J Math Chem (2009) 45:287–294
DOI 10.1007/s10910-008-9405-4

ORIGINAL PAPER

QSAR modeling of human catechol *O*-methyltransferase enzyme kinetics

Adina Costescu · Cristina D. Moldovan · Gabriel Katona · Mircea V. Diudea

Published online: 7 August 2008

Abstract QSAR models based on two different similarity methods were developed to predict two enzyme kinetic parameters K_m, and V_{max} for catecholic substrates of human soluble catechol *O*-methyl-transferase (S-COMT). The similarity of the data set is measured using: (a) the 3D molecular structure, by TOPOCLUJ software, and (b) classical topological descriptors, by SIMIL program. The measured properties of 45 substrates were correlated with the topological and electronic parameters of the graphs associated to molecular structures. Clustering methods based on similarity descriptors succeeded a better prediction than the method using 3D structure similarity. All models were statistically significant and showed good stability to data variation in leave-one-out (LOO) cross-validation experiment.

Keywords QSAR · Topological descriptors · Catechol *O*-methyltransferase enzyme kinetics

1 Introduction

The catechol group, two vicinal hydroxyls in an aromatic ring, is present in the structure of catecholamine neurotransmitters and hormones, and many of their metabolites, and is also found in metabolites of phenolic hormones such as estrogens. The group is an important pharmacophore, present in drugs with varying pharmacological action, e.g., the dopamine agonist apomorphine, the β_1-receptor agonist dobutamine, the antihypertensive α-methyldopa, the dopadecarboxylase inhibitor carbidopa, and the catechol *O*-methyltransferase inhibitor entacapone. Catechol groups are also present in numerous dietary compounds (coumarins, flavonoids, anthocyanins, etc.) [1].

A. Costescu (✉) · C. D. Moldovan · G. Katona · M. V. Diudea
Faculty of Chemistry and Chemical Engineering, Babes-Bolyai University, Cluj-Napoca, Romania
e-mail: acostescu@chem.ubbcluj.ro

COMT is an enzyme that catalyzes the transfer of a methyl group from a methyl donor to one hydroxyl moiety of the catechol ring of a substrate. Cell fractionation and immunological studies have shown that the enzyme occurs in two distinct forms, in the cytoplasm as a soluble protein (S-COMT) and in association with membranes as a membrane-bound [2].

S-COMT was shown to have a high K_m value for dopamine but a very high capacity (V_{max} from 50 pmol/min × mg protein in skeletal muscle to values as high as 14, 690 pmol/min × mg protein in the liver). The V_{max} values are strongly dependent on the enzyme activities in various tissues rather than on the basic kinetic constants of these enzymes [3].

There are a few published two-dimensional QSAR studies of S-COMT inhibition [4–6]. In this study we present a predictive three-dimensional QSAR models for the enzyme kinetic parameters K_m and V_{max}, on a data set of 45 different catecholic substrates.

2 Data set

A data set of 45 structures was taken from the publications by Julius Sipilä and Jyrki Taskinen [7].

Catechol target structure herein investigated is shown in Fig. 1 while the description of each molecule with the kinetic parameters K_m, and V_{max} is given in Table 1.

3 Data analysis

The QSAR analysis consists of the following steps: (1) structure optimization using a semiempirical method; (2) calculation of molecular descriptors; (3) variance analysis; (4) principal component analysis (*PCA*); (5) splitting the data set by using TOPOCLUJ [8] and SIMIL [9] similarity software packages; (6) finding a regression function for the model; (7) validation by leave-one-out procedure; (8) interpretation of the model.

All the 45 catechols were modeled and optimized by using the semiempirical AM1 method available in HyperChem [10]. Topological and electronic indices were calculated using the TOPOCLUJ software.

Fig. 1 Catechol target structure

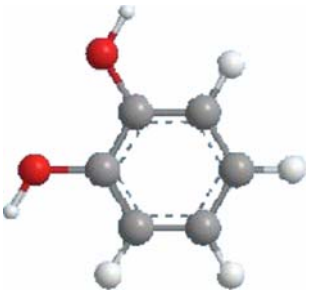


Table 1 Catechol derivatives, kinetic parameters and the set clustering[a]

	Compound	$\log K_m^{-1}$	$\log V_{max}$	SIMIL clustering	TOPOCLUJ clustering
1	2,3-Dihydroxybenzoic acid	3.42	–	0	1
2	2,3-Dihydroxynaphthalene	5.82	1.48	1	1
3	2-Hydroxyestradiol	5.43	1.66	1	1
4	3,4-Dihydroxyacetophenone	5.85	0.99	1	1
5	3,4-Dihydroxybenzoic acid	4.70	1.64	1	1
6	3,4-Dihydroxymandelic acid	3.88	1.64	1	0
7	3,4-Dihydroxyphenylacetic acid	4.14	1.58	0	0
8	3,4-Dihydroxyphenylglycol	4.33	1.57	1	1
9	3-Fluorocatechol	5.11	1.49	1	1
10	3-Methoxy-5-bromocatechol	5.59	1.24	1	1
11	3-Methoxycatechol	4.61	1.60	1	1
12	3-Nitrocatechol	7.70	−0.82	1	1
13	4-Chlorocatechol	4.93	1.56	1	1
14	4-Hydroxyestradiol	4.91	1.50	0	0
15	4-Isopropylcatechol	4.58	1.64	1	1
16	4-Methoxy-5-tert-butylcatechol	4.73	1.52	0	1
17	4-Methylcatechol	4.54	1.68	1	1
18	4-Nitrocatechol	6.92	0.19	1	1
19	4-Tert-butylcatechol	4.50	1.62	1	1
20	5-Hydroxydopamine	4.21	1.67	1	1
21	6,7-Dihydroxycoumarin	6.52	0.52	1	1
22	6-Hydroxydopamine	3.61	1.40	1	1
23	Adrenaline	3.88	1.64	1	1
24	Amedopa	2.75	1.38	0	1
25	Apomorphine	3.62	–	0	0
26	Benserazide	4.3	1.62	0	1
27	Caffeic acid	5.49	1.62	1	1
28	Carbidopa	3.22	1.16	0	1
29	Catechol	4.30	1.69	1	1
30	Dihydrexidine	4.60	1.31	1	0
31	Dobutamine	4.64	1.56	0	0
32	Dopamine	3.73	1.56	1	1
33	Entacapone	9.50	−1.12	1	0
34	Ethyl-3,4-dihydroxybenzoate	6.18	1.03	1	1
35	Hydrocaffeic acid	4.60	1.64	1	0
36	Isoprenaline	3.84	1.46	0	1
37	Levodopa	3.25	1.47	0	1
38	Levodopa metylester	4.29	1.65	0	0
39	Methylgallate	6.00	1.07	1	1
40	Noradrenaline	3.59	1.52	1	1
41	Pyrogallol	4.98	1.73	1	1
42	Salsolinol	4.53	1.46	0	0
43	skf38393	4.14	1.19	0	1
44	Tetrachlorocatechol	7.52	−0.21	1	1

[a] Structures with 1 show the similarity index >0.8

TOPOCLUJ program was used to develop 939 topological and electronic descriptors. Some of this descriptors, with the variance values 0 or $<10^{-6}$ were eliminated. The topological descriptors are derived from topological matrices or polynomials by

Table 2 Statistics for the modeling of log K_m^{-1} and log V_{max} kinetic parameter, in the more similar subset (by SIMIL clustering method)

Statistics for log K_m^{-1}	*b*	*s*	*t*(21)	*p*-Level
Intercept	3.7309	0.412368	9.04760	1.08E−08
Lumo	−1.5521	0.186119	−8.33915	4.21E−08
HF(kcal/mol)	0.0104	0.001663	6.24158	3.43E−06
X[Sh[CjMax[covalent radius]]]	4.7320	1.646251	2.87440	9.08E−03
PDS3[LM[mass]]	0.0012	0.000281	4.27012	3.41E−04
IE[CfMax[charge]]	−28.6612	5.884384	−4.87072	8.14E−05
Charges	−2.6943	0.708999	−3.80013	1.05E−03
Statistics for log V_{max}	*b*	*s*	*t*(22)	*p*-Level
Intercept	1.839804	0.053629	34.30614	1.35E−20
IP[CfMin[charge]]	4.919851	0.444612	11.06550	1.86E−10
PDS4[LM[charge]]	−0.335913	0.036058	−9.31593	4.31E−09
X[Sh[CjMax[atomic radius]]]	1.149936	0.949175	1.21151	2.39E−01

using some weighting schemes for the molecular graphs, with atomic partial charges, electronegativities and molecular fragments weights.

All possible combinations of the remaining descriptor pairs were examined, taking a threshold of 0.95 for intercorrelation: above this value, half of them have been discarded. The decision of which one to retain was based on the possible physical interpretation of the descriptor, ease of calculation, or usefulness in the past studies. The result of this analysis is a reduced pool of information-rich descriptors which can then be screened by using multiple linear regression analysis.

Principal component analysis (PCA) was used for data reduction. Loading factors from the first five PC were used to evaluate the relevant descriptors.

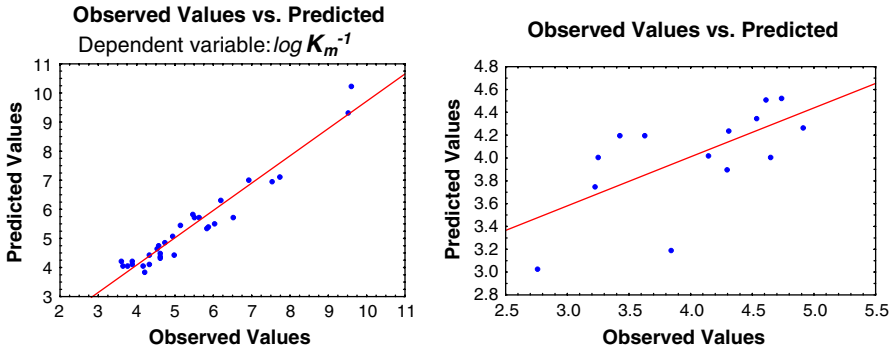
A high loading value indicates that the PC is aligned in a direction close to the main descriptor response. Selection of descriptors, contributing highly to the data variance, can be accomplished by examination of loading plots or tables. Additionally, the relation of the descriptors to each other (i.e., co-linearity) can be explored.

Clustering of the data set was performed using similarity coefficient values evaluated by two different software packages: (a) TOPOCLUJ software, which operates on the 3D molecular structure, and (b) SIMIL program, which considers the classical topological descriptors.

All the 45 structures were arranged in decreasing order of similarity coefficient with respect to the catechol target structure. The set was split in two subsets, one including molecules with similarity coefficient values >0.8 and the remaining molecules were inserted in the second subset, to justify the importance of the clustering by similarity in the data set (Table 1).

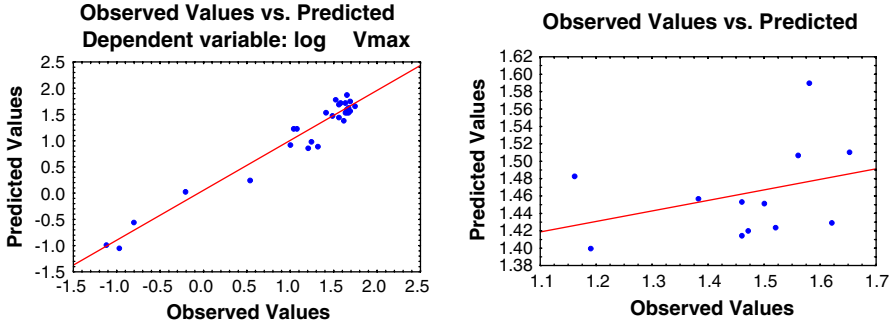
For our similarity study by SIMIL we choose the union of two topological descriptors: a theoretical one (the Randić index [11,12]) and another one based on electronegativities (DSI [13,14]). In case of SIMIL clustering, we classified 31 similar structures. Descriptors and kinetic parameters for the discussed set of structures, used in multi linear regression MLR, are presented in Table 2.

The quality of the models was estimated by: squared correlation coefficient (R^2), standard error of estimate (s), Fischer test (F) probability of error (p) and cross-



(a) n = 31	**(b)** n = 14
R^2 = 0.938, s =0.30 , F =445.158 , p =1.11E-16 R^2_{CV} = 0.902	R^2 = 0.422

Fig. 2 Plots of observed versus predicted values of log K_m^{-1} kinetic parameter. (SIMIL clustering method); (**a**) similarity subset with similarity coefficient values >0.8; (**b**) the remaining molecules



(a) n = 30	**(b)** n = 12
R^2 = 0.951, s =0.13 , F = 553.91.76 , p =1.0E-3 R^2_{CV} = 0.931	R^2 = 0.12

Fig. 3 Plots of observed versus predicted values of log V_{max} kinetic parameter. (SIMIL clustering method); (**a**) similarity subset with similarity coefficient values >0.8; (**b**) the remaining molecules

validated correlation coefficient (R^2_{CV}). In Table 2, b is the regression coefficient for each descriptor.

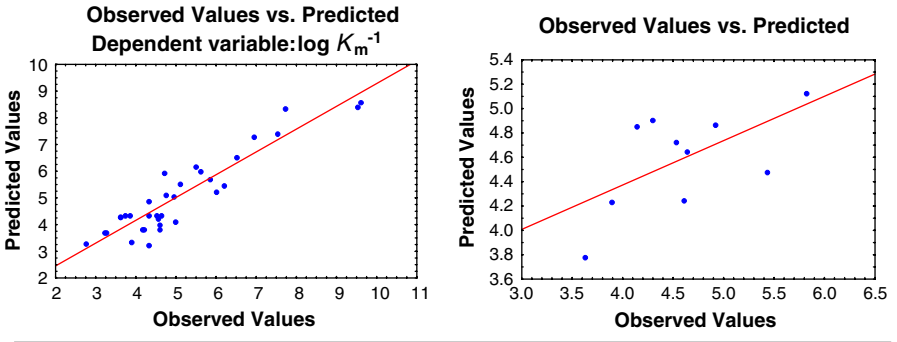
The predictive ability of the model still needs to be tested using a statistical procedure, called cross-validation (CV). In this paper, a leave-one-out (LOO) method, one of many CV techniques, is employed. Each compound is systematically excluded from the data set and its property predicted by a model that is derived from the remaining compounds.

The plots of observed vs. predicted values of kinetic parameters are represented in Figs. 2 and 3. Predictability and stability of the obtained models are carried out by means of LOO cross-validation.

Modeling, by similarity clustering and topological descriptors, of kinetic parameters log K_m^{-1} and log V_{max} for similar catecholic substrates transformed by human soluble

Table 3 Statistics for the modeling of log K_m^{-1} and log V_{max} kinetic parameter, in the more similar subset (by TOPOCLUJ clustering method)

Statistics for log K_m^{-1}	*b*	*s*	*t*(25)	*p*-Level
Intercept	4.86078	0.354889	13.69662	4.01E−13
Lumo	−1.70686	0.230241	−7.41338	9.15E−08
HF(kcal/mol)	0.00436	0.002771	1.57446	1.28E−01
PDS4[LM[charge]]	0.4723	0.107801	4.36176	1.85E−04
Statistics for log V_{max}	*b*	*s*	*t*(22)	*p*-Level
Intercept	1.428709	0.162010	8.81862	7.75E−09
IP[CfMax[charge]]	−0.994569	1.257938	−0.79063	4.37E−01
IP[CfMin[charge]]	5.200204	0.929865	5.59243	1.09E−05
PDS4[Sh[CfMax[charge]]]	−0.139491	0.026811	−5.20283	2.82E−05
HF/N	−0.079058	0.027714	−2.85260	9.01E−03



(a) n = 34 | **(b)** n = 10

$R^2 = 0.853$, $s = 0.4$, $F = 185.58$, $p = 9.15E\text{-}13$
$R^2_{CV} = 0.820$

$R^2 = 0.36$

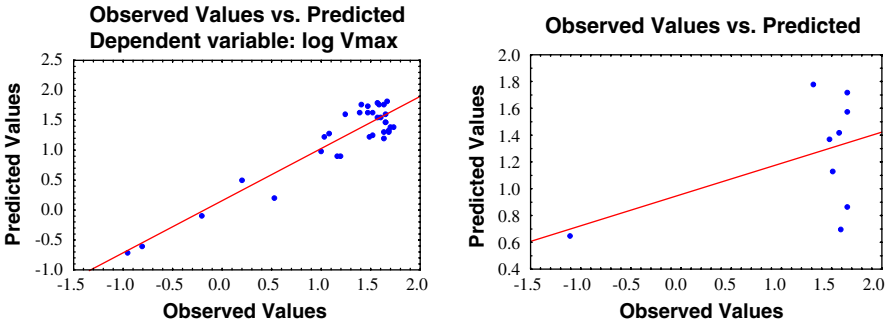
Fig. 4 Plots of observed versus predicted values of log K_m^{-1} kinetic parameter (TOPOCLUJ clustering method): (**a**) similarity subsets with similarity coefficient values >0.8: (**b**) remaining molecules

catechol *O*-methyltransferase (S-COMT), show good quality. Six descriptors explain more than 93% and predict more than 90% of the variance, with a low standard error ($s = 0.3$) for log K_m^{-1} kinetic parameter. Even better result was obtained for log V_{max} kinetic parameter. Three descriptors describes more than 95% and predict more than 93% of the variance with a lower standard error of estimate ($s = 13\%$).

The less similar structure subsets (with the similarity coefficient values <0.8, Figs. 2b and 3b) show very low regression coefficients, which demonstrate the importance of similarity clustering in the data sets.

In case of clustering by TOPOCLUJ program, we classified 34 structures with similarity coefficients >0.8. Data are presented in Table 3.

The plots of observed vs. predicted values of kinetic parameters are represented in Figs. 4 and 5.



(a) n = 33	**(b)** n = 9
$R^2 = 0.871$, s =0.23, F = 183.009 , p = 2.81E-13 $R^2_{CV} = 0.845$	$R^2 = 0.232$

Fig. 5 Plots of observed versus predicted values of log V_{max} kinetic parameter (TOPOCLUJ clustering method): (**a**) similarity subsets with similarity coefficient values >0.8: (**b**) remaining molecules

Modeling log K_m^{-1} and log V_{max} kinetic parameters, by using TOPOCLUJ similarity as a clustering method, showed good prediction ability (85% and 87%, respectively, Figs. 4a and 5a).

4 Conclusions

The most important aspect considered in the present QSAR study was the clustering data set by similarity as measured by SIMIL and TOPOCLUJ methods. Similarity plays an important role in making a set of structures congeneric-like.

The modeling of two enzyme kinetic parameters K_m, and V_{max} for catecholic substrates of human soluble catechol *O*-methyltransferase (S-COMT) by the clustering based on similarity of descriptors (the SIMIL method) showed a better prediction ability than that using the clustering based on 3D optimized geometries (TOPOCLUJ method). In the SIMIL approach, the model explains more than 95% and predicts 93.4% of the variance in S-COMT enzyme response. The results are comparable to others already published in the literature.

References

1. J. Taskinen, B.T. Ethell, P. Pihlavisto, A.M. Hood, B. Burchell, M.W.H. Coughtrie, Drug Metab. Dispos. **31**, 1187 (2003)
2. S. Dawling, N. Roodi, R.L. Mernaugh, X. Wang, F.F. Parl, Cancer Res. **61**, 6716 (2001)
3. P.T. Männistö, S. Kaakkola, Pharmacol. Rev. **51**, 593 (1999)
4. J. Taskinen, J. Vidgren, M. Ovaska, R. Backstrom, A. Pippuri, E. Nissinen, Quant. Struct.-Act. Relat. **8**, 210 (1989)
5. T. Lotta, J. Taskinen, R. Backstrom, E. Nissinen, J. Comput.-Aided Mol. Design **6**, 253 (1992)
6. P Lautala, I. Ulmanen, J. Taskinen, Mol. Pharmacol. **59**, 393 (2001)
7. J. Sipilä, J. Taskinen, J. Chem. Inf. Comput. Sci. **44**, 97 (2004)
8. M.V. Diudea, O. Ursu, *TOPOCLUJ* (Copyright Babes-Bolyai Univ. Cluj, 2002)
9. M.V. Diudea, *SIMIL* (Copyright by Topo Group Cluj, 2000).

10. HyperChem [TM], release 4.5 for SGI, © 1991–1995, Hypercube, Inc.
11. M. Randic, C.L. Wilkins, J. Phys. Chem. **83**, 1525 (1979)
12. M.V. Diudea, M. Topan, A. Graovac, J. Quant. Chem. **24**, 435 (1983)
13. M.V. Diudea, O. Ivanciuc, *Ed. Compex* (Cluj-Napoca, 1995)
14. M.V. Diudea, I. Silaghi-Dumitrescu, Rev. Roumainie Chim. **34**, 1175 (1989)