Quantitative Structure−Activity Relationships for the Enantioselectivity of Oxirane Ring-Opening Catalyzed by Epoxide Hydrolases

Simona Funar-Timofei,[†] Takahiro Suzuki,[‡] Joachim A. Paier,[§] Andreas Steinreiber,[§] Kurt Faber,[§] and Walter M. F. Fabian*[,§]

Institute of Chemistry, Romanian Academy, Bul. Mihai Viteazul 24, 1900 Timisoara, Romania, Faculty of Economics, Toyo University, 2-11-10 Oka, Asaka, Saitama 351-8510, Japan, and Institut für Chemie (IfC), Karl-Franzens Universität Graz, Heinrichstrasse 28, A-8010 Graz, Austria

Received August 2, 2002

The enantioselective ring-opening catalyzed by epoxide hydrolases originating from seven different sources of a series of 2,2-disubstituted oxiranes containing alkyl chains of different lengths, unsaturated (alkenyl, alkinyl) and aromatic groups as well as electronegative heteroatoms at various positions within the side chain was analyzed by quantitative structure−activity relationships. Models for the enantioselectivity were derived with the aid of multiple linear regression analysis (MLR) using several steric and electronic (quantum chemical) descriptors. On the basis of the models derived by MLR nonlinear modeling with artificial neural networks (ANN) was also done. Good predictive performance was observed for both modeling approaches. The models also indicate that different steric and/or electronic features account for the enantioselectivities observed for the individual epoxide hydrolases.

INTRODUCTION

Epoxide hydrolases catalyze the hydrolysis of epoxides,[1] allowing easy access to *vicinal* diols in enantiopure form.[2] Preparative-scale application of these catalysts became only feasible after several microbial sources[3−5] were identified, which ensured their supply in sufficient amounts by fermentation. Among them, bacterial epoxide hydrolases proved to be extremely flexible by exhibiting exquisite stereoselectivities over a wide substrate pattern, particularly on 2,2-disubstituted oxiranes.[6−14] This synthetic strategy has been successfully employed for the asymmetric total synthesis of various bioactive compounds in nonracemic form.[14b] To facilitate the choice of the optimal epoxide hydrolase for a given substrate, i.e., a 2,2-disubstituted oxirane, we have undertaken a quantitative structure−activity relationship (QSAR) study on a series of 2,2-disubstituted epoxides (see Chart 1 for structures) spanning a wide range of alkyl, alkenyl, and aryl moieties as well as heterosubstituted derivatives thereof.

MATERIALS AND METHODS

Target Property. Experimental enantioselectivities (enantiomeric ratio) *E* (for definition of the enantiomeric ratio *E*, see eq 1[15]) of oxirane ring-opening to *vicinal* diols were previously measured by us[6−14] with the aid of gas chromatography or HPLC for seven different organisms, namely *Rhodococcus ruber* NCIMB 11216 (**A**), *Rhodococcus ruber* DSM 43338 (**B**), *Rhodococcus ruber* DSM 44541 (**C**), *Rhodococcus equi* IFO 3730 (**D**), *Mycobacterium paraffinicum* NCIMB 10420 (**E**), *Rhodococcus ruber* DSM 44540 (**F**), and *Rhodococcus ruber* DSM 44539 (**G**), and are listed together with the predicted values in Table 4 (see below).

These latter *E*-values were determined by using eq 1. This method shows the advantage that easily accessible racemic starting materials can be used for the determination of enantioselectivities. For a biochemical description, *E*-values can formally be expressed as ratio k_{cat}/K_M of the rate constants k_{cat} for catalysis and the Michaelis−Menten constants K_M of the two enantiomers (eq 2[16]). It has been shown that the molecular reasons for enantioselectivity in biocatalyzed reactions may in principle originate either from differences of enantiomers in binding (K_M-values) or in catalysis (k_{cat}-values).[16b−e] In practice, however, both effects play a role to a certain extent.[17] In any case, log *E* corresponds to differences in catalysis free energies $\Delta\Delta G^{\ddagger}$ between the (*R*)- and (*S*)-enantiomers. Consequently, log *E*-values served as the target property used to derive quantitative structure−activity relationship.[18]

$$E = \frac{\ln[1 - c(1 + ee_p)]}{\ln[1 - c(1 - ee_p)]} \quad (1)$$

$$E = \frac{(k_{cat}/K_M)_R}{(k_{cat}/K_M)_S} \quad (2)$$

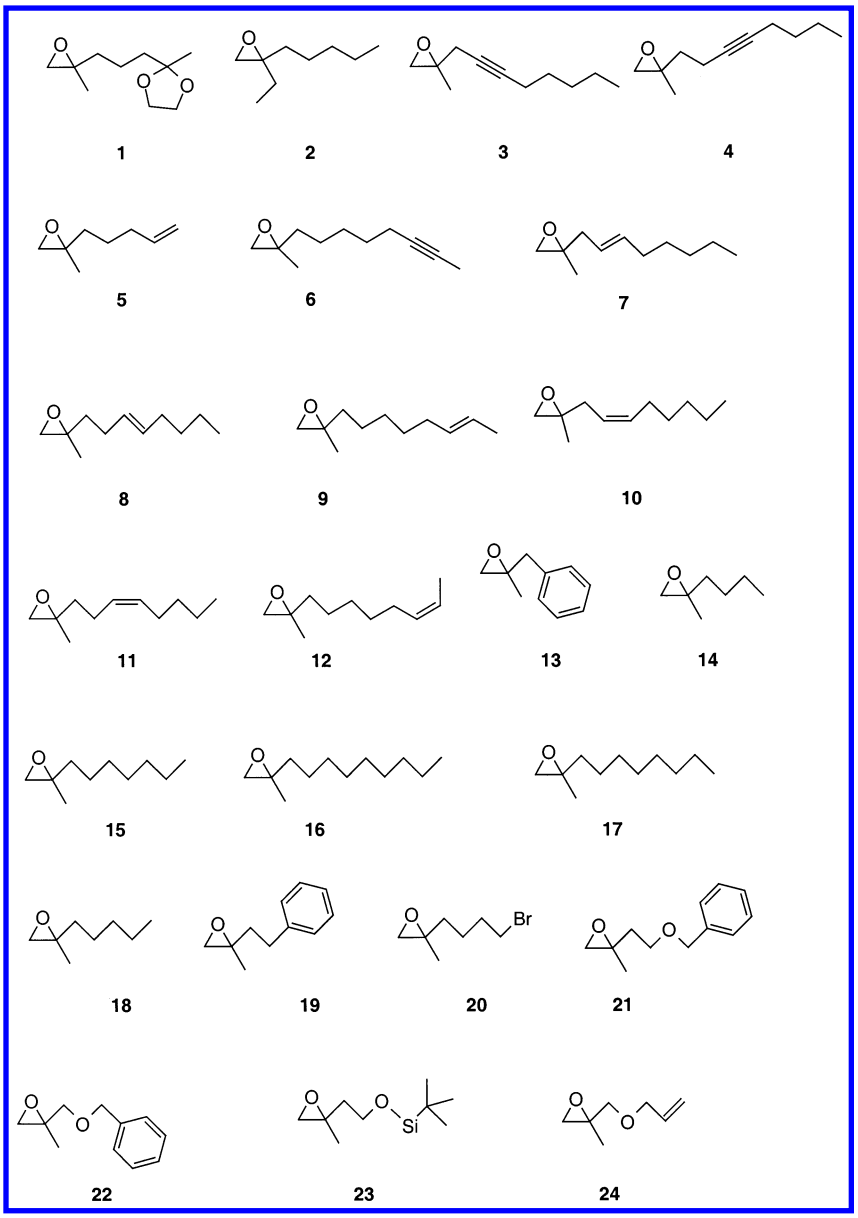
Here, E = enantiomeric ratio, c = conversion, ee_p = enantiomeric excess of product, k_{cat} = rate constant for catalysis, and K_M = Michaelis−Menten constant.

Generally, experimental data were determined under substrate saturation condition by gas chromatography and/or HPLC and are accurate to ±10%. However, due to the nature of eq 1 even quite small variations in high values of ee_p lead to large fluctuations in the calculated *E*-values. Thus, rather high enantioselectivities are difficult to measure. Consequently, *E*-values >200 are associated with consider-

* Corresponding author phone: ++43-316-380-8636; fax: ++43-316-380-9840; e-mail: walter.fabian@uni-graz.at.
[†] Romanian Academy.
[‡] Toyo University.
[§] Karl-Franzens Universität Graz.

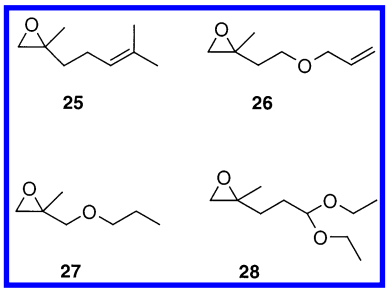
10.1021/ci020047z CCC: $25.00
Published on Web 04/26/2003

Chart 1



ably larger errors. Despite possible ambiguities introduced thereby in the statistical models, we decided not to exclude compounds showing essentially 100% enantioselective (i.e., $E > 200$) oxirane ring-opening. Furthermore, not all compounds depicted in Chart 1 could be reacted with all of the investigated hydrolases. As a consequence, the data sets for the individual enzymes are slightly different, according to their individual substrate tolerance. Four additional substrates (for structures, see Chart 2) were used as test set for prediction of log E values by the QSAR models described below.

Chart 2



Molecular Structures. Molecular modeling was done with the aid of the Sybyl program package.[19] The molecular structures of epoxides **1**−**24** were completely optimized using MOPAC[20] with the semiempirical AM1 Hamiltonian[21] (eigenvector following routine (EF[22]), keyword PRECISE).

Molecular Descriptors. The computer software DRAGON[23] was used to prepare a set of >800 molecular descriptors ranging from 1D to 3D descriptors such as WHIM.[24] In addition, a variety of other molecular properties, e.g., those derived from charged partial surface areas (CPSA[25]) as well as various quantum chemical indices, e.g., frontier orbital energies E_{HOMO} and E_{LUMO}, hardness $H = -(E_{HOMO} - E_{LUMO})/2$, dipole moment μ, minimum and maximum charges, partial charges of selected atoms, polarizability α, hyperpolarizability β, were used. Furthermore, several other steric (e.g., the Connolly Solvent Accessible Area (SAS),[26] Connolly Molecular Surface Area, Connolly Solvent Excluded Volume (SEV) and molecular weight), hydrophobicity (ClogP) and thermodynamic descriptors [e.g., the dipole−dipole energy (E_d), torsion energy (E_t) and the

Table 1. Molecular Descriptors of Epoxides Included in the Final MLR and ANN Models[a]

compd	SAS	SEV	ClogP	MR	Ed	Ev	tor-O3C*
1	396.88	184.29	1.54	4.92	1.51	−2.71	−86
2	376.59	161.6	3.41	4.33	−1.02	−0.99	−6.6
3	438.63	176.15	3.61	5.13	−0.21	−1.21	81.67
4	440.16	175.72	3.26	5.13	−0.10	−1.58	146.83
5	343.61	134.51	2.40	3.84	−0.97	−0.5	−4.44
6	434.95	175.01	3.11	5.13	−0.42	−0.46	47.83
7	440.64	185.01	3.99	5.23	−0.73	−0.59	42.59
8	441.33	185.4	3.99	5.23	−0.83	−1.55	152.09
9	440.35	184.86	3.99	5.23	−0.94	−0.8	2.39
10	434.9	184.84	3.99	5.23	−0.6	0.17	37.31
11	434.2	185.75	3.99	5.23	−0.77	−0.86	149.98
12	433.91	185.09	3.99	5.23	−0.9	−0.51	7.93
13	340.09	139.68	2.33	4.52	−0.94	−1.76	149.66
14	320.44	125.21	2.35	3.4	−1.03	0.07	47.83
15	414.04	177.61	3.94	4.79	−1.03	−0.88	3.53
16	475.32	212.03	4.99	5.72	−1.03	−1.21	3.29
17	444.71	194.84	4.47	5.26	−1.03	−1.04	3.4
18	351.08	142.41	2.88	3.86	−1.03	−0.11	47.94
19	375.69	155.67	2.71	4.98	−0.95	−0.52	46.49
20	353.77	146.78	2.21	4.18	−0.84	−1.31	151.7
21	425.97	178.37	2.45	5.6	−0.5	−0.39	62.39
22	397.54	162.51	2.16	5.14	−0.16	−0.75	160.77
23	402.32	190.6	1.72	5.3	−0.29	−1.39	140.91
24	339.12	124.01	0.95	3.53	−0.18	−1.37	−136.68
25	368.11	149.82	2.80	4.30	−0.81	−0.96	−61.07
26	358.13	139.46	1.19	3.99	−0.50	−0.10	−63.35
27	343.81	132.75	1.23	3.55	−0.26	−0.38	−102.46
28	431.32	194.71	1.00	5.09	−0.12	−1.31	−60.94

compd	nDB	EHOMO	H	PPSA1	FPSA1	PV	AR
1	0	−10.31	6.38	374.06	0.93	41.95	404.48
2	0	−10.9	6.74	366.99	0.96	14.48	381.77
3	0	−10.23	5.93	416.33	0.94	34.27	444.26
4	0	−10.25	5.98	442.21	0.94	106.9	469.53
5	1	−9.99	5.68	321.44	0.94	99.65	341.65
6	0	−10.13	5.97	412.6	0.94	0.49	439.44
7	0	−10.93	6.75	461.06	0.96	17.77	479.68
8	0	−10.94	6.75	428.8	0.96	14.75	447.42
9	0	−9.56	4.94	309.94	0.88	112.09	350.89
10	0	−10.99	6.79	306.82	0.94	17.22	325.66
11	1	−9.66	5.41	412.39	0.95	18.03	434.3
12	0	−10.98	6.79	337.65	0.95	13.86	356.53
13	0	−9.43	4.94	336.35	0.91	92.34	371.18
14	1	−9.63	5.41	416.37	0.95	19.83	438.68
15	1	−9.72	5.44	389.55	0.95	78.15	412.29
16	1	−9.53	5.41	415.84	0.95	20.17	438.04
17	1	−9.7	5.43	431.08	0.95	118.34	453.71
18	1	−9.54	5.42	422.83	0.96	8.37	442.37
19	0	−10.94	6.76	399.85	0.96	13.03	418.73
20	0	−10.78	5.75	272.75	0.77	113.29	356.62
21	0	−9.35	4.92	407.6	0.89	134.41	455.96
22	0	−9.47	4.93	370.74	0.89	121.24	419.15
23	0	−7.54	3.53	376.29	0.9	45.49	417.33
24	1	−10.12	5.62	319.1	0.91	43.4	349.17
25	1	−9.25	5.22	355.144	0.943	23.185	376.475
26	1	−9.89	5.63	339.292	0.927	39.329	366.067
27	0	−10.56	6.44	329.599	0.931	35.685	354.210
28	0	−10.55	6.51	424.249	0.940	75.189	451.284

[a] SAS represents the Connolly Solvent Accessible Area; SEV − the Connolly Solvent Excluded Volume; ClogP − hydrophobicity parameter (the logarithm of octanol/water coefficient); MR − molecular refractivity; E_d − the dipole−dipole energy; E_v − the sum of pairwise van der Waals interaction energy terms for atoms separated by more than three chemical bonds; tor-O3C* − the torsion angles of the side group with respect to the oxirane ring; nDB − number of double bonds; E_{HOMO} − the HOMO molecular orbital energy; H − hardness; PPSA1 − the sum of the positively charged solvent-accessible atomic surface areas in the molecule; FPSA1 − the fractional partial positive surface area; PV − the polar volume; AR − total surface area.

sum of pairwise van der Waals interaction energy terms for atoms separated by more than 3 chemical bonds (E_v)] derived from the optimized 3D molecular structures were calculated by the Chem3D Ultra software.[27] Other steric descriptors obtainable from the optimized 3D molecular structure included the torsion angles of the side group with respect to the oxirane ring (tor_O3C*) and the distance between the chiral carbon atom and the last (most distant) heavy atom of the chain attached to the chiral center. Structural descriptors used in the final MLR/ANN models as well as those obtained for the test molecules **25**−**28** are summarized in Table 1.

Multiple Linear Regression (MLR) Modeling. The multiple linear regression (MLR) analysis[28] has been performed by the STATISTICA software.[29] The leave-one-out[30] cross-validation procedure was applied to test the reliability of the models. Starting from the entire set of structural descriptors, variable selection by a stepwise regression procedure based on the Fischer test was performed. In the variable selection procedure starting variables with high dependence to the epoxide enantioselectivities were considered. Intercorrelations between variables were also checked (see Table S1 of the Supporting Information for details.)

Because of the statistical quality of the obtained models, outliers have been tested by the standard residuals.[28] All the statistical tests were done at a significance level of 5% or less. Statistical results are presented in Table 2.

Nonlinear Modeling by ANN. The theory and practical applications of artificial neural networks (ANNs) in chemistry have been reviewed in depth.[31] A fully connected three-layer network was used to relate *E*-values for various epoxide hydrolases to a selected set of molecular descriptors as derived by the MLR modeling described above. The hidden layer contained variable nodes, and the input and hidden variables each had a bias neuron. A sigmoid transfer function was used for each neuron, and weights were adjusted iteratively by back-propagation using the generalized delta rule to minimize mean square errors between observed and calculated outputs. Input and output data were normalized between 0.1 and 0.9, and models were evaluated on the basis of correlation coefficient (r) and root-mean-square error (RMSE).[32] To avoid network overfitting, each network was limited to a maximum of 1000 training iterations. The predictive ability of the network was also evaluated by the leave-one-out cross-validation procedure as with MLR. To highlight any "outliers" or "strong points" and to examine the model for chance correlations, the "jackknifed r" values[33,34] was used. In this method, each observation was in turn deleted from the analysis just as the leave-one-out cross validation, and the resulting correlation coefficient (r_j) was noted as the jackknifed r. Thus, cases with unduly high r_j values could be suspected to be "outliers", and those with low values could be considered "influential points". The architectures of the nets and relevant statistical parameters of the best ANN-derived model corresponding to the MLR-QSARs derived using the same molecular descriptors are listed in Table 3. For the determination of the number of hidden neurons, the values (ρ) for the ratio of the number of data points (training compounds) to the number of model parameters or weights should be around 2.[35] The ρ values in this study are in the recommended range (1.89−3.00) as shown in Table 3. All computations were performed using

Table 2. Results of the MLR Models for Epoxide Hydrolases from *Rhodococcus ruber* NCIMB 11216 (**A**), *Rhodococcus ruber* DSM 43338 (**B**), *Rhodococcus ruber* DSM 44541 (**C**), *Rhodococcus equi* IFO 3730 (**D**), *Mycobacterium paraffinicum* NCIMB 10420 (**E**), *Rhodococcus ruber* DSM 44540 (**F**), and *Rhodococcus ruber* DSM 44539 (**G**)[a]

log *E*	descriptor[b]	*n*	*r*	*s*	*F*	q^2_{LOO}	outliers[c]
A	PPSA1 (−0.45), Ed (−0.76)	15	0.897	0.27	24.8	0.688	**5**, ***6***, **19**, **21**, **22**
B	H (−0.69), Ev (0.74)	13	0.863	0.31	14.7	0.601	**11**, **12**
C	FPSA1 (−0.35), SAS (−0.26), Ed (−0.63)	17	0.808	0.41	8.1	0.348	**6**
D	nDB (0.86), SEV (0.64)	18	0.909	0.24	36.1	0.705	**2**, **11**, **17**, **21**
E	E_{HOMO} (0.78), PV (−0.49), ClogP (0.49), tor-O3C* (−0.60)	20	0.821	0.39	7.7	0.422	***2***, **20**, **22**
F	PPSA1 (−0.24), ClogP (−0.37), Ed (−0.78)	16	0.953	0.23	39.6	0.823	**6**, **19**, **22**
G	AR (−0.21), MR (−0.65), tor-O3C* (−0.28)	18	0.822	0.41	9.8	0.472	

[a] *n* − number of compounds, *r* − correlation coefficient, *s* − standard error, *F*-test, and cross-validated correlation coefficients q^2_{LOO}. [b] Relative contributions in parentheses (for explanation of descriptors, see text). [c] Epoxides with experimental *E*-values >200 are indicated in italics.

Table 3. Results of ANN Models Using the Same Set of Descriptors as In the MLR Models

Log *E*	ANN's configuration	*n*	ρ[a]	*r*	RMS	q^2_{LOO}	outliers[b]
A	2−2−1	17	1.89	0.892	0.277	0.502	**5**, ***6***, **22**
B	2−1−1	13	2.60	0.870	0.266	0.508	**11**, **12**
C	3−1−1	17	2.83	0.884	0.284	0.623	**6**
D	2−2−1	20	2.22	0.832	0.298	0.506	**2**, **11**
E	4−1−1	20	2.86	0.890	0.272	0.598	***2***, **19**, **20**
F	3−1−1	16	2.67	0.967	0.170	0.748	**6**, **19**, **22**
G	3−1−1	18	3.00	0.841	0.340	0.528	

[a] ρ = number of data points in the training set/sum of the number of connections in the ANN. [b] Epoxides with experimental *E*-values >200 are indicated in italics.

our own programs, written in C language, on a microcomputer running Windows NT as its operating system.

RESULTS AND DISCUSSION

Starting from the total set of potential descriptors and the variable selection procedure described above, the MLR analysis has been applied to model the enantioselectivity of the oxirane ring-opening catalyzed by epoxide hydrolases of seven different organisms. The resulting best MLR models are collected in Table 2.

Notably, at most four, or in several instances (**A**, **B**, and **D**) even just two, of the initial set of descriptors proved to be sufficient to provide statistically meaningful models. In all cases, except for the hydrolase from *Rhodococcus ruber* DSM 44539 (**G**), a few outliers were detected in the modeling. At least some of these outliers (indicated in italics in Tables 2 and 3) are caused by very high and, thus, less accurate experimental *E*-values. Among all organisms investigated, the best statistical results were noticed for QSAR models of the *Rhodococcus ruber* DSM 44540 epoxide hydrolase **F**. The enantioselectivity observed for oxirane ring cleavage catalyzed by this enzyme can be modeled by just three descriptors: the partial positive surface area (PPSA1), the epoxide hydrophobicity (ClogP), and the dipole−dipole energy of the epoxide (E_d). The most important contribution to the enantioselectivity is given by E_d which can be related to intramolecular electrostatic interactions within the epoxide. Notably, a decreasing influence of this parameter on log *E*-values is evident (Table 2). The other descriptors play a less significant role according to their relative contribution (see Table 2). PPSA1, one of the charged partial surface area (CPSA) descriptors introduced by Jurs,[25] is defined as the sum of the positively charged solvent-accessible atomic surface areas in the molecule. This descriptor is expected to encode the features responsible for polar substrate−receptor interactions.

A similar model, also based on the PPSA1 and E_d descriptors, was obtained for log *E* of the hydrolase from *Rhodococcus ruber* NCIMB 11216 (**A**). The statistical results—although still rather good—obtained for this hydrolase, however, are less satisfactory than that for log *E* of **F**. Similar outliers have been observed in both models: compounds **6**, **19,** and **22**. In addition, two other outliers were noticed for the model having log *E* (**A**) as target variable: **5** and **21**. Two descriptors are indicative for the observed enantioselectivity with the hydrolase of *Rhodococcus equi* IFO 3730 (**D**): the number of double bonds present in the epoxides (nDB) and the Connolly Solvent Excluded Volume (SEV). Both increase the epoxide enantioselectivity and a higher contribution of the first descriptor was noticed. Four outliers were found: epoxides **2**, **11**, **17**, and **21**. The enantioselectivity for the epoxide reaction with the *Rhodococcus ruber* DSM 43338 hydrolase **B** is also described by only two descriptors: epoxide hardness (*H*), accounting for the electron donor ability of the epoxide oxygen atom and the sum of pairwise van der Waals interaction energy terms for atoms separated by more than three chemical bonds (E_v). Epoxide hardness *H* has the highest correlation coefficient (see Table S1 of the Supporting Information) for *Rhodococcus ruber* DSM 43338 but a rather low value for other organisms, e.g., *Rhodococcus ruber* DSM 44539 (**G**). This descriptor (*H*), describing the tendency of low epoxide polarizability, has a lower contribution in comparison with E_v, which describes the van der Waals epoxide intramolecular interactions. Two outliers were observed for this target variable: compounds **11** and **12**. Similar statistical results were noticed in MLR models in which the target variables are log *E* of *Mycobacterium paraffinicum* NCIMB 10420 (**E**) and *Rhodococcus ruber* DSM 44539 (**G**). The log *E* (**E**) variable is described by four descriptors: the HOMO molecular orbital energy (E_{HOMO}), the polar volume (MolProp-PV), hydrophobicity (ClogP), and the torsion angle tor_O3C*. The highest contribution is given by E_{HOMO}, accounting for the electron donor ability of the epoxide. The higher this donor property the higher is the respective enantioselectivity. Three outliers were observed for this target variable: **2**, **20**, and **22**. The model describing log *E* of the *Rhodococcus ruber* DSM 44539 (**G**) epoxide hydrolase includes three descriptors: total surface area (AR), molecular refractivity (MR), and the torsion angle tor-O3C*. The highest contribution to the enantioselectivity is given by the

Table 4. Experimental and Predicted by Multiple Linear Regression Analysis (MLR) and Artificial Neural Network (ANN) Modeling Enantioselectivities (log *E*-Values) for Oxirane Ring Opening of Substrates **1**−**28** Catalyzed by Epoxide Hydrolases from Organisms **A**−**G**

	Rh. ruber NCIMB 11216 (**A**)			*Rh. ruber* DSM 43338 (**B**)			*Rh. ruber* DSM 44541 (**C**)			*Rh. equi* IFO 3730 (**D**)		
	Exp	MLR	ANN	Exp	MLR	ANN	Exp	MLR	ANN	Exp	MLR	ANN
1												
2				0.95	0.80	0.69	2.05	1.72	1.76	0.08		
3	1.08	0.86	0.72	1.51	1.09	1.05	0.23	0.66	0.34	0.61	1.05	0.98
4	0.43	0.60	0.66	0.23	0.80	0.73	0.20	0.51	0.15	1.00	1.04	0.97
5	0.85									1.59	1.45	1.48
6	2.30			1.66	1.60	1.67	2.23			1.45	1.03	0.96
7	1.15	1.35	0.97	1.00	1.08	0.97	1.00	1.15	0.93	1.15	1.18	1.16
8	1.15	1.62	1.24	0.59	0.39	0.46	1.28	1.28	1.21	1.26	1.19	1.17
9	2.10	2.28	2.16	1.62	1.93	1.95	1.88	1.74	1.99	1.23	1.18	1.16
10	1.69	1.85	1.74	1.51	1.60	1.62	1.15	1.11	1.09	1.00	1.18	1.16
11	1.43	1.62	1.12	0.93			0.92	1.27	1.28	0.80		
12	2.15	2.11	1.87	1.73			1.68	1.44	1.55	1.46	1.18	1.16
13	2.05	2.17	1.99	1.15	1.25	1.33	1.49	2.01	2.02	0.23	0.50	0.36
14							1.88	1.99	1.95			
15	2.10	2.05	1.98							2.30	2.10	1.89
16	2.30	1.94	1.94							2.30	2.62	1.99
17	2.30	1.88	1.94	1.90	1.49	1.58	1.98	1.53	1.64	1.30		1.96
18	2.02	1.92	1.95	2.30	2.16	2.04	2.30	1.85	1.86	1.69	1.57	1.59
19	0.98		1.69	0.84	1.12	1.03	1.08	1.66	1.74	0.75	0.74	0.59
20	2.30	2.31	2.22	1.18	1.12	1.09	2.30	2.47	2.10	0.73	0.60	0.45
21	0.48		0.80							0.45		1.03
22	2.30						1.65	0.99	1.48	0.59	0.84	0.71
23										1.52	1.27	1.27
24	1.56	1.24	1.56				1.40	1.09	1.37	1.15	1.29	1.29
25	0.63	0.87	0.68	0.20	0.73	0.75	0.28	0.69	0.71	0.45	0.72	0.72
26	0.72	0.72	0.58				0.46	0.58	0.63	0.67	0.65	0.66
27	1.00	0.60	0.63	0.96	0.62	0.60	0.99	0.47	0.44	0.96	0.16	0.14
28	0.08	0.36	0.35				0.00	0.25	0.09	0.15	0.56	0.57

	Mycobact. paraff. NCIMB 10420 (**E**)			*Rh. ruber* DSM 44540 (**F**)			*Rh. ruber* DSM 44539 (**G**)		
	Exp	MLR	ANN	Exp	MLR	ANN	Exp	MLR	ANN
1	1.00	1.24	0.99						
2	2.30			2.10	2.01	1.97	2.30	2.13	2.23
3	1.23	1.13	0.78	0.36	0.38	0.31	0.64	1.15	1.16
4	0.23	0.21	0.66	0.15	0.22	0.23	0.18	0.89	0.90
5	0.54	0.88	0.69	2.30	2.34	2.12	2.30	2.58	2.37
6	1.51	1.44	1.39	2.16			1.82	1.27	1.31
7	1.48	1.12	0.86	1.08	1.05	0.93	1.40	1.10	1.15
8	0.72	0.62	0.67	1.11	1.32	1.29	0.73	0.87	0.85
9	1.56	1.48	1.40	1.97	1.86	1.86	1.82	1.58	1.59
10	0.38	1.12	0.86	1.20	1.27	1.27	1.00	1.55	1.49
11	0.96	1.36	0.94	0.90	1.27	1.23	1.08	0.91	0.87
12	1.00	1.28	1.25	1.75	1.71	1.74	1.61	1.55	1.55
13	0.85	0.52	0.67	2.09	2.27	2.10	1.56	1.58	1.62
14				1.91	2.20	2.07	2.30	2.45	2.37
15	2.30	1.59	2.01						
16	2.30	2.41	2.56						
17	1.38	1.48	1.50	1.89	1.52	1.50	1.32	1.27	1.34
18	2.29	1.67	2.12	2.30	2.02	1.98	2.30	2.12	2.27
19	1.00	0.77	0.69	1.08		1.97	1.11	1.43	1.50
20	0.93			2.30	2.31	2.12	2.30	1.85	1.96
21	0.54	0.72	0.67						
22	1.26			2.03					
23	1.51	1.82	1.59	1.53	1.20	1.34	1.45	0.94	0.87
24	1.51	1.41	1.52						
25	0.26	0.97	1.15	0.34	0.78	0.86	0.38	1.02	0.83
26	0.63	0.56	0.45	0.49			0.50	1.13	0.89
27	0.98	0.49	0.41	0.94					
28	0.23	0.26	0.34	0.04			0.04	0.70	0.51

MR term, accounting for the bulkiness of epoxides. No outliers were found for this particular system. The lowest statistical results were noticed in the model with log *E* (**C**) as target variable. Three variables were included therein: the fractional partial positive surface area (FPSA1), the Connolly Surface Accessible Area (SAS), and the dipole−dipole energy (E_d). The enantioselectivity is decreased by an increase of E_d. Compound **6** was observed as outlier for this model.

Although all MLR models had statistically good *r* values ($r > 0.8$), unfortunately, only four out of the seven QSARs had acceptable predictivities ($q^2 \geq 0.50$).[36] This suggests the need of nonlinear modeling, e.g., by artificial neural networks (ANN). Relevant results obtained thereby are collected in Table 3.

As expected, the resulting ANN-QSARs have improved values of *r* and all q^2 scores exceed the threshold for acceptability ($q^2 \geq 0.50$), indicating a nonlinear dependence

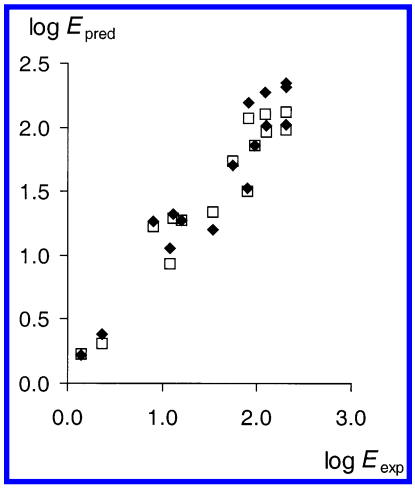


Figure 1. Plot of experimental log *E*-values vs those predicted by MLR (filled diamonds) and ANN (open squares) modeling for *Rhodococcus ruber* DSM 44540 (**F**).

of log *E* on the descriptors used. This is especially evident for log *E* (**C**) and/or log *E* (**E**). As found by MLR, for all seven epoxide hydrolases several outliers were identified in the respective QSAR models. Gratifyingly, by and large, the same outliers were found by MLR and ANN modeling, giving us confidence in the QSARs derived thereby. For log *E* (**A**), log *E* (**B**), and log *E* (**F**), the respective MLR models give better predictions than those obtained by ANN, although two of the outliers (**19** and **21**) present in the MLR model for log *E* (**A**) can be included in the corresponding ANN model. Apparently then, in case of these epoxide hydrolases, log *E* is essentially linearly dependent on the selected descriptors.

Experimental log *E*-values as well as those predicted by MLR and ANN modeling for the seven epoxide hydrolases are provided in Table 4. A plot of actual vs predicted enantioselectivities for *Rhodococcus ruber* DSM 44540 (**F**) is shown in Figure 1.

Rather than discuss all the individual data we will present some selected examples to highlight the results obtained by the statistical models presented in Table 2. Experimentally, a quite diverse behavior of the various substrates toward the individual organisms (or cell extracts) is found: *E*-values for epoxide **2** differ substantially among **A**−**G** ($E = 7$ (**A**), 8.9 (**B**), 113 (**C**), 1.2 (**D**), > 200 (**E**), 126 (**F**), > 200 (**G**)); in contrast, compound **18** shows considerably less variation ($E = 105$ (**A**), > 200 (**B**), > 200 (**C**), 49 (**D**), 194 (**E**), > 200 (**F**), > 200 (**G**)). *Rhodococcus ruber* DSM 44540 (**F**) and *Rhodococcus ruber* DSM 44539 (**G**) appear to be insensitive to hydrogenation of the terminal double bond (compare **5** vs **18**: both substrates are hydrolyzed by **F** as well as **G** with $E > 200$). In striking contrast, with *Mycobacterium paraffinicum* NCIMB 10420 (**E**) the enantioselectivity significantly increases for the saturated derivative **18** ($E = 194$) as compared to the terminal olefin **5** ($E = 3.5$). Replacement of the terminal carbon in **18** by the electronegative bromine to give **20** has little impact on the *E*-values obtained by *Rhodococcus ruber* DSM 44541 (**C**, $E > 200$), *Rhodococcus ruber* DSM 44540 (**F**, $E > 200$), and *Rhodococcus ruber* DSM 44539 (**G**, $E > 200$), whereas for *Rhodococcus ruber* DSM 43338 (**B**) a drop from $E > 200$ to $E = 15$ is observed. As can be seen from the data presented in Table 2, these effects are quite satisfactorily modeled by both the MLR and ANN equations. An ethyl group instead of methyl in position 2 of the oxirane ring (**2** vs **18**) has only a much smaller effect on *E*-values for *Rhodococcus ruber* DSM 44539 (**G**, $E > 200$ for both **2** and **18**), *Rhodococcus ruber* DSM 44540 (**F**, $E = 126$ (**2**) and >200 (**18**)), and *Rhodococcus ruber* DSM 44541 (**C**, $E = 113$ (**2**) and >200 (**18**)) than for *Rhodococcus ruber* DSM 43338 (**B**, $E = 8.9$ (**2**) and >200 (**18**)). With respect to the influence of the position of the double bond, among the C8 derivatives, the oct-6-enyl group generally leads to the highest *E*-values (see, for instance **7** vs **9** and **10** vs **12**), the only exception being *Rhodococcus equi* IFO 3730 (**D**). As can be seen from the data presented in Table 2, all these effects are quite satisfactorily modeled by both the MLR and ANN equations.

However, our results also indicate that quite subtle changes of epoxide structure can—but not necessarily need—have rather large effects on enantioselectivities obtainable from several of the enzymes used. This is even more evident for the enantioselectivities predicted for the test molecules **25**−**28** which are also summarized in Table 4. First, in case of *Rhodococcus ruber* DSM 44540 (**F**), ClogP-values for substrates **26**−**28** were outside the range of the training set. Consequently, predictions for these molecules are not possible. Otherwise, with the exception of *Rhodococcus ruber* DSM 44539 (**G**), log *E*-values predicted for **26** are in excellent agreement with experimental data. The less reliable prediction in case of **G** can be attributed to the lack of a training structure of type **26** in the series of substrates of *Rhodococcus ruber* DSM 44539. Obviously, the effect of replacing a carbon atom within a **saturated** chain is not properly described by the training set and, consequently predictions for **27** show rather large errors. We also note that the training set for *Mycobacterium paraffinicum* NCIMB 10420 (**E**) is the only one containing a molecule of type **28**. Not surprisingly then, for this epoxide hydrolase the predicted log *E*-value of compound **28** is in very good agreement with the experimental one whereas for the other enzymes less satisfactory results are obtained.

CONCLUSIONS

We have presented a quantitative structure−activity relationship study concerning the enantioselectivity (log *E*) observed for the enzymatic ring-opening reactions of epoxides for a series of epoxide hydrolases from seven different organisms. To the best of our knowledge, besides QSAR studies on the inhibition of soluble epoxide hydrolases by urea-like compounds,[37,38] no such modeling has been done so far. Models for log *E* were derived by multiple linear regression and artificial neural networks. Generally, a statistical meaningful modeling of log *E* values for all hydrolases can be obtained by as few as two or three descriptors. Notably, the enantioselectivity of the various hydrolases is modeled by different descriptors, indicating different steric and/or electronic requirements for the enantioselectivity in substrate binding.

Supporting Information Available: Correlation coefficients between log *E*-values (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Thomas, H.; Oesch, F. Functions of epoxide hydrolases. *ISI Atlas Sci.: Biochem.* **1988**, *1*, 287−291.

(2) (a) For the asymmetric chemo-hydrolysis: Ready, J. M.; Jacobsen, E. N. A practical oligomeric [(salen)Co] catalyst for asymmetric epoxide ring-opening reactions. *Angew. Chem., Int. Ed. Engl.* **2002**, *41*, 1374−1377. (b) Schaus, S. E.; Brandes, B. D.; Larrow, J. F.; Tokunaga, M.; Hansen, K. B.; Gould, A. E.; Furrow, M. E.; Jacobsen, E. N. Highly Selective Hydrolytic Kinetic Resolution of Terminal Epoxides Catalyzed by Chiral (salen)CoIII Complexes. Practical Synthesis of Enantioenriched Terminal Epoxides and 1,2-Diols. *J. Am. Chem. Soc.* **2002**, *124*, 1307−1315. (c) For the biohydrolysis see: Faber, K.; Orru, R. V. A. In *Enzyme Catalysis in Organic Synthesis*, 2nd ed.; Drauz, K., Waldmann, H., Eds.; Wiley-VCH: Weinheim, 2002; Vol. II, pp 579−608.

(3) Archelas, A. Fungal epoxide hydrolases: new tools for the synthesis of enantiopure epoxides and diols. *J. Mol. Catal. B* **1998**, *5*, 79−85.

(4) Weijers, C. A. G. M.; de Bont, J. A. M. Epoxide hydrolases from yeasts and other sources: versatile tools in biocatalysis. *J. Mol. Catal. B* **1999**, *6*, 199−214.

(5) Steinreiber, A.; Faber, K. Microbial epoxide hydrolases for preparative biotransformations. *Curr. Opin. Biotechnol.* **2001**, *12*, 552−558.

(6) (a) Wandel, U.; Mischitz, M.; Kroutil, W.; Faber, K. Highly selective asymmetric hydrolysis of 2,2-disubstituted epoxides using lyophilized cells of *Rhodococcus sp.* NCIMB 11216. *J. Chem. Soc., Perkin Trans. 1* **1995**, 735−736. (b) Mischitz, M.; Kroutil, W.; Wandel, U.; Faber, K. Asymmetric microbial hydrolysis of epoxides. *Tetrahedron: Asymmetry* **1995**, *6*, 1261−1272.

(7) (a) Kroutil, W.; Osprian, I.; Mischitz, M.; Faber, K. Chemoenzymic synthesis of (S)−(−)-frontalin using bacterial epoxide hydrolases. *Synthesis* **1997**, 156−158. (b) Osprian, I.; Kroutil, W.; Mischitz, M.; Faber, K. Biocatalytic resolution of 2-methyl-2-(aryl)alkyloxiranes using novel bacterial epoxide hydrolases. *Tetrahedron: Asymmetry* **1997**, *8*, 65−71.

(8) Orru, R. V. A.; Mayer, S. F.; Kroutil, W.; Faber, K. Chemoenzymic deracemization of (±)-2,2-disubstituted oxiranes. *Tetrahedron* **1998**, *54*, 859−874.

(9) Krenn, W.; Osprian, I.; Kroutil, W.; Braunegg, G.; Faber, K. Bacterial epoxide hydrolases of opposite enantiopreference. *Biotechnol. Lett.* **1999**, *21*, 687−690.

(10) Steinreiber, A.; Osprian, I.; Mayer, S. F.; Orru, R. V. A.; Faber, K. Enantioselective hydrolysis of functionalized 2,2-disubstituted oxiranes with bacterial epoxide hydrolases. *Eur. J. Org. Chem.* **2000**, 3703−3711.

(11) Steinreiber, A.; Hellström, H.; Mayer, S. F.; Orru, R. V. A.; Faber, K. Chemo-enzymatic enantioconvergent synthesis of C4-building blocks containing a fully substituted chiral carbon center using bacterial epoxide hydrolases. *Synlett* **2001**, 111−113.

(12) Hellström, H.; Steinreiber, A.; Mayer, S. F.; Faber, K. Bacterial epoxide hydrolase-catalyzed resolution of a 2,2-disubstituted oxirane: optimization and upscaling. *Biotechnol. Lett.* **2001**, *23*, 169−173.

(13) Osprian, I.; Stampfer, W.; Faber, K. Selectivity enhancement of epoxide hydrolase catalyzed resolution of 2,2-disubstituted oxiranes by substrate modification. *J. Chem. Soc., Perkin Trans. 1* **2000**, 3779−3785.

(14) (a) Hult, K.; Faber, K. In *Encyclopedia of Catalysis*; Horvath, I. T., Ed.; Wiley: New York, 2002; in press. (b) For a review, see: Orru, R. V. A.; Archelas, A.; Furstoss, R.; Faber, K. Epoxide hydrolases and their synthetic applications. *Adv. Biochem. Eng. Biotechnol.* **1999**, *63*, 145−167.

(15) Chen, C. S.; Fujimoto, Y.; Girdaukas, G.; Sih, C. J. Quantitative Analysis of Biochemical Kinetic Resolutions of Enantiomers. *J. Am. Chem. Soc.* **1982**, *104*, 7294−7299.

(16) (a) Fersht A. Enzyme−Substrate Complementarity and the use of Binding Energy in Catalysis and Specifity and Editing Mechanisms. In *Enzyme Structure and Mechanisms*; W. H. Freeman & Co.: New York, 1977; pp 98−367. (b) Pieters, R. J.; Lutje Spelberg, J. H.; Kellogg, R. M.; Janssen, D. B. The enantioselectivity of haloalkane dehalogenases. *Tetrahedron Lett.* **2001**, *42*, 469−471. (c) van der Lugt, J. P.; Elfrink, H.; Evenaar, J.; Doddema, H. J. Kinetics and validation of mathematical models of enantioselective transesterification in organic solvents. In *Microbial reagents in organic synthesis*; Servi, S., Ed.; Nato ASI Series C; Kluwer Acad. Publ.: Dordrecht, 1992; Vol. 381, pp 261−272. (d) Kazlauskas, R. J. Molecular modeling and biocatalysis: explanations, predictions, limitations, and opportunities. *Curr. Opin. Chem. Biol.* **2000**, *4*, 81−88. (e) Utsumi, R.; Izumi, S.; Hirata, T. Enantioselectivity in the hydrolysis by lipase: a study of MALDI TOF-MS analysis. *Chem. Lett.* **2001**, 892−893.

(17) For instance, epoxide hydrolases from fungi, e.g. from *Beauveria bassiana* and *Aspergillus niger* have been found to be most selective on styrene-oxide-type substrates.[39] Due to the presence of a benzylic carbon atom, however, mixed regioselectivities have been observed (i.e. the biohydrolysis proceeds with *retention* and/or *inversion* of configuration) and, as a consequence, the use of *E*-values for the description of the enantioselectivity is inappropriate in these cases.[40] Thus, these data could not be incorporated into this study.

(18) Faber, K.; Griengl, H.; Hönig, H.; Zuegg, J. On the prediction of the enantioselectivity of *Candida rugosa* lipase by comparative molecular field analysis. *Biocatalysis* **1994**, *9*, 227−239.

(19) Sybyl 6.8, Tripos Inc., St. Louis, MO.

(20) Stewart, J. J. P. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aid. Mol. Des.* **1990**, *4*, 1−105.

(21) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(22) Baker, J. An algorithm for the location of transition states. *J. Comput. Chem.* **1986**, *7*, 385−395.

(23) DRAGON v1.1: Chemometrics and QSAR Research Group, Dipartimento di Scienze dell'Ambiente e del Territorio, Università degli Studi di Milano-Bicocca, Italy; available from http://www.disat.unimib.it/chum/Dragon.html.

(24) Di Marzio, W.; Galassi, S.; Todeschini, R.; Consolaro, F. Traditional versus WHIM molecular descriptors in QSAR approaches applied to fish toxicity studies. *Chemosphere* **2001**, *44*, 401−406.

(25) Stanton, D. T.; Jurs, P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure−property relationship studies. *Anal. Chem.* **1990**, *62*, 2323−2329.

(26) Connolly, M. L. The molecular surface package. *J. Mol. Graphics* **1993**, *11*, 139−141.

(27) *Chem3D Ultra* v 6.0; CambridgeSoft. Com, Cambridge, MA.

(28) Wold, S.; Dunn, W. J. Multivariate quantitative structure−activity relationships (QSAR): conditions for their applicability. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 6−13.

(29) *STATISTICA for Windows*, v. 5.5; StatSoft, Inc.: Tulsa, OK, 1995.

(30) Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Cross-Validation, Bootstrapping, and Partial Least-Squares Compared with Multiple-Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18−25.

(31) Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design: An Introduction*, 2nd ed.; Wiley-VCH: Weinheim, 1999.

(32) Suzuki, T.; Ebert, R.-U.; Schüürmann, G. Development of Both Linear and Nonlinear Methods To Predict the Liquid Viscosity at 20 °C of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1122−1128.

(33) Dietrich, W. S.; Dreyer, N. D.; Hansch, C. Confidence interval estimators for parameters associated with quantitative structure−activity relationships. *J. Med. Chem.* **1980**, *23*, 1201−1205.

(34) Cornish-Bowden, A.; Wong, J. T. Evaluation of rate constants for enzyme-catalyzed reactions by the jackknife technique. Application to liver alcohol dehydrogenase. *Biochem. J.* **1978**, *175*, 969−976.

(35) Andrea, T. A.; Kalayeh, H. Applications of neural networks in quantitative structure−activity relationships of dihydrofolate reductase inhibitors. *J. Med. Chem.* **1991**, *34*, 2824−2836.

(36) Reynolds, C. H., Holloway, M. K., Cox, H. K., Eds.; *Computer-Aided Molecular Design: Applications in Agrochemicals, Materials, and Pharmaceuticals*; ACS Symposium Series 589; American Chemical Society: Washington, DC, 1995; pp 64−81.

(37) McElroy, N. R.; Jurs, P. C. QSAR of Inhibition of Soluble Epoxide Hydrolases by Urea-like Compounds. *Abstr. Pap. Am. Chem. Soc.* **2001**, *222*, 144.

(38) Nakagawa, Y.; Wheelock, C. E.; Morisseau, C.; Goodrow, M. H.; Hammock, B. G.; Hammock, B. D. 3-D QSAR Analysis of Inhibition of Murine Soluble Epoxide Hydrolase (Mseh) by Benzoylureas, Arylureas, and Their Analogues. *Bioorg. Med. Chem.* **2000**, *8*, 2663−2673; Erratum: *Bioorg. Med. Chem.* **2001**, *9*, 1941.

(39) Pedragosa-Moreau, S.; Archelas, A.; Furstoss, R. Microbiological transformations-XXIX. Enantioselective hydrolysis of epoxides using microorganisms: a mechanistic study. *Bioorg. Med. Chem.* **1994**, *2*, 609−616.

(40) Lu, A. Y. H.; Miwa, G. T. Molecular properties and biological functions of microsomal epoxide hydrase. *Annu. Rev. Pharmacol. Toxicol.* **1980**, *20*, 513−531.

CI020047Z