Environmental Science
Processes & Impacts

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

rsc.li/process-impacts

Environmental impact statement

Many of anthropogenic chemicals of environmental concern are chiral compounds. Tracing their alteration in the environment is of high importance for the assessment of enantiomer-specific environmental toxicity. The enantiomeric enrichment is often used to identify the sources and effects of microbial degradation on chiral chemicals in the environment. Recently, it was demonstrated that the Rayleigh equation is valid to describe the enantioselective behavior and the enantiomeric enrichment factor (ε_{ER}) can be used as an identifying tool for a specific enzymatic reaction. Application of Rayleigh equation for assessing the transformations of chiral compounds in real environmental systems requires the knowledge of ε_{ER}, which is specific for each compound. The present study demonstrates that quantitative structure-activity relationship model (QSAR) describes well the dependence of ε_{ER} on molecular structure and can be used for the evaluation of ε_{ER} for unstudied chiral compounds belonging to a well-studied homologous series.

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.

KEYWORDS

Rayleigh equation, Hansch equation, QSAR, LFER, QSER, enantiomeric enrichment, enzymatic degradation.

ABSTRACT

It was recently demonstrated that under environmentally relevant conditions the Rayleigh equation is valid to describe the enantiometric enrichment - conversion relationship, yielding a proportional constant called the enantiometric enrichment factor, ε_{ER}. In the present study we demonstrate a quantitative structure-activity relationship model (QSAR) that describes well the dependence of ε_{ER} on molecular structure. The enantiomeric enrichment factor can be predicted by the linear Hansch model, which correlates biological activity with physicochemical properties. Enantioselective

hydrolysis of sixteen derivatives of 2-(phenoxy)propionate (PPMs) have been analyzed during enzymatic degradation by lipases from *Pseudomonas fluorescens* (PFL), *Pseudomonas cepacia* (PCL), and *Candida rugosa* (CRL). In all cases the QSAR relationships were significant with R^2 values of 0.90-0.93, and showed high predictive abilities with internal and external validations providing Q^2_{LOO} values of 0.85 - 0.87 and Q^2_{Ext} values of 0.8-0.91. Moreover, it is demonstrated that this model enables differentiation between enzymes with different binding site shapes. The enantioselectivity of PFL and PCL was dictated by the electronic properties, whereas the enantioselectivity of CRL was determined by lipophilicity and steric factors. The predictive ability of the QSAR model demonstrated in the present study may serve as a helpful tool in environmental studies, assisting in source tracking of unstudied chiral compounds belonging to a well-studied homologous series.

INTRODUCTION

Enantioselective degradation of micropollutant stereoisomers (as chiral pesticides and drugs) in polluted aquifers received growing research attention in recent years,[1] demonstrating that shifts in the enantiomeric enrichment of micropollutants in effluents and contaminated streams can be used for source tracking[2] and elucidation of the degradation mechanisms.[3, 4] In compound-specific isotope analysis the Rayleigh equation is used to describe the relation between changes in isotopic composition vs. contaminant concentration during the degradation process. The isotope enrichment factor, ε, derived from the Rayleigh equation may serve as a parameter for the specific reaction pathway[5] and as an assessment for source tracking[6], since it does not depend on the conversion. Recently,[7,8,9,10] it was established that the Rayleigh equation is also effective in describing the

enantioselective behavior by deriving the enantiomeric enrichment factor (ε_{ER}), which can be used as an identifying tool for a specific enzymatic reaction (eq 1).

$$ln\frac{ER_t}{ER_0} = \varepsilon_{ER} \times ln\, f \quad (1)$$

ER_t and ER_0 represent the initial and conversion-dependent enantiomeric enrichments (ER=[R]/[S]), f is the residual fraction (C_t/C_0), and ε_{ER} represents the enantiomeric enrichment factor.

Taking the correlation between the enantiomeric enrichment and the kinetic degradation a step forward, reveals an interesting phenomenon: the enantiomeric enrichment factor derived from enzymatic degradation reactions is in correlation to the molecular structure. Therefore the Rayleigh equation can be used for prediction of the enzymatic kinetics and enantioselective enrichments of molecules belonging to homologous series, by linear free energy relationships (LFER). The ability of modeling the fate of enantioselective degradation and biodegradation, as a whole, can save the need of multiple lab work to predict biodegradation of chemicals in natural systems. To this end, The United States Environmental Protection Agency (USEPA) is investing considerable effort in research aimed at reliable structure-activity relationships (SAR) and models are needed to understand the mechanisms of biodegradation, to classify chemicals according to relative biodegradability, and to develop reliable biodegradation estimation methods for new chemicals.[11,12,13]

LFER have served a fundamental role in physical organic chemistry by providing a quantitative correlation between structure and reactivity.[14,15] Presently, extended forms of LFERs, namely, quantitative structure activity relationships (QSARs), are commonly used for formulating mathematical relationships which describe the structural dependence of biological activities.[16,17,18] These predictive models are derived based on the correlation between experimental data and biological features and can lead to identify the bioavailability, toxicity and biological activities of compounds as the dependent variables.[19,20] In the last decade, these methods have been applied to

chemical catalysis with respect to catalyst activity and selectivity,[21] developing quantitative structure enantioselective relationship models (QSER) for asymmetric chemical[22,23] and enzymatic[24] reactions. These models can predict the outcome of asymmetric reactions,[25] describe the enantioselective mechanism[26] and design improved catalysts.[27]

In this work, the multiple linear regression (MLR) method was applied to build the QSAR based on the Linear Hansch model.[28] In the classical Hansch approach[29], substituent constants like the Hansch lipophilicity parameter[30] (π), Hammett's electronic parameter (σ) and Taft's steric parameter (Es) are employed as structural descriptors for the variation in the test set and are correlated with the dependent variable C, the concentrations of the compound producing the biological response being measured (eq 2).

$$\log 1/C = a\sigma + bEs + c\pi + d \quad (2)$$

In 1965[31] Hansch et al. applied the steric parameter Es, to reactions occurring on enzymes, governing equation 3 by using $1/k_M = k_f/k_r$ (k_M is the Michaelis–Menten saturation constant, k_f and k_r are the enzyme-substrate complexation and dissociation rate constants, respectively).

$$\log \frac{k_f}{k_r} = k_a\pi + k_b\sigma + k_cEs + k_d \quad (3)$$

In a previous article[7] we have defined the enantiomeric enrichment factor in equation 1 by equation 4.

$$\varepsilon_{ER} = \frac{-\overline{k}}{k_c} \quad (4)$$

k_c is the observed overall first order rate constant of both enantiomers, and $\overline{k}$ is the difference between the individual first order rate constants of each enantiomer undergoing the enzymatic degradation. Replacing k_f/k_r in equation 3 by ε_{ER}, leads to equation 5 which was used for building our QSAR model.

$$\log \varepsilon_{ER} = k_a\pi + k_b\sigma + k_c Es^k + k_d \quad (5)$$

In order to build the model, series of structural analogs that adhere to the same binding/degradation mechanism, are to be analyzed for obtaining the enantiomeric enrichment data series. Herein, we analyzed the enantioselective hydrolysis of sixteen derivatives of 2-(phenoxy)propionate (PPMs) (Figure 1, Table 1), some of them are common herbicides[32] that are ubiquitous water contaminants.[33,34] The enantioselective degradation has been carried out with three lipase enzymes from three species: *Pseudomonas fluorescens* (PFL), *Pseudomonas cepacia* (PCL) and *Candida rugosa* (CRL).

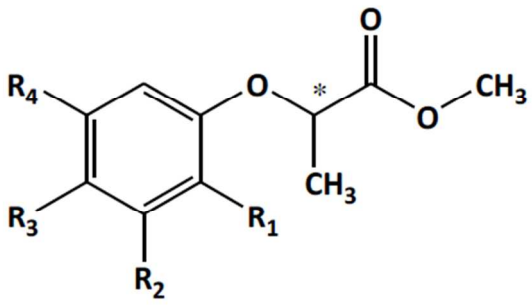

Figure 1. Structure of PPMs. R_1, R_2, R_3 and R_4 represent different substituents on PPM as described in Table 1; chiral center is denoted by an asterisk (*).

MATERIALS AND METHODS

Materials and reagents are described in the Supplementary Information, SI.

All the studied enzymatic reactions were carried out at 21±2ºC. The kinetic tracking of the transformations was carried out in parallel separate vials, and the whole content of each vial was used for a single analysis. Detailed reaction and extraction procedures can be found in the SI.

Enantiomeric enrichment analysis

The chiral reactants **1-4, 8-11** and **13-16** (Table 1) were analyzed by GC-SMB-QQQ-MS (a combined instrument comprised of GC, Agilent 7890A, Aviv Analytical supermolecular beam ion source[35] and Agilent 7000A triple quadruple mass spectrometer), equipped with a chiral column (β-cyclodextrin, 13105Rt-bDEXsm, 30 m x 250 µm x 0.25 µm; Restek). Compounds **5-7** and **12** (Table 1) where analyzed by HPLC-UV (Finnigan TSP 4000 series) equipped with a chiral column (Cellulose-Tris-(3,5-dimethylphenyl)-carbamate, Reprosil Chiral-OM, 5 µm x 250 mm x 4.6 mm ID, Dr. Maisch (Germany)). Detailed analytical method is presented in the SI.

Statistical procedures

Multiple linear regression (MLR) analyses and statistical analysis were performed using SPSS 8.0 and Microsoft Excel 2010 software. Detailed model development and statistical validation for the significance confirmation of the QSAR model is provided in the SI.

RESULTS AND DISCUSSION

Enantioslective enzymatic degradation analysis

The kinetic tracking of all the CPPMs, detailed in Table 1, gave first order kinetic fits with overall rate constants in the range of 0.009-1.002 hr^{-1} (detailed in Table S1 in the SI). The enantioselective degradation of all compounds followed the Rayleigh dependence (R^2 = 0.94-0.99), obtaining the enantiomeric enrichment factors detailed in Table 1.

Table 1. Structures, Hansch fitted parameters and Rayleigh enrichment factors, for the PPMs degraded by different lipases.

		Substituents [a]				Hansch fitted parameters			Rayleigh enrichment factors[e]		
no.	analytes	R_1	R_2	R_3	R_4	π^b	σ^c	$Es^{k\,d}$	ε_{ER} PCL	ε_{ER} PFL	ε_{ER} CRL

1	PPM	H	H	H	H	0	0	-3.05	-23±2	-41±5	-21±12
2	CPPM	Cl	H	H	H	0.76	0.23	-2.43	-75±9	-83±19	-72±11
3	DCPPM	Cl	H	Cl	H	1.46	0.46	-1.81	-89±15	-113±21	-168±14
4	MCPPM	CH_3	H	Cl	H	1.54	0.06	-1.83	-46±24	-66±10	-194±4
5	TCPPM	Cl	H	Cl	Cl	2.22	0.83	-1.19	-234±7	-266±19	-268±15
6	HYPPM	H	H	OH	H	-0.61	-0.37	-2.45	-15±3	-19±4	-159±12
7	BPPM	H	H	Ph	H	2.69	-0.01	-1.93	-27±3	-62±5	-76±5
8	DBrPPM	Br	H	Br	H	1.77	0.46	-1.55	-50±7	-63±3	-282±21
9	OCH_3PPM	OCH_3	H	H	H	-0.33	-0.27	-2.32	-20±3	-23±4	-125±3
10	DMPPM	H	CH_3	H	CH_3	1.35	-0.239	-1.85	-19±4	-24±5	-130±8
11	OCF_3PPM	H	H	OCF_3	H	1.21	0.35	-1.64	-66±17	-101±15	-267±6
12	NPPM	H	H	NO_2	H	0.24	0. 78	-1.60	-156±19	-149±6	-299±11
13	IPPM	H	H	I	H	1.26	0.18	-2.19	-46±10	-61±2	-86±13
14	BrPPM	H	H	Br	H	1.02	0.23	-2.3	-41±3	-67±6	-63±10
15	FPPM	F	H	H	H	0.01	0.06	-2.55	-30±2	-48±1	-45±4
16	NaphPPM	H	H	C_2H_2	C_2H_2	1.24	0.04	n.d.	-28±3	-46±3	-324±21

a Positions of R_1, R_2, R_3 and R_4 groups are labeled on the common structure in Figure 1.; *b* Hansch's lipophilicity parameter, taken from ref. 30 (from the phenoxyacetic acid system) and ref. 36; *c* Hammett's electronic parameter, taken from ref. 15 and 37; *d* Taft's steric parameter calculated from Kier's kappa values[38], taken from ref.39; n.d -not determined in the literature; *e* in percent units (%). The ± sign indicates the 95% confidence interval of the slope of the regression line in the Rayleigh plots. The enzymatic hydrolysis reactions were carried at pH 7.4.

QSAR analysis

Using the data in table 1, multiple linear regressions analysis (MLR) was performed to build the QSAR models from the various descriptors. Interestingly, when we tried using all three descriptors for PCL and PFL the coefficients of π and Es^k were insignificant (-0.02(±0.04)π, 0.08(±0.1)Es^k and

0.08(±0.04)π, -0.09(±0.09)Es^k, respectively, P-values>0.001). Therefore, for the two *Pseudomonas* enzymes (PCL and PFL) the model utilized only one descriptor- the electronic structural parameter, σ, obtaining equations 6 and 7. However, this does not hold true for the *Candida* enzyme (eq 8), where σ was not significant at all (-0.18(±0.1)σ, P-value>0.001) (Figure 2) and the model was built based on the lipophilicity and steric parameters. These different relations were reported previously when studying the effect of substituents on the enantioselectivity of lipase catalyzed reactions (these reports used values that are conversion dependent (ER) whereas we use ε_{ER} that is a more general value). Y. Kawanami et al.[40] showed that the electron-withdrawing character might be the main factor to enhance the enantioselectivity in PCL. On the other hand Ueji et al.[41] reported the absence of Hammet correlation between the enantioselectivity and the electronic effect in CRL as in our case.

$$\log(\varepsilon_{ER}) = 0.98(\pm 0.10)\,\sigma + 1.47(\pm 0.04) \quad (6)$$

$$\log(\varepsilon_{ER}) = 0.87(\pm 0.10)\,\sigma + 1.63(\pm 0.03) \quad (7)$$

$$\log(\varepsilon_{ER}) = -0.27(\pm 0.06)\,\pi + 1.04(\pm 0.12)\,Es^k + 4.5(\pm 0.30) \quad (8)$$

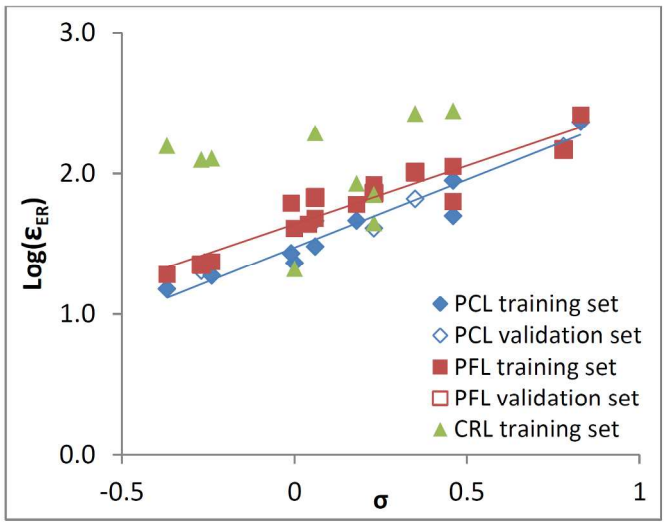

Figure 2. Correlations between the enantiomeric enrichment factor, ε_{ER}, and Hammett's electronic parameter, σ, for lipase from and *Pseudomonas cepacia* (PCL), *Pseudomonas fluorescens* (PFL) and lipase from *Candida rugosa* (CRL).

Table 2 details the statistical results obtained for the three QSAR model equations. As can be seen the sixteen compounds were divided into training and external validation sets (the splitting method is detailed in the SI), while the ratio between the number of descriptors and training compounds (1:11) in models 6 and 7 is two times higher than the minimum Toppliss and Costello criterion[42] and in model 8 the ratio is compatible with the criterion (should be at least 1:5).

An appropriate QSAR model is indicated by large F, small STD, small sig F, small P value, small RMSE and R^2 value close to 1.[43] Frequently, P value (<0.001) and sig F.(<0.01) are used as a criterion for the significance of the regression model and $Q^2_{LOO\text{-}cv}/_{Ext} > 0.5$ and small root mean square errors of prediction (Table 3) are used as a criterion of both robustness and predictive ability of the model[44] (see detailed explanation and analysis in the SI). Thus, the values in Tables 2,3 and the trend line of the training and validation sets (the red lines in Figure 3) that is close to the y = x line (the black solid line in Figure 3), demonstrate that the model is performing with high correlation and predictive ability. Additionally, the difference between the R^2 and R^2adj values, as well as between the R^2 and Q^2 values, is less than 0.3, indicating that the number of descriptors involved in the model is acceptable and the model is not over-fitted.[44] When analyzing the cross validated residuals for the training set and from the predictions for the validation set, we did not identify any significantly outlying results i.e. the residuals are not differing by more than 2.5 standard deviations from zero (Figure S2 in the SI).

Table 2. Statistical results of the MLR for the three different enzymes.

Eq. no (Enzyme)	n	R	R^2	R^2 adj	STD	F	sig. F	P-value	RMSEC
6 (PCL)	11 T 5 V	0.95	0.91	0.90	0.11	86.35	$6x10^{-6}$	$2x10^{-11}$	0.10

7 (PFL)	11 T 5 V	0.95	0.90	0.89	0.10	79.55	$9x10^{-6}$	$4x10^{-12}$	0.09
8 (CRL)	10 T 5 V	0.96	0.93	0.91	0.11	45.15	$1x10^{-6}$	$1x10^{-6}$	0.09

n -number of compounds; T -training set V- validation set ; R - coefficient of correlation; R^2- coefficient of determination; R^2 adj- adjusted coefficient of determination; STD - Standard Deviation; F -sequential Fischer test value; sig. F-significance F; P value- calculated probability; RMSEC-Root Mean Square Error of Calibration.

Table 3. Statistical results for the internal and external validations.

Eq. no (Enzyme)	Q^2_{LOO-CV}	RMSECV	Q^2_{Ext}	RMSEP
6 (PCL)	0.85	0.13	0.91	0.09
7 (PFL)	0.85	0.12	0.87	0.10
8 (CRL)	0.87	0.12	0.80	0.12

$Q2_{LOO}$-cv- cross-validated correlation coefficient; RMSECV-cross-validated root mean square error of prediction; Q^2_{Ext}- externally validated determination coefficient; RMSEP - root mean square error of prediction.

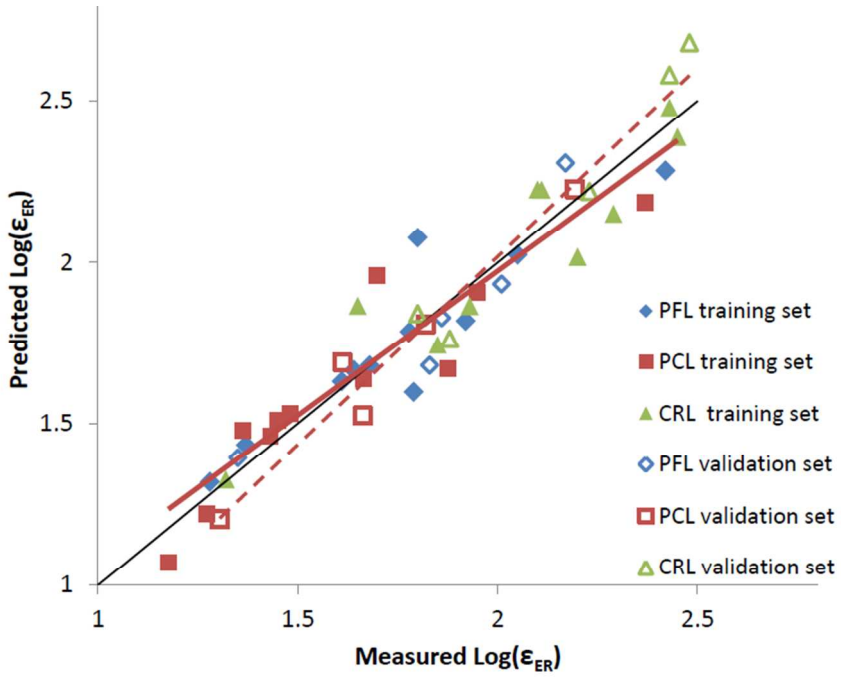

Figure 3. Experimental vs. predicted ε_{ER} values by LOO-CV MLR for the training set and external validation for the validation set. The solid and dashed red lines are the trend lines of the training and validation sets with all three enzymes, respectively. The black solid line is the y = x line.

There are three ways to describe enantiomeric enrichments by the Rayleigh dependence[45, 46, 7] (eqs 1,9,10). In a previous article[7] we have demonstrated that all three forms are equivalent, connected by the relationship: $\boldsymbol{\varepsilon_{ER} = \varepsilon_1 \times k_1/k_c = \varepsilon_2 \times k_2/k_c}$ (k_1, k_2 and k_c are the individual and overall first order rate coefficients, respectively). Thus we have performed the MLR on the data of ε_1 and ε_2, obtaining, as expected, the same dependence of the QSAR model as described for eqs 6-8 (eqs S16-S21 in the SI).

$$ln\frac{ER_t}{ER_0} = \varepsilon_1 \times ln\, f\left[\frac{(1+1/ER_0)}{(1+1/ER_t)}\right] \quad (9)$$

$$\boldsymbol{ln\frac{ER_t}{ER_0} = \varepsilon_2 \times ln\, f\left[\frac{(1+ER_0)}{(1+ER_t)}\right]} \quad (10)$$

According to PCL[47,48] and CRL[49] X-ray crystal structures, the binding site contains three main arms (Figure 4): (a) an esterase site (ES) with the catalytic triad Ser, His and Asp/Glu that comprise the active site which attacks the ester carbonyl group of the substrate, operating the hydrolysis reaction; (b) an oxyanion hole (OA), Gly, Leu/Ala which stabilizes the tetrahedral intermediate. And (c) an acyl chain binding site (ACS), which binds the acyl chain of the substrate. In addition, there is the stereospecific packet (SSP),[50] a hydrophobic zone that binds the more hydrophobic part of the remaining groups in the stereocenter,[51] (in our case the methyl versus the hydrogen). In the accepted mechanism[50, 52] for esters hydrolysis, there is a nucleophilic attack of the serine oxygen on the carbon of the ester once it binds to the active site. This attack forms a tetrahedral intermediate; hydrogen bonds from two amide N-H bonds stabilize the oxyanion in this intermediate. Breakdown of the tetrahedral intermediate releases the alcohol and forms an acyl enzyme intermediate .The selectivity of the lipase depends on the stability and reactivity of the tetrahedral intermediate, which depends on the binding of the substrate to the active site. However, the affinity of the acyl chain to the ACS determines the possible configuration for the bond of the methyl/hydrogen to the SSP, affecting the level of selectivity, namely the enantiomeric enrichment factor (ε_{ER}). So in order to understand the relationship between ε_{ER} and σ we have to refer to the acyl chain binding site. The ACS is a hydrophobic zone[48] which can be described as the groove with the following amino acids on its walls (the exact amino acid numbering is for PCL): Leu17, Leu167 and Leu164 on one side; Val266, Val267 and Phe119 on the other with Pro113 closing the groove.[48,53] the common assumption is that the acyl chain bound via van der Waals' hydrophobic interactions. The Pro113 at the end of the groove (Figure 4) can participate in electrostatic interactions,[54] which can impact the enrichment factor, thus being responsible for the observed strong electronic dependence.

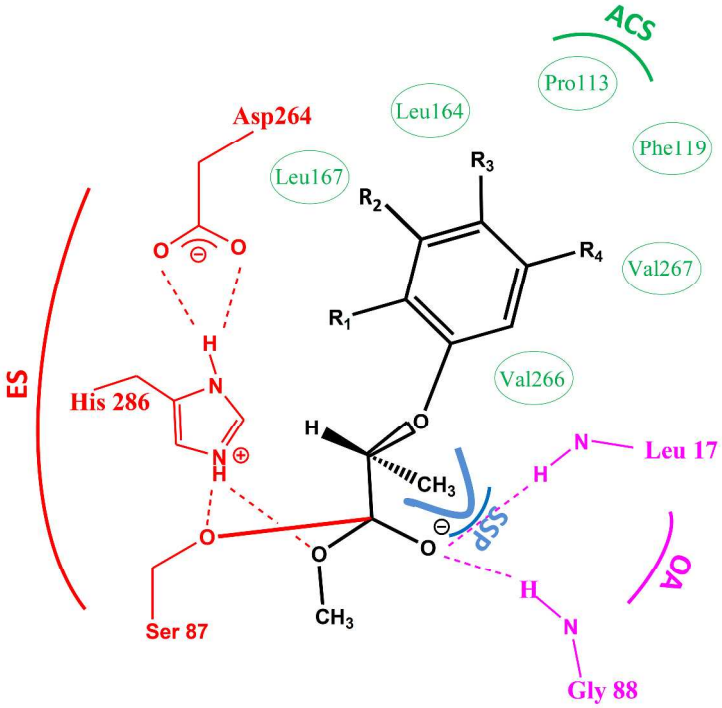

Figure 4. Scheme of the possible position of an enantiomer of PPM in the active site of lipase from *Pseudomonas cepacia.* ACS- acyl binding site (green); ES- esterase site - the catalytic trade: Asp264, His286 and Ser87 (red); SSP-Stereo-specific packet (blue); OA- oxyanion hole (magenta). The catalytic and stabilizing bonds are marked in solid and dashed lines, respectively. The structure of the binding site is according to references 48, 50.

Considering the main forces operating in the active site, we suggest that the difference in the correlation between ε_{ER} and σ is due to the different structure of the lipase binding site of the *Pseudomonas* lipases verses the *Candida rugosa* lipase. Pleiss et al.[53] analyzed and compared the shape of the binding site of six lipases and subdivided them into three sub groups: (1) lipases with a crevice-like binding site (lipases from Rhizomucor and Rhizopus); (2) lipases with a funnel-like binding site (lipases from Candida antarctica, *Pseudomonas* and mammalian pancreas and cutinase); and (3) lipases with a tunnel-like binding site (lipase from *Candida rugosa*). Illustrating the shapes of the binding sites as Pleiss et al. in Figure 5A,B shows that the binding pocket of *Pseudomonas*

cepacia lipase (PCL) is an elliptical funnel (the length is 17 Å and the width at the base is 4.5 Å that increases to 10.5 Å at the entrance to the binding site) and the catalytic active site lies in the base of the funnel. Although Pleiss et al. did not investigate the *Pseudomonas fluorescens* lipase, this illustration is likely suitable for PFL due to their significant structure similarity.[47, 55] Figure 5C,D illustrates the binding pocket of *Candida rugose* lipase (CRL) as a tunnel with a wide entrance at the right hand side (the tunnel is at least 22 Å long with a diameter of about 4 Å) and the catalytic active site lies just behind the entrance to the tunnel. We hypothesize that in the "open" structure of the binding sites of PFL and PCL, the forces operating in the active site exclusively determine the selectivity (and activity) of the processes (σ). Whereas in the tunnel-like structure of the binding site of CRL, additional accessibility factors (e.g. hydrophobic and steric parameters) may affect the reaction's selectivity.

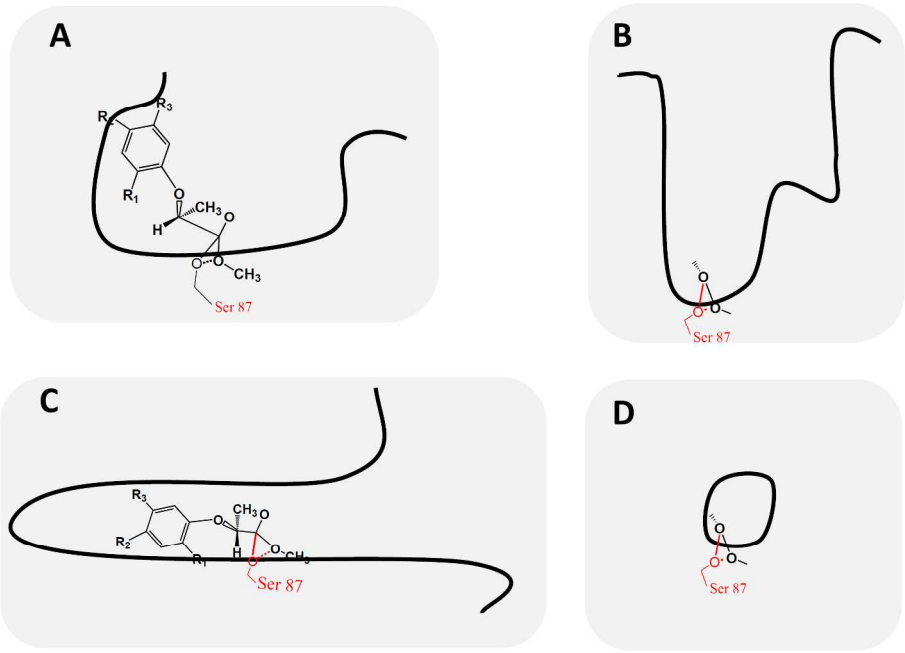

Figure 5. Shape of the binding site of PCL (A,B) and CRL(C,D) in side view(A,C) and front view (B,D). The catalytic active site is marked by the serine (in red) and the PPM molecule binds to the serine inside the binding pocket. According to reference 53.

The kinetics of chemical[56,57] and enzymatic[58,59] processes are frequently described by QSAR relationships. When preforming the linear Hansch model for the first order kinetic constants (Table S1 in the SI), significant models were obtained for the overall kinetic constants, $k_{c,}$ of PCL and PFL only when relying on π, Es^k and σ (eqs S22,S23 in the SI). Whereas the individual enantiomer rate coefficients- k_1 and k_2 as well as all the kinetic constants of CRL did not correlate with these descriptors (eqs S24-S30 in the SI). The stronger correlation between the structural parameters and selectivity compared with the dependence of the rate on these parameters may derive from the fact that the selectivity coefficient (ε_{ER}) is a ratio between the rate coefficients (eq 4); It has been shown that in some cases[60,61] of statistical interpretation a ratio between two parameters depends less on the individual characteristics than each of the parameters, because these effects cancel each other. The success of QSAR in describing enantiomeric enrichment compared to its ability to predict individual kinetics emphasizes the importance of using the Rayleigh equation for describing enantiomeric enrichments.

In conclusion, this study not only demonstrated for the first time the predictive power of QSAR and Hansch modeling for analysis of the structural dependence of the chiral enrichment factor, but also revealed that, at times, the QSAR fit of the enrichment factors are much more significant and better predictive tools than the QSAR fit of the underlying individual kinetic parameters. We have shown that chiral analysis using the Rayleigh equation and QSAR modeling uncover the latent binding site similarity between the two *Pseudomonas* lipases, as well as their difference from the *Candida* lipase which are not readily observed based on QSAR analysis of the individual kinetic coefficients. This ability, to predict enantioselective- conversion dependencies by the Rayleigh equation, can present a powerful tracer tool in environmental studies

ACKNOWLEDGMENT

The authors are grateful to the Excellence Center of the Israel Ministry of Agriculture, the Israel Water Authority, the Scientific Infrastructure Program of the Israel Ministry of Science and TEVA Pharmaceutical Industries Ltd. for the financial support of this study.

ASSOCIATED CONTENT

Supplementary Information (ESI) available: Materials and reagents, experimental details, QSAR data, details on the QSAR modeling procedure and confirming the significance of the QSAR model, Summary of fundamental equations leading to some constants listed in the article, kinetic data and more QSAR equations. See DOI: 10.1039/x0xx00000x

AUTHOR INFORMATION

*** Ovadia Lev**

The Hebrew University of Jerusalem

Jerusalem, 91904, Israel

E-mail: ovadia@mail.huji.ac.il

Office: +972(0)26584191; fax: +972(0)26586155

***Faina Gelman**

The Geological Survey of Israel,

Malkhei Israel 30, Jerusalem 95501, Israel

E-mail: faina@gsi.gov.il

Office: +972(0)25314208

REFERENCES

1. C. S. Wong, *Anal Bioanal Chem*, 2006, **386**, 544-558.
2. L. Fono and D. L. Sedlak, *Abstr Pap Am Chem S*, 2005, **230**, U1534-U1535.
3. V. Matamoros, M. Hijosa and J. M. Bayona, *Chemosphere*, 2009, **75**, 200-205.
4. C. Zipper, M. J. F. Suter, S. B. Haderlein, M. Gruhl and H. P. E. Kohler, *Environ Sci Technol*, 1998, **32**, 2070-2076.
5. M. Thullner, F. Centler, H. H. Richnow and A. Fischer, *Org Geochem*, 2012, **42**, 1440-1460.
6. A. Cincinelli, F. Pieri, Y. Zhang, M. Seed and K. C. Jones, *Environ Pollut*, 2012, **169**, 112-127.
7. S. Jammer, A. Voloshenko, F. Gelman and O. Lev, *Environmental Science & Technology*, 2014, **48**, 3310-3318.
8. G. Gasser, I. Pankratov, S. Elhanany, P. Werner, J. Gun, F. Gelman and O. Lev, *Chemosphere*, 2012, **88**, 98-105.
9. S. Bashir, A. Fischer, I. Nijenhuis and H. H. Richnow, *Environmental science & technology*, 2013, **47**, 11432-11439.
10. S. Qiu, E. Gozdereliler, P. Weyrauch, E. C. Lopez, H. P. Kohler, S. R. Sorensen, R. U. Meckenstock and M. Elsner, *Environmental science & technology*, 2014, **48**, 5501-5511.
11. J. C. Dearden, *J Brazil Chem Soc*, 2002, **13**, 754-762.
12. M. T. D. Cronin, J. D. Walker, J. S. Jaworska, M. H. I. Comber, C. D. Watts and A. P. Worth, *Environ Health Persp*, 2003, **111**, 1376-1390.
13. J. W. Raymond, T. N. Rogers, D. R. Shonnard and A. A. Kline, *J Hazard Mater*, 2001, **84**, 189-215.
14. P. R. Wells, *Linear free energy relationships*, Academic P., 1968.
15. R. W. Taft, *Progress in Physical Organic Chemistry*, John Wiley & Sons, 1976.
16. A. N. Choudhary, A. Kumar and V. Juyal, *Mini Rev Med Chem*, 2010, **10**, 705-714.
17. M. Iman and A. Davood, *Med Chem Res*, 2013, **22**, 5029-5035.
18. A. Bajpai, N. Agarwal and S. P. Gupta, *Indian J Biochem Biophys*, 2014, **51**, 244-252.
19. M. Yuan, B. Liu, E. M. Liu, W. Sheng, Y. Zhang, A. Crossan, I. Kennedy and S. Wang, *Anal Chem*, 2011, **83**, 4767-4774.
20. N. A. Al-Masoudi, D. S. Ali, B. Saeed, R. W. Hartmann, M. Engel, S. Rashid and A. Saeed, *Arch Pharm (Weinheim)*, 2014.
21. M. S. Sigman and J. J. Miller, *J Org Chem*, 2009, **74**, 7633-7643.
22. A. Milo, E. N. Bess and M. S. Sigman, *Nature*, 2014, **507**, 210-214.
23. J. P. Wolbach, D. K. Lloyd and I. W. Wainer, *J Chromatogr A*, 2001, **914**, 299-314.
24. S. Funar-Timofei, T. Suzuki, J. A. Paier, A. Steinreiber, K. Faber and W. M. F. Fabian, *J Chem Inf Comp Sci*, 2003, **43**, 934-940.
25. K. C. Harper and M. S. Sigman, *P Natl Acad Sci USA*, 2011, **108**, 2179-2183.
26. K. C. Harper and M. S. Sigman, *Science*, 2011, **333**, 1875-1878.
27. K. C. Harper and M. S. Sigman, *J Org Chem*, 2013, **78**, 2813-2818.
28. C. Hansch, *Cc/Life Sci*, 1982, 18-18.
29. C. F. Hansch, T., *J Am Chem Soc*, 1964, **86**, 1616-1626.

30. T. I. Fujita, J Hansch, C., *Journal of the American Chemical Society*, 1964, **86** 5175–5180.
31. C. Hansch, E. W. Deutsch and R. N. Smith, *Journal of the American Chemical Society*, 1965, **87**, 2738-2742.
32. K. Kirkland and J. D. Fryer, *Weed Res*, 1972, **12**, 90-&.
33. C. Zipper, M. Bunk, A. J. B. Zehnder and H. P. E. Kohler, *J Bacteriol*, 1998, **180**, 3368-3374.
34. N. H. Spliid and B. Koppen, *Chemosphere*, 1998, **37**, 1307-1316.
35. A. Amirav, *Org Mass Spectrom*, 1991, **26**, 1-17.
36. A. Leo, C. Hansch and D. Elkins, *Chem Rev*, 1971, **71**, 525-+.
37. C. Hansch and A. Leo, *Substituent constants for correlation analysis in chemistry and biology*, Wiley, 1979.
38. L. B. Kier, *Quant Struct-Act Rel*, 1987, **6**, 8-12.
39. H. van de Waterbeemd, N. el Tayar, P. A. Carrupt and B. Testa, *J Comput Aided Mol Des*, 1989, **3**, 111-132.
40. Y. Kawanami, A. Honnma, K. Ohta and N. Matsumoto, *Tetrahedron*, 2005, **61**, 693-697.
41. S. Ueji, K. Watanabe, T. Koshiba, M. Nakamura, K. Oh-ishi, Y. Yasufuku and T. Miyazawa, *Biotechnol Lett*, 1999, **21**, 865-868.
42. J. G. Topliss and R. J. Costello, *J Med Chem*, 1972, **15**, 1066-&.
43. P. Gramatica, *Qsar Comb Sci*, 2007, **26**, 694-701.
44. 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.
45. B. Morasch, H. H. Richnow, B. Schink, A. Vieth and R. U. Meckenstock, *Appl Environ Microb*, 2002, **68**, 5191-5194.
46. D. Hunkeler, *Appl Environ Microb*, 2002, **68**, 5205-5206.
47. J. D. Schrag, Y. G. Li, M. Cygler, D. M. Lang, T. Burgdorf, H. J. Hecht, R. Schmid, D. Schomburg, T. J. Rydel, J. D. Oliver, L. C. Strickland, C. M. Dunaway, S. B. Larson, J. Day and A. McPherson, *Structure*, 1997, **5**, 187-202.
48. D. A. Lang, M. L. M. Mannesse, G. H. De Haas, H. M. Verheij and B. W. Dijkstra, *Eur J Biochem*, 1998, **254**, 333-340.
49. P. Grochulski, F. Bouthillier, R. J. Kazlauskas, A. N. Serreqi, J. D. Schrag, E. Ziomek and M. Cygler, *Biochemistry-Us*, 1994, **33**, 3494-3500.
50. F. Haeffner and T. Norin, *Chem Pharm Bull*, 1999, **47**, 591-600.
51. W. V. Tuomi and R. J. Kazlauskas, *J Org Chem*, 1999, **64**, 2638-2647.
52. A. Mezzetti, J. D. Schrag, C. S. Cheong and R. J. Kazlauskas, *Chem Biol*, 2005, **12**, 427-437.
53. J. Pleiss, M. Fischer and R. D. Schmid, *Chem Phys Lipids*, 1998, **93**, 67-80.
54. N. J. Zondlo, *Acc Chem Res*, 2013, **46**, 1039-1049.
55. S. Larson, J. Day, A. Greenwood, J. Oliver, D. Rubingh and A. Mcpherson, *J Mol Biol*, 1991, **222**, 21-22.
56. A. Hatipoglu and Z. Cinar, *J Mol Struc-Theochem*, 2003, **631**, 189-207.
57. S. Vanderhoeven, J. Lindon, J. Troke, G. Tranter, I. Wilson and J. Nicholson, *Xenobiotica*, 2004, **34**, 73-85.
58. W. J. G. M. Peijnenburg, K. G. M. Debeer, H. A. Denhollander, M. H. L. Stegeman and H. Verboom, *Environ Toxicol Chem*, 1993, **12**, 1149-1161.
59. S. Masunaga, N. Wolfe and L. Camera, *Water Science & Technology*, 1993, **28**, 123-132.
60. R. A. Fisher, *Biometrics*, 1947, **3**, 65-68.
61. G. V. Fuguitt and S. Lieberson, *CDE working papers*, 1972, **72**, 18.