

AFRL-RH-WP-TR-2012-0089

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

22 September 11

**Christopher Ruark
& Kyung Yu**

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



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.					
1. REPORT DATE (DD-MM-YYYY) 22-09-2011		2. REPORT TYPE Interim		3. DATES COVERED (From - To) September 10 – September 12	
4. TITLE AND SUBTITLE Quantitative Structure-Activity Relationships for Organophosphate Enzyme Inhibition (Briefing Charts)				5a. CONTRACT NUMBER NA	
				5b. GRANT NUMBER NA	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Christopher Ruark* & Kyung Yu**				5d. PROJECT NUMBER ODT	
				5e. TASK NUMBER WP	
				5f. WORK UNIT NUMBER ODTWP004	
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				8. PERFORMING ORGANIZATION REPORT NUMBER	
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				10. SPONSOR/MONITOR'S ACRONYM(S) 711 HPW/RHDJ	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RH-WP-TR-2012-0089	
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution A: Approved for public release; distribution unlimited.					
13. SUPPLEMENTARY NOTES					
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:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kyung Yu
U	U	U	SAR	26	19b. TELEPHONE NUMBER (include area code) NA



Outline

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

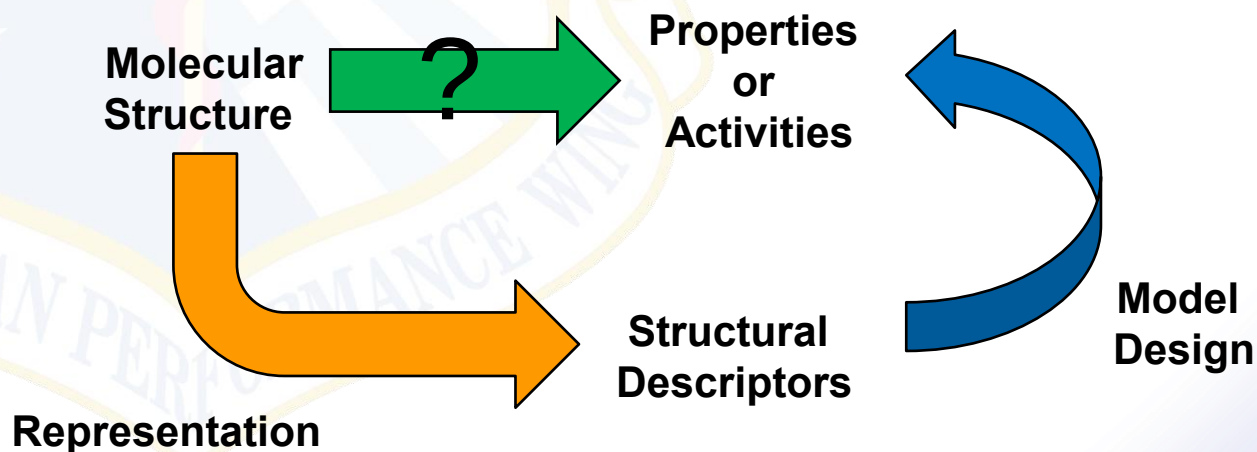




What is QSAR?

Quantitative Structure-Activity Relationship

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



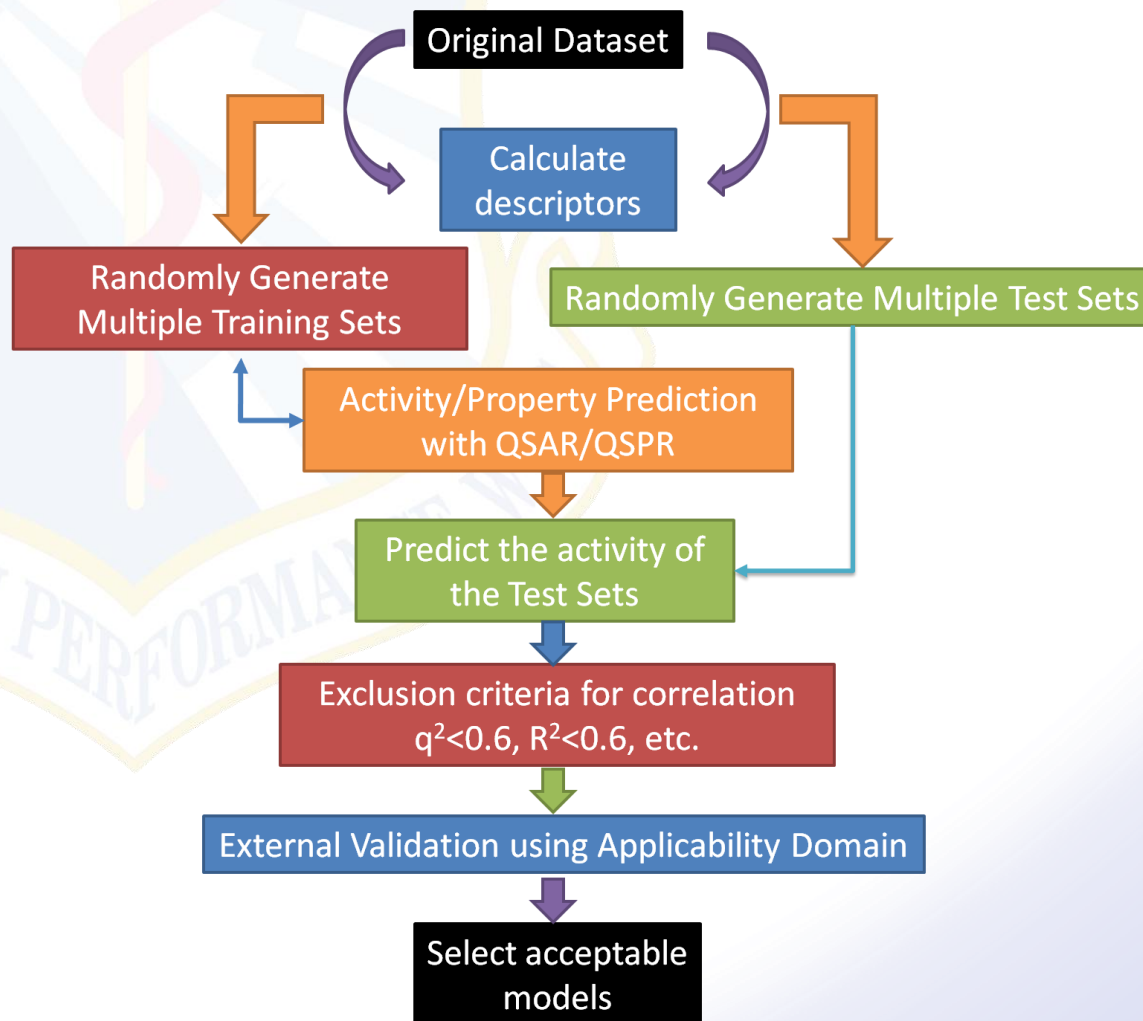
There are guidelines/rules to this approach

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)





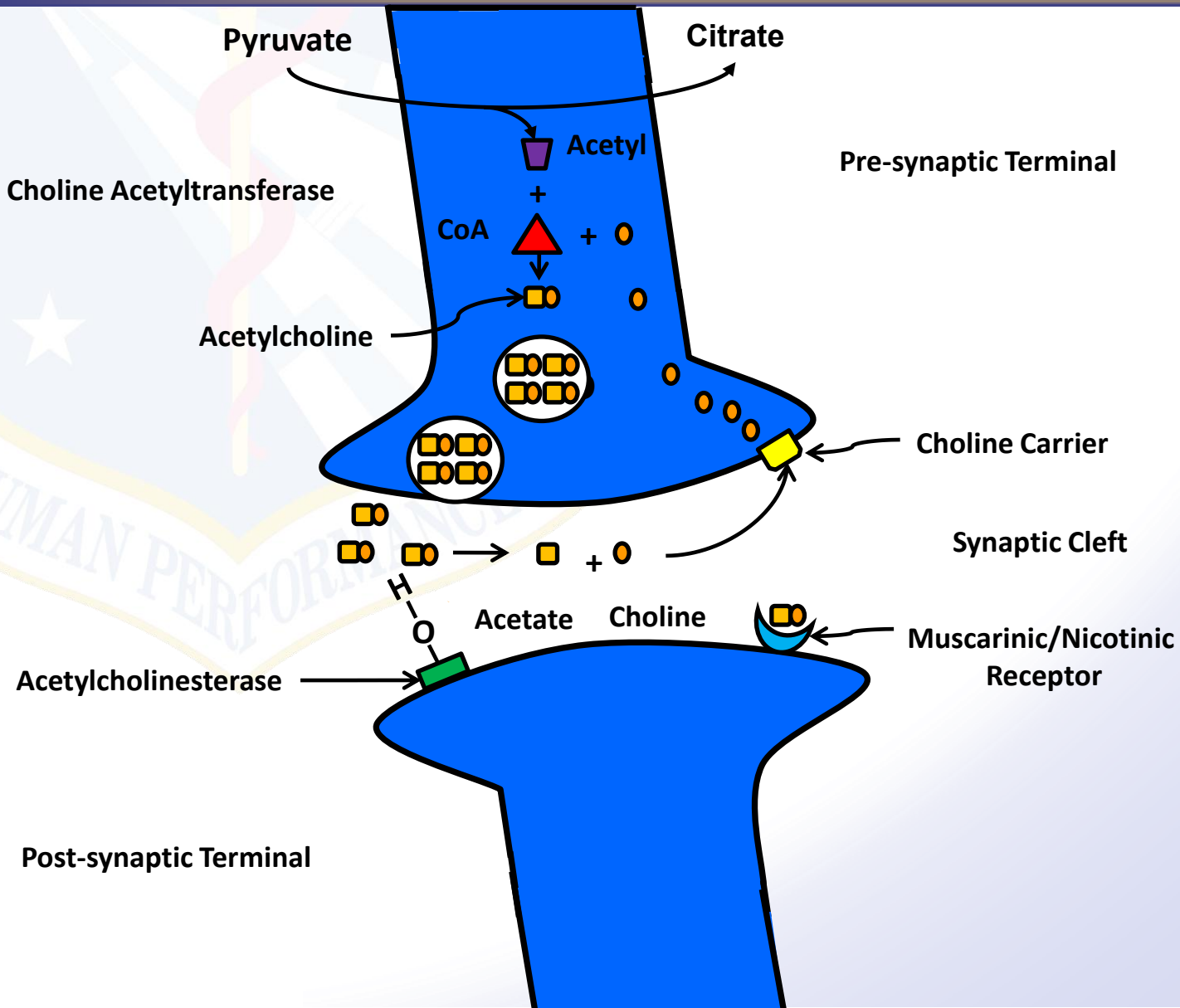
QSAR Overview





Cholinergic Nervous System

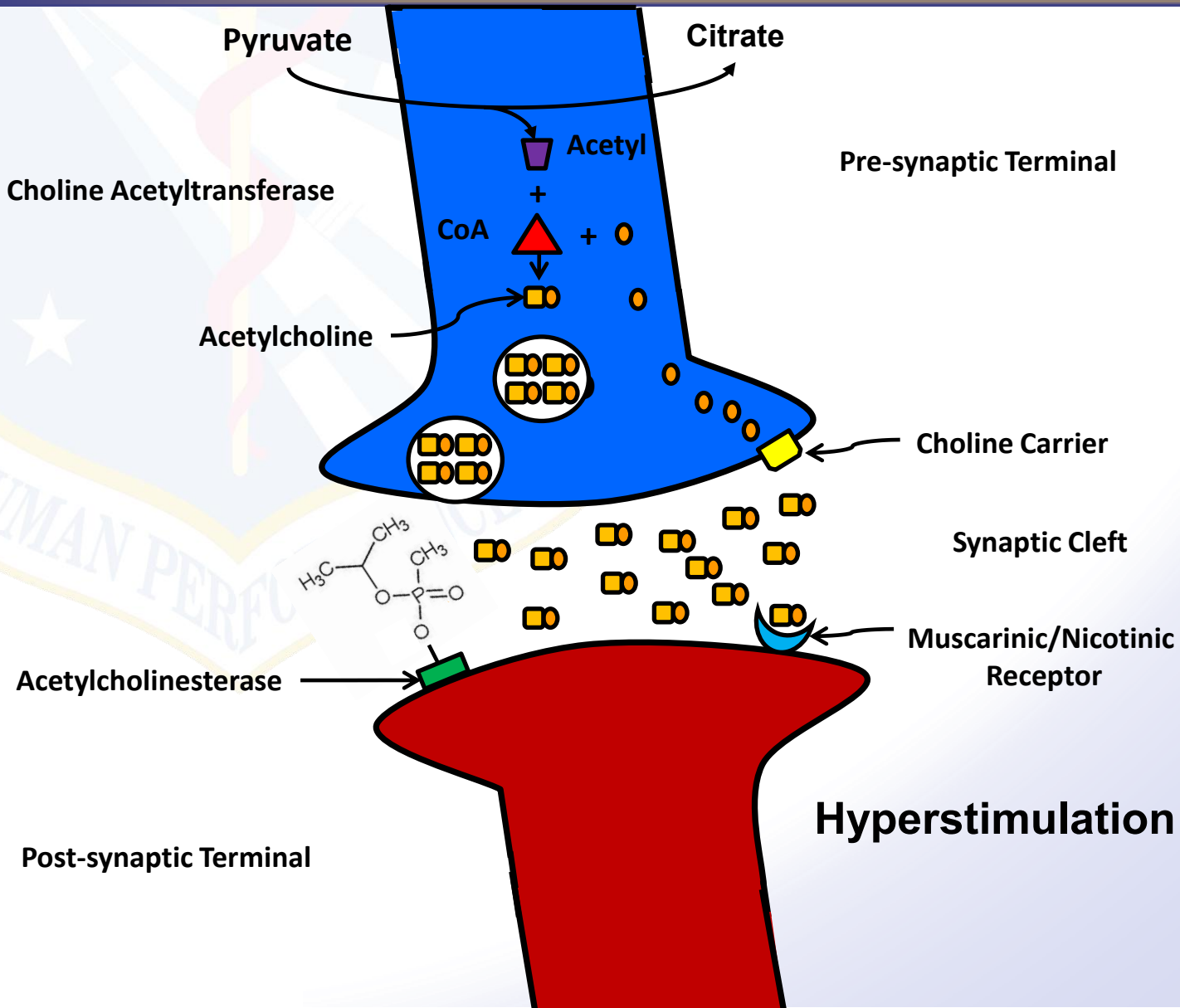
“Normal Mechanism of Action”



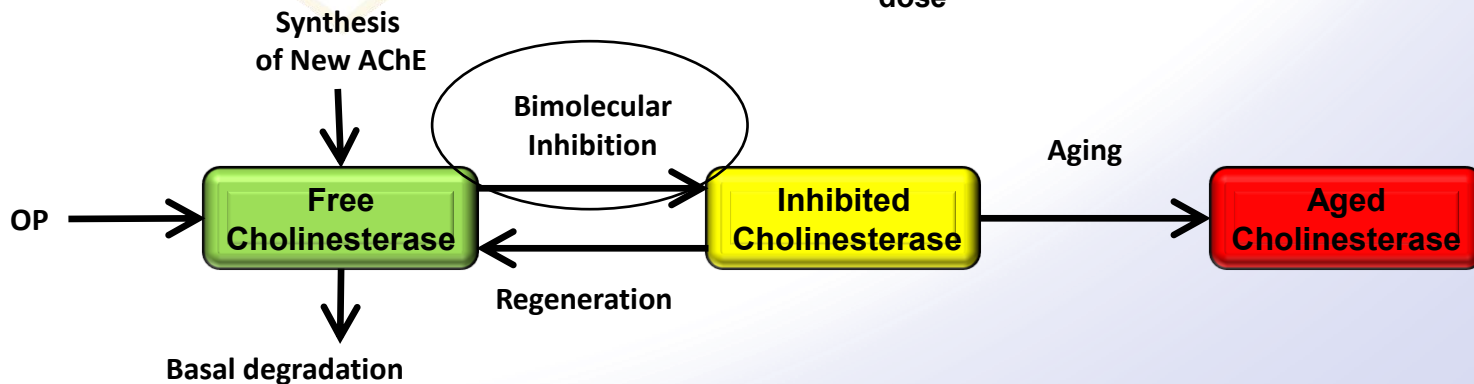
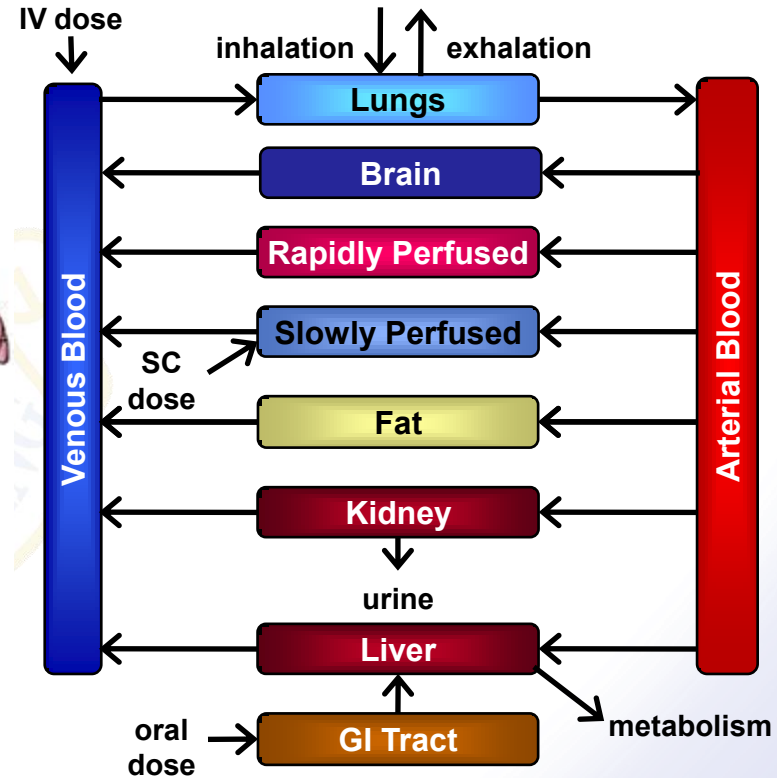
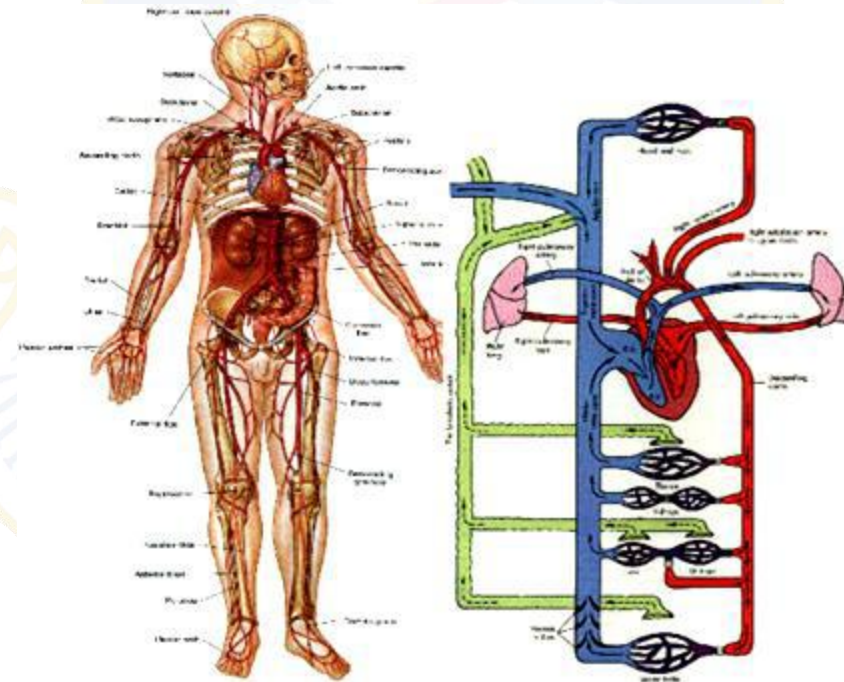


Cholinergic Nervous System

“OP Mechanism of Action”

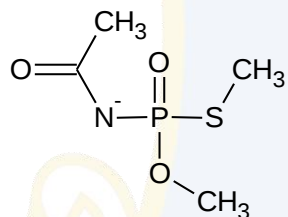


Physiologically Based Pharmacokinetic/Pharmacodynamic Modeling

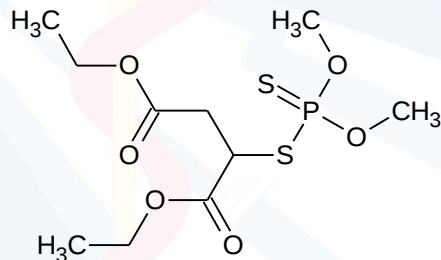




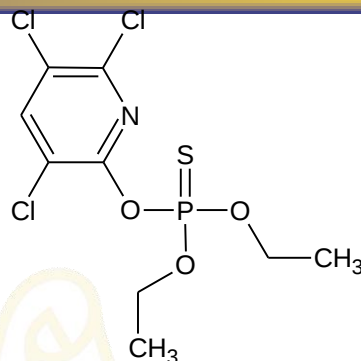
Organophosphate Structures



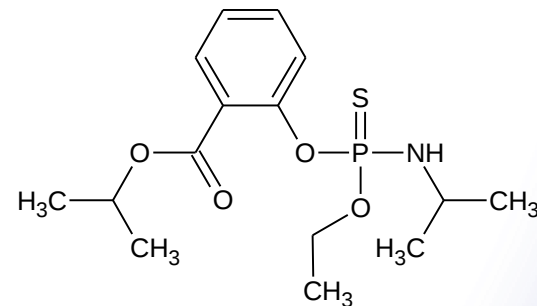
Acephate



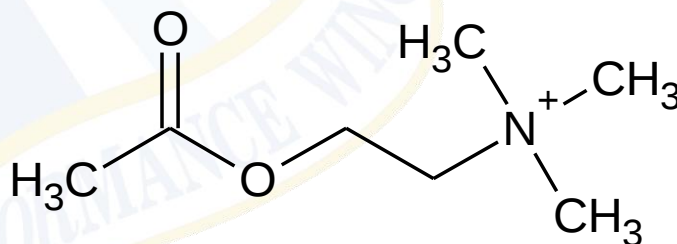
Malathion



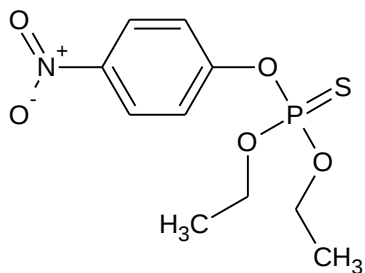
Chlorpyrifos



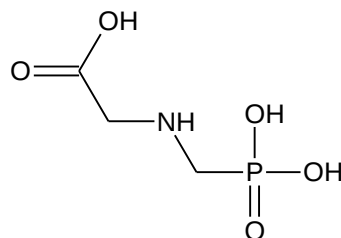
Isofenphos



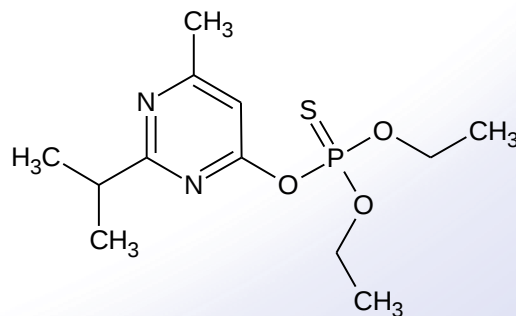
Acetylcholine



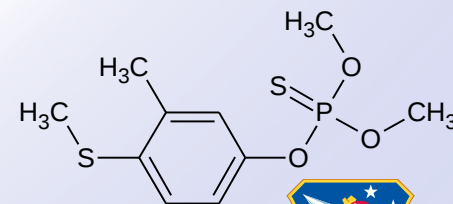
Parathion



Glyphosate



Diazinon

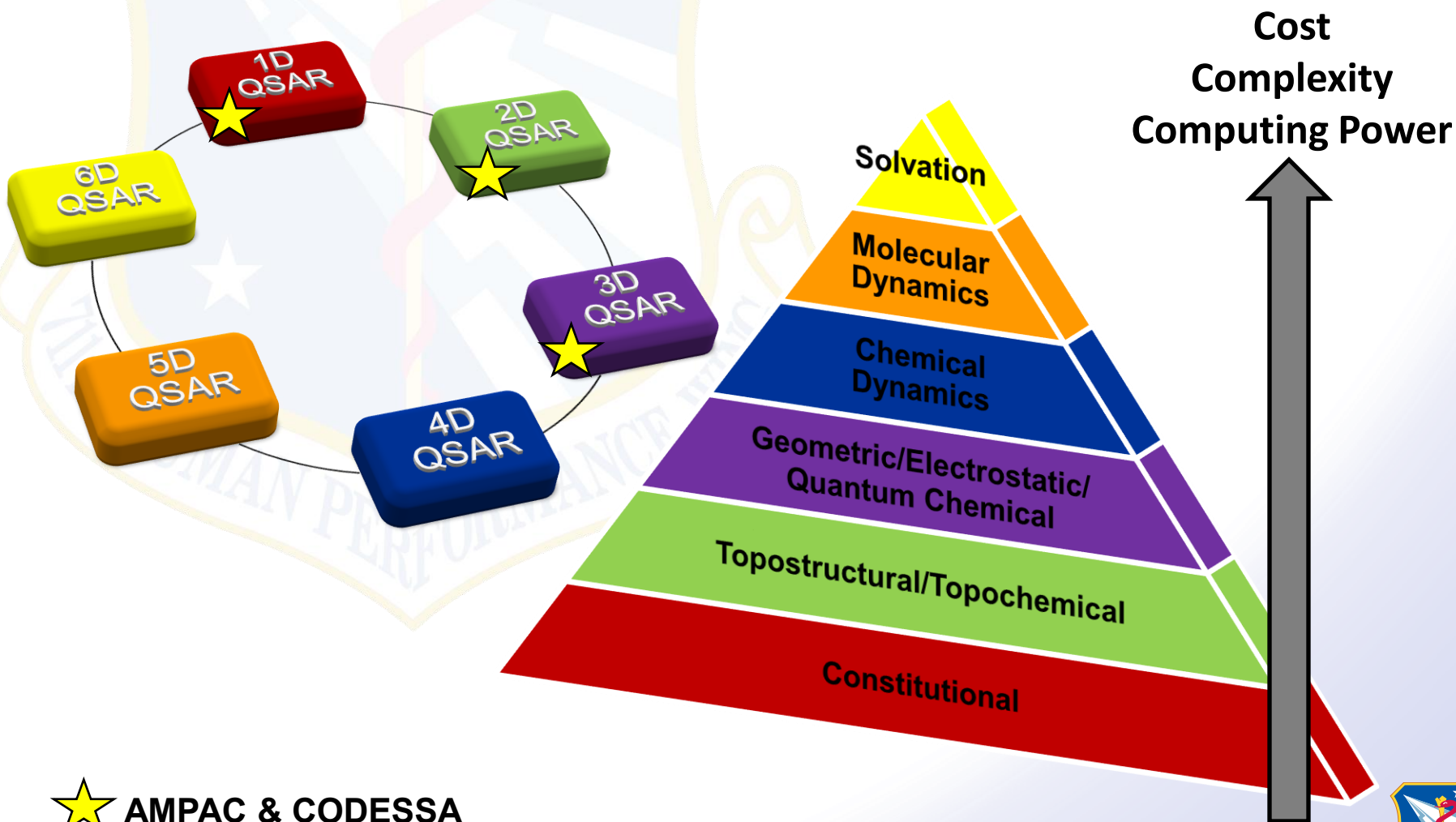


Fenthion





Physiochemical Descriptors



★ AMPAC & CODESSA





Constitutional Descriptors



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

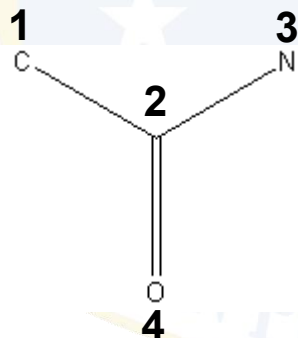
- Number of atoms
 - Absolute and relative numbers of C, H, O, S, N, F, Cl, Br, I, P atoms
- Number of bonds
 - Absolute and relative numbers of single, double, triple and aromatic bonds
- Number of rings
 - Number of rings divided by the number of atoms, number of benzene rings, number of benzene rings divided by the number of atoms
- Molecular and average atomic weight





Topostructural Descriptors

A molecular graph is made up of Edges and Vertices



(G)

Adjacency matrix (A)

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

Distance matrix (D)

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

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

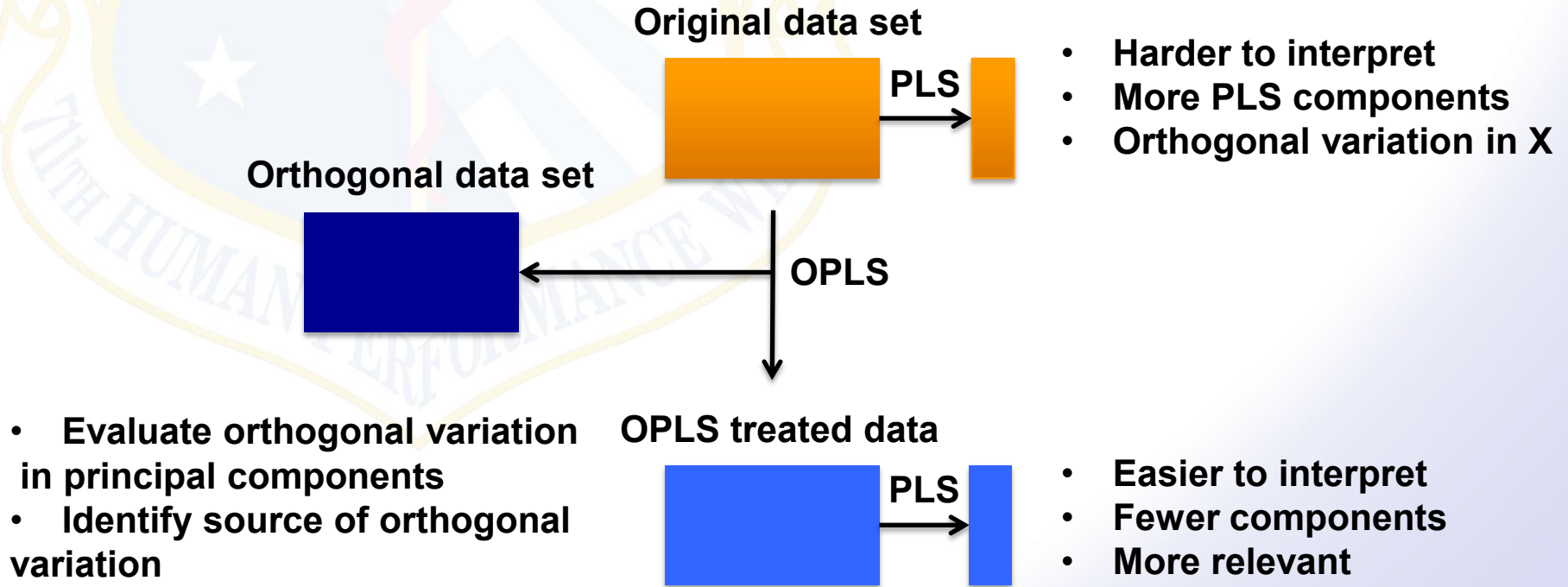
- k-nearest neighbor
- Random Forest





Regression Techniques

Overview of orthogonal projection to latent structures (O-PLS)



Acetylcholinesterase Bimolecular Rate Constants ($M^{-1}min^{-1}$)



OP Library

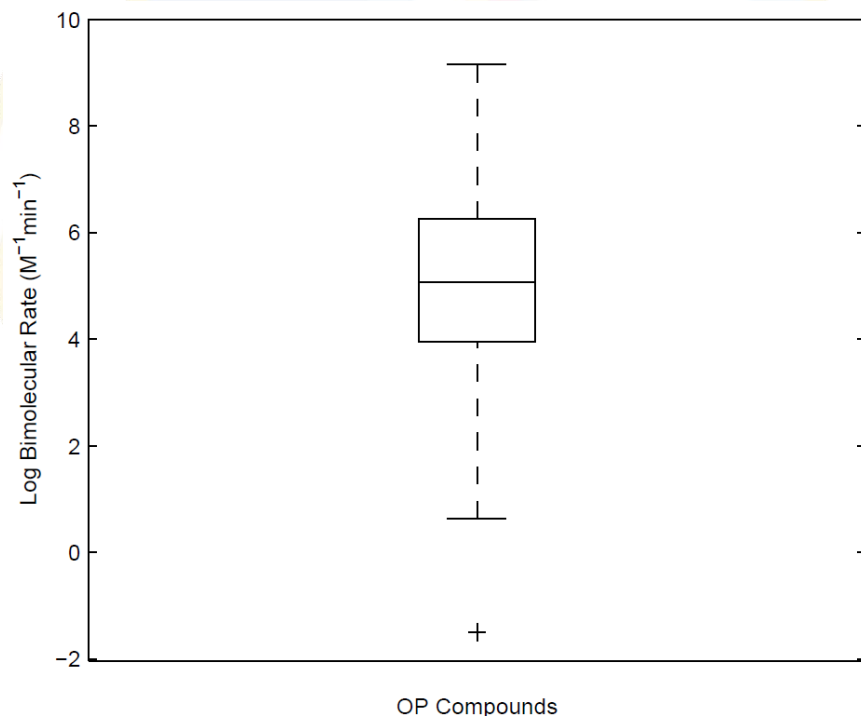


Table 1. Percentage of database that represents a particular temperature and species.

Species	Temperature ($^{\circ}C$)							Total Percent
	25	37	Unknown	5	30	22	27	
Human	28.46	6.93	8.15	0.00	0.00	0.00	0.19	43.72
Bovine	21.44	3.56	0.09	0.09	1.22	0.00	0.00	26.40
Unknown	7.77	0.28	1.50	0.00	0.00	0.00	0.00	9.55
Fly	2.25	0.56	4.87	0.00	1.03	0.00	0.00	8.71
Rat	0.00	1.22	0.47	0.66	0.00	0.47	0.00	2.81
Hen	0.56	0.47	0.00	0.00	0.00	0.47	0.00	1.50
Rabbit	0.47	0.00	0.00	0.00	1.03	0.00	0.00	1.50
Eel	0.75	0.75	0.00	0.00	0.00	0.00	0.00	1.50
Cricket	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.66
Guinea Pig	0.19	0.00	0.00	0.00	0.00	0.47	0.00	0.66
Pig	0.00	0.66	0.00	0.00	0.00	0.00	0.00	0.66
Mouse	0.19	0.00	0.00	0.00	0.09	0.28	0.00	0.56
NHP	0.00	0.00	0.00	0.00	0.00	0.47	0.00	0.47
Catfish	0.00	0.00	0.00	0.00	0.00	0.47	0.00	0.47
Frog	0.00	0.00	0.00	0.00	0.00	0.47	0.00	0.47
Minipig	0.00	0.37	0.00	0.00	0.00	0.00	0.00	0.37
Total Percent	62.73	14.79	15.07	0.75	3.37	3.09	0.19	100.00

Data collected from 69
peer-reviewed journal articles.

1068 observations

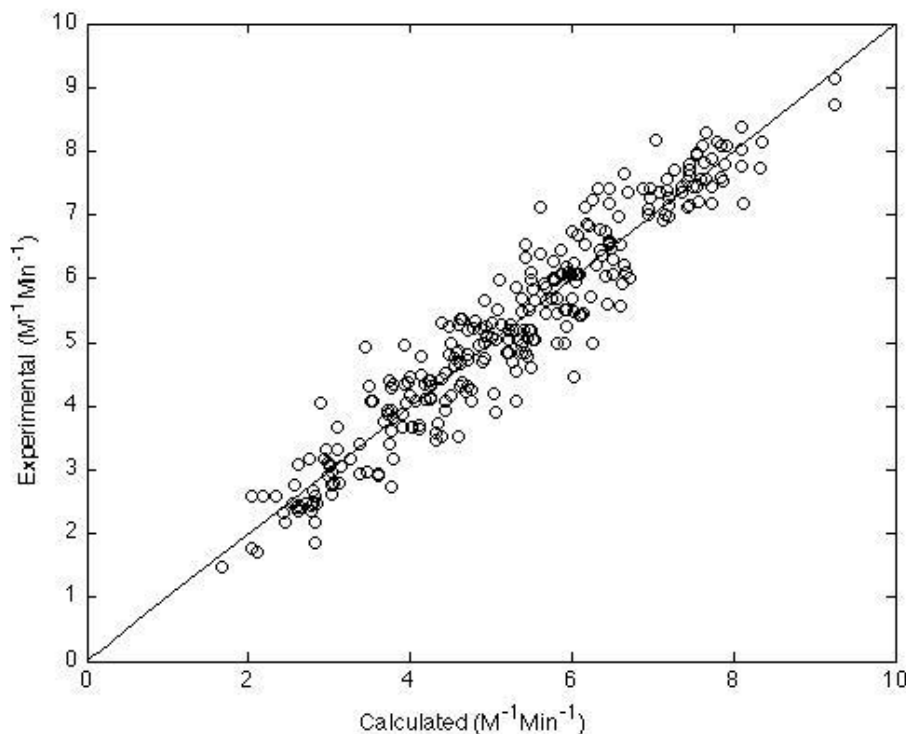




Acetylcholinesterase Bimolecular Rate Constants ($M^{-1}min^{-1}$)

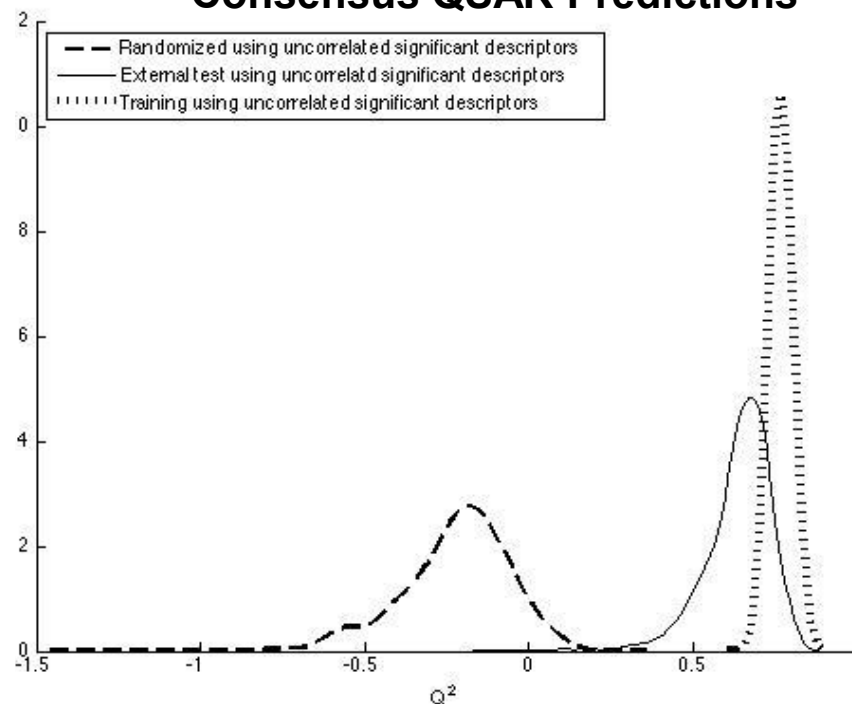


Global Human Orthogonal-PLS



Global training $R^2=0.91$ using 74 significant and uncorrelated descriptors.

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 $Q^2 = -0.23 \pm 0.18$.



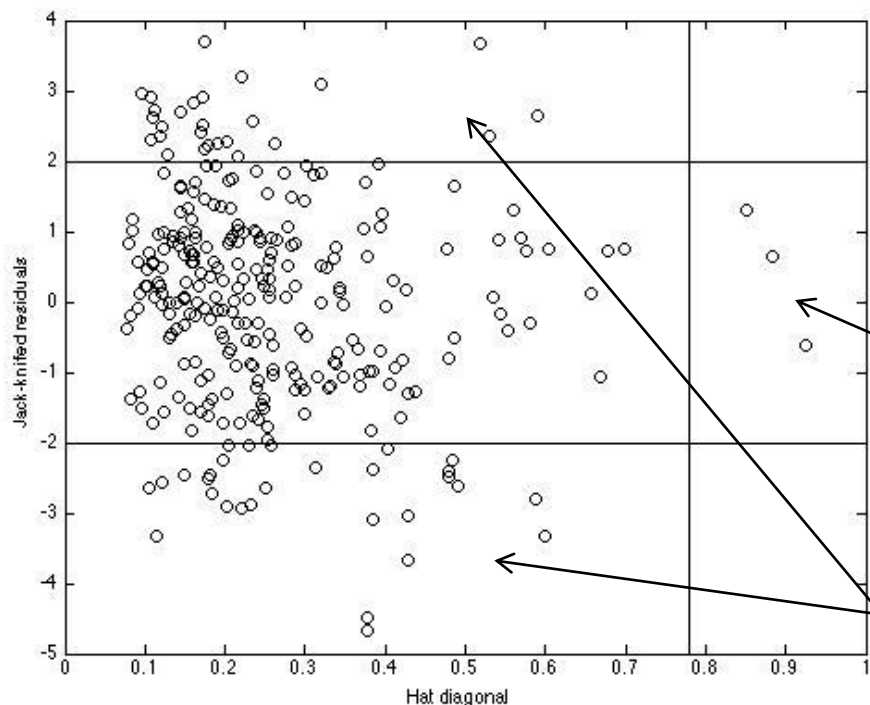


Domain of Applicability

- 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



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





Regression Techniques



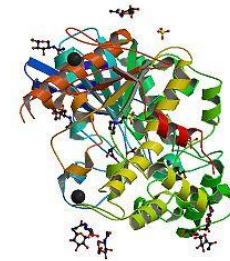
- **Heuristic (Built into CODESSA 2.51)**

- 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.

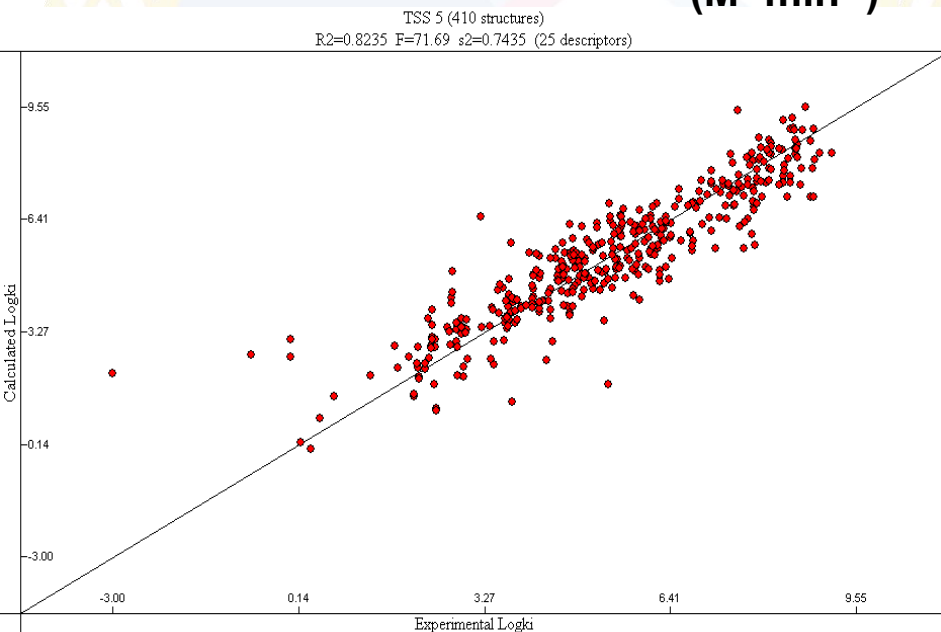




Butyrylcholinesterase “Serum Cholinesterase”



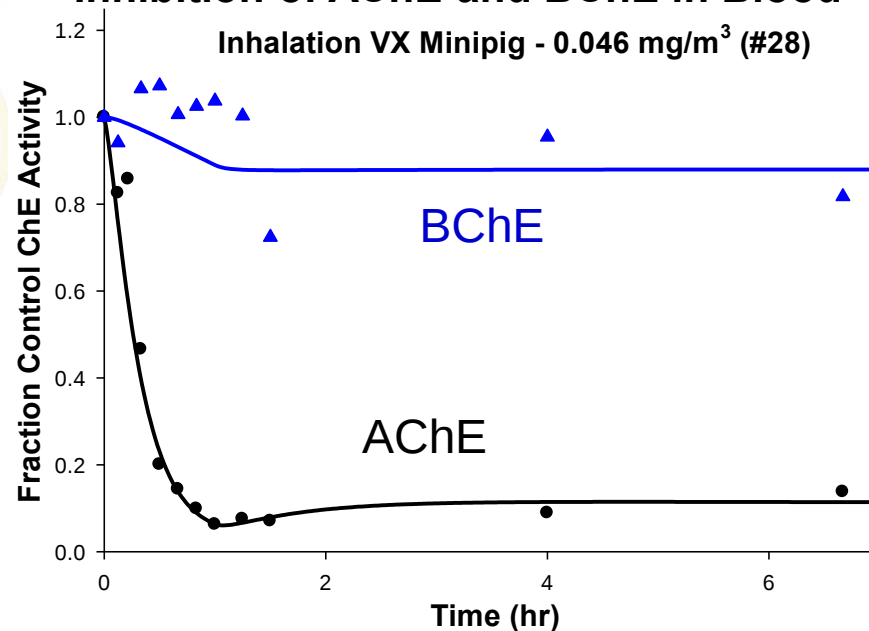
**Bimolecular Rate
($M^{-1}min^{-1}$)**



**$R^2=0.82$, 25 descriptors,
 $F=71.69$, 410 compounds**

Inhibition of AChE and BChE in Blood

Inhalation VX Minipig - 0.046 mg/m^3 (#28)

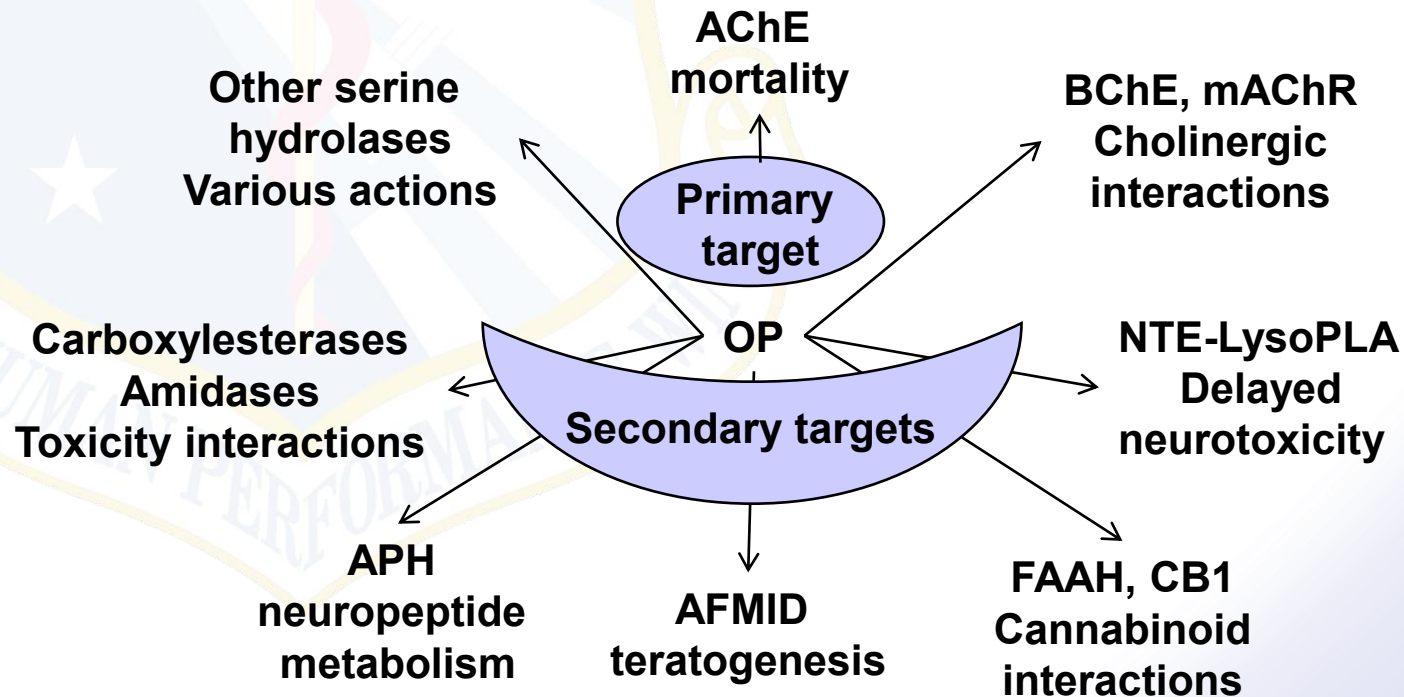


Data taken from literature.





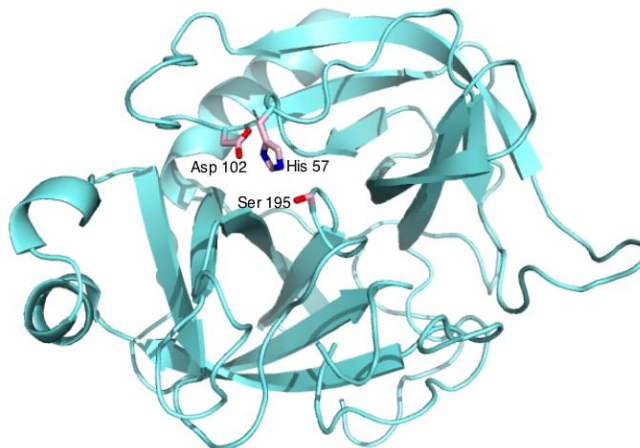
Noncholinergic Targets



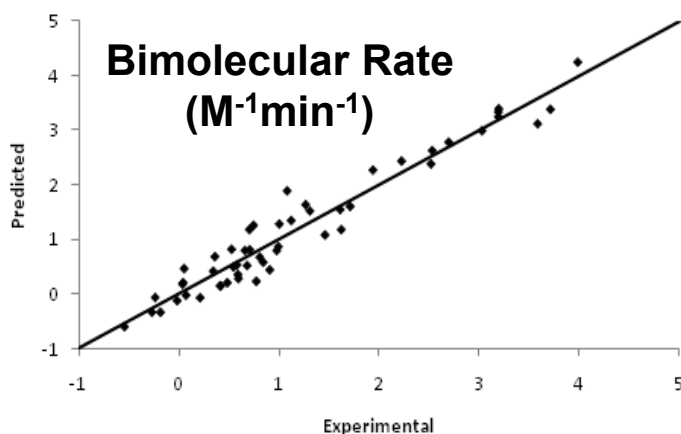


Noncholinergic Targets

Digestive Proteases

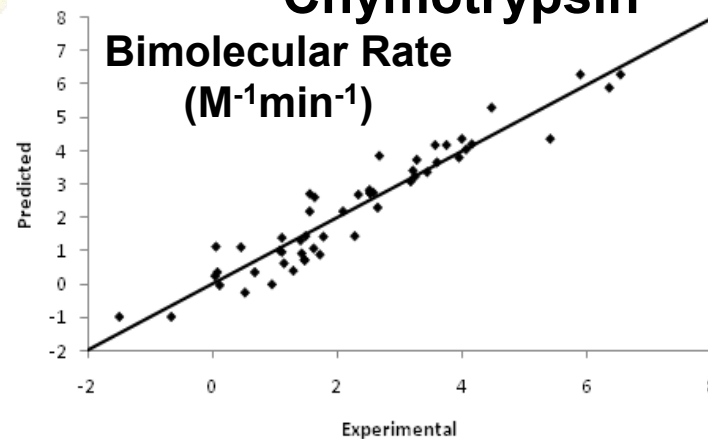


Trypsin



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

Chymotrypsin



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





External Validation

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

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.





Trypsin Descriptors



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	FPSSA-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





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.**

Cleared for public release: ABW-2011-5001, 9/20/2011

