Quantitative Structure-Activity Relationships for Organophosphate Enzyme Inhibition (Briefing Charts)

Christopher Ruark & Kyung Yu

711th Human Performance Wing
Human Effectiveness Directorate
Air Force Research Laboratory
“Biological Modeling Group”

22 September 11
REPORT DOCUMENTATION PAGE
Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

<table>
<thead>
<tr>
<th>REPORT DATE (DD-MM-YYYY)</th>
<th>22-09-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>REPORT TYPE</td>
<td>Interim</td>
</tr>
<tr>
<td>DATES COVERED (From - To)</td>
<td>September 10 – September 12</td>
</tr>
</tbody>
</table>

**4. TITLE AND SUBTITLE**
Quantitative Structure-Activity Relationships for Organophosphate Enzyme Inhibition (Briefing Charts)

**6. AUTHOR(S)**
Christopher Ruark* & Kyung Yu**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
*Henry M. Jackson Foundation for the Advancement of Military Medicine, Wright-Patterson AFB OH 45433
**AFRL, 711 HPW, Human Effectiveness Directorate, Bioeffects Division, Wright-Patterson AFB OH 45433

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Air Force Materiel Command
Air Force Research Laboratory
711 Human Performance Wing
Human Effectiveness Directorate
Bioeffects Division
Molecular Bioeffects Branch
Wright-Patterson AFB OH 45433-5707

**13. SUPPLEMENTARY NOTES**
Distribution A: Approved for public release; distribution unlimited.

**14. ABSTRACT**
Organophosphates (OPs) are a group of pesticides that inhibit enzymes such as acetylcholinesterase. Numerous OP structural variants exist and toxicity data can be difficult to quickly obtain. To address this concern, quantitative structure-activity relationship (QSAR) models were developed to predict acetylcholinesterase, butyrylcholinesterase, trypsin and chymotrypsin inhibition, key components in biologically-based dose-response (BBDR) models. The acetylcholinesterase database consisted of 747 structures developed from 69 peer reviewed publications. AMPAC and CODESSA descriptors (SemiChem, Inc.) were calculated for each compound. The acetylcholinesterase results show that the average nucleophilic reactive index for a carbon atom contributed most significantly to binding. A training $R^2$ of 0.73±0.01 and an external test set $Q^2$ of 0.62±0.06 was achieved. The QSAR models discussed in this seminar will complement OP BBDR modeling by filling critical data gaps for key parameter values, leading to better risk assessment and prioritization of animal and human toxicity studies especially for OPs lacking experimental data.

**15. SUBJECT TERMS**
QSAR, Organophosphates, Acetylcholinesterase, BBDR

**16. SECURITY CLASSIFICATION OF:**

<table>
<thead>
<tr>
<th>a. REPORT</th>
<th>b. ABSTRACT</th>
<th>c. THIS PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
</tbody>
</table>

**17. LIMITATION OF ABSTRACT**
SAR

**18. NUMBER OF PAGES**
26

**19a. NAME OF RESPONSIBLE PERSON**
Kyung Yu

**19b. TELEPHONE NUMBER**
NA

Standard Form 298 (Rev. 8-98) Prescribed by ANSI Std. 239.18
1. Introduction
   a) What is QSAR?
   b) Organophosphate structure and mechanism of toxicity
   c) Linking QSAR and OP PBPK/PD

2. Methods
   a) Physiochemical Descriptors
   b) Regression Techniques

3. Results
   a) Bimolecular rate constants
      i. Acetylcholinesterase & Butyrylcholinesterase
      ii. Trypsin & Chymotrypsin

4. Discussion
**What is QSAR?**

**Quantitative Structure-Activity Relationship**

A technique used to quantify differences between biological activity and that of a molecular structure.

There are guidelines/rules to this approach:

1. Choose well-defined Activity endpoints
2. Choose plausible molecular descriptors
3. Explore the data with statistics
4. Test hypotheses with new data (ie. iterate)
QSAR Overview

1. Original Dataset
2. Calculate descriptors
3. Randomly Generate Multiple Training Sets
4. Activity/Property Prediction with QSAR/QSPR
5. Predict the activity of the Test Sets
6. Exclusion criteria for correlation $q^2 < 0.6$, $R^2 < 0.6$, etc.
7. External Validation using Applicability Domain
8. Select acceptable models
Cholinergic Nervous System

“Normal Mechanism of Action”

Pyruvate → Citrate → Acetyl CoA → Acetylcholine 

Choline Acetyltransferase → Acetylcholine 

Acetylcholinesterase → Acetate + Choline 

Pre-synaptic Terminal → Synaptic Cleft → Post-synaptic Terminal 

Muscarinic/Nicotinic Receptor 

Choline Carrier
Cholinergic Nervous System
“OP Mechanism of Action”

Pre-synaptic Terminal
Choline Carrier
Synaptic Cleft
Muscarinic/Nicotinic Receptor

Choline Acetyltransferase
CoA
Acetyl
Citrate
Pyruvate

Acetylcholine

Acetylcholinesterase

Post-synaptic Terminal

Hyperstimulation

Acetyl + CoA

Citrate

Acetylcholine

Pyruvate

Hyperstimulation

Muscarinic/Nicotinic Receptor

Choline Carrier

Pre-synaptic Terminal

Choline Acetyltransferase

Acetylcholine

Acetyl + CoA

Citrate

Pyruvate
Physiologically Based Pharmacokinetic/Pharmacodynamic Modeling

Free Cholinesterase

Inhibited Cholinesterase

OP

Bimolecular Inhibition

Synthesis of New AChE

Regeneration

Basal degradation

IV dose

inhalation

exhalation

Lungs

Brain

Rapidly Perfused

Slowly Perfused

Fat

Kidney

Liver

GI Tract

Urine

oral dose

metabolism

Aging

Aged Cholinesterase
Organophosphate Structures

Acephate

Malathion

Chlorpyrifos

Isifenphos

Acetylcholine

Parathion

Glyphosate

Diazinon

Fenthion
Constitutional Descriptors

Reflect molecular composition of compound without using geometry or electronic structure of molecule:

- Number of atoms
  - Absolute and relative numbers of C, H, O, S, N, F, Cl, Br, I, P atoms

- Number of bonds
  - Absolute and relative numbers of single, double, triple and aromatic bonds

- Number of rings
  - Number of rings divided by the number of atoms, number of benzene rings, number of benzene rings divided by the number of atoms

- Molecular and average atomic weight
Topostructural Descriptors

A molecular graph is made up of Edges and Vertices

Adjacency matrix (A)

$$A(G) = \begin{bmatrix}
(1) & (2) & (3) & (4) \\
(1) & 0 & 1 & 0 & 0 \\
(2) & 1 & 0 & 1 & 1 \\
(3) & 0 & 1 & 0 & 0 \\
(4) & 0 & 1 & 0 & 0 \\
\end{bmatrix}$$

Distance matrix (D)

$$D(G) = \begin{bmatrix}
(1) & (2) & (3) & (4) \\
(1) & 0 & 1 & 2 & 2 \\
(2) & 1 & 0 & 1 & 1 \\
(3) & 2 & 1 & 0 & 2 \\
(4) & 2 & 1 & 2 & 0 \\
\end{bmatrix}$$

Many topostructural indices can be derived from matrices A and D
Regression Techniques

Linear Regression Examples
- Heuristic
- Partial Least Squares (PLS)
- Principle Component Regression (PCR)
- Orthogonal Projection to Latent Structures (OPLS)
- Ridge Regression

Non-Linear Regression Examples
- Support Vector Machines (SVM)
- Neural Networks (NN)
- Kernel Orthogonal Projection to Latent Structures (KOPLS)
- Kernel Partial Least Squares (KPLS)

Clustering Regression Examples
- k-nearest neighbor
- Random Forest
Overview of orthogonal projection to latent structures (O-PLS)

- Original data set
  - PLS
  - OPLS treated data
  - Orthogonal data set
- Original data set
  - PLS
  - OPLS treated data
  - Orthogonal data set

- Harder to interpret
- More PLS components
- Orthogonal variation in X

• Evaluate orthogonal variation in principal components
• Identify source of orthogonal variation

• Easier to interpret
• Fewer components
• More relevant

Adapted from Trygg and Wold 2002
Acetylcholinesterase Bimolecular Rate Constants (M⁻¹min⁻¹)

Table 1. Percentage of database that represents a particular temperature and species.

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>25</th>
<th>37</th>
<th>Unknown</th>
<th>5</th>
<th>30</th>
<th>22</th>
<th>27</th>
<th>Total Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>28.46</td>
<td>6.93</td>
<td>8.15</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
<td>43.72</td>
</tr>
<tr>
<td>Bovine</td>
<td>21.44</td>
<td>3.56</td>
<td>0.09</td>
<td>0.09</td>
<td>1.22</td>
<td>0.00</td>
<td>0.00</td>
<td>26.40</td>
</tr>
<tr>
<td>Unknown</td>
<td>7.77</td>
<td>0.28</td>
<td>1.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9.55</td>
</tr>
<tr>
<td>Fly</td>
<td>2.25</td>
<td>0.56</td>
<td>4.87</td>
<td>0.00</td>
<td>1.03</td>
<td>0.00</td>
<td>0.00</td>
<td>8.71</td>
</tr>
<tr>
<td>Rat</td>
<td>0.00</td>
<td>1.22</td>
<td>0.47</td>
<td>0.66</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>2.81</td>
</tr>
<tr>
<td>Hen</td>
<td>0.56</td>
<td>0.47</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>0.00</td>
<td>1.50</td>
</tr>
<tr>
<td>Rabbit</td>
<td>0.47</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.03</td>
<td>0.00</td>
<td>0.00</td>
<td>1.50</td>
</tr>
<tr>
<td>Eel</td>
<td>0.75</td>
<td>0.75</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.50</td>
</tr>
<tr>
<td>Cricket</td>
<td>0.66</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.66</td>
</tr>
<tr>
<td>Guinea Pig</td>
<td>0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>0.66</td>
</tr>
<tr>
<td>Pig</td>
<td>0.00</td>
<td>0.66</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.66</td>
</tr>
<tr>
<td>Mouse</td>
<td>0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>0.00</td>
<td>0.56</td>
</tr>
<tr>
<td>NHP</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>0.47</td>
</tr>
<tr>
<td>Catfish</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>0.47</td>
</tr>
<tr>
<td>Frog</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>0.47</td>
</tr>
<tr>
<td>Minipig</td>
<td>0.00</td>
<td>0.37</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.37</td>
</tr>
<tr>
<td>Total Percent</td>
<td>62.73</td>
<td>14.79</td>
<td>15.07</td>
<td>0.75</td>
<td>3.37</td>
<td>3.09</td>
<td>0.19</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Data collected from 69 peer-reviewed journal articles.

1068 observations
Acetylcholinesterase Bimolecular Rate Constants (M\(^{-1}\) min\(^{-1}\))

Global Human Orthogonal-PLS

Monte Carlo/Bootstrap Cross-Validation
"Leave-random-number-out" Consensus QSAR Predictions

A mean training $R^2$ of 0.77±0.02 and an external test set $Q^2$ of 0.64±0.10 was achieved using the significant uncorrelated descriptors. Y-randomization $Q^2=-0.23±0.18$. 

Global training $R^2=0.91$ using 74 significant and uncorrelated descriptors.
A number of techniques exist to quantify the Domain of Applicability. QSAR model predictions are only valid within the applicability domain. If your test compound falls within the DOA then you can expect a reliable prediction. Compounds with high leverage can heavily influence a model. Predicted responses outside of the warning leverage may not be reliable. Possible outliers.
## AChE Descriptor Significance

<table>
<thead>
<tr>
<th>Descriptor Name</th>
<th>Normalized P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg nucleoph. react. index for a C atom</td>
<td>1.000</td>
</tr>
<tr>
<td>HOMO energy</td>
<td>0.995</td>
</tr>
<tr>
<td>Min nucleoph. react. index for a O atom</td>
<td>0.991</td>
</tr>
<tr>
<td>Max nucleoph. react. index for a C atom</td>
<td>0.947</td>
</tr>
<tr>
<td>Max n-n repulsion for a C-H bond</td>
<td>0.889</td>
</tr>
<tr>
<td>(1/6)X GAMMA polarizability (DIP)</td>
<td>0.883</td>
</tr>
<tr>
<td>1X GAMMA polarizability (DIP)</td>
<td>0.883</td>
</tr>
<tr>
<td>Max e-n attraction for a C-H bond</td>
<td>0.831</td>
</tr>
<tr>
<td>HOMO - LUMO energy gap</td>
<td>-0.827</td>
</tr>
<tr>
<td>ESP-Max net atomic charge for a F atom</td>
<td>-0.827</td>
</tr>
</tbody>
</table>
Regression Techniques

• Heuristic (Built into CODESSA 2.51)
  – Pre-selection of descriptors based upon a series of criteria cutoffs
    • Variation in descriptors
    • F-test
    • $R^2$
    • T-value
    • Inter-correlation
  – F-test measures significance of the whole model, t-test reflects significance of the parameter.
Butyrylcholinesterase
“Serum Cholinesterase”

Inhalation VX Minipig - 0.046 mg/m³

Time (hr)
0 2 4 6

Fraction Control ChE Activity
0.0 0.2 0.4 0.6 0.8 1.0 1.2

Bimolecular Rate
(M⁻¹min⁻¹)

Inhibition of AChE and BChE in Blood

Inhalation VX Minipig - 0.046 mg/m³ (#28)

AChE
BChE

R²=0.82, 25 descriptors,
F=71.69, 410 compounds

Data taken from literature.
Noncholinergic Targets

Primary target

Secondary targets

ACH mortality

BChE, mAChR Cholinergic interactions

NTE-LysoPLA Delayed neurotoxicity

FAAH, CB1 Cannabinoid interactions

Carboxylesterases Amidases Toxicity interactions

APH neuropeptide metabolism

AFMID teratogenesis

Other serine hydrolases Various actions
Noncholinergic Targets
Digestive Proteases

Trypsin

Bimolecular Rate (M⁻¹min⁻¹)

$R^2=0.94$, $Q^2=0.90$, 52 structures, 10 descriptors

Chymotrypsin

Bimolecular Rate (M⁻¹min⁻¹)

$R^2=0.92$, $Q^2=0.87$, 62 structures, 10 descriptors

Ruark et al. 2011
### External Validation

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

<table>
<thead>
<tr>
<th>Training set</th>
<th>Number of compounds</th>
<th>R²</th>
<th>Q²</th>
<th>F</th>
<th>s²</th>
<th>Test set</th>
<th>Number of compounds</th>
<th>R²&lt;sub&gt;test&lt;/sub&gt;</th>
<th>RMSE&lt;sub&gt;test&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A + B</td>
<td>35</td>
<td>0.95</td>
<td>0.88</td>
<td>48.46</td>
<td>0.10</td>
<td>C</td>
<td>17</td>
<td>0.85</td>
<td>0.46</td>
</tr>
<tr>
<td>A + C</td>
<td>35</td>
<td>0.94</td>
<td>0.91</td>
<td>77.02</td>
<td>0.11</td>
<td>B</td>
<td>17</td>
<td>0.59</td>
<td>0.70</td>
</tr>
<tr>
<td>B + C</td>
<td>34</td>
<td>0.91</td>
<td>0.84</td>
<td>30.89</td>
<td>0.15</td>
<td>A</td>
<td>18</td>
<td>0.82</td>
<td>0.56</td>
</tr>
<tr>
<td>Average</td>
<td>34.67</td>
<td>0.93</td>
<td>0.88</td>
<td>52.12</td>
<td>0.12</td>
<td>Average</td>
<td>17.33</td>
<td>0.75</td>
<td>0.57</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Number of Compounds</th>
<th>R²</th>
<th>Q²</th>
<th>F</th>
<th>s²</th>
<th>Test Set</th>
<th>Number of Compounds</th>
<th>R²&lt;sub&gt;test&lt;/sub&gt;</th>
<th>RMSE&lt;sub&gt;test&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A + B</td>
<td>42</td>
<td>0.86</td>
<td>0.79</td>
<td>26.24</td>
<td>0.66</td>
<td>C</td>
<td>20</td>
<td>0.86</td>
<td>0.91</td>
</tr>
<tr>
<td>A + C</td>
<td>41</td>
<td>0.90</td>
<td>0.63</td>
<td>34.21</td>
<td>0.49</td>
<td>B</td>
<td>21</td>
<td>0.81</td>
<td>1.13</td>
</tr>
<tr>
<td>B + C</td>
<td>41</td>
<td>0.68</td>
<td>0.58</td>
<td>14.55</td>
<td>1.36</td>
<td>A</td>
<td>21</td>
<td>0.16</td>
<td>1.89</td>
</tr>
<tr>
<td>Average</td>
<td>41.33</td>
<td>0.81</td>
<td>0.67</td>
<td>25.00</td>
<td>0.84</td>
<td>Average</td>
<td>20.67</td>
<td>0.61</td>
<td>1.31</td>
</tr>
</tbody>
</table>

R² = Coefficient of determination.

Q² = Cross-validated LOO R².

F = Fisher F-test.

s² = Mean squared error. \( s^2 = \sum \limits_{i=1}^{N_s} ((Y_{ic} - Y_{io}) + (Y_{ic} - Y_{io})) / (N_s - N_d - 1) \) where Yic is the ith calculated/predicted property value, Yio is the ith 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.
## Trypsin Descriptors

<table>
<thead>
<tr>
<th>Descriptor Code</th>
<th>Descriptor Name</th>
<th>T-test (Global training set)</th>
<th>T-test (AB training set)</th>
<th>T-test (AC training set)</th>
<th>T-test (BC training set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA²</td>
<td>Error</td>
<td>4.40</td>
<td>3.23</td>
<td>-6.87</td>
<td>0.77</td>
</tr>
<tr>
<td>D₁</td>
<td>Number of F atoms</td>
<td>7.82</td>
<td>6.63</td>
<td>12.65</td>
<td>4.92</td>
</tr>
<tr>
<td>D₂</td>
<td>Kier shape index (order 2)</td>
<td>9.83</td>
<td>8.09</td>
<td>8.12</td>
<td>4.08</td>
</tr>
<tr>
<td>D₃</td>
<td>RNCG Relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]</td>
<td>-0.84</td>
<td>0.10</td>
<td>8.33</td>
<td>1.80</td>
</tr>
<tr>
<td>D₄</td>
<td>Kier&amp;Hall index (order 3)</td>
<td>-7.49</td>
<td>-5.06</td>
<td>-6.66</td>
<td>-3.80</td>
</tr>
<tr>
<td>D₅</td>
<td>Balaban index</td>
<td>-3.37</td>
<td>b</td>
<td>-2.65</td>
<td>-3.40</td>
</tr>
<tr>
<td>D₆</td>
<td>PPSA-3 Atomic charge weighted PPSA [Zefirov's PC]</td>
<td>-5.81</td>
<td>-4.75</td>
<td>b</td>
<td>-3.40</td>
</tr>
<tr>
<td>D₇</td>
<td>Number of O atoms</td>
<td>-5.24</td>
<td>-3.92</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>D₈</td>
<td>Relative number of H atoms</td>
<td>-5.11</td>
<td>-5.17</td>
<td>b</td>
<td>1.03</td>
</tr>
<tr>
<td>D₉</td>
<td>FPSA-1 Fractional PPSA (PPSA-1/TMSA) [Zefirov's PC]</td>
<td>3.66</td>
<td>3.10</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>D₁₀</td>
<td>Kier shape index (order 3)</td>
<td>2.00</td>
<td>2.08</td>
<td>-1.23</td>
<td>-0.49</td>
</tr>
</tbody>
</table>

Ruark et al. 2011
Conclusion

1. QSAR can be used to predict organophosphate oxon bimolecular rate constants for AChE, BChE, trypsin and chymotrypsin.

2. Approach can be applied to other PBPK/PD modeling parameters.

3. QSAR descriptors can provide a mechanistic description of the enzymatic reactions.
   Steric hindrance, connectivity, lipophilicity, electrophilicity, electrostatics, hydrogen bonding, van der Waals
Acknowledgements

Dr. Jeff Gearhart
Dr. Peter Robinson
Eric Hack
Dr. Paul Anderson

ACS Conference Organizers

This project received support from the Defense Threat Reduction Agency – Joint Science and Technology Office, Basic and Supporting Sciences Division.