

Wright State University

CORE Scholar

[Browse all Theses and Dissertations](#)

[Theses and Dissertations](#)

2010

Quantitative Structure-Activity Relationships for Organophosphates Binding to Trypsin and Chymotrypsin

Christopher Daniel Ruark
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all



Part of the [Pharmacology, Toxicology and Environmental Health Commons](#)

Repository Citation

Ruark, Christopher Daniel, "Quantitative Structure-Activity Relationships for Organophosphates Binding to Trypsin and Chymotrypsin" (2010). *Browse all Theses and Dissertations*. 350.
https://corescholar.libraries.wright.edu/etd_all/350

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS FOR
ORGANOPHOSPHATES BINDING TO TRYPSIN AND
CHYMOTRYPSIN

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

By

CHRISTOPHER DANIEL RUARK
B.S., Miami University, 2007

2010
Wright State University

WRIGHT STATE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

June 28, 2010

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Christopher D. Ruark ENTITLED Quantitative Structure-Activity Relationships for Organophosphates Binding to Trypsin and Chymotrypsin BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

Jeffery Gearhart, Ph.D.
Thesis Director

James Lucot, Ph.D.
Thesis Co-Director

Mariana Morris, Ph.D.
Department Chair
Pharmacology and Toxicology

Committee on
Final Examination:

Jeffery Gearhart, Ph.D.
Thesis Director

James Lucot, Ph.D.
Thesis Co-Director

Peter Robinson, Ph.D.

John A. Bantle, Ph.D.
Vice President for Research and
Graduate Studies and Interim Dean
Of Graduate Studies

ABSTRACT

Ruark, Christopher, D. M.S., Department of Pharmacology and Toxicology, Wright State University, 2010. Quantitative Structure-Activity Relationships for Organophosphates Binding to Trypsin and Chymotrypsin.

Organophosphate (OP) nerve agents such as sarin, soman, tabun, and O-ethyl S-[2-(diisopropylamino) ethyl] methylphosphonothioate (VX) do not react solely with acetylcholinesterase (AChE). Evidence suggests that a wide range of cholinergic-independent pathways are also targeted, including serine proteases. These proteases comprise nearly one-third of all known proteases and play major roles in synaptic plasticity, learning, memory, neuroprotection, wound healing, cell signaling, inflammation, blood coagulation and protein processing. Inhibition of these proteases by OPs was found to exert a wide range of noncholinergic effects depending on the type of OP, the dose, and the duration of exposure. Consequently, in order to understand these differences, *in silico* biologically-based dose-response and quantitative structure-activity relationship (QSAR) methodologies need to be integrated. Here, QSARs were used to predict OP bimolecular rate constants for trypsin and α -chymotrypsin. A heuristic regression of over 500 topological/constitutional, geometric, thermodynamic, electrostatic, and quantum mechanical descriptors, using the software Ampac 8.0 and Codessa 2.51 (SemiChem, Inc., Shawnee, KS), was developed to obtain statistically verified equations for the models. General models, using all data subsets, resulted in R^2

values of 0.94 and 0.92 and leave-one-out Q^2 values of 0.9 and 0.87 for trypsin and α -chymotrypsin. To validate the general model, training sets were split into independent subsets for test set evaluation. A y-randomization procedure, used to estimate chance correlation, was performed 10,000 times resulting in mean R^2 values of 0.24 and 0.3 for trypsin and α -chymotrypsin. The results show that these models are highly predictive and capable of delineating the complex mechanism of action between OPs and serine proteases, and ultimately, by applying this approach to other OP enzyme reactions such as AChE, facilitate the development of biologically based dose response models.

TABLE OF CONTENTS

	Page
I. INTRODUCTION.....	1
a. Biologically Based Dose Response Modeling.....	1
b. Bimolecular Rate Constants.....	2
c. Organophosphate Biologically Based Dose Response Modeling.....	2
d. Organophosphate Structure.....	3
e. Acetylcholinesterase.....	3
f. History of Quantitative Structure-Activity Relationships.....	4
g. Quantitative Structure-Activity Relationships in Biologically Based Dose Response Modeling.....	4
h. Noncholinergic Targets.....	5
i. Trypsin and Chymotrypsin.....	5
j. Conclusion.....	6
k. Hypothesis and Specific Aims.....	7
II. METHODS.....	8
a. CODESSA and AMPAC Background.....	8
b. Datasets.....	9
c. OP Descriptor Calculations.....	10
d. QSAR Regression Technique.....	10
e. QSAR Model Selection.....	11
f. The Break Point.....	11

TABLE OF CONTENTS CONTINUED

	Page
g. Training Set Validation.....	12
III. RESULTS.....	13
a. Database Distributions.....	13
b. Comparison of Databases.....	14
c. Statistical Analysis of Heuristic Regression.....	14
d. ABC and Y-Randomization Cross-Validation.....	15
IV. DISCUSSION.....	17
a. The QSAR Models.....	17
b. BBDR Model Application.....	19
c. Physical Interpretation of Descriptors.....	20
d. Conclusion.....	25
e. Acknowledgment	26
V. REFERENCES.....	27

LIST OF FIGURES OR ILLUSTRATIONS

Figure	Page
1. Generic chemical structure of OP compounds. R ₁ , R ₂ , and R ₃ are side chains.....	37
2. Schematic for noncholinergic enzyme inhibition and aging by OP nerve agents (Adapted from Gearhart <i>et al.</i> , 1990).....	38
3. Log ₁₀ OP bimolecular rate constant box plots for trypsin and α-chymotrypsin global datasets.....	39
4. R ² vs. number of descriptors for ABC trypsin bimolecular rate constant regression.....	40
5. R ² vs. number of descriptors for ABC α-chymotrypsin bimolecular rate constant regression.....	41
6. Experimental vs. predicted trypsin bimolecular rate constants for the ABC QSAR.....	42
7. Experimental vs. predicted α-chymotrypsin bimolecular rate constants for the ABC QSAR.....	43

LIST OF TABLES

Table	Page
1. Descriptor codes and t-test values for parameters used in the trypsin bimolecular rate constant regressions.....	44
2. Descriptor codes and t-test values for parameters used in the α -chymotrypsin bimolecular rate constant regressions.....	45
3. Inter-correlation between descriptors in global trypsin bimolecular rate constant QSAR.....	46
4. Inter-correlation between descriptors in global trypsin bimolecular rate constant QSAR.....	47
5. Trypsin results from the external validation using the ABC approach.....	48
6. α -Chymotrypsin results from the external validation using the ABC approach.....	49
7. Bimolecular rate constant ($M^{-1}min^{-1}$) trypsin and α -chymotrypsin QSAR models for the ABC, AB, AC, and BC training sets as a function of a compound's descriptors (D_1, D_2, \dots, D_{20}).....	50
8. Trypsin tissue distribution in rat and human tissues compared with that of AChE.....	55

APPENDIX TABLES

Table	Page
A. Listing of OP compounds and their respective trypsin, α -chymotrypsin, and AChE bimolecular rate constants.....	56

LIST OF ABBREVIATIONS

ACh: acetylcholine

AChE: acetylcholinesterase

BBDR: biologically-based dose-response

CNS: central nervous system

CODESSA: COmprehensive DEscriptors for Structural and Statistical Analyses

DFP: diisopropyl fluorophosphate

LOO: leave-one-out

ML: multi-linear

NIPALS: non-linear iterative partial least squares

OP: organophosphates

PLS: partial least squares

Q^2 : cross-validated leave-one-out correlation coefficient

QM: quantum mechanical

QSAR: quantitative structure-activity relationship

R^2 : correlation coefficient

SEM: standard error of the mean

TSAME: p-toluenesulfonyl arginine methylester

VX: o-ethyl s-[2-(diisopropylamino) ethyl] methylphosphonothioate

ACKNOWLEDGEMENT

I would like to thank everyone who has lent support, academic or otherwise, as I have pursued this degree. Without your encouragement, love and support I would not have been able to attain this honor. Thank you very much.

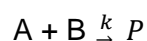
INTRODUCTION

For years researchers have relied on experimental animal models to obtain human toxicity estimates. However, error associated with these estimates has been linked with an inability to extrapolate across various species, doses, dose routes, and durations of exposure. As a result, *in silico*, biologically-based dose-response (BBDR) models were developed to improve the investigator's ability to interpret and extrapolate animal toxicity data to human predictions (Setzer *et al.*, 2001; Conolly and Andersen, 1991; Kavlock and Setzer, 1996).

Specifically, BBDR models are compartmental models that strive to be mechanistic by mathematically describing the physiology of the organism. Mathematical compartments correspond to predefined organs or tissues for which the interconnections correspond to blood or lymph flows. Systems of differential equations are then written which include physiological parameters such as blood flows, pulmonary ventilation rates, organ volumes, etc., for which information is available in scientific publications. The description of the body is simple but a balance is struck between complexity and simplicity depending upon the modeled compound.

These models have a few advantages over the "classical" pharmacokinetic models, which are based less on physiology. They can predict internal tissue concentrations of chemicals and/or metabolites and help interpolate/extrapolate between doses, exposure durations, routes of administration, species, and individuals. While

toxicology has greatly benefited from the success of these approaches, such mechanistically-based models rely on biochemical data not ordinarily collected in simple animal toxicity studies. Examples of these data not ordinarily collected include: tissue:blood partition coefficients and Michaelis-Menten enzyme kinetic parameters, V_{Max} and K_M . Therefore, there are often data gaps in the literature when one begins to develop a BBDR model. For organophosphates (OP) in particular, these models require quantitative estimates of their bimolecular rate constants describing the binding reaction kinetics between the OP and its target enzyme. This binding rate is proportional to the inhibitor concentration and the rate of inhibition is equal to the rate of loss of the OP in the system. Therefore, bimolecular rate constants have units of $\text{M}^{-1}\text{min}^{-1}$ and can be characterized in the following reaction scheme:



where A and B are reactants, P is a product and $d[A] / dt = d[B] / dt = -k [A] [B]$ and $d[P] / dt = k [A] [B]$.

Organophosphates are widely used as insecticides and have the potential to be used as chemical warfare agents. Therefore, there is a potential for human exposure. These pesticides constitute a problem for public health and, according to the World Health Organization, it is estimated that there are about three million cases of acute intoxication and 220,000 deaths each year (Pancetti *et al.*, 2007). For these reasons, a variety of OP BBDR models were developed with the primary aim of predicting cholinergic toxicity, such as acetylcholinesterase (AChE) inhibition (Gearhart *et al.* 1990, 1994; Langenberg *et al.*, 1997; Poet *et al.*, 2004; Sultatos, 1990; Sweeney *et al.*, 2006; Timchalk *et al.*, 2002, 2007; van der Merwe *et al.*, 2006). These models have relied on experimentally measured model parameters and, while these measurements are

necessary to validate the model predictions, they are time consuming, expensive, and not available for numerous OP. Noncholinergic targets are also rapidly gaining acceptance in the literature (Holmuhamedov *et al.*, 1996; Lockridge, 2008; Pancetti *et al.*, 2007; Richards *et al.*, 2000) and there may be a need to incorporate these new targets into future BBDR models. Consequently, with an ever growing number of cholinergic and noncholinergic parameters needed for model validation, *in silico* techniques to predict these parameters will become crucial tools for the future of OP BBDR model development.

Agricultural companies have placed an enormous amount of effort into developing structure-activity relationships for OP insecticides, fine tuning their products by utilizing a common phosphoryl-oxygen and varying the leaving group (R_1) and/or substituents (R_2 - R_3) (Figure 1). In particular, they have been designed to reversibly inhibit the active site of serine esterases (Vilceanu *et al.*, 1977). Once inhibited, many of these serine esterases promote a dealkylation (aging) reaction, such as trypsin with diisopropyl fluorophosphate (DFP) (Kossiakoff, 1984) and AChE complexed with sarin, soman, or DFP (Millard *et al.*, 1999) (Figure 2), which results in the formation of a covalent bond. This has provided a wide range of toxicities for OP compounds and made BBDR model development for OP lacking toxicokinetic and toxicodynamic data a challenging endeavor.

Acetylcholinesterase, above all other enzymes, is believed to be the most significant target for OP toxicity (Bajgar *et al.*, 2008). This enzyme is responsible for the rapid hydrolysis of acetylcholine (ACh), a neurotransmitter involved in the numerous cholinergic pathways in the body. Inhibition of this enzyme leads to the accumulation of

ACh in synapses and neuromuscular junctions, leading to hyper-stimulation of cholinergic pathways (Aroniadou-Anderjaska *et al.*, 2009). AChE, like many esterases, belongs to an esterase superfamily which has highly conserved active site residues with similar mechanisms of action (Pen *et al.*, 1990). Because of this, all serine esterase superfamily members should have the capacity to react with OP compounds (Pancetti *et al.*, 2007). Thus, characterizing members of this class of enzymes through the integration of BBDR and quantitative structure-activity relationship (QSAR) modeling will provide a useful resource for the identification of novel OP targets and aid our understanding of the beneficial and adverse effects these chemicals have on the body.

QSAR has been used in drug design for decades and only more recently become a tool to estimate BBDR model parameters (Barratt, 1998; Blaauboer, 2003). Specifically, QSAR relies on developing a statistical algorithm that quantitatively relates differences in activity with that of changes in molecular descriptors for each compound. These descriptors can be described as useful numbers that are transformed from the molecular structure and are commonly used by medicinal chemists (Hansch *et al.*, 2004). Descriptors have evolved over the years from simple global properties like LogP (capturing hydrophobicity), solubility, and ionization, to today, where there are more than 3,000 descriptors encompassing six different dimensions (Sumathy, 2007). QSAR methods have successfully reduced cost and streamlined the pharmaceutical industry and with increasing demand for OP toxicity estimates, the application of QSAR may be necessary to assess the risk and threat of OP compounds to human health.

Examples of QSAR models predicting biochemical parameters such as skin permeability coefficients, skin/water and tissue/blood partition coefficients, bimolecular

rate constants, and metabolic parameters such as V_{\max} and K_m , for the development of BBDR models and/or for human health risk assessment are not a new endeavor (Hansch and Deutsch, 1966; Kamgang *et al.*, 2008; Knaak *et al.*, 2004; Lowe *et al.*, 2006; Poulin and Krishnan, 1995; Raevskii *et al.*, 1990; Yang *et al.*, 1998). Specifically, AMPAC and CODESSA methodology has shown promising results in medicinal chemistry applications including predicting BBDR specific parameters (Katritzky *et al.*, 1998, 2005). However, the relationship between OP structure with that of trypsin and α -chymotrypsin bimolecular rate constants has, to my knowledge, not yet been evaluated.

Little attention has been paid to the QSAR between serine proteases and OP toxicity but these enzymes play essential roles in a wide range of biological processes (Bhongade *et al.*, 2005). Serine proteases are found in a variety of tissues throughout the body including the central nervous system (CNS), blood, and digestive tract (Davies *et al.*, 2001; Gingrich and Traynelis, 2000) and in every organelle and compartment of most, if not all, eukaryotic cells (Ekici *et al.*, 2008). There have been over 500 proteases identified in the human genome (Puente *et al.*, 2005), one-third of which are serine proteases, and are also associated with a number of diseases, including cardiovascular and Alzheimer's disease, cancer, autism, autoimmune diseases, inflammation, and hypertension (Ekici *et al.*, 2008).

Trypsin and chymotrypsin are the best-known serine proteases which are exclusively expressed in the digestive tract and responsible for hydrolyzing proteins which ultimately provide nutrients to the body (Rawlings and Barrett, 1994). While these proteases were shown to be inhibited by OP (Schaffer *et al.*, 1957), they are not believed to be toxicologically significant at high dose short term exposures, but rather

may play a role in chronic low-level toxicity. For example, cognitive deficits produced from these exposure conditions (Joosen *et al.*, 2009; Raveh *et al.*, 2002) may result from nutrient deprivation in the brain. Other structurally similar serine proteases, many of which are found in the CNS and blood, are involved in regulating neurotoxicity, learning, memory, synaptic plasticity, inflammation, cell signaling, and wound healing responses (Bhongade and Gadad, 2004; Bhongade *et al.*, 2005; Davies *et al.*, 2001; Gingrich and Traynelis, 2000; Movsesyan *et al.*, 2001; Richards *et al.*, 2000) and potentially may be involved in OP induced toxicity.

Therefore, quantifying their activity can be performed by elucidating small differences at a molecular level through the use of various descriptors. By utilizing a step-wise or tiered approach, beginning with a screening level tool such as this, scientists will be able to make initial assessments of the rate at which OP inhibit trypsin and α -chymotrypsin. Furthermore, the formation of these QSAR models will assist in elucidating a mechanism of interaction for members of the serine esterase superfamily, potentially allowing comparisons with other cholinergic and noncholinergic targets. Incorporating these QSAR into BBDR models will allow for assessment of the dose-response, help identify potential or likely health effects (cholinergic and noncholinergic), and by linking multiple BBDR models, have the potential to aid in the investigation of various therapeutic interactions even in the absence of experimental data on the specific compound. Thus, the integrated QSAR-BBDR modeling approach described in this study should be useful for predicting *in vivo* OP trypsin and α -chymotrypsin toxicodynamics in animals and humans under data-poor situations and set the stage for future QSAR predicting OP bimolecular rate constants with other members of the serine esterase superfamily.

HYPOTHESIS

The chemical structure of organophosphate nerve agents can be used to quantify their toxicity through physiochemical descriptors, thereby eliminating data gaps in biologically based dose response model development.

SPECIFIC AIMS

Test the hypothesis that quantitative structure-activity relationships can be used to predict bimolecular rate constants of organophosphates binding to trypsin.

Test the hypothesis that quantitative structure-activity relationships can be used to predict bimolecular rate constants of organophosphates binding to α -chymotrypsin.

Test the hypothesis that α -chymotrypsin will have a higher affinity than trypsin for organophosphates.

METHODS

CODESSA and AMPAC Background

CODESSA (Comprehensive Descriptors for Structural and Statistical Analyses), developed by SemiChem Inc. (Shawnee, KS), was used for developing the trypsin and α -chymotrypsin QSAR models. This program takes a database of chemical structures for which the biochemical property of interest is known, and calculates descriptors based on the chemical structure. It then performs a regression of the biochemical property on the descriptors to derive an equation. A wide variety of chemical descriptors can be calculated, including topostructural/topochemical, geometric, thermodynamic, electrostatic, and quantum mechanical (QM) properties (Ivanciuc, 1997) when used with the semi-empirical quantum mechanics features in the software AMPAC (Shawnee, KS). All of the descriptors used in the equations are theoretical; therefore, the trypsin and α -chymotrypsin bimolecular rates can be predicted for novel chemical structures. It includes regression techniques such as stepwise partial least squares (PLS), heuristic, multi-linear (ML), and non-linear iterative partial least squares (NIPALS) that can be used for searching the best multi-descriptor correlation (Equation 1).

Equation 1. Description of QSAR using chemical structure descriptors.

Biological Activity = f (Chemical Structure) + Error

where f is some function of the chemical structure descriptor.

Datasets

The database of OP bimolecular rate constants for binding with trypsin and α -chymotrypsin in their active form (phosphoryl-oxygen) was collected from the literature (Ooms, 1961) and can be found in the appendix. It is believed that this dissertation, while the experimental techniques and approaches may be out dated, is the only available large dataset for trypsin and α -chymotrypsin OP bimolecular rate constants. The OP were obtained from the TNO Prins Maurits Laboratory while chymotrypsin and trypsin were obtained from General Biochemical Inc. (Chagrin Falls, OH, U.S.A.) and The British Drug Houses Ltd. (Poole, England) respectively.

A titrimetric method using p-toluenesulfonyl arginine methylester (TSAME) as the substrate was used to quantify the trypsin reaction progress and the non-crystalline commercial trypsin preparation was purified according to Anson and Mirsky (1934). It had a specific activity of 8.6×10^{-4} U/mg egg-white Nitrogen, whereby a unit of trypsin was defined as the enzyme quantity which frees so much acid from TSAME that 1 μ l of 0.01 N NaOH/min is necessary to keep the pH at 8.0. The chymotrypsin commercial preparation, on the other hand, had an egg white N-content of 14.6% and a specific activity of 1.4×10^5 U/mg egg-white Nitrogen. A unit of chymotrypsin was defined as that quantity of enzyme which under the experimental conditions (Balls and Jansen 1952) frees so much acid from N acetyl-L-tyrosine ethyl-ester that 1 μ L of 0.01 N NaOH/min is necessary to keep the pH at 7.8. The proteolytic hemoglobin method of Anson (Northrop et al. 1948) based on the freeing of tyrosine and tryptophan out of denatured hemoglobin by α -chymotrypsin, was selected for quantifying the reaction progress. The egg white was cast down by trichlor acetic acid and tyrosine and tryptophan were precipitated with Folinphenol-reactors and spectrophotometry quantified at 280nm

(Ooms, 1961). Descriptive statistics (skewness, normality, inter-quartile range, etc.) was also performed with Prism 5 (GraphPad Software, San Diego, CA) to identify similarities and differences between the two enzymes.

OP Descriptor Calculations

The chemical structures were drawn, three-dimensionally optimized, and saved as .MOL files by using ChemSketch (Advanced Chemistry Developments, Inc., Toronto, Ontario, Canada). 3D optimization was based on the CHARMM force field parameterization (Brooks *et al.* 1983) and the stereo bonds on the 3D structure were ambiguous. Each .MOL file contained chemical information necessary to compute CODESSA topological, topochemical, and geometric descriptors while the output files (.OUT) carried the quantum chemical, electrostatic, and thermodynamic descriptors computed in AMPAC. To generate the AMPAC .OUT files, each .MOL file was loaded into AMPAC's Graphic User Interface (AGUI). The CODESSA output scheme (Job Type: Opt+Fre, Minimize Energy, Using TRUST, IR Frequencies and Thermodynamic Properties; Method: Model AM1, Wavefunction Restricted HF; Properties: Calculate bond orders, Calculate ESP charges, Generate output for CODESSA; Solvent: Solvation None; General: Default; Comment: default settings; Title: default settings) was used in the AMPAC calculation setup and each .MOL file was submitted and saved as a .OUT file. The .OUT and .MOL files were then loaded into the CODESSA program to calculate the molecular descriptors.

QSAR Regression Technique

The heuristic regression procedure was used to find the best correlated models from the selected non-co-linear descriptors. This regression selected the best two-

parameter regression equation, the best three parameter regression equation, etc., by eliminating variables from consideration on the basis of the R^2 value, F-test, standard deviation, and a t-test for each parameter in the stepwise regression procedure. During the heuristic procedure, the descriptor scales were normalized and centered, and the final result was given in natural scales. Inter-correlation among the descriptors was evaluated prior to descriptor inclusion in the model (Katritzky et al. 2001; Coi et al. 2006).

QSAR Model Selection

Default values for control parameters and criteria were used: one-parameter R^2 test for significance=0.01, high inter-correlation level=0.99, significant inter-correlation level=0.80, one-parameter t-test for significance=0.10, multi-parameter t-test for significance=3.0, branching criteria=3.0, maximum number of saved correlations=10.0, minimum one-parameter ANOVA F-test=1.0. Evaluation of the best-correlated models was carried out by validation of the stability of each regression model by a cross-validation technique, leave-one-out (i.e., the sensitivity of the model to the elimination of any single datum). The QSAR model was selected on the basis of the best statistical parameters (i.e., the highest squared correlation coefficient (R^2) and the highest the F-value).

The Breakpoint

During the heuristic regression procedure, it was important to know when to stop adding descriptors. An excessive number of descriptors can lead to over-correlated equations that poorly predict anything but the available training set. In general, a simple procedure following the principle of Occam's razor is available to prevent over-correlation (Hawkins, 2004). This procedure relies on not exceeding the number of

descriptors that crosses the break point. The break point can be identified from analyses of the plot of the number of descriptors involved versus the squared correlation coefficient (R^2), and the cross-validated square correlation coefficient (Q^2). As the numbers of descriptors are increased the R^2 will eventually plateau. The number of descriptors corresponding to the point where the R^2 is nearly at its plateau is called the break point, and the model corresponding to the break point has the optimum number of descriptors to be used (Katritzky *et al.*, 2005).

Training Set Validation

General QSAR models using all the data were first developed. These models were designed to meet the breakpoint criteria. The data sets were then split into three subsets (denoted as A, B, and C) by selection of every third datum point. Three new datasets, (A+B, A+C, and B+C) were then constructed. The descriptors chosen from each of the general models were then used in developing QSAR models for the A+B, A+C, and B+C datasets. The counterparts to each of these three datasets (C, B, and A) were used as external validation datasets (Katritzky *et al.*, 2009).

The sensitivity of the general model to chance correlations was also estimated by applying a y-randomization procedure (Rucker *et al.*, 2007) involving 10,000 randomizations using MatLab 7.9.0 (R2009b; Natick, MA). The bimolecular rate constants were scrambled while leaving chemical descriptors constant. The heuristic regression was then applied to the scrambled data set and the mean R^2 was calculated.

RESULTS

The OP compounds identified from the literature were grouped according to \log_{10} trypsin and $\log_{10}\alpha$ -chymotrypsin bimolecular rate constants. Both datasets, containing 52 and 62 structures respectively, were found to be normally distributed based upon the D'Agostino and Pearson omnibus normality test at $\alpha=0.05$. It was found that both the trypsin and α -chymotrypsin databases were positively skewed (trypsin=0.83 and α -chymotrypsin=0.29) and that the trypsin database had more of its variance as the result of infrequent extreme deviations from the mean based upon the calculated kurtosis values (trypsin=-0.24 and α -chymotrypsin=-0.68). Finally, the sum of the values within the trypsin database was found to be $\log_{10}60.68 \text{ M}^{-1}\text{min}^{-1}$ while the sum of the values within the α -chymotrypsin database was found to be $\log_{10}122.80 \text{ M}^{-1}\text{min}^{-1}$ (GraphPad Prism 5, La Jolla, CA).

The mean and standard deviation of the trypsin database were found to be $\log_{10}1.17 \pm \log_{10}1.18 \text{ M}^{-1}\text{min}^{-1}$ with a standard error of the mean (SEM) of $\log_{10}0.16 \text{ M}^{-1}\text{min}^{-1}$. From this sample it can be said, with 95% confidence, that the true mean of the trypsin population for OP inhibition lies between $\log_{10}0.84$ and $\log_{10}1.50 \text{ M}^{-1}\text{min}^{-1}$. The inter-quartile range for the database was found to be $\log_{10}1.58 \text{ M}^{-1}\text{min}^{-1}$ while the minimum, maximum, and median values were found to be $\log_{10}-0.60$, $\log_{10}4.240$, and $\log_{10}0.80 \text{ M}^{-1}\text{min}^{-1}$ (Figure 3) (GraphPad Prism 5).

On the other hand, the mean and standard deviation of the α -chymotrypsin database was found to be $\log_{10}1.98 \pm \log_{10}1.95 \text{ M}^{-1}\text{min}^{-1}$ with a SEM of $\log_{10}0.25 \text{ M}^{-1}\text{min}^{-1}$. From this sample it can be said, with 95% confidence, that the true mean of the α -chymotrypsin population for OP inhibition lies between $\log_{10}1.49$ and $\log_{10}2.48 \text{ M}^{-1}\text{min}^{-1}$. The inter-quartile range for the α -chymotrypsin database was found to be $\log_{10}3.09 \text{ M}^{-1}\text{min}^{-1}$ while the minimum, maximum, and median values were found to be $\log_{10}-1.00$, $\log_{10}6.30$, and $\log_{10}1.67 \text{ M}^{-1}\text{min}^{-1}$ (Figure 3). Lastly, the difference between the means of these two databases (-0.81 ± 0.31) was found to be significantly different and it can be said, with 95% confidence, that the difference between the two population means lies between -1.43 and -0.20 (i.e. α -chymotrypsin is more active than trypsin; GraphPad Prism 5).

It was found that by comparing and contrasting the trypsin and α -chymotrypsin database with that of our in-house AChE bimolecular rate constant database that AChE has a much higher mean binding affinity for OP. Sarin, for example, had measured bimolecular rate constants of $\text{Log}3.24$, $\text{Log}4.36$ and $\text{Log}7.15$ for trypsin, α -chymotrypsin and AChE respectively (Appendix). QSAR delineating these complex reaction mechanisms for AChE are in development and will be published at a later date. The sample mean of this AChE database was found to be $\text{Log}5.16 \text{ M}^{-1}\text{min}^{-1}$ (unpublished data). Chymotrypsin and trypsin sample means were $\log1.98$ and $\log1.17 \text{ M}^{-1}\text{min}^{-1}$ by comparison. Reasons for these differences can be attributed to the physiochemical properties of the OP as well as the characteristics of the enzyme active site.

The heuristic regression technique, applied to the entire training set of OP compounds, identified a breakpoint of 10 descriptors for both enzymes (Figures 4-5).

The global trypsin R^2 was found to be 0.94 and the global α -chymotrypsin R^2 was found to be 0.92 (Figures 6-7). The largest global F-values, 66.96 for trypsin and 62.37 for α -chymotrypsin, were found for the 10 parameter equations in CODESSA. The global variances of the populations (s^2) were found to be 0.01 and 0.34 and the global leave-one-out (LOO) Q^2 was noted to be 0.90 and 0.87 for trypsin and α -chymotrypsin respectively. The descriptors utilized in the rate constant regressions can be found in Tables 1 and 2. Seventy % of the descriptors utilized in the global QSAR models were constitutional/topological while the remaining 30% were quantum chemical descriptors. A t-test was performed on the descriptors incorporated in the models and the Kier shape index (order 2) and the count of H-donors sites [Zefirov's PC] were identified as the most important contributors to the trypsin and α -chymotrypsin global QSAR models, respectively. It was also shown that the number of fluorine and oxygen atoms were the highest correlated descriptor pairs ($R^2=-0.79$) for the global trypsin rate constant regression (Table 3). Contrasting that of trypsin, the highest correlated descriptor pairs for α -chymotrypsin resulted in a tie between the count of H-donors sites [Zefirov's PC] and the Kier flexibility index as well as the relative number of aromatic bonds and the number of rings ($R^2=0.82$) (Table 4).

The three subsets (A, B, and C) constructed for the trypsin QSAR validation contained 18, 17, and 17 compounds. The α -chymotrypsin subsets (A, B, and C) contained 21, 21, and 20 compounds. An average training set R^2 of 0.93 (Table 5) and 0.81 (Table 6) was found for the trypsin and α -chymotrypsin models. The average test set R^2 values were shown to be 0.75 and 0.61 for trypsin and α -chymotrypsin. The descriptors and t-test values for each model validation can be found in Tables 1 and 2 and the equations associated with those QSAR models are shown in table 7. Finally,

the y-randomization procedure yielded a mean and standard deviation trypsin $R^2=0.24\pm0.08$ and a mean and standard deviation α -chymotrypsin $R^2=0.3\pm0.07$.

DISCUSSION

The QSAR Models

The biological data for these QSAR models were found to be skewed but by performing a \log_{10} transformation of the data the distribution was moved to normal. The QSAR models produced R^2 values of 0.94 and 0.92 for the trypsin and α -chymotrypsin global training sets indicating that the degree of relationship between the predicted and observed rates was significant (an R^2 of 1.0 is indicative of a perfect correlation). The F-test characterized the ratio of the variance explained by the model to that of the variance not explained by the model and was dependent upon the number of parameters included in the model. The F-test for each of these QSAR models was large indicating that the appropriate number of descriptors was added to each model and this was confirmed by the breakpoint. It has also been shown that a high LOO cross-validated Q^2 is necessary but not sufficient for QSAR models to have high predictive power (Golbraikh and Tropsha 2002). The LOO cross-validated global Q^2 did not differ significantly from the global R^2 , indicating that the predictive power for both models was retained after removing any one datum.

With regard to the QSAR submodels (A+B, A+C, and B+C), the descriptors used were similar to the global models and the external validation data set R^2 values remained close to the training subsets, thus demonstrating the robustness of the generated sub-models. Data suggest that no one compound in the training set was responsible for a large portion of the regression's predictive capabilities. The y-

randomization procedure generated R^2 values that averaged 0.24 and 0.30 for trypsin and α -chymotrypsin. This suggests that the descriptors in each model were not chosen by random chance. It was also found that $\frac{3}{4}$ of the models were over-parameterized by one or more descriptors making it clear that the break point can be difficult to identify from visual inspection. For example, in the trypsin AC training model the standard error of the regression coefficient of D_3 is greater than the regression coefficient itself (2.04 +/- 2.45). From these results it is recommended that a more stringent heuristic regression approach be taken by additionally comparing the standard error with the regression coefficient before accepting the breakpoint.

QSAR regressions require sufficient variation in activity and in descriptors within the training set (Yang *et al.*, 2009). These QSAR models spanned 4 trypsin and 6 α -chymotrypsin \log_{10} units of activity and utilized over 500 initial descriptors from three different dimensions in the selection process meeting this requirement. However, due to the vast chemical space that accompanies OP compounds, the limited datasets of 52 and 62 OP reduced the domain of applicability. All together, 66 OP combinations were used in this database and it was divided as follows: 35 p-nitrophenyl combinations, 2 p-nitrothiophenyl combinations, 4 o-nitrophenyl combinations, 4 m-nitrophenyl combinations, 4 m-dimethylamino-ethanthiol combinations, 10 fluoridates, 2 diethylamino-ethanthiol combinations, and 5 not further classified combinations. Utilizing these QSAR for chemicals that fall outside this domain (i.e., compounds not of the OP class or OP with different structural features not included in the model) will give erroneous results. An example of a model giving flawed results can be shown in the α -chymotrypsin B+C model. The test set R^2 was found to be 0.16, significantly lower than the training set R^2 of 0.68. It is believed that these QSAR models can be improved by:

increasing the number of experimental compounds used in the training set, utilizing higher dimensional descriptors that are better correlated with the experimental data (Vedani *et al.*, 2005), and/or use different regression techniques (support vector machine, neural network, etc.). It would also be recommended that the over-parameterized models be re-parameterized by eliminating those descriptors with standard errors greater than its coefficient.

BBDR Model Application

While these QSAR models predict bimolecular rates for trypsin and α -chymotrypsin, data is lacking on OP trypsin and α -chymotrypsin regeneration and aging rate constants. Therefore, more QSAR approaches need to be developed to complete the toolkit. Once completed, this toolkit could help determine the target tissue dose in which a response, such as altered digestion, is seen in an exposed animal. Also, extrapolating this technique to other serine proteases could help determine if these enzymes sequester OP at physiologically relevant doses. If true, serine proteases could play a vital role in OP toxicity and even cause symptoms such as impaired synaptic plasticity, learning, memory, neuroprotection, wound healing, cell signaling, inflammation, and blood coagulation.

Other necessary parameters for incorporation into an OP BBDR model include the relative tissue distribution of trypsin, α -chymotrypsin, and other serine proteases in a variety of species. Trypsin, for example, has been found in the serum of Wistar rats at a concentration greater than that of blood AChE. The enzyme is also found at a concentration 100 times higher in the pancreas than that of AChE in the blood (Table 8). Due to such a high concentration in the pancreas, inhibition of these enzymes could lead

to acute pancreatitis. Finally, it is believed that in order to identify other tissue concentrations for a variety of serine proteases, a more thorough literature review will need to be conducted.

Physical Interpretation of Descriptors

Electrophilicity

Constitutional/topological descriptors, despite their conceptual simplicity, provided important one-dimensional structural characteristics for both global QSAR models. The numbers of fluorine (D_1), oxygen (D_7 , D_{16}), and sulfur (D_{15}) atoms in the models were believed to partially characterize OP electrophilicity, which is essential for OP inhibition of esterases (Ahmad, 1970). The electron withdrawing capability of each substituent is altered in the presence of these elements and consequently plays a crucial role in breaking the P=O bond in the phosphorylation of the enzyme (Jinsong *et al.*, 2004). The number of fluorine atoms (D_1) was found to be positively correlated while the number of oxygen atoms (D_7) was found to be negatively correlated with trypsin OP bimolecular rate constants. On the other hand, the number of oxygen atoms (D_{16}) was negatively correlated and number of sulfur atoms (D_{15}) was positively correlated with α -chymotrypsin OP bimolecular rate constants. Other features, such as induction or resonance effects, formed through the addition or loss of electron density could also be characterized by these descriptors. They are driven by the compounds electronegativity, number of lone pair electrons, and the presence of multiple bonds.

The number of aromatic bonds (D_{11}) and number of rings (D_{18}), which are also constitutional descriptors, are believed to characterize the highly mobile and polarizable electrons, creating resonance effects (the movement of electron density through π

bonds) which make for good covalent catalysts. The number of aromatic bonds (D_{11}) was found to be negatively correlated while the number of rings (D_{18}) was found to be positively correlated with α -chymotrypsin OP rate constants. α -Chymotrypsin substrates were all found to contain rings with aromatic bonds (Voet and Voet, 2004) and it is believed that this property influences the ability of OP to penetrate the gorge, interact at the hydrophobic active site residues, and offer extra long range electronic stability which is necessary during the inversion of configuration when the leaving group is excised.

van der Waals Forces

van der Waals interactions were found to be important contributors of general enzyme activity and it is believed that OP use these same interactions to stabilize the transition state. This force is highly dependent upon the size of atoms and the distance between them and these QSAR models chose to characterize the property through the atomic radii. This three-dimensional feature determines the area of contact between the two molecules: the greater the area, the stronger the interaction. Fluorine (D_1), oxygen (D_7 , D_{16}), and sulfur (D_{15}) atoms, for example, have very small radii, giving them a larger surface area in which to interact through van der Waals forces. OP with elements such as these should produce desirable pharmacodynamic attributes (Makhaeva *et al.*, 2009). The number of fluorine atoms (D_1) was positively correlated while the number of oxygen atoms (D_7) was negatively correlated with trypsin OP bimolecular rate constants. On the other hand, the number of oxygen atoms (D_{16}) was negatively correlated and number of sulfur atoms (D_{15}) was positively correlated with α -chymotrypsin OP bimolecular rate constants.

London Dispersion Forces and Lipophilicity

London dispersion forces also become stronger as the surface area is increased. This produces an uneven distribution of electrons on OPs which could produce temporary dipoles. This property is believed to be characterized by the topographic electronic index (all bonds) (D_{19}), FPSA-1 Fractional PPSA (PPSA-1/TMSA) [Zefirov's PC] (D_9), and PPSA-3 Atomic charge weighted PPSA [Zefirov's PC] (D_6). The topographic electronic index (all bonds) (D_{19}) was negatively correlated with α -chymotrypsin, the FPSA-1 Fractional PPSA (PPSA-1/TMSA) [Zefirov's PC] (D_9) was positively correlated with trypsin and the PPSA-3 Atomic charge weighted PPSA [Zefirov's PC] (D_6) was negatively correlated with trypsin. Dipoles such as these could describe the tendency of OP QSAR models to show strong descriptor correlations with lipophilicity and/or hydrophobicity. Some examples of QSAR descriptors characterizing OP lipophilicity include the linear free energy polar parameter (σ^*) (Metcalf and Metcalf, 1984), the π constant (Hansch and Leo, 1979), and in some cases the octanol/water partition coefficient (Leo *et al.*, 1971). Lipophilic compounds generally have little or no capacity to form hydrogen bonds, another important requirement in OP inhibition of serine proteases; therefore an optimum must be found to create a quality catalytic environment for bond formation.

Hydrogen Bonds

The conformational distortion that occurs with the formation of the substrate tetrahedral intermediate, and, in the case of OP, the trigonal bipyrimidal intermediate, creates a hydrogen bond network within the oxyanion hole (Taylor *et al.*, 1999). These hydrogen bonds come in two forms: strong and weak. The strong hydrogen bonds, also known as low barrier hydrogen bonds, in the active site of serine esterases can create a charge relay system responsible for the catalytic activity in the triad. The weak hydrogen

bonds in the Michaelis complex are converted to a strong hydrogen bond in the transition state, facilitating proton transfer from aspartate to histidine, to the active site serine residue, creating an environment that is safe for OP covalent bond formation (Voet and Voet, 2004). In fact, a distinguishing difference between trypsin, α -chymotrypsin, and AChE are the hydrogen bond lengths in the catalytic triad. The His⁴⁴⁰ - Asp³²⁷ bond distance is 2.5 Å (2.8 Å resolution) in AChE and 0.2 Å shorter than in the corresponding hydrogen bonds of trypsin and α -chymotrypsin (1.5 Å resolution) (Chambers and Stroud, 1977). Appearing in the QSAR models were the HA dependent HDSA-1 [Zefirov's PC] (D_{20}), count of hydrogen-donor sites [Zefirov's PC] (D_{12}), number of oxygen atoms (D_7), and the relative number of hydrogen atoms (D_8). They were respectively found to be positively, negatively, negatively, and negatively correlated with their enzyme bimolecular rates. It is believed that changes in these hydrogen bonding properties alter the charge relay system for these serine proteases.

Electrostatic Forces

Serine esterases such as AChE, BChE, trypsin, and α -chymotrypsin all contain a negatively charged oxyanion hole which is generally believed to stabilize substrates and preferentially bind the transition state, lowering the activation energy of the reaction (Zhang *et al.*, 2002). This fundamental property, electrostatic attraction or repulsion between two elements, follows Coulomb's law and becomes highly important with these esterases as both short range and long range electrostatic "steering forces" are utilized, most likely to maintain a high catalytic turnover of substrates. In doing so, water molecules are rapidly pumped in and out of the active site gorge to maintain the appropriate pH and in providing the correct environment for catalytic activity. This property is believed responsible for much of the catalytic efficiency of these enzymes

and Wang et al (2006) concluded that these enzymatic reactions are strongly influenced by the solvent environment generated by these electrostatic forces. Descriptors that characterize these properties included the RPCG relative positive charge (QMPOS/QTPLUS) (D_{14}) (negatively correlated), RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC] (D_3) (negatively correlated), PPSA-3 Atomic charge weighted PPSA [Zefirov's PC] (D_6) (negatively correlated), and the FPSA-1 Fractional PPSA (PPSA-1/TMSA) [Zefirov's PC] (D_9) (positively correlated).

Steric Hindrance and Connectivity

Topological descriptors are two-dimensional descriptors that reflect the molecular connectivity without geometric information. Multiple topological descriptors were found in each of the two models characterizing the importance of connectivity in predicting bimolecular rates. α -Chymotrypsin and trypsin bimolecular rate constants appear to be driven by the Balaban index (D_5) and the Kier & Hall (1986) descriptors. Examples include the Kier & Hall (1986) index (order 3) (D_{13}), Kier (1985) flexibility index (D_{17}), and the Kier (1985) shape index (order 2 and 3) (D_2 , D_{10}). These Kier indices, commonly denoted by 1k , 2k , and 3k , are 2D topological indices based on graph theory concepts. Index 1k quantifies the cyclicity of a molecule, index 2k quantifies the star-like attributes of a molecule, and index 3k quantifies the place in the chain where the branching occurs. The equations for calculating these Kier indices are given elsewhere (Kier, 1985, Kier and Hall 1986). The Balaban index (D_5) and the Kier & Hall (1986) index (order 3) (D_{13}) was found to be negatively correlated with trypsin and α -chymotrypsin OP bimolecular rates. On the other hand, the Kier (1985) flexibility index (D_{17}) was found to be positively correlated with α -chymotrypsin OP bimolecular rates while the Kier (1985) shape index (order 2 and 3) (D_2 , D_{10}) were both found to be positively correlated with trypsin OP

bimolecular rates. This suggests that the active site of trypsin and α -chymotrypsin for OP agents is driven, at least in part, by the branching characteristics of the chemical structure. This indicates that isomeric structures will differ in their affinity towards the active site and this is confirmed by the fact that serine hydrolase enzymes show stereoselectivity for the levorotatory isomers of soman; bimolecular rates of $10^4 \text{ M}^{-1}\text{min}^{-1}$ for AChE and $10^2 \text{ M}^{-1}\text{min}^{-1}$ for trypsin and α -chymotrypsin (Kovach, 1988; Ooms and van Dijk, 1966).

Conclusion

QSAR models were developed for the prediction of trypsin and α -chymotrypsin bimolecular rate constants using descriptors calculated from OP structures alone. It was found from these models that α -chymotrypsin was the more reactive of the two digestive proteases and descriptors that characterize electrophilicity, van der Waals interactions, London dispersion forces, lipophilicity, hydrogen bonding, electrostatic charges, steric hindrance, and connectivity were the most useful properties in predicting their bimolecular rates. These QSAR provided important mechanistic features about the interaction of serine proteases with OP compounds as well as an approach towards the development and validation of BBDR models for OP that lack toxicokinetic and toxicodynamic data.

While not necessarily toxicologically significant under all exposure circumstances, trypsin and α -chymotrypsin may play a role in noncholinergic OP induced toxicity. They represent two members of a highly conserved protease superfamily that comprises nearly one-third of all known proteases identified to date, which play crucial roles in a wide variety of cellular and extracellular functions of the body. This demonstrates the

potential that OP could disrupt vital processes such as learning, memory, synaptic plasticity, cell signaling, immunomodulation, inflammation, and digestion which can not be attributed to cholinergic toxicity. In fact, a recent study published in Pediatrics, has linked the insecticide chlorpyrifos, which is used on some fruits and vegetables, with delays in learning rates, reduced physical coordination, and behavioral problems in children, especially attention deficit hyperactivity disorder (Bouchard et al. 2010). Inhibition of these enzymes by OP could also be linked with pancreatitis (Somogyi et al. 2001). This evidence, and along with others, clearly indicates that a wide variety of secondary proteins are targets for OP (Murray *et al.*, 2005; Quistad and Casida, 2000; Richardson, 1992) however without further knowledge of their aging and regeneration rates it is difficult to determine a dose at which these effects could be seen. Thus, characterizing members of this class of enzymes through the integration of biologically based mathematical models such as BBDR and QSAR modeling can provide a useful resource for the identification of novel OP targets and aid our understanding of the beneficial and adverse effects these chemicals exert on the body. While these QSAR models can not obviously create new data, they provide an economic and less cumbersome alternative in the screening and priority selection of large inventories of hazardous OP compounds. They may be used to better interpret experimental results and inform future QSAR models for OP targets such as AChE.

Acknowledgements

This project received support from the Defense Threat Reduction Agency - Joint Science and Technology Office, Basic and Supporting Sciences Division (grant number 2G 806 08 AHB C).

REFERENCES

- Ahmad, S. 1970. The recovery of esterases in organophosphate-treated housefly (*Musca Domestica L.*). *Comp. Biochem. Physiol.* 33:579-586.
- Anson, M.L. and Mirsky, A.E. 1934. The equilibrium between active native trypsin and inactive denatured trypsin. *J. Gen. Physiol.* 17:393-398.
- Aroniadou-Anderjaska, V., Figueiredo, T. H., Aplan, J. P., Qashu, F., and Braga, M. F. M. 2009. Primary brain targets of nerve agents: The role of the amygdale in comparison to the hippocampus. *Neurotoxicology* 30:772-776.
- Bajgar, J., Fusek, J., Kassa, J., Jun, D., Kuca, K., and Hajek, P. 2008. An attempt to assess functionally minimal acetylcholinesterase activity necessary for survival of rats intoxicated with nerve agents. *Chem. Biol. Interact.* 175: 281-285.
- Balls, A.K. and Jansen, E.F. 1952. Stoichiometric inhibition of chymotrypsin. *Adv. Enzymol.* 13:321-343.
- Barratt, M. D. 1998. Integrating computer prediction systems with *in vitro* methods towards a better understanding of toxicology. *Toxicol. Lett.* 102:102-103.
- Bhongade, B. and Gadad, A. 2004. 3D-QSAR CoMFA/CoMSIA studies on urokinase plasminogen activator (uPA) inhibitors: a strategic design in novel anticancer agents. *Bioorg. Med. Chem.* 12:2797-2805.
- Bhongade, B., Gouripur, V. V., and Gadad, A. K. 2005. 3D-QSAR CoMFA studies on trypsin-like serine protease inhibitors: a comparative selectivity analysis. *Bioorg. Med. Chem.* 13: 2773-2782.

- Blaauboer, B. J. 2003. The integration of data on physico-chemical properties, *in vitro*-derived toxicity data and physiologically based kinetic and dynamic as modelling a tool in hazard and risk assessment. A commentary. *Toxicol. Lett.* 138:161-171.
- Bouchard, M.F., Bellinger, D.C., Wright, R.O., and Weisskopf, M.G. 2010. Attention-Deficit/Hyperactivity disorder and urinary metabolites of organophosphate pesticides. *Pediatrics.* 125:e1270-e1277.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4: 187-217.
- Chambers, J. L. and Stroud, R. M. 1977. Difference fourier refinement of the structure of DIP-trypsin at 1.5Å with a minicomputer technique. *Acta Crystallogr., Sect. B.* 33:1824-1837.
- Coi, A., Massarella, I., Murgia, L., Saraceno, M., Calderone, V. and Bianucci, A.M. 2006. Prediction of hERG potassium channel affinity by the CODESSA approach. *Bioorg. Med. Chem.* 14:3153-3159.
- Conolly, R. B. and Andersen, M. E. 1991. Biologically based pharmacodynamic models: Tools for toxicological research and risk assessment. *Annu. Rev. Pharmacol. Toxicol.* 31:503-523.
- Davies, B., Kearns, I. R., Ure, J., Davies, C. H., Lathe, R. 2001. Loss of hippocampal serine protease BSP1/Neurosin predisposes to global seizure activity. *J. Neurosci.* 21: 6993-7000.
- Ekici, O. D., Paetzel, M., and Dalbey, R. E. 2008. Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci.* 17:2023-2037.

- Frier, B.M., Adrian, T.E., Saunders, J.H.B., and Bloom, S.R. 1980. Serum trypsin concentration and pancreatic trypsin secretion in insulin-dependent diabetes mellitus. *Clinica Chimica Acta*. 105: 297-300.
- Gearhart, J. M., Jepson, G. W., Clewell III, H. J., Andersen, M. E., and Conolly, R. B. 1990. Physiologically based pharmacokinetic and pharmacodynamic model for the inhibition of acetylcholinesterase by diisopropylfluorophosphate. *Toxicol. Appl. Pharmacol.* 106:295-310.
- Gearhart, J. M., Jepson, G. W., Clewell III, H. J., Andersen, M. E., and Conolly, R. B. 1994. Physiologically based pharmacokinetic model for the inhibition of acetylcholinesterase by organophosphate esters. *Environ. Health Perspect.* 102:Suppl. 11: 51-60
- Gingrich, M. and Traynelis, S. 2000. Serine proteases and brain damage-Is there a link? *Trends Neurosci.* 3:399-407.
- Golbraikh, A. and Tropsha, A. 2002. Beware of q²!. *J. Mol. Graphics Modell.* 20:269-276.
- Hansch, C. and Deutsch, E. W. 1966. The use of substituent constants in the study of structure-activity relationships in cholinesterase inhibitors. *Biochim. Biophys. Acta.* 126:117-128.
- Hansch, C. and Leo, A. 1979. Substituent constants for correlation analysis in chemistry and biology. New York: John Wiley & Sons Ltd.
- Hansch, C., Leo, A., Mekapati, S. B., and Kurup, A. 2004. QSAR and ADME. *Bioorg. Med. Chem.* 12:3391-3400.
- Hawkins, D.M. 2004. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* 44:1-12.

- Holmuamedov, E. L., Kholmoukhamedova, G. L., and Baimuradov, T. B. 1996. Non-cholinergic Toxicity of Organophosphates in Mammals: Interaction of Ethaphos with Mitochondrial Functions. *J. Appl. Toxicol.* 16: 475-481.
- Ivanciuc, O. 1997. CODESSA Version 2.13 for Windows. *J. Chem. Inf. Comput. Sci.* 37: 405-406.
- Jinsong, Z., Bin, W., Zhaoxia, D., Xiaodong, W., and Lingren, W. 2004. 3D-quantitative structure-activity relationship study of organophosphate compounds. *Chin. Sci. Bull.* 49:240-245.
- Joosen, M. J., Jousma, E., van den Boom, T. M., Kuijpers, W. C., Smit, A. B., Lucassen, P. J., and van Helden, H. P. 2009. Long-term cognitive deficits accompanied by reduced neurogenesis after soman poisoning. *Neurotoxicology.* 30:72-80.
- Kamgang, E., Peyret, T., and Krishnan, K. 2008. An integrated QSPR-PBPK modelling approach for *in vitro- in vivo* extrapolation of pharmacokinetics in rats. *SAR and QSAR in Environmental Research.* 19:669-680.
- Katritzky, A. R., Wang, Y., Sild, S., and Tamm, T. 1998. QSPR studies on vapor pressure, aqueous solubility, and the prediction of water-air partition coefficients. *J. Chem. Inf. Comput. Sci.* 38:720-725.
- (Katritzky, A. R., Perumal, S., Petrukhin, R., and Kleinpeter, E. 2001. CODESSA-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities. *J. Chem. Inf. Comput. Sci.* 41:569-574.
- Katritzky, A. R., Kuanar, M., Fara, D. C., Karelson, M., Acree Jr., W. E., Solov'ev, V. P., and Varnek, A. 2005. QSAR modeling of blood:air and tissue:air partition coefficients using theoretical descriptors. *Bioorg. Med. Chem.* 13:6450-6463.

- Katritzky, A. R. Slavov, S. H., Stoyanova-Slavova, I. S., Kahn, I., and Karelson, M. 2009. Quantitative structure-activity relationship (QSAR) modeling of EC₅₀ of aquatic toxicities for *Daphnia magna*. *J. Toxicol. Environ. Health A*. 72:1181-1190.
- Kavlock, R. J. and Setzer, R. W. 1996. The road to embryologically based dose-response models. *Environ. Health Perspect.* 104(Suppl. 1): 107-121.
- Kier, L. 1985. A shape index from molecular graphs. *Quant. Struct. Act. Relat.* 4: 109-116.
- Kier, L. and Hall, L. 1986. Molecular Cconnectivity in Structure-Activity Analysis. John Wiley Publ, London.
- Knaak, J. B., Dary, C.C., Power, F., Thomson, C. B., Blancato, J. N. 2004. Physicochemical and biological data for the development of predictive organophosphorus pesticide QSARs and PBPK/PD models for human risk assessment. *Crit. Rev. Toxicol.* 34:143-207.
- Kossiakoff, A.1984. Use of the neutron diffraction h/d exchange technique to determine the conformational dynamics of trypsin. *Basic Life Sci.* 27:281-304.
- Kovach, I. 1988. Structure and dynamics of serine hydrolase-organophosphate adducts. *Enzyme Inhib.* 2:199-208.
- Langenberg, J. P., Dijk, C. V., Sweeney, R. E., Maxwell, D. M., De Jong, L. P. A., Benschop, H. P. 1997. Development of a physiologically based model for the toxicokinetics of C(+/-)P(+/-)-soman in the atropinized guinea pig. *Arch. Toxicol.* 71:320-331.
- Leo, A., Hansch, C., and Elkins, D. 1971. Partition coefficients and their uses. *Chem. Rev.* 71: 525-616.

- Lockridge, O. 2008. Mass spectrometry to identify new biomarkers of nerve agent exposure. U.S. Army Medical Research and Materiel Command, Fort Detrick, MD. ADA483399.
- Lowe, E. R., Rick, D. L., West, R. J., and Bartels, M. J. 2006. The evaluation of quantitative structure-property relationship predictions of tissue: Blood partition coefficients of highly lipophilic chemicals. *Drug Metab. Rev.* 38:78.
- Makhaeva, G. F., Aksinenko, A. Y., Sokolov, V. B., Serebryakova, O. G., and Richardson, R. J. 2009. Synthesis of organophosphates with fluorine-containing leaving groups as serine esterase inhibitors with potential for Alzheimer disease therapeutics. *Bioorg. Med. Chem. Lett.* 19: 5528-5530.
- Metcalf, R. and Metcalf, R. 1984. Steric, electronic, and polar parameters that affect the toxic actions of O-alkyl, O-phenyl phosphonothionate insecticides. *Pestic. Biochem. Physiol.* 22: 169-177.
- Millard, C. B., Kryger, G., Ordentlich, A., Greenblatt, H. M., Harel, M., Raves, M. L., Segall, Y., Barak, D., Shafferman, A., Silman, I., Sussman, J. L. 1999. Crystal structures of aged phosphorylated acetylcholinesterase: nerve agent reaction products at the atomic level. *Biochemistry* 38:7032-7039.
- Movsesyan, V. A., Yakovlev, A. G., Fan, L., and Faden, A. I. 2001. Effect of serine protease inhibitors on posttraumatic brain injury and neuronal apoptosis. *Exp. Neurol.* 167:366-375.
- Murray, A., Rathbone, A. J., and Ray, D. E. 2005. Novel protein targets for organophosphorus pesticides in rat brain. *Environ. Toxicol. Pharmacol.* 19:451-454.
- Northrop, J. H., Kunitz, M., and Herriott, R.M. 1948. Crystalline Enzymes. Columbia University Press, New York. 307.

- Ooms, A. J. J. 1961. The reactivity of organic phosphor combinations in regards to a number of esterases. Rijks Universiteit, Groningen, Netherlands.
- Ooms, A. J. J. and van Dijk, C. 1966. The reaction of organophosphorus compounds with hydrolytic enzymes. III. The inhibition of α -chymotrypsin and trypsin. *Biochem. Pharmacol.* 15: 1361-1377.
- Pancetti, F., Olmos, C., Dagnino-Subiabre, A., Rozas, C., Morales, B. 2007. Noncholinesterase effects induced by organophosphate pesticides and their relationship to cognitive processes: Implication for the action of acylpeptide hydrolase. *J. Toxicol. Environ. Health B.* 10:623-630.
- Pen, J., Bolks, G. J., Hoeksema-Du Pui, M. L. L., and Beintema, J. J. 1990. Serine esterases: structural conservation during animal evolution and variability in enzymic properties in the genus *Drosophila*. *Genetica* 81: 125-131.
- Poet, T. S., Kousba, A. A., Dennison, S.L., and Timchalk, C. 2004. Physiologically based pharmacokinetic/pharmacodynamic model for the organophosphorus pesticide diazinon. *Neurotoxicology* 25:1013-1030.
- Poulin, P. and Krishnan, K. 1995. An algorithm for predicting tissue: blood partition coefficients of organic chemicals from n-octanol: water partition coefficient data. *J Toxicol. Environ. Health.* 46:117-129.
- Puente, X. S., Sanchez, L. M., Gutierrez-Fernandez, A., Velasco, G., and Lopez-Otin, C. 2005. A genomic view of the complexity of mammalian proteolytic systems. *Biochem. Soc. Trans.* 33: 331-334.
- Quistad, G. and Casida, J. 2000. Sensitivity of blood-clotting factors and digestive enzymes to inhibition by organophosphorus pesticides. *J. Biochem. Mol. Toxicol.* 14: 51-56.

- Raevsky, O. A., Chistyakov, V. V., Agabekyan, R. S., Sapegin, A. M., and Zefirov, N. S. 1990. Formation of a model of the interrelationship between the structure of organophosphorus compounds and their capacity for inhibiting cholinesterases. *Bioorg Khim.* 16: 1509-1522.
- Rawlings, N. and Barrett, A. 1994. Families of serine peptidases. *Meth. Enzymol.* 244: 19–61.
- Raveh, L., Weissman, B. A., Cohen, G., Alkalay, D., Rabinovitz, I., Sonogo, H., and Brandeis, R. 2002. Caramiphen and scopolamine prevent soman-induced brain damage and cognitive dysfunction. *Neurotoxicology.* 23:7-17.
- Reddy, S., Bibby, N.J., Smith, P.A., and Elliott, R.B. 1985. Rat trypsin: Purification, radioimmunoassay and age-related serum levels in normal and spontaneously diabetic BB Wistar rats. *Aust. J. Exp. Biol. Med. Sci.* 63: 667-681.
- Richards, P. G., Johnson, M. K., and Ray, D. E. 2000. Identification of acylpeptide hydrolase as a sensitive site for reaction with organophosphorus compounds and a potential for cognitive enhancing drugs. *Mol. Pharmacol.* 58:577-583.
- Richardson, R. 1992. Interactions of organophosphorus compounds with neurotoxic esterase. In *Organophosphates—Chemistry, Fate, and Effects* (J.E. Chambers and P.E. Levi, Eds.), pp. 299-323. Academic Press, San Diego.
- Rucker, C., Rucker, G., and Meringer, M. 2007. Y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.* 47: 2345-2357.
- Schaffer, N. K., Lang, R. P., Simet, L., and Drisko, R. W. 1957. Phosphopeptides from acid-hydrolyzed P³²-labeled isopropyl methylphosphonofluoridate-inactivated trypsin. *J. Biol. Chem.* 1: 185-192.
- Setzer, W. R., Lau, C., Mole, M. L., Copeland, M. F., Rogers, J. M., and Kavlock, R. J. 2001. Toward a biologically based dose-response model for developmental

- toxicity of 5-fluorouracil in the rat: A mathematical construct. *Toxicol. Sci.* 59:49-58.
- Somogyi, L., Martin, S. P., Venkatesan, T., Ulrich, C. D. 2001. Recurrent acute pancreatitis: an algorithmic approach to identification and elimination of inciting factors. *Gastroenterology*. 120: 708-17.
- Sultatos, L. 1990. A physiologically based pharmacokinetic model of parathion based on chemical-specific parameters determined *in vitro*. *J. Am. Coll. Toxicol.* 9:611-619.
- Sumathy, K. "Chemical Descriptors" From QSARWorld – A strand life sciences web resource. <http://www.qsarworld.com/insilico-chemistry-chemical-descriptors.php>.
- Sweeney, R. E., Langenberg, J. P., Maxwell, D. M. 2006. A physiologically based pharmacokinetic (PB/PK) model for multiple exposure routes of soman in multiple species. *Arch. Toxicol.* 80:719-731.
- Taylor, P., Wong, L., Radic, Z., Tsigelny, I., Bruggemann, R., Hosea, N. A., Berman, H. A. 1999. Analysis of cholinesterase inactivation and reactivation by systematic structural modification and enantiomeric selectivity. *Chem.-Biol. Interact.* 119-120:3-15.
- Timchalk, C., Nolan, R. J., Mendrala, A. L., Dittenber, D. A., Brzak, K. A., and Mattsson, J. L. 2002. A physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model for the organophosphate insecticide chlorpyrifos in rats and humans. *Toxicol. Sci.* 66:34-53.
- Timchalk, C., Kousba, A., and Poet, T. S. 2007. An age-dependent physiologically based pharmacokinetic/pharmacodynamic model for the organophosphorus insecticide chlorpyrifos in the preweanling rat. *Toxicol. Sci.* 98:348-365.

- van der Merwe, D., Brooks, J. D., Gehring, R., Baynes, R. E., Monteiro-Riviere, N. A., Riviere, J. E. 2006. A physiologically based pharmacokinetic model of organophosphate dermal absorption. *Toxicol. Sci.* 89:188-204.
- Vedani, A., Dobler, M., and Lill, M. A. 2005. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands at the estrogen receptor. *J. Med. Chem.* 48: 3700-3703.
- Vilceanu, R., Chiriac, A., Szabadai, Z., Simon, Z. 1977. Structure-activity correlations for the toxicity of organic phosphorus compounds of the Schrader type. *Seria Stiinte Fizice-Chim.* 15: 61-68.
- Voet, D. and Voet, J. 2004 *Biochemistry*, 3rd Edition. John Wiley & Sons, Inc.,
- Wang, J., Gu, J., and Leszczynski, J. 2006. Phosphonylation mechanisms of sarin and acetylcholinesterase: A model DFT study. *J. Phys. Chem. B.* 110(14):7567-7573.
- Yang, R. S., Thomas, R. S., Gustafson, D. L., Campain, J., Benjamin, S. A., Verhaar, H. J., and Mumtaz, M. M. 1998. Approaches to developing alternative and predictive toxicology based on *PBPK/PD* and QSAR modeling. *Environ. Health Perspect.* 106:1385-1393
- Yang, M. Zhou, L., Zuo, Z., Mancera, R., Song, H., Tang, X., Ma, X. 2009. Docking study and three-dimensional quantitative structure-activity relationship (3D-QSAR) analyses of transforming growth factor-B type I receptor kinase inhibitors. *QSAR Comb. Sci.* 28:: 1300-1308.
- Zhang, Y., Kua, J., and McCammon, J. A. 2002. Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: An ab initio QM/MM study. *J. Am. Chem. Soc.* 124: 10572-10577.

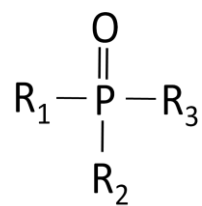


Figure 1. Generic chemical structure of OP compounds. R₁, R₂, and R₃ are side chains.

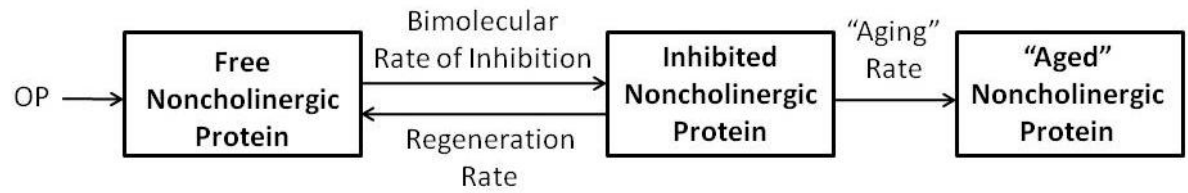


Figure 2. Schematic for noncholinergic enzyme inhibition and aging by OP nerve agents
(Adapted from Gearhart *et al.*, 1990).

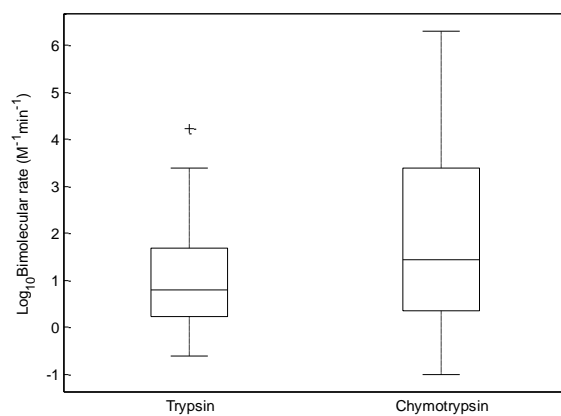


Figure 3. Log_{10} OP bimolecular rate constant box plots for trypsin and α -chymotrypsin global datasets.

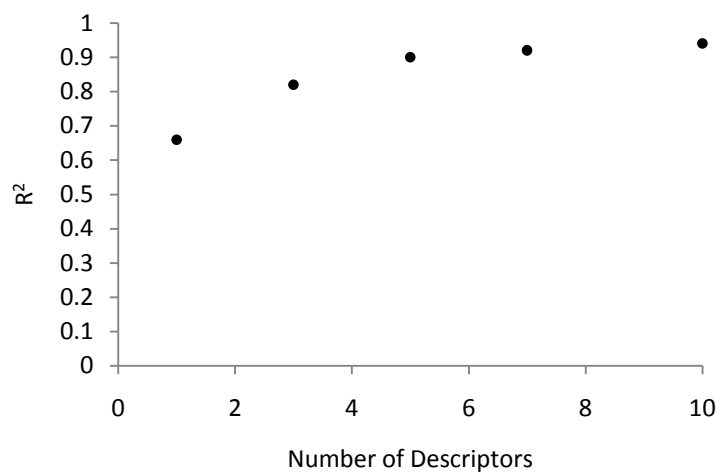


Figure 4. R^2 vs. number of descriptors for ABC trypsin bimolecular rate constant regression.

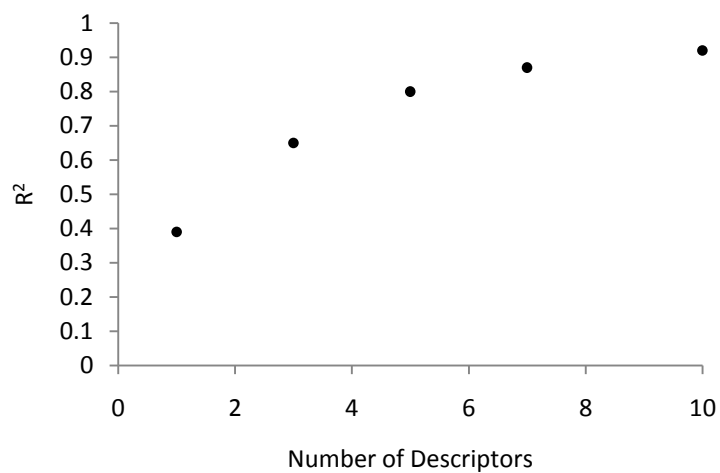


Figure 5. R^2 vs. number of descriptors for ABC α -chymotrypsin bimolecular rate constant regression.

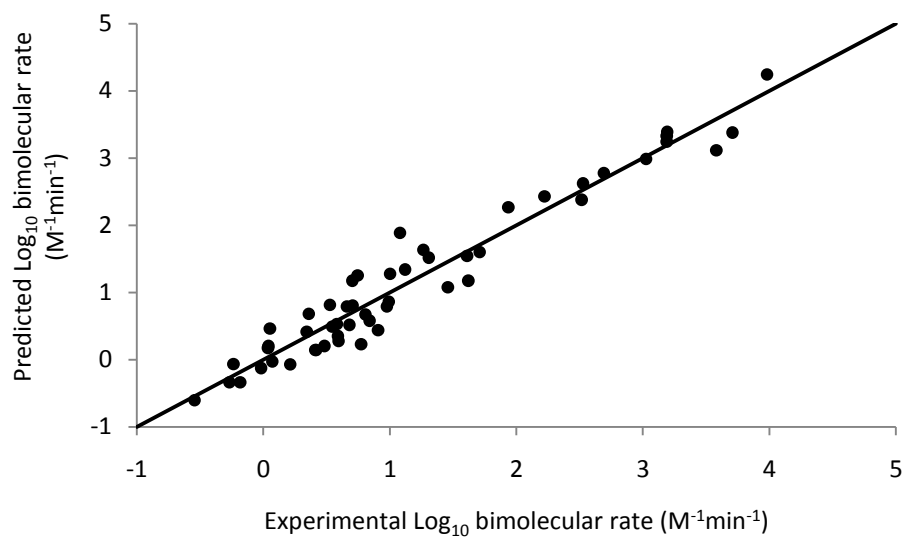


Figure 6. Experimental vs. predicted trypsin bimolecular rate constants for the ABC QSAR.

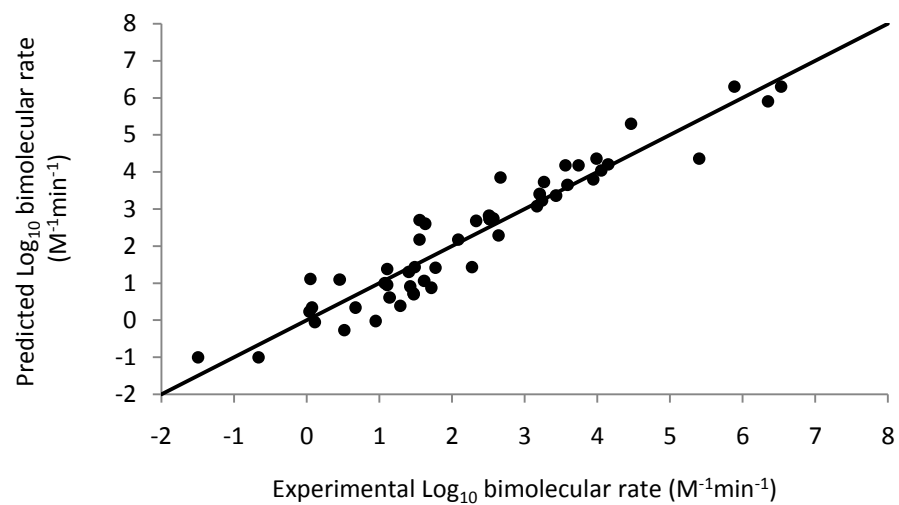


Figure 7. Experimental vs. predicted α -chymotrypsin bimolecular rate constants for the ABC QSAR.

Table 1. Descriptor codes and t-test values for parameters used in the trypsin bimolecular rate constant regressions.

Descriptor Code	Descriptor Name	T-test (Global training set)	T-test (AB training set)	T-test (AC training set)	T-test (BC training set)
NA ^a	Error	4.40	3.23	-6.87	0.77
D ₁	Number of F atoms	7.82	6.63	12.65	4.92
D ₂	Kier shape index (order 2)	9.83	8.09	8.12	4.08
D ₃	RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]	-0.84	0.10	8.33	1.80
D ₄	Kier&Hall index (order 3)	-7.49	-5.06	-6.66	-3.80
D ₅	Balaban index	-3.37	b	-2.65	-3.40
D ₆	PPSA-3 Atomic charge weighted PPSA [Zefirov's PC]	-5.81	-4.75	b	-3.40
D ₇	Number of O atoms	-5.24	-3.92	b	b
D ₈	Relative number of H atoms	-5.11	-5.17	b	1.03
D ₉	FPSA-1 Fractional PPSA (PPSA-1/TMSA) [Zefirov's PC]	3.66	3.10	b	b
D ₁₀	Kier shape index (order 3)	2.00	2.08	-1.23	-0.49

a. Not applicable. b. Not included in model.

Table 2. Descriptor codes and t-test values for parameters used in the α -chymotrypsin bimolecular rate constant regressions.

Descriptor Code	Descriptor Name	T-test (Global training set)	T-test (AB training set)	T-test (AC training set)	T-test (BC training set)
NA ^a	Error	9.78	7.14	6.59	2.90
D ₁₁	Relative number of aromatic bonds	-6.29	-2.90	-4.60	b
D ₁₂	Count of H-donors sites [Zefirov's PC]	-12.91	-9.26	-7.62	-5.26
D ₁₃	Kier&Hall index (order 3)	-13.13	-8.37	-10.10	b
D ₁₄	RPCG Relative positive charge (QMPOS/QTPLUS) [Zefirov's PC]	-5.49	-5.26	-4.42	b
D ₁₅	Number of S atoms	3.79	3.59	b	-1.23
D ₁₆	Number of O atoms	-4.35	-3.02	-1.78	-2.25
D ₁₇	Kier flexibility index	7.29	4.05	3.16	b
D ₁₈	Number of rings	6.65	b	4.59	-2.77
D ₁₉	Topographic electronic index (all bonds) [Zefirov's PC]	-4.81	-1.46	-2.65	2.62
D ₂₀	HA dependent HDSA-1 [Zefirov's PC]	2.83	b	b	b

a. Not applicable.

b. Not included in the model.

Table 3. Inter-correlation between descriptors in global trypsin bimolecular rate constant QSAR.

	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆	D ₇	D ₈	D ₉	D ₁₀
D ₁	1	-0.64	0.43	-0.67	0.67	-0.43	-0.79	0.63	0.22	0.05
D ₂		1	-0.73	0.73	-0.59	0.60	0.63	-0.06	0.30	0.59
D ₃			1	-0.27	0.42	-0.76	0.22	0.30	-0.18	-0.44
D ₄				1	-0.54	0.17	0.05	0.59	-0.44	0.21
D ₅					1	-0.47	0.67	-0.59	0.42	-0.54
D ₆						1	-0.43	0.60	-0.76	0.17
D ₇							1	-0.70	-0.25	0.10
D ₈								1	0.72	0.52
D ₉									1	0.58
D ₁₀										1

Table 4. Inter-correlation between descriptors in global α -chymotrypsin bimolecular rate constant QSAR.

	D ₁₁	D ₁₂	D ₁₃	D ₁₄	D ₁₅	D ₁₆	D ₁₇	D ₁₈	D ₁₉	D ₂₀
D ₁₁	1	-0.34	0.44	-0.71	-0.06	0.81	0.30	0.82	0.34	-0.32
D ₁₂		1	0.14	-0.16	0.20	-0.42	0.82	-0.19	0.32	0.64
D ₁₃			1	-0.77	0.46	0.18	0.34	0.32	0.43	-0.04
D ₁₄				1	-0.08	-0.49	-0.32	0.64	-0.04	-0.07
D ₁₅					1	-0.14	-0.06	0.20	0.46	-0.08
D ₁₆						1	0.81	-0.42	0.18	-0.49
D ₁₇							1	0.21	0.78	0.00
D ₁₈								1	0.39	-0.30
D ₁₉									1	0.25
D ₂₀										1

Table 5. Trypsin results from the external validation using the ABC approach.

Training set	Number of compounds	R^2	Q^2	F	s^2	Test set	Number of compounds	R^2_{test}	$\text{RMSE}_{\text{test}}$
A + B	35	0.95	0.88	48.46	0.10	C	17	0.85	0.46
A + C	35	0.94	0.91	77.02	0.11	B	17	0.59	0.70
B + C	34	0.91	0.84	30.89	0.15	A	18	0.82	0.56
Average	34.67	0.93	0.88	52.12	0.12	Average	17.33	0.75	0.57

R^2 =Coefficient of determination.

Q^2 =Cross-validated LOO R^2 .

F=Fisher F-test.

s^2 =Mean squared error. $s^2 = \sum_{i=1}^{Ns} ((Y_{ic} - Y_{io}) * (Y_{ic} - Y_{io})) / (Ns - Nd - 1)$ where Y_{ic} is the i th calculated/predicted property value, Y_{io} is the i th observed/input property value, Ns is the number of training structures, Nd is the number of descriptors and the sum runs from 1 to Ns .

RMSE: Root mean standard error.

Table 6. α -Chymotrypsin results from the external validation using the ABC approach.

Training Set	Number of Compounds	R^2	Q^2	F	s^2	Test Set	Number of Compounds	R^2_{test}	$\text{RMSE}_{\text{test}}$
A + B	42	0.86	0.79	26.24	0.66	C	20	0.86	0.91
A + C	41	0.90	0.63	34.21	0.49	B	21	0.81	1.13
B + C	41	0.68	0.58	14.55	1.36	A	21	0.16	1.89
Average	41.33	0.81	0.67	25.00	0.84	Average	20.67	0.61	1.31

R^2 =Coefficient of determination.

Q^2 =Cross-validated LOO R^2 .

F=Fisher F-test.

s^2 =Mean squared error. $s^2 = \sum_{i=1}^{Ns} ((Y_{ic} - Y_{io}) * (Y_{ic} - Y_{io})) / (Ns - Nd - 1)$ where Y_{ic} is the i th calculated/predicted property value, Y_{io} is the i th observed/input property value, Ns is the number of training structures, Nd is the number of descriptors and the sum runs from 1 to Ns .

RMSE: Root mean standard error.

Table 7. Bimolecular rate constant ($\text{M}^{-1}\text{min}^{-1}$) trypsin and α -chymotrypsin QSAR models for the ABC, AB, AC, and BC training sets as a function of a compound's descriptors (D_1, D_2, \dots, D_{20}).

Enzyme	Training Set	Descriptor	Coefficient	Standard Error
Trypsin	ABC	Error	17.31	3.93
		D1	2.45	0.31
		D2	0.96	0.098
		D3	-3.57	4.24
		D4	-0.63	0.084
		D5	-0.55	0.16
		D6	-0.85	0.15
		D7	-1.22	0.23
		D8	-27.42	5.36
		D9	9.9	2.71
		D10	0.15	0.072
Trypsin	AB	Error	14.78	4.57
		D1	2.81	0.42

	Trypsin	AC	D2	1.04	0.13
			D3	0.53	5.53
			D4	-0.53	0.1
			D6	-0.96	0.2
			D8	-32.74	6.34
			D7	-1.12	0.29
			D9	12.21	3.93
			D10	0.19	0.091
			Error	-6.75	0.98
			D1	3.36	0.27
	Trypsin	BC	D2	1.02	0.13
			D3	2.04	2.45
			D4	-0.64	0.1
			D5	-0.42	0.16
			D10	-0.1	0.08
			Error	1.624	2.1
			D1	2.52	0.51
			D10	-0.051	0.1

α -chymotrypsin	ABC	D5	-0.83	0.24
		D6	-0.48	0.14
		D2	0.63	0.15
		D4	-0.6	0.16
		D3	7.28	4.04
		D8	3.41	3.3
		Error	21.03	2.15
		D11	-19.81	3.15
		D12	-0.28	2.13
		D13	-1.88	0.14
		D14	-24.44	4.45
		D15	2.18	0.57
		D16	-0.81	0.19
		D17	1.12	0.15
		D18	2.51	0.38
		D19	-5.18	1.08
α -chymotrypsin	AB	D20	0.039	0.01
		Error	25.09	3.51

			D11	-14.85	5.12
			D12	-0.33	0.036
			D13	-1.96	0.23
			D14	-36.12	6.86
			D15	3.4	0.95
			D16	-0.97	0.32
			D17	0.72	0.18
			D19	-2.19	1.5
			Error	20.27	3.07
α-chymotrypsin	AC		D11	-21.66	4.7
			D12	-0.23	0.03
			D13	-1.71	0.17
			D14	-25.46	5.76
			D15	-3.71	1.4
			D16	1.89	0.41
			D17	0.71	0.23
			D19	-0.51	0.29
			Error	3.45	1.19
α-chymotrypsin	BC				

D18	-1.9	0.69
D19	-0.24	0.46
D12	3.08	1.18
D16	-0.68	0.29
D15	-0.92	0.75

Table 8. Trypsin tissue distribution in rat and human tissues compared with that of AChE.

Tissue	Rat AChE (nmoles/g tissue)	Rat Trypsin (nmoles/g tissue)	Human Trypsin (nmoles/g tissue)
Blood	0.0097 ^b	N/A ^a	N/A ^a
Serum	N/A ^a	0.0154 ^c	0.002 ^d
Pancreas	N/A ^a	5.069 ^c	N/A ^a

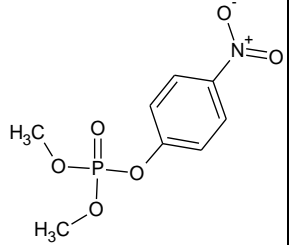
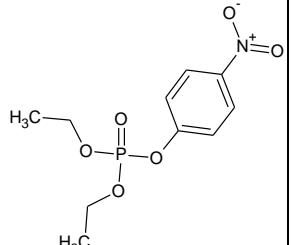
a. N/A=Not available.

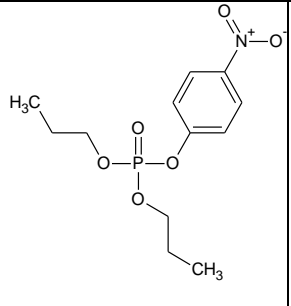
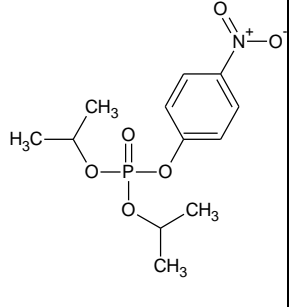
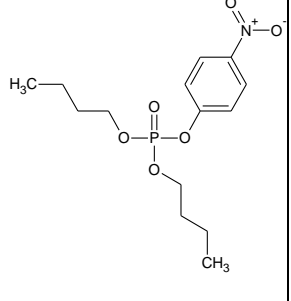
b. Sweeney et al., 2006.

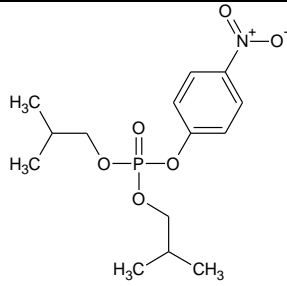
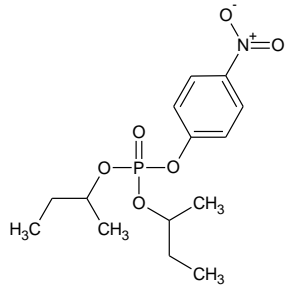
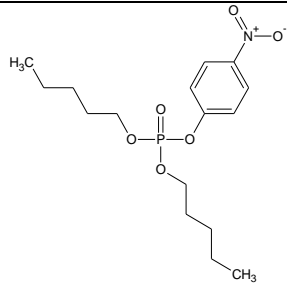
c. Reddy et al., 1985.

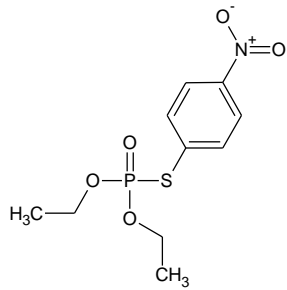
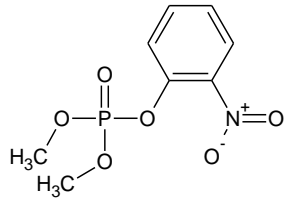
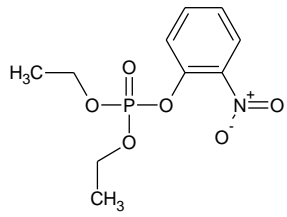
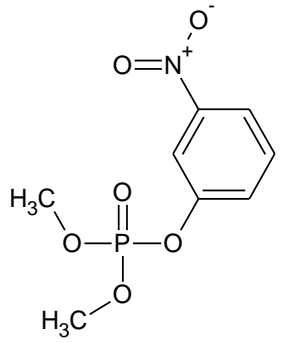
d. Frier et al., 1980.

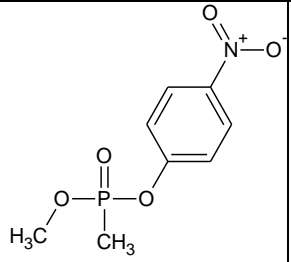
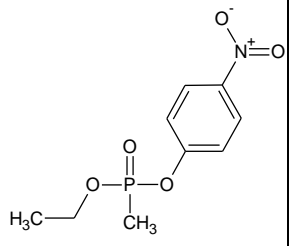
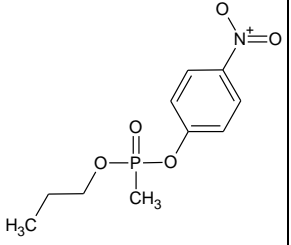
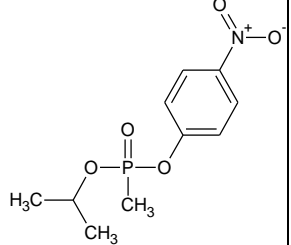
Appendix. Listing of OP compounds and their respective trypsin, α -chymotrypsin, and AChE bimolecular rate constants.

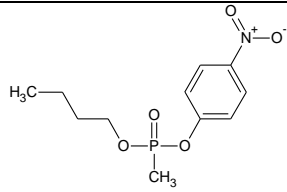
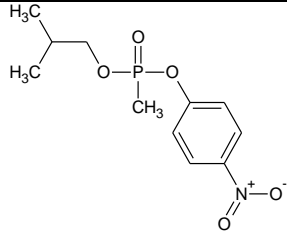
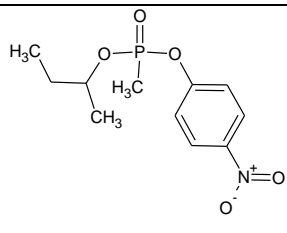
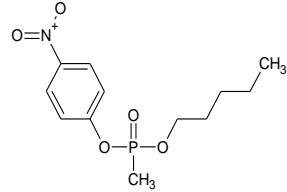
Label	Structure	IUPAC Name ^d Smiles notation ^c	Log ₁₀ trypsin rate constant (M ⁻¹ min ⁻¹) ^a	Log ₁₀ α -chymotrypsin rate constant (M ⁻¹ min ⁻¹) ^a	Log ₁₀ AChE rate constant (M ⁻¹ min ⁻¹) ^a
1		dimethyl 4-nitrophenyl phosphate (methyl paraoxon) <chem>[O-][N+](=O)c1ccc(OP(=O)(OC)OC)cc1</chem>	-0.70	0.34	5.08
2		diethyl 4-nitrophenyl phosphate (ethyl paraoxon) <chem>[O-][N+](=O)c1ccc(OP(=O)(OCC)OCC)cc1</chem>	0.20	0.95	5.65

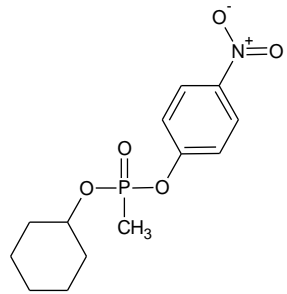
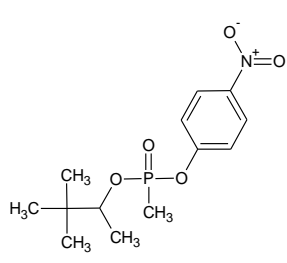
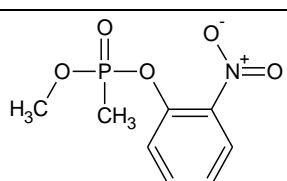
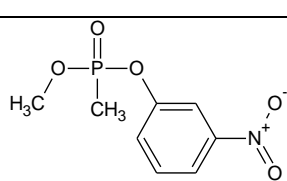
3		<p>4-nitrophenyl dipropyl phosphate (propyl paraoxon)</p> <p>[O-][N+](=O)c1ccc(OP(=O)(OCCC)OCCC)cc1</p> <p>1</p>	1.34	2.74	5.92
4		<p>4-nitrophenyl dipropan-2-yl phosphate (iso-propyl paraoxon)</p> <p>CC(C)OP(=O)(OC(C)C)Oc1ccc(cc1)[N+](O-)=O</p>	0.46	-0.02	4.32
5		<p>dibutyl 4-nitrophenyl phosphate (butyl paraoxon)</p> <p>[O-][N+](=O)c1ccc(OP(=O)(OCCCC)OCCCC)cc1</p>	1.08	3.41	6.11

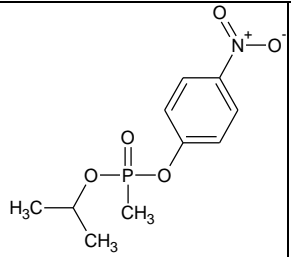
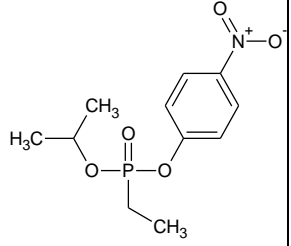
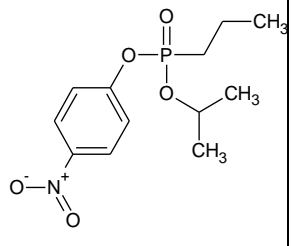
6		bis(2-methylpropyl) 4-nitrophenyl phosphate (iso-butyl paraoxon) <chem>CC(C)COP(=O)(OCC(C)C)Oc1ccc(cc1)[N+](=O)[O-]</chem>	1.26	3.85	5.88
7		dibutan-2-yl 4-nitrophenyl phosphate (sec-butyl paraoxon) <chem>CC(CC)OP(=O)(OC(C)CC)Oc1ccc(cc1)[N+](=O)[O-]</chem>	0.15	0.34	4.68
8		4-nitrophenyl dipentyl phosphate (pentyl paraoxon) <chem>[O-]P(=O)(OCCCCC)OCCCCc1ccc([N+](=O)[O-])cc1</chem>	1.18	3.80	6.26

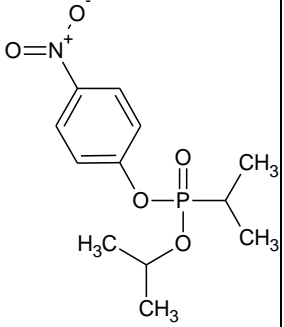
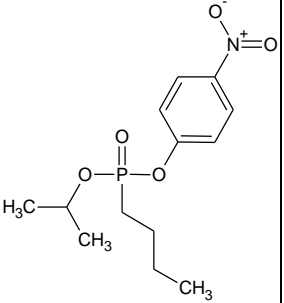
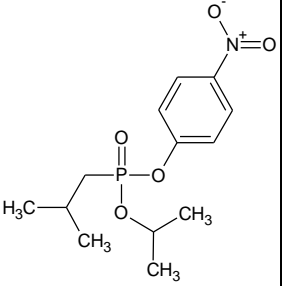
9		O,O-diethyl S-(4-nitrophenyl) phosphorothioate [O-] <chem>[O-][N+](=O)c1ccc(SP(=O)(OCC)OCC)cc1</chem>	0.20	0.39	5.67
10		dimethyl 2-nitrophenyl phosphate <chem>[O-][N+](=O)c1ccccc1OP(=O)(OC)OC</chem>	-0.12	1.11	3.71
11		diethyl 2-nitrophenyl phosphate <chem>[O-][N+](=O)c1ccccc1OP(=O)(OCC)OCC</chem>	-0.07	1.00	4.23
12		dimethyl 3-nitrophenyl phosphate <chem>[O-][N+](=O)c1cccc(OP(=O)(OC)OC)c1</chem>	-0.07	-0.05	2.72

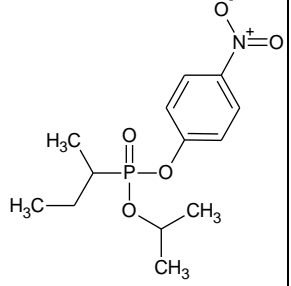
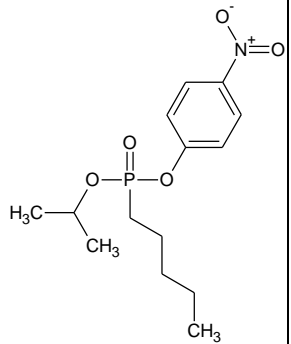
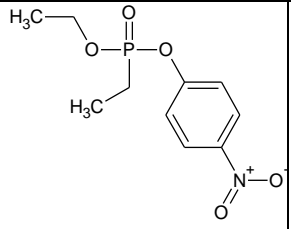
13		methyl 4-nitrophenyl methyl-phosphonate <chem>[O-][N+](=O)c1ccc(OP(C)(=O)OC)cc1</chem>	0.82	2.70	5.91
14		ethyl 4-nitrophenyl methyl-phosphonate <chem>[O-][N+](=O)c1ccc(OP(C)(=O)OCC)cc1</chem>	0.86	2.68	6.71
15		4-nitrophenyl propyl methyl-phosphonate <chem>[O-][N+](=O)c1ccc(OP(C)(=O)OCCC)cc1</chem>	1.52	3.08	7.08
16		4-nitrophenyl propan-2-yl methyl-phosphonate <chem>[O-][N+](=O)c1ccc(OP(C)(=O)OC(C)C)cc1</chem>	0.79	2.73	5.84

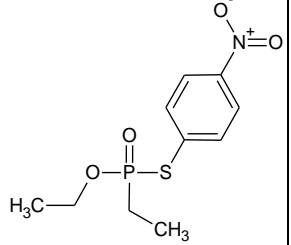
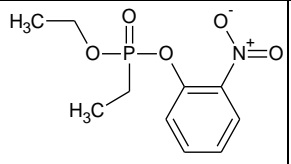
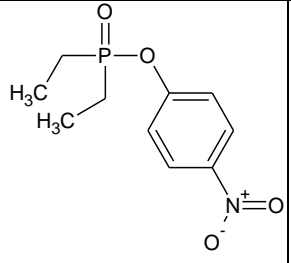
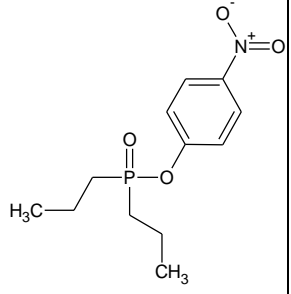
17		butyl 4-nitrophenyl methyl-phosphonate [O-][N+](=O)c1ccc(OP(C)(=O)OCCCC)cc1	1.54	3.65	7.32
18		2-methylpropyl 4-nitrophenyl methyl-phosphonate [O-][N+](=O)c1ccc(OP(C)(=O)OCC(C)C)cc1	1.28	3.36	7.96
19		butan-2-yl 4-nitrophenyl methyl-phosphonate [O-][N+](=O)c1ccc(OP(C)(=O)OC(C)CC)cc1	0.52	2.82	6.36
20		4-nitrophenyl pentyl methyl-phosphonate [O-][N+](=O)c1ccc(OP(C)(=O)OCCCCC)cc1	1.60	4.04	7.04

21		cyclohexyl 4-nitrophenyl methyl- phosphonate [O-][N+](=O)c2ccc(OP(C)(=O)OC1CCCCC1) cc2	1.18	4.18	6.66
22		3,3-dimethylbutan-2-yl 4-nitrophenyl methyl-phosphonate [O-][N+](=O)c1ccc(OP(C)(=O)OC(C)C(C)(C) C)cc1	-0.03	2.60	4.45
23		methyl 2-nitrophenyl methyl-phosphonate [O-][N+](=O)c1ccccc1OP(C)(=O)OC	0.81	2.18	4.54
24		methyl 3-nitrophenyl methyl-phosphonate [O-][N+](=O)c1cccc(OP(C)(=O)OC)c1	0.15	1.06	4.23

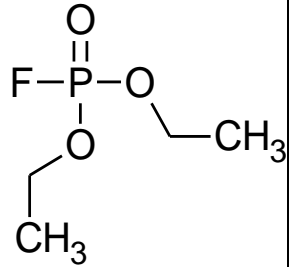
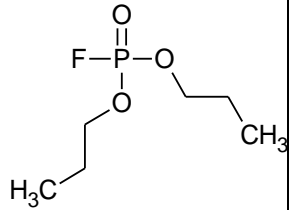
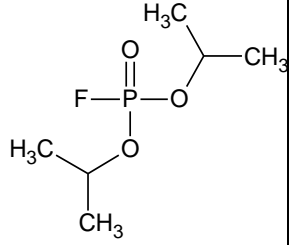
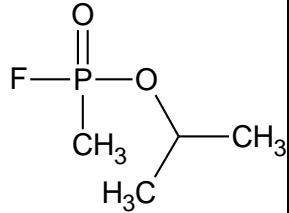
25		4-nitrophenyl propan-2-yl methyl- phosphonate [O-][N+](=O)c1ccc(OP(C)(=O)OC(C)C)cc1	0.79	2.73	5.85
26		4-nitrophenyl propan-2-yl ethyl- phosphonate [O-][N+](=O)c1ccc(OP(=O)(OC(C)C)CC)cc1	0.23	0.72	5.46
27		4-nitrophenyl propan-2-yl propyl- phosphonate [O-][N+](=O)c1ccc(OP(=O)(OC(C)C)CCC)cc 1	0.49	0.88	4.90

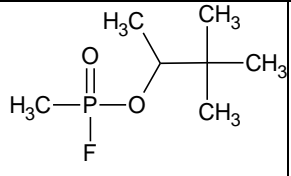
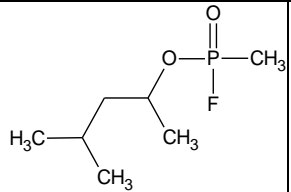
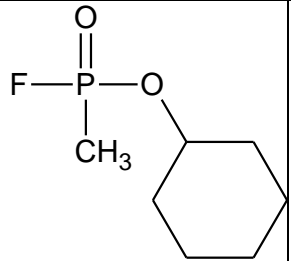
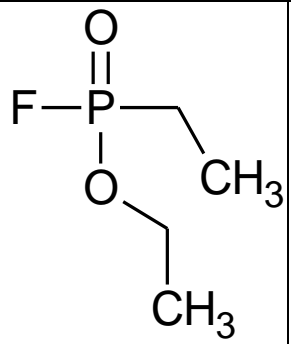
28		<p>4-nitrophenyl propan-2-yl propan-2-ylphosphonate</p> <chem>CC(C)P(=O)(OC(C)C)Oc1ccc(cc1)[N+](=[O-])=O</chem>	-0.34	-0.27	2.93
29		<p>4-nitrophenyl propan-2-yl butylphosphonate</p> <chem>[O-][N+](=O)c1ccc(OP(=O)(OC(C)C)CCCC)c1</chem>	0.67	1.43	5.10
30		<p>4-nitrophenyl propan-2-yl (2-methylpropyl)phosphonate</p> <chem>CC(C)CP(=O)(OC(C)C)Oc1ccc(cc1)[N+](=[O-])=O</chem>	0.53	1.43	4.32

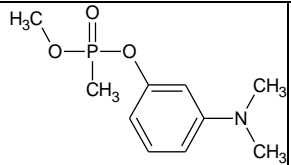
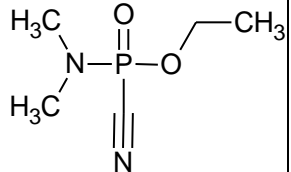
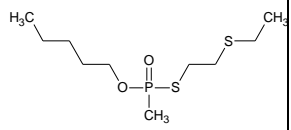
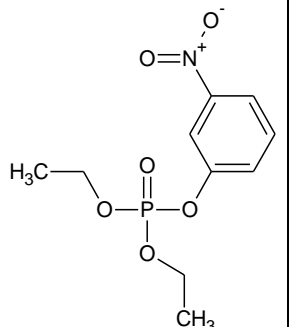
31		<p>4-nitrophenyl propan-2-yl butan-2-ylphosphonate</p> <chem>CC(CC)P(=O)(OC(C)C)Oc1ccc(cc1)[N+](O-)=O</chem>	0.18	0.23	3.11
32		<p>4-nitrophenyl propan-2-yl pentylphosphonate</p> <chem>[O-][N+](=O)c1ccc(OP(=O)(OC(C)C)CCCC)cc1</chem>	1.89	2.29	5.11
33		<p>ethyl 4-nitrophenyl ethylphosphonate</p> <chem>[O-][N+](=O)c1ccc(OP(=O)(OCC)CC)cc1</chem>	0.28	0.91	6.26

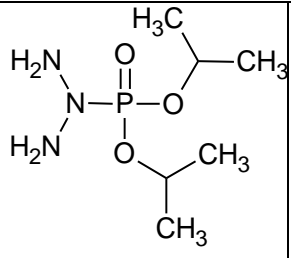
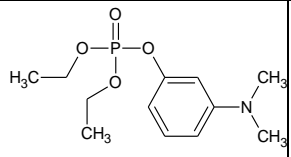
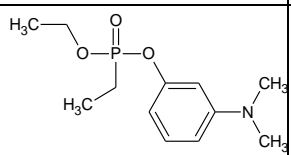
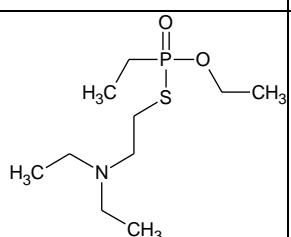
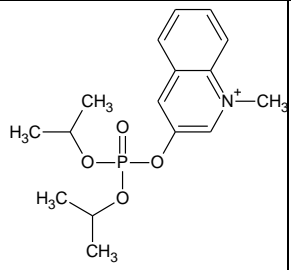
34		O-ethyl S-(4-nitrophenyl) ethyl-phosphonothioate <chem>[O-][N+](=O)c1ccc(SP(=O)(OCC)CC)cc1</chem>	0.68	1.10	6.59
35		ethyl 2-nitrophenyl ethyl-phosphonate <chem>[O-][N+](=O)c1ccccc1OP(=O)(OCC)CC</chem>	0.35	1.30	4.04
36		4-nitrophenyl diethyl-phosphinate <chem>O=P(Oc1ccc(cc1)[N+](O-)=O)(CC)CC</chem>	0.58	1.38	3.54
37		4-nitrophenyl dipropyl-phosphinate <chem>O=P(Oc1ccc(cc1)[N+](O-)=O)(CCC)CCC</chem>	1.63	2.18	4.26

38		4-nitrophenyl dibutylphosphinate <chem>O=P(Oc1ccc(cc1)[N+](=[O-])=O)(CCCC)CCCC</chem>	2.27	3.73	4.82
39		4-nitrophenyl bis(2-methylpropyl) phosphinate <chem>O=P(Oc1ccc(cc1)[N+](=[O-])=O)(CC(C)C)CC(C)C</chem>	0.44	1.41	2.60
40		4-nitrophenyl dipentylphosphinate <chem>O=P(Oc1ccc(cc1)[N+](=[O-])=O)(CCCCC)CCCCC</chem>	2.78	4.36	4.89
41		dimethyl phosphorofluoridate <chem>FP(=O)(OC)OC</chem>	2.62	3.23	4.86

42		diethyl phosphorofluoridate <chem>FP(=O)(OCC)OCC</chem>	3.39	4.20	5.31
43		dipropyl phosphorofluoridate <chem>CCCOP(F)(=O)OCCC</chem>	4.24	6.30	5.96
44		dipropan-2-yl phosphorofluoridate <chem>FP(=O)(OC(C)C)OC(C)C</chem>	2.99	4.18	4.66
45		propan-2-yl methyl-phosphonofluoridate (Sarin (GB)) <chem>CC(C)OP(C)(F)=O</chem>	3.24	4.36	7.15

46		3,3-dimethylbutan-2-yl methyl- phosphonofluoridate (Soman (GD)) <chem>CC(OP(C)(F)=O)C(C)(C)C</chem>	2.43	5.30	7.78
47		4-methylpentan-2-yl methyl- phosphonofluoridate <chem>CP(F)(=O)OC(C)CC(C)C</chem>	3.11	5.90	8.30
48		cyclohexyl methyl-phosphonofluoridate (Cyclosarin (GF)) <chem>O=P(C)(F)OC1CCCCC1</chem>	3.38	6.30	8.52
49		ethyl ethyl-phosphonofluoridate <chem>CCP(F)(=O)OCC</chem>	3.33	4.53	6.49

50		3-(dimethylamino)phenyl methyl methylphosphonate <chem>CN(C)c1cccc(OP(C)(=O)OC)c1</chem>	-0.34	-0.51	3.04
51		ethyl dimethyl-phosphoramido-cyanidate (Tabun (GA)) <chem>N#CP(=O)(OCC)N(C)C</chem>	0.41	3.20	6.56
52		S-[2-(ethylsulfanyl)ethyl] O-pentyl methyl- phosphonothioate <chem>CCCCCOP(C)(=O)SCCSCC</chem>	2.38	N/A ^b	N/A ^b
53		diethyl 3-nitrophenyl phosphate [O-][N+](=O)c1cccc(OP(=O)(OCC)OCC)c1	N/A ^b	0.61	4.04

54		dipropan-2-yl triazan-2-ylphosphonate <chem>O=P(OC(C)C)(OC(C)C)N(N)N</chem>	N/A ^b	1.90	5.15
55		3-(dimethylamino)phenyl diethyl phosphate <chem>CN(C)c1cccc(OP(=O)(OCC)OCC)c1</chem>	N/A ^b	-1.00	3.56
56		3-(dimethylamino)phenyl ethyl ethyl- phosphonate <chem>CN(C)c1cccc(OP(=O)(OCC)CC)c1</chem>	N/A ^b	-1.00	2.11
57		S-[2-(diethylamino)ethyl] O-ethyl ethyl- phosphonothioate <chem>CCOP(=O)(CC)SCCN(CC)CC</chem>	N/A ^b	-0.12	7.61
58		3-[[bis(propan-2-yloxy)phosphoryl]oxy]-1- methylquinolinium <chem>CC(C)OP(=O)(OC(C)C)Oc2cc1cccc1[n+](C)c2</chem>	N/A ^b	-1.00	6.26

59		3-(dimethylamino)phenyl dimethyl phosphate <chem>CN(C)c1cccc(OP(=O)(OC)OC)c1</chem>	N/A ^b	-1.00	3.34
60		4-nitrophenyl dibutan-2-ylphosphinate <chem>O=P(Oc1ccc(cc1)[N+](=[O-])=O)(C(C)CC)C(C)CC</chem>	N/A ^b	-1.00	3.34
61		4-nitrophenyl dipropan-2-ylphosphinate <chem>O=P(Oc1ccc(cc1)[N+](=[O-])=O)(C(C)C)C(C)C</chem>	N/A ^b	-1.00	2.94
62		ethyl 3-nitrophenyl ethyl-phosphonate <chem>[O-][N+](=O)c1cccc(OP(=O)(OCC)CC)c1</chem>	N/A ^b	0.70	4.41
63		2-(diethylamino) ethyl ethyl ethyl- phosphonate <chem>CCN(CCOP(=O)(OCC)CC)CC</chem>	N/A ^b	0.0	N/A ^b

- a. Data taken from Ooms (1961).
- b. N/A = No published data available.
- c. SMILES = Simplified Molecular Input Line Entry Specification
(<http://www.daylight.com>) used to represent the chemical structure.
- d. IUPAC Name = International Union of Pure and Applied Chemistry nomenclature.