Electronic Supplementary Material (ESI) for Environmental Science: Processes & Impacts.

Supplementary Information

Quantitative Structure-Activity Relationship

Correlation between Molecular Structure and the

Rayleigh Enantiomeric Enrichment Factor

S. Jammer[+]*,D. Rizkov*[+]*, F. Gelman*[‡*] *and O. Lev*[+*]

[+]The Casali Center of Applied Chemistry, The Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; [‡]The Geological Survey of Israel, Jerusalem 95501 , Israel.

MATERIALS AND METHODS

Materials and reagents

Lipase from *Pseudomonas cepacia*, lipase from *Candida rugos*, and Amano lipase from *Pseudomonas fluorescens* were supplied by Sigma-Aldrich Co. Ltd. with a reported activity of 37.8 U/mg, 1020 U/mg and 20,000 U/mg respectively. (*R,S*)-2-Methyl-(2,4-dichloro-phenoxy)propionate (DCPPM) (99.9%), (*R,S*)-2-Methyl-(4-chloro-2-methyl-phenoxy)propanoate (MCPPM) (99.2%), (*R,S*)-2-(4-Hydroxy-phenoxy)propionic acid (HYCPP) (97%), (*R,S*)-2,4,5-TP methyl ester (TCPPM) (99.5%), (*R,S*)-2-(2-Chloro-phenoxy)propionic acid, (CPP), (*R,S*)-2-(4-Biphenylyloxy)propionic acid (BPP), (*R,S*)-2-(4-Trifluoromethoxy-phenoxy)propionic acid (OCF_3PP), (*R,S*)-2-(2-methoxy-phenoxy)propionic acid (OCH_3PP), (*R,S*)-2-(2,4-dibromo-phenoxy)propionic acid (DBrPP), (*R,S*)-2-(4-bromo-phenoxy)propionic acid (BrPP), (*R,S*)-2-(2-flouro-phenoxy)propionic acid (FPP), (*R,S*)-2-(2,1-naphthyloxy)propionic acid (NaphPP), (*R,S*)-diethyl 2-methylsuccinate (DEMS), (*R,S*)-2-(3,5-dimethyl-phenoxy)propionic acid (DMPP) and (*R,S*)-dimethyl 2-methylsuccinate (DMMS) (98%), were supplied by Sigma-Aldrich Co. Ltd. (*R,S*)-2-(methylphenoxy)proranoate (PPM) (98%), (*R,S*)-2-(4-idophenoxy)proranoate (IPPM) (98%) and (*R,S*)-2-(4-nitrophenoxy)proranoate (NPPM) (98%) was supplied by Tzamal D-Chem Laboratories Ltd. 0.5 M potassium phosphate buffer pH 7.4 (0.2 μM filtered) was supplied by BD Gentest. All organic solvents were HPLC grade and were supplied by Bio-Lab Ltd.

Experimental Details

Preparation of esters:

CPP, HYCPP, BPP, OCH_3PP, DBrPP, DMPP, BrPP, FPP, NaphPP and OCF_3PP acids overcame the flowing methylation procedure in order to produce the corresponding esters CPPM, HYCPPM, BPPM, OCH_3PPM, DBrPPM, DMPPM, BrPPM, FPPM, NaphPPM and OCF_3PPM: The methylation reaction was carried out using 100x13mm test tubes with Teflon-lined screw caps. Five milliliter of an aqueous solution containing the desired acid (2 mg/mL) was added to 2 mL methanol and 1 mL Sulfuric acid (50%). The tubes were capped and kept in room temperature o/n. upon completion of the reaction, 2 mL of H_2O and 0.5 mL of chloroform were added and the mixture was briefly vortexed and centrifuged. The chloroform extract was dried by gentle nitrogen evaporation and re-diluted with 0.5 mL of methanol.

Stock solutions of the PPMs were prepared in methanol at 1 mg/mL and stored in 4°C before use. Working standard solutions were prepared by diluting the stock solutions with the proper reaction solvent.

Enzyme-catalyzed hydrolysis reactions

The enzymatic degradation reactions were carried out in 20 mL vials. Five milliliter of 0.1M potassium phosphate buffer (pH 7.4) was placed in a vial and then 0.2 mL substrate stock solutions (1 mg/mL) and 227U lipase enzymes were added. At time intervals each reaction mix was extracted by liquid-liquid extraction with dichloromethane (four 2 mL washes) while 0.15 mL DMMS was added as extraction internal standard and followed by drying over anhydrous Na_2SO_4. The extraction solutions were diluted to 10 mL and samples were taken for chiral analysis, adding DEMS as an injection internal standard. These enzymatic hydrolysis reactions were done in duplicates.

Analytical methods

Enantiomer enrichment analysis:

PPM, CPPM, DCPPM , MCPPM, DBrPPM, OCH_3PPM, DMPPM, OCF_3PPM, IPPM, BrPPM, and NaphPPM (**1-4**,**8-11 and 13-16** in Table S1) enantiomers, were separated by GC-SMB-QQQ-MS with the following methods: PPM, DCPPM , IPPM and MCPPM: Injection volume was 2 µL in splitless mode, flow rate 2 mL/min for 16 min and oven program: 60 °C, then 4 °C/min to 140 °C, then 40 °C/min to 230 °C for 4 min, total run time was 26.25 min ; CPPM: Injection volume was 2 µL in splitless mode, flow rate 1.5 mL/min for 16 min and oven program: 60 °C, then 4 °C/min to 140 °C, then 25 °C/min to 230 °C for 0.5 min, total run time was 24.1 min; DBrPPM and BrPPM: Injection volume was 2 µL in splitless mode, flow rate 3.5 mL/min for 16 min and oven program: 40 °C, then 4 °C/min to 210 °C, total run time was 42.5 min; OCH_3PPM: Injection volume was 2 µL in splitless mode, flow rate 3.5 mL/min for 16 min and oven program: 30 °C, then 3 °C/min to 160 °C, total run time was 43.3 min; DMPPM: Injection volume was 2 µL in splitless mode, flow rate 3.5 mL/min for 16 min and oven program: 30 °C, then 2.5 °C/min to 153 °C, total run time was 49.2 min; OCF_3PPM: Injection volume was 2 µL in splitless mode, flow rate 2 mL/min for 16 min and oven program: 60 °C, then 4 °C/min to 140 °C, then 40 °C/min to 220 °C for 5 min, total run time was 27 min. FPPM: Injection volume was 2 µL in splitless mode, flow rate 3.5 mL/min for 16 min and oven program: 40 °C, then 4°C/min to 150 °C, total run time was 27.5 min. NaphPPM: Injection volume was 2 µL in splitless mode, flow rate 3.5 mL/min for 16 min and oven program: 30 °C,

then 3 °C/min to 200 °C, total run time was 54 min.

TCPPM, HYPPM BPPM and NPPM (**5-7 and 12** in Table S1) enantiomers were separated by HPLC at ambient temperature, isocratically: TCPPM was analyzed with n-hexane/isopropanol (99: 1) at a flow rate of 0.5 mL/min and detected at 230 nm; while HYPPM and NPPM were analyzed with n-hexane/isopropanol (95: 5) at a flow rate of 1.5 mL/min and detected at 225 nm and 290nm, respectively. BPPM was analyzed with n-hexane/isopropanol (98: 2) at a flow rate of 0.5 mL/min and detected at 258 nm.

The enantiomeric ratios were expressed as ER – the peak area of one enantiomer (the more abundant) divided by that of the other enantiomer.

Summary of the definitions of the fundamental equations (equations 1, 9,10) (see reference 1) :

$$A_1 + A_2 = C \quad \text{(S1)}$$

A_1 and A_2 are the concentrations of two enantiomers.

$$\ln \frac{A_{i,t}}{A_{i,0}} = - k_i \times t \quad \text{(S2)}$$

A_i is A_1,A_2 or C and the subscript "$_0$"denotes initial conditions. As a convention we assign the more reactive enantiomer as A2, i.e.

$$k_1 < k_2 \quad \text{(S3)}$$

k_1 and k_2 are the individual first order rate constants of each enantiomer..

$$ln\frac{A_{1,t}/A_{2,t}}{A_{1,0}/A_{2,0}} = -(k_1 - k_2)t = \bar{k} \times t \quad \text{(S4)}$$

$$ER = A_1/A_2 \quad \text{(S5)}$$

$$ln\frac{ER_t}{ER_0} = \bar{k} \times t \quad \text{(S6)}$$

Expressing t as a function of C or $A_{1,t}/A_{1,0}$ or $A_{2,t}/A_{2,0}$ by eq S2 and introducing it to eq S6 leads to equations S7,S8 and S9:

$$ln\frac{ER_t}{ER_0} = \varepsilon_{ER} \times ln\frac{C_t}{C_0} = -\frac{\bar{k}}{k_c} \times ln\frac{C_t}{C_0} \quad \text{(S7)}$$

$$ln\frac{ER_t}{ER_0} = \varepsilon_1 \times ln\frac{A_{1,t}}{A_{1,0}} = \varepsilon_1 \times ln\frac{C_t\,(1 + {}^{1}/_{ER_0})}{C_0(1 + {}^{1}/_{ER_t})} = -\frac{\bar{k}}{k_1} \times ln\frac{A_{1,t}}{A_{1,0}} \quad \text{(S8)}$$

$$ln\frac{ER_t}{ER_0} = \varepsilon_2 \times ln\frac{A_{2,t}}{A_{2,0}} = \varepsilon_2 \times ln\frac{C_t(1 + ER_0)}{C_0(1 + ER_t)} = -\frac{\bar{k}}{k_2} \times ln\frac{A_2,t}{A_{2.0}} \quad \text{(S9)}$$

Thus, there are three ways to define the Rayleigh enrichment factors and they can be easily calculated based on one of them if the first order kinetics is known. Table S1 depicts all the three sets of kinetic coefficients and Table S2 shows that each of the enrichment factors can be predicted by QSAR modeling, therefore it is not important which of the three conventions to define the enrichment factor is followed.

Table S1. Kinetic data for the PPMs degraded by the different lipase enzymes (PCL,PFL,CRL):

no.	analytes	k_1 PCL (hr^{-1})	k_2 PCL (hr^{-1})	k_c PCL (hr^{-1})	k_1 PFL (hr^{-1})	k_2 PFL (hr^{-1})	k_c PFL (hr^{-1})	k_1 CRL (hr^{-1})	k_2 CRL (hr^{-1})	k_c CRL (hr^{-1})
1	PPM	0.006	0.014	0.009	0.009	0.017	0.013	0.281	0.642	0.389
2	CPPM	0.020	0.036	0.020	0.022	0.036	0.042	0.250	0.840	0.283
3	DCPPM	0.077	0.162	0.095	0.088	0.205	0.092	0.822	2.507	1.002
4	MCPPM	0.038	0.058	0.043	0.030	0.051	0.036	0.599	1.922	0.664
5	TCPPM	0.106	0.110	0.136	0.052	0.173	0.159	0.080	0.383	0.470
6	HYPPM	0.007	0.014	0.008	0.009	0.012	0.002	0.007	0.208	0.084
7	BPPM	0.056	0.072	0.062	0.022	0.038	0.027	0.230	0.258	0.243
8	DBrPPM	0.022	0.042	0.169	0.018	0.053	0.196	0.007	0.118	0.059
9	OCH_3PPM	0.006	0.013	0.005	0.008	0.013	0.008	0.053	0.163	0.146
10	DMPPM	0.015	0.027	0.018	0.011	0.016	0.058	0.421	1.322	0.681
11	OCF_3PPM	0.012	0.027	0.082	0.013	0.051	0.097	0.636	0.942	0.761
12	NPPM	0.027	0.044	0.074	0.008	0.013	0.039	0.226	1.071	0.435
13	IPPM	0.057	0.04	0.038	0.07	0.06	0.056	0.973	0.875	0.83
14	BrPPM	0.042	0.072	0.076	0.077	0.119	0.052	0.896	1.372	1.077
15	FPPM	0.035	0.04	0.023	0.088	0.102	0.033	0.419	0.509	0.611
16	NaphPPM	0.020	0.032	0.016	0.058	0.069	0.049	0.897	1.372	1.077

k_1 - Observed linear rate coefficients of enantiomer A_1; k_2 -Observed linear rate coefficients of enantiomer A_2; k_c –Observed overall linear rate coefficients of both enantiomers.

Table S2. QSAR data for the PPMs degraded lipase from *Pseudomonas cepacia* (PCL)

No.	Analytes	Observed log (ε_{ER}) (%)	Predicted log (ε_{ER}) (%)	Residuals	Leverage value, h
1	PPM	1.36	1.47	-0.10	0.02

2	CPPM	1.88	1.69	0.18	0.01
3	DCPPM	1.95	1.92	0.03	0.08
4	MCPPM	1.66	1.52	0.14	0.01
5	TCPPM	2.37	2.28	0.09	0.40
6	HYPPM	1.18	1.11	0.07	0.23
7	BPPM	1.43	1.46	-0.03	0.02
8	DBrPPM	1.70	1.92	-0.22	0.08
9	OCH_3PPM	1.30	1.20	0.10	0.11
10	DMPPM	1.27	1.23	0.04	0.13
11	OCF_3PPM	1.82	1.81	0.01	0.02
12	NPPM	2.19	2.23	-0.03	0.21
13	IPPM	1.67	1.64	0.02	0.00
14	BrPPM	1.61	1.69	-0.08	0.00
15	FPPM	1.48	1.52	-0.04	0.01
16	NaphPPM	1.45	1.51	-0.06	0.01

Table S3. QSAR data for the PPMs degraded lipase from *Pseudomonas fluorescens* (PFL)

No	Analytes	Observed log (ε_{ER}) (%)	Predicted log (ε_{ER}) (%)	Residuals	Leverage value, h
1	PPM	1.61	1.63	-0.02	0.02
2	CPPM	1.92	1.83	0.09	0.01
3	DCPPM	2.05	2.03	0.02	0.08
4	MCPPM	1.83	1.68	0.15	0.01
5	TCPPM	2.42	2.35	0.07	0.40
6	HYPPM	1.28	1.31	-0.03	0.23
7	BPPM	1.79	1.62	0.17	0.02
8	DBrPPM	1.8	2.03	-0.23	0.08
9	OCH_3PPM	1.35	1.39	-0.04	0.11
10	DMPPM	1.37	1.42	-0.05	0.13
11	OCF_3PPM	2.01	1.93	0.08	0.02
12	NPPM	2.17	2.31	-0.14	0.21
13	IPPM	1.78	1.79	-0.01	0.00
14	BrPPM	1.86	1.83	0.03	0.00
15	FPPM	1.68	1.68	0.00	0.01
16	NaphPPM	1.64	1.66	-0.02	0.01

Table S4. QSAR data for the PPMs degraded lipase from *Candida rugosa* (CRL)

No	Analytes	Observed log (ε_{ER}) (%)	Predicted log (ε_{ER}) (%)	Residuals	Leverage value, h
1	PPM	1.32	1.32	0.00	0.49
2	CPPM	1.85	1.76	0.09	0.08
3	DCPPM	2.23	2.22	0.01	0.05
4	MCPPM	2.29	2.18	0.11	0.09
5	TCPPM	2.43	2.58	-0.15	0.29
6	HYPPM	2.20	2.11	0.09	0.42
7	BPPM	1.88	1.76	0.12	0.40
8	DBrPPM	2.45	2.41	0.04	0.21
9	OCH_3PPM	2.10	2.17	-0.08	0.31
10	DMPPM	2.11	2.21	-0.09	0.06
11	OCF_3PPM	2.43	2.46	-0.04	0.18
12	NPPM	2.48	2.68	-0.20	0.06
13	IPPM	1.93	1.88	0.06	0.09
14	BrPPM	1.65	1.83	-0.18	0.07
15	FPPM	1.80	1.84	-0.04	0.04

Model development and Statistical validation

Selection of training and validation sets

In order to perform an appropriate validation of the model we divided the analyzed compounds into training and external validation sets. The selection of the external validation set was based on the K-Means Clustering Technique[2] -a non-hierarchical procedure where the number of cluster, k, is fixed (in our case k=5). Using the SPSS 8.0 software, five clusters were derived based on the descriptors space (σ for PCL and PFL; π, Es^k for CRL), the divisions of the compounds according to the clusters are presented in Tables S5 and S6.

Table S5. K-means clustering using the descriptor space of σ .

Cluster no.	Analytes in cluster (table S1)	Number of compounds in cluster	Number of compounds in validation set
1	6,9,10	3	1

2	2,13,14	3	1
3	3,8,11	3	1
4	1,4,15,16,7	5	1
5	5,12	2	1

Table S6. K-means clustering using the descriptor space of π and Es^k.

Cluster no.	Analytes in cluster (table S1)	Number of compounds in cluster	Number of compounds in validation set
1	7,14	2	1
2	12,10	2	1
3	2,3,4,11,13	5	1
4	1,6,9,15	4	1
5	5,8	2	1

Validation of the QSAR model

The calculated determination coefficient R^2 and the Root Mean Square Error of Calibration (RMSEC) were used to measure the goodness-of-fit of the model (eq S10[3])

$$RMSEC = \sqrt{\sum (Y_{Tact} - Y_{Tpred})^2 / n_T} \quad \text{(S10)}$$

Y_{Tact} and Y_{Tprd} are the actual (measured) and predicted values of $\log(\varepsilon_{ER})$ in the training set, respectively; n_T is the number of compounds in the training set.

Internal validation

For internally validating the QSAR model we used the leave-one-out cross-validation method (LOO-CV) that is used mainly in small datasets, to determine how large a model can be used for a given data set and as a diagnostic tool to evaluate the predictive power of an equation.[3] The LOO-CV process repeats the regression many times on subsets of data; each molecule was left

out once, in turn, and the regression was computed using the predicted values of the missing molecules. The cross-validated correlation coefficient Q^2_{CV} and the cross-validated root mean square error of prediction (RMSECV) were used to reduce the probability of the model's over-fitting and to measure the robustness of the model. (eqs S11, S12[3])

$$Q^2_{cv} = 1 - \sum (Y_{act} - Y_{predcv})^2 / \sum (Y_{act} - Y_{mean})^2 \quad \text{(S11)}$$

$$RMSECV = \sqrt{\sum (Y_{Tact} - Y_{Tpredcv})^2 / n_T} \quad \text{(S12)}$$

Where Y_{predcv}, Y_{act} and Y_{mean} are predicted, actual and mean values of log (ε_{ER}) in the training set, respectively; n_T is the number of compounds in the training set. In an appropriate QSAR model $Q^2>0.5$.[3]

External validation

To confirm the predictive ability of the models, we carried out the external validation by applying the models to the validation set i.e. determining the predicted log (ε_{ER}) by the model equations obtained for the training set (eqs 6-8) . The externally validated determination coefficient Q^2_{Ext} and the root mean square error of prediction RMSEP were used to measures the predictive ability. (eqs S13,S14[3])

$$Q^2_{Ext} = 1 - \sum (Y_{Vact} - Y_{Vpred})^2 / \sum (Y_{Vact} - Y_{Vmean})^2 \quad \text{(S13)}$$

$$RMSECP = \sqrt{\sum (Y_{Vact} - Y_{Vpred})^2 / n_V} \quad \text{(S14)}$$

Where $Y_{V\text{pred}}$, $Y_{V\text{act}}$ and $Y_{V\text{mean}}$ are predicted, actual and mean values of log(ε_{ER}) in the validation set, respectively; n_V is the number of compounds in the validation set. In an appropriate QSAR model $Q^2>0.5$.[3]

Avoiding chance correlation

To avoid the correlation by chance and to confirm significance of the QSAR model, we performed Y-scrambling of the descriptors.[4]

We built about 100 random 'models' by using the same descriptors (σ for PCL and PFL; π, Es^k for CRL) and correlated it with the log(ε_{ER}) data that were randomly shuffled every time. By plotting the RMSEC vs. RMSECV values for these 'random' models we were able to determine the level of noise – the minimal error that can be calculated without presence of any model.

Since the values of both RMSEC and RMSECV for our 'true' QSAR model were more than two times lower than those for the randomly obtained models (Figure S1). This clearly confirms that the model has not been obtained by a chance correlation.

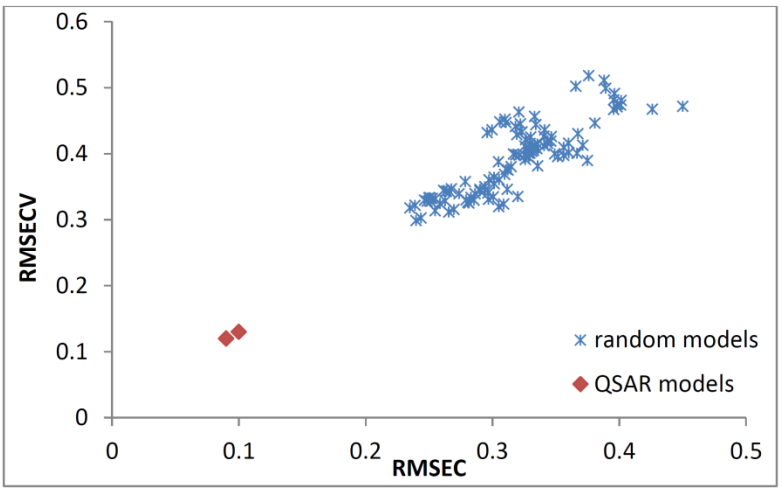

Figure S1. Y-scrambling test.

Evaluation of the optimum prediction space (Applicability Domain) of the model

In order to define the optimum prediction space we employed the leverage approach and Williams plot to visualize the results[5]. When the leverage value of a compound is greater than

the critical h^* value (eq S15) it indicates that the structure of the compound substantially differs from those used for the calibration thus the compound is located outside the optimum prediction space[6].

$$h^* = 3(P + 1)/n \qquad \text{(S15)}$$

Where p is the number of variables used in the model and n is the number of compounds.

The applicability domain (AD) of the 3 models was defined by using the SPPS 8.0 software that provided the standardized residuals and leverage values. The AD is limited by the critical h^* values of 0.54 for PCL and PFL and 0.9 for CRL (eq S15). By plotting the standardized residuals vs. leverage values (Williams plot) it is clearly visualized that neither of the training or validation compounds from our study did exceed the critical h^* value (Figure S2). Thus, the model can be applied for predicting enantiomeric enrichment factors of any other PPM's, if their structures are not substantially different from the training set (i.e. their calculated leverage values are not higher than the above critical values of h^*).

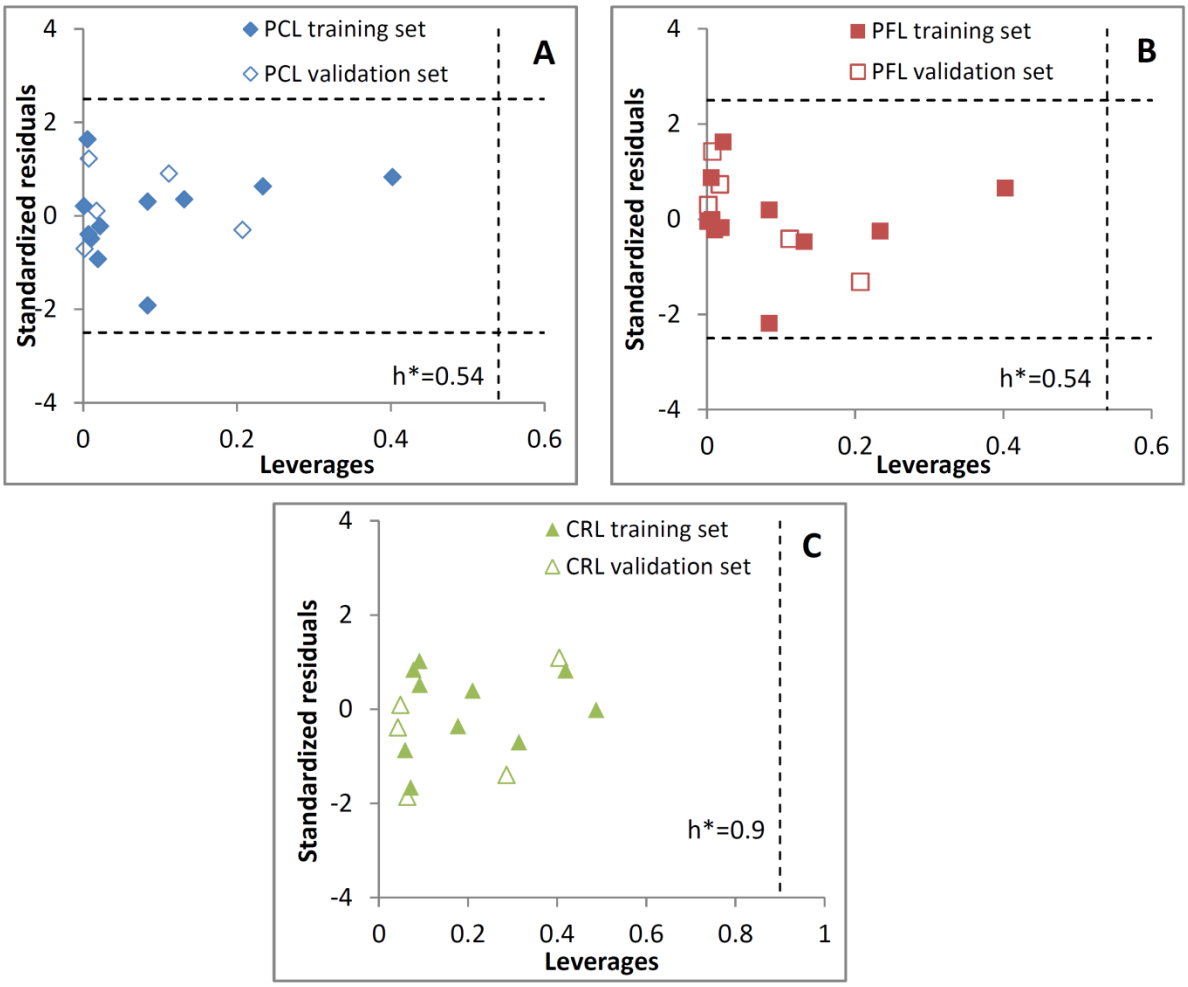

Figure S2. The Williams plot- visualization of the applicability domain for: (A) PCL training and validation set; (B) PFL training and validation set and (C) CRL training and validation set.

Table S7. Equations and statistical results of the MLR for the three different enzymes; Equations S24-S30 are insignificant and have no meaning.

eq. no (enzyme)		**R^2**
S16 (PCL)	$log(\varepsilon_1) = 1.34(\pm 0.18)\ \sigma 1.57(\pm 0.07)$	0.90
S17 (PFL)	$log(\varepsilon_1) = 1.53(\pm 0.29)\ \sigma + 1.7(\pm 0.11)$	0.89
S18 (CRL)	$log(\varepsilon_1) = -\ 0.40(\pm 0.14)\ \pi + 1.30(\pm 0.27) Es^k + 5.42(\pm 0.672)$	0.92
S19 (PCL)	$log(\varepsilon_2) = 1.43(\pm 0.15)\ \sigma + 1.33(\pm 0.05)$	0.90

S20 (PFL)	$log(\varepsilon_2) = 1.27(\pm 0.1125)\,\sigma + 1.50(\pm 0.09)$	0.90
S21 (CRL)	$log(\varepsilon_2) = -0.12(\pm 0.06)\,\pi - 0.77(\pm 0.12)Es^k - 3.57(\pm 0.30)$	0.91
S22 (PCL)	$\log\left(1/k_c\right) = -0.21(\pm 0.07)\pi - 0.73(\pm 0.18)\sigma + 0.3(\pm 0.16)Es^k + 1.52(\pm 0.41)$	0.88
S23 (PFL)	$\log\left(1/k_c\right) = -0.27(\pm 0.1)\pi - 0.75(\pm 0.28)\sigma + 0.28(\pm 0.15)Es^k + 1.82(\pm 0.86)$	0.89
S24(CRL)	$\log\left(1/k_c\right) = -0.13(\pm 0.15)\pi - 0.4(\pm 0.4)\sigma + 0.3(\pm 0.35)Es^k + 1.24(\pm 0.86)$	0.30
S25(PCL)	$\log\left(1/k_1\right) = -0.27(\pm 0.1)\pi - 0.5(\pm 0.26)\sigma + 0.15(\pm 0.23)Es^k + 2.3(\pm 0.6)$	0.63
S26 (PFL)	$\log\left(1/k_1\right) = -0.24(\pm 0.14)\pi - 0.5(\pm 0.36)\sigma + 0.48(\pm 0.32)Es^k + 3(\pm 0.8)$	0.31
S27 (CRL)	$\log\left(1/k_1\right) = -0.35(\pm 0.26)\pi - 0.62(\pm 0.68)\sigma + 0.88(\pm 0.61)Es^k + 3(\pm 1.5)$	0.2
S28 (PCL)	$\log\left(1/k_2\right) = -0.2(\pm 0.07)\pi - 0.43(\pm 0.2)\sigma + 0.09(\pm 0.18)Es^k + 1.86(\pm 0.44)$	0.67
S29 (PFCL)	$\log\left(1/k_2\right) = -0.25(\pm 0.12)\pi - 0.62(\pm 0.33)\sigma + 0.34(\pm 0.29)Es^k + 2.44(\pm 0.72)$	0.46
S30 (CRL)	$\log\left(1/k_2\right) = -0.07(\pm 0.16)\pi - 0.36(\pm 0.42)\sigma + 0.22(\pm 0.38)Es^k + 0.8(\pm 0.92)$	0.07

References:

1. S. Jammer, A. Voloshenko, F. Gelman and O. Lev, *Environ Sci Technol*, 2014, **48**, 3310-3318.
2. J. T. Leonard and K. Roy, *Qsar Comb Sci*, 2006, **25**, 235-251.
3. R. R. Veerasamy, H.; Jain, A.; Sivadasan, S.; Varghese, C.P;. Agrawal, R.K., *International Journal of Drug Design and Discovery*, 2011, **2**, 511-519.
4. OECD, in *Series on Testing and Assessment* 2007, vol. 69.
5. A. Tropsha, P. Gramatica and V. K. Gombar, *Qsar Comb Sci*, 2003, **22**, 69-77.
6. P. Gramatica, *Qsar Comb Sci*, 2007, **26**, 694-701.