Monte Carlo optimization. *APP* is atom pair's proportions self-organized vector.

Table 1 gives examples of *APP*. S_k is SMILES atoms, i.e. one symbol or a group of symbols that cannot be examined separately ('Br', '%11', '@@). SS_k is fragments of SMILES containing two SMILES atoms. *CW(APP)*, $CW(S_k)$, and $CW(SS_k)$ are correlation weights for these SMILES attributes.

The *T** is the threshold, i.e. a value to discriminate molecular features extracted from SMILES into two categories rare (frequency in the active training set is less than *T**) and active (frequency in the active training set is equal or better than *T**). The *N** is the number of epochs of the Monte Carlo optimization.

[Table 1 around here]

Monte Carlo optimization

QSAR-models were calculated with the Monte Carlo optimization based on the target functions:

$$TF = r_{AT} + r_{PT} - \left|r_{AT} - r_{PT}\right| * 0.1$$

2

r_{AT} and r_{PT} are correlation coefficients between observed and predicted endpoint for the active training and passive training sets, respectively. The maximum given by Eq. 2 is the target of the Monte Carlo calculations. Having the corresponding numerical data in the correlation weights, the endpoint is calculated as follows:

$$humanAChEK_1 = C_0 + C_1 \times DCW_{SMILES}\left(T^*, N^*\right)$$

3

the C_0 and C_1 are regression coefficients.

Results And Discussion

Domain of applicability

The domain of applicability for these compounds is defined according to so-called SMILES defects (Toropova et al. 2020). The number of potential outliers in the validation set is 6, 7, and 5 respectively for splits 1, 2, and 3. The SMILES defect is not a rigid categorization of compounds but only indicates that the representability of fragments of the corresponding SMILES in the active training set (Toropova et al. 2020) is not enough. Figure 1 gives examples of the statistical outliers. The outliers are not all similar for different splits though some are the same (e.g. #1 for splits 1 and 2; #198 for splits 2 and 3).

[Figure 1 around here]

Mechanistic interpretation

[Table 2 around here]

Models

Computational experiments with three random splits give the following models:

Split 1

$humanAChEK_1$ = -0.9042(± 0.0411) + 0.6609(± 0.0043) * DCW(1,15) (4)

Split 2

$humanAChEK_1$ = 1.0303(± 0.0275) + 1.0200(± 0.0063) * DCW(1,15) (5)

Split 3

$humanAChEK_1$ = 0.9458(± 0.0239) + 0.8805(± 0.0050) * DCW(1,15) (6)

Table 3 contains the statistical characteristics of models for three random splits. There is dispersion of the statistical characteristics, but not too much. In other words, the models suggested are robust and can be used for practical applications.

[Table 3 around here]

Comparison with models from the literature

Table 4 compares the statistical quality of models suggested here with some other models suggested (Ruark et al. 2013). Despite the interest in OPCs (Devillers 2000; Mudasir et al. 2013; Ruark et al. 2013; Bora et al. 2020) it is impossible to find any QSAR studies of the rate constants of human AChE k1. Therefore the only comparison can be the results here and those from the original article (Ruark et al. 2013). In our case, the original set of compounds was split into four sets: active training, passive training, calibration, and validation, while Ruark et al. (2013) report the results on three subsets. However, it is possible to make a comparison. The results of the active training are quite similar. No results on the passive training are available in the original paper, because the authors applied a different approach; however, the values for the passive training used only in our case are higher than those on the training set in the original paper. The values for the calibration and validation sets are always better in our case, particularly for the validation set. The validation set values are of particular interest because they demonstrate that our model is not affected by overtraining and quite probably can be applied for substances not used in the modeling phase.

Besides their good statistical results, our model is simpler, does not require the calculation of molecular descriptors as in Ruark's paper (2013), but only needs the SMILES to represent the molecular structure.

[Table 4 around here]

Supplementary materials section gives technical details of the suggested models.

Conclusions

We have developed new models for AChE binding for OPC that are robust and predictive. These models are accompanied by the domain of applicability and a mechanistic interpretation. The statistical quality of the models examined here is better than in models for the same endpoint obtained with the CODESSA software (Ruark et al. 2013). The approach we adopt is quite simple and uses the freely available CORAL software.

Declarations

Acknowledgments

The authors are grateful for the contribution of the project LIFE-CONCERT contract (LIFE17 GIE/IT/000461) and Optitox (OC/EFSA/SCER/2018/01) for support.

Author contributions

APT was involved in formal analysis, data collection, and curation, investigation, methodology incl. statistics, software, visualization, writing of original draft, and revision. AAT was involved in conceptualization, data curation, formal analysis, investigation, methodology incl. statistics, software, visualization, writing of original draft, review, and editing. AR was involved in conceptualization, formal analysis, visualization, writing of original draft, review, and editing. EB was involved in conceptualization, formal analysis, visualization, review, and editing.

ETHICS DECLARATIONS

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval

The manuscript is not submitted to any other journal.

Consent to participate

The paper did not involve any human participants.

Consent to publish

The submitted work is original.

Funding

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this article (and its supplementary information files).

References

1. Bora A, Crisan L, Borota A, Funar-Timofei S, Ilia G (2020) Ecotoxicological QSAR Modeling of Organophosphorus and Neonicotinoid Pesticides. In: Roy K, Ed (eds) Ecotoxicological QSARs.

Methods in Pharmacology and Toxicology. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0150-1_21

2. Devillers J (2000) Prediction of toxicity of organophosphorus insecticides against the midge, Chironomus riparius, via a QSAR neural network model integrating environmental variables. Toxicol Methods 10(1):69–79. DOI: 10.1080/105172300242562
3. Islam MA, Pillay TS (2016) Simplified molecular input line entry system-based descriptors in QSAR modeling for HIV-protease inhibitors. Chemom Intell Lab Syst 153:67–74. DOI: 10.1016/j.chemolab.2016.02.008
4. Kumar A, Chauhan S (2017) QSAR Differential Model for Prediction of SIRT1 Modulation using Monte Carlo Method. Drug Res 67(3):156–162. DOI: 10.1055/s-0042-119725
5. Kumar P, Kumar A, Sindhu J (2019) Design and development of novel focal adhesion kinase (FAK) inhibitors using Monte Carlo method with index of ideality of correlation to validate QSAR. SAR QSAR Environ Res 30(2):63–80. DOI: 10.1080/1062936X.2018.1564067
6. Lee S, Barron MG (2016) A mechanism-based 3D-QSAR approach for classification and prediction of acetylcholinesterase inhibitory potency of organophosphate and carbamate analogs. J Comput Aided Mol Des 30(4):347–363. DOI: 10.1007/s10822-016-9910-7
7. Marimuthu P, Lee Y-J, Kim B, Seo SS (2019) In silico approaches to evaluate the molecular properties of organophosphate compounds to inhibit acetylcholinesterase activity in housefly. J Biomol Struct Dyn 37(2):307–320. DOI: 10.1080/07391102.2018.1426046
8. Maxwell DM, Brecht KM (1992) Quantitative Structure-Activity Analysis of Acetylcholinesterase Inhibition by Oxono and Thiono Analogues of Organophosphorus Compounds. Chem Res Toxicol 5(1):66–71. DOI: 10.1021/tx00025a011
9. Mendonca ML, Ray D, Cramer CJ, Snurr RQ (2020) Exploring the Effects of Node Topology, Connectivity, and Metal Identity on the Binding of Nerve Agents and Their Hydrolysis Products in Metal-Organic Frameworks. ACS Appl Mater Interfaces 12(31):35657–35675. DOI: 10.1021/acsami.0c08417
10. Morrill JA, Topczewski JJ, Lodge AM, Yasapala N, Quinn DM (2015) Development of quantitative structure activity relationships for the binding affinity of methoxypyridinium cations for human acetylcholinesterase. J Mol Graph Model 62:181–189. DOI: 10.1016/j.jmgm.2015.09.016
11. Mudasir M, Wibowo YM, Pranowo HD (2013) Design of new potent insecticides of organophosphate derivatives based on QSAR analysis. Indones J Chem 13(1):86–93. DOI: 10.22146/ijc.21331
12. Naik PK, Sindhura, Singh T, Singh H (2009) Quantitative structure–activity relationship (QSAR) for insecticides: development of predictive *in vivo* insecticide activity models. SAR QSAR Environ Res 20(5–6):551–566. DOI: 10.1080/10629360903278735
13. Niraj RRK, Saini V, Kumar A (2015) QSAR analyses of organophosphates for insecticidal activity and its in-silico validation using molecular docking study. Environ Toxicol Pharmacol 40(3):886–894. DOI: 10.1016/j.etap.2015.09.021

14. Ruark CD, Hack CE, Robinson PJ, Gearhart JM (2011) Quantitative structure-activity relationships for organophosphates binding to trypsin and chymotrypsin. J Toxicol Environ Health A 74(1):1–23. DOI: 10.1080/15287394.2010.501716
15. Ruark CD, Hack CE, Robinson PJ, Anderson PE, Gearhart JM (2013) Quantitative structure-activity relationships for organophosphates binding to acetylcholinesterase. Arch Toxicol 87(2):281–289. DOI: 10.1007/s00204-012-0934-z
16. Torres-Palma RA, Serna-Galvis EA (2018) Sonolysis. In: Ameta SC, Ameta R (Eds) Advanced Oxidation Processes for Wastewater Treatment: Emerging Green Chemical Technology, Academic Press, pp 177–213. DOI: 10.1016/B978-0-12-810499-6.00007-3
17. Toropov AA, Toropova AP, Benfenati E (2020) QSAR model for pesticides toxicity to Rainbow Trout based on “ideal correlations”. Aquat Toxicol 227:105589. DOI: 10.1016/j.aquatox.2020.105589
18. Toropov AA, Toropova AP, Benfenati E, Dorne JL (2018) SAR for gastro-intestinal absorption and blood-brain barrier permeation of pesticides. Chem Biol Interact 290:1–5. DOI: 10.1016/j.cbi.2018.04.030
19. Toropov AA, Toropova AP, Marzo M, Dorne JL, Georgiadis N, Benfenati E (2017) QSAR models for predicting acute toxicity of pesticides in rainbow trout using the CORAL software and EFSA’s OpenFoodTox database. Environ Toxicol Pharmacol 53:158–163. DOI: 10.1016/j.etap.2017.05.011
20. Toropova AP, Toropov AA, Carnesecchi E, Benfenati E, Dorne JL (2020) The using of the Index of Ideality of Correlation (IIC) to improve predictive potential of models of water solubility for pesticides. Environ Sci Pollut Res 27(12):13339–13347. DOI: 10.1007/s11356-020-07820-6
21. Veselinović AM, Veselinović JB, Živković JV, Nikolić GM (2015) Application of smiles notation based optimal descriptors in drug discovery and design. Curr Top Med Chem 15(18):1768–1779. DOI: 10.2174/1568026615666150506151533
22. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28(1):31–36. https://doi.org/10.1021/ci00057a005
23. Zhao J, Yu S (2013) Quantitative structure-activity relationship of organophosphate compounds based on molecular interaction fields descriptors. Environ Toxicol Pharmacol 35(2):228–234. DOI: 10.1016/j.etap.2012.11.018

Tables

Tables 1 to 4 are available in the Supplementary Files section

Figures

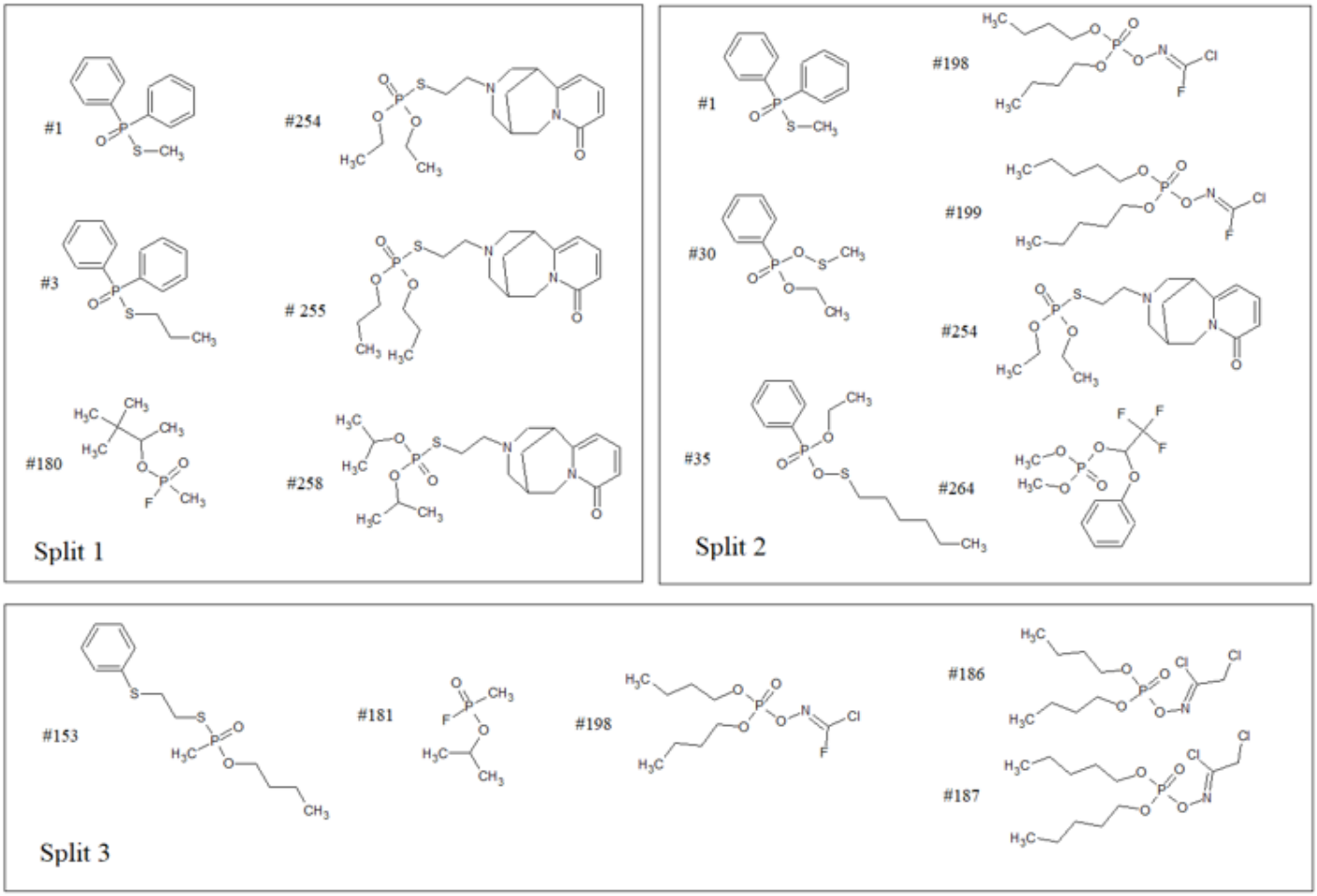

Figure 1

Statistical outliers for the validation set in the cases split 1, 2, and 3

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx
- SupplememntaryMaterials.xlsx