

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/257035293>

CORAL: Predictions of rate constants of hydroxyl radical reaction using representation of the molecular structure...

Article in *Chemometrics and Intelligent Laboratory Systems* · March 2012

DOI: 10.1016/j.chemolab.2011.12.003

CITATIONS

13

READS

84

7 authors, including:



Andrey Toropov

Mario Negri Institute for Pharmacological Res...

262 PUBLICATIONS 3,301 CITATIONS

SEE PROFILE



Alla Toropova

Mario Negri Institute for Pharmacological Res...

174 PUBLICATIONS 1,832 CITATIONS

SEE PROFILE



Giuseppina Carla Gini

Politecnico di Milano

184 PUBLICATIONS 1,674 CITATIONS

SEE PROFILE



Danuta Leszczynska

Jackson State University

129 PUBLICATIONS 2,022 CITATIONS

SEE PROFILE

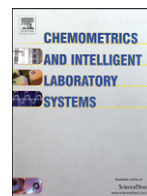
Some of the authors of this publication are also working on these related projects:



Development and use of in silico methods [View project](#)



MOL2NET Conference [View project](#)



Short Communication

CORAL: Predictions of rate constants of hydroxyl radical reaction using representation of the molecular structure obtained by combination of SMILES and Graph approaches

A.A. Toropov^{a,*}, A.P. Toropova^a, S.E. Martyanov^b, E. Benfenati^a, G.Gini^c, D. Leszczynska^d, J. Leszczynski^e^a Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy^b Teleca OOO, 603093, 23, Rodionova st, Nizhny Novgorod, Russia^c Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy^d Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA^e Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA

ARTICLE INFO

Article history:

Received 4 November 2011

Received in revised form 28 November 2011

Accepted 17 December 2011

Available online 24 December 2011

Keywords:

QSAR

Rate Constants of Hydroxyl Radical Reaction

CORAL software

ABSTRACT

New version of CORAL software is now available on the Internet (<http://www.insilico.eu/CORAL/>). Previous versions of this software built up a quantitative structure-property/activity relationships (QSPR/QSAR) based on representation of the molecular structure by the simplified molecular input line entry system (SMILES). The present study has shown that the novel 'hybrid' representation of molecular structure by combination of SMILES and the molecular graph can improve the predictive potential of CORAL models. The evaluation of this approach has been performed for a large ($n = 460$) set of heterogeneous organic compounds. The tested endpoint represent rate constants of hydroxyl radical reaction of polybrominated diphenyl ethers and (benzo-)triazoles.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Quantitative structure-property / activity relationships (QSPR/QSAR) represent an efficient approach to predict unknown parameters of various compounds. Such approaches correlate physicochemical or biochemical parameters with so called molecular descriptors. Such correlations could be used for estimation of analogous parameters (endpoints) for substances which have not been examined in experiment. QSAR methodology has been widely applied in scientific research as well as industrial R&D laboratories [1–10].

CORAL software is designed to provide users an option to construct one-variable QSPR/QSAR models built up by the Monte Carlo method [10–17]. This is combined with the representation of the molecular structure by SMILES [18–21]. However there are various ways to develop one-variable QSPR/QSARs. Such task can be also achieved using molecular graph [22–25]. Recently, the predictability of the graph-based and SMILES-based models has been compared [11].

A logical way to improve predictions is to combine the SMILES and molecular graph approaches to take advantage of the unique features of both methods. The aim of the present study is the evaluation of the hybrid models where the molecular structure is represented by both SMILES and graph. This is carried out for large set of experimental data for the OH radical degradation rate constants of 460 heterogeneous organic compounds [26].

2. Method

2.1. Data

Data of the OH radical degradation rate constants of 460 heterogeneous organic compounds were taken from literature [26]. The data was collected for reactions at 25 °C and 1 atm; all the rate constants, represented in $\text{cm}^3 \text{s}^{-1}$ per molecule, are transformed to logarithmic units and multiplied by -1 (i.e. high value of $-\log(\text{OH})$ means the low value of reactivity). SMILES notations have been taken from the literature [26], except SMILES for cis-1,3-Dichloropropene (CAS 10061-01-5): instead of $[\text{Cl}:5][\text{CH}_2:4][\text{CH}:1]=[\text{CH}:2]/[\text{Cl}:3]$ we have used $\text{C}(\backslash\text{Cl})=\text{C}/\text{Cl}$. Three random splits into the sub-training, calibration, and test sets have been examined. The splits have been done according to the following rules: the first, the number of compounds in the sub-training and calibration sets is approximately equal to the number of compounds in the test set; and second, the ranges of the endpoint are approximately identical for all three sets.

2.2. Optimal descriptor

The CORAL software [27] can generate three kinds of optimal descriptors: graph-based, SMILES-based, and hybrid descriptors which are calculated with both graph and SMILES. In addition, using CORAL program one can generate three kinds of molecular graphs: the above-mentioned HSG, hydrogen filled graph (HFG), and graph of atomic orbitals (GAO).

* Corresponding author.

E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

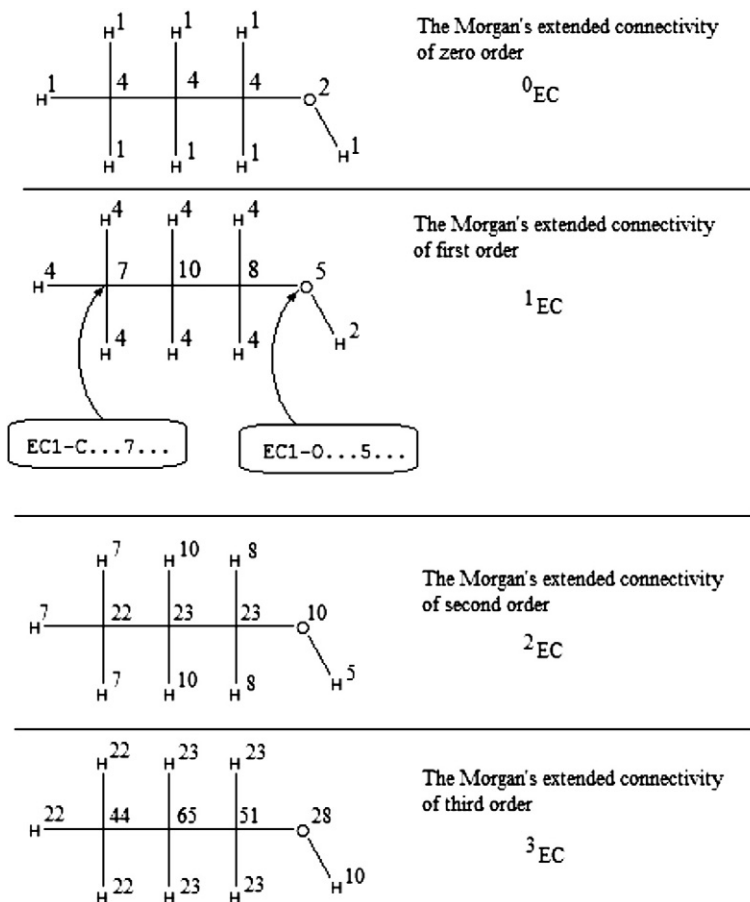


Fig. 1. Calculation of the Morgan's extended connectivity with the recurrence formula: ${}^{m+1}EC_k = \sum_{(kj)}^m EC_j$, (kj) is edge.

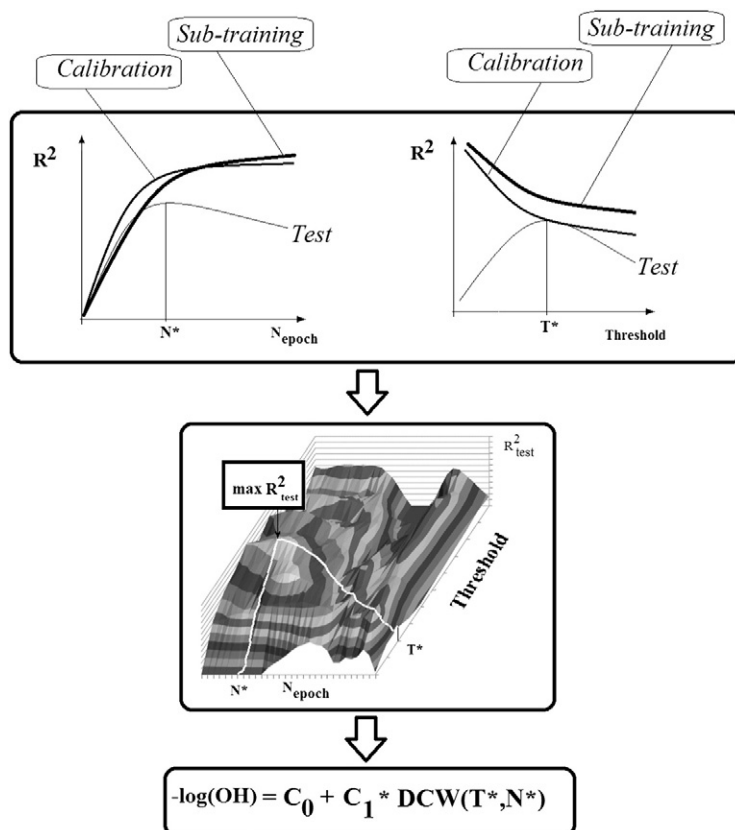


Fig. 2. Preferable values of the threshold (T^*) and of the number of epochs of the Monte Carlo method optimization (N^*) which give best model for the test set.

The graph-based optimal descriptors are calculated as the following:

$$\begin{aligned} \text{Graph DCW(Threshold, } N_{\text{epoch}}) = & \sum CW(A_k) + \alpha \sum CW({}^0\text{EC}_k) \\ & + \beta \sum CW({}^1\text{EC}_k) + \gamma \sum CW({}^2\text{EC}_k) \\ & + \delta \sum CW({}^3\text{EC}_k) \end{aligned} \quad (1)$$

where A_k is chemical element, such as, C, N, O, etc., for HSG and HFG; or atomic orbitals, such as, $1s^1$, $2p^3$, $3d^{10}$, etc. for GAO; ${}^0\text{EC}_k$, ${}^1\text{EC}_k$, ... ${}^3\text{EC}_k$ is the hierarchy of the Morgan's extended connectivity (Fig. 1) [25]; α , β , γ , and δ can be 1 or 0: the combination of their values gives possibility to define various versions of the graph-based optimal descriptor; $CW(x)$ represents the correlation weight of a molecular feature (encoded by A_k or ${}^x\text{EC}_k$).

The SMILES-based optimal descriptors are calculated as the following:

$$\begin{aligned} \text{SMILES DCW(Threshold, } N_{\text{epoch}}) = & \alpha \sum CW(S_k) + \beta \sum CW(SS_k) \\ & + \gamma \sum CW(SSS_k) + x \cdot CW(\text{NOSP}) \\ & + y \cdot CW(\text{HALO}) + z \cdot CW(\text{BOND}) \\ & + \lambda \cdot CW(\text{ATOMPAIR}) \end{aligned} \quad (2)$$

where S_k , SS_k , and SSS_k are one-, two-, and three-components SMILES attributes, respectively; the component of SMILES is one symbol (e.g. C, c, N, n, =, ., etc.) or two symbols which cannot be separated (e.g. Cl,

Br, @@, etc.); NOSP, HALO, BOND and ATOMP AIR represent indices [11] calculated according to the presence or absence of chemical elements: nitrogen, oxygen, sulphur, and phosphorus (NOSP); fluorine, chlorine, and bromine (HALO). The BOND is a mathematical function of the presence or absence of double (=), triple (#), or stereo chemical bonds (@ or @@). The ATOMP AIR is a mathematical function of presence of seven chemical elements: F, Cl, Br, N, O, S, and P. The coefficients α , β , γ , x , y , and z can be 1 or 0: combinations of their values provide an option to define various versions of the SMILES-based optimal descriptor. $CW(x)$ is the correlation weight of a molecular feature (encoded by S_k , SS_k , SSS_k , NOSP, HALO, BOND, ATOMP AIR, and ${}^x\text{EC}_k$).

The hybrid optimal descriptors are calculated taking into account both representations of the molecular structure by graph and by SMILES:

$$\begin{aligned} \text{Hybrid D CW(Threshold, } N_{\text{epoch}}) = & \text{SMILES DCW(Threshold, } N_{\text{epoch}}) \\ & + \text{Graph DCW(Threshold, } N_{\text{epoch}}) \end{aligned} \quad (3)$$

Threshold and N_{epoch} (in Eqs. (1)–(3)) are parameters of the Monte Carlo optimization. Threshold is the criterion for the classification of components of the representation of the molecular structure into two classes: rare and active (not rare). The correlation weight of a rare component is fixed as zero, because this component brings

Table 1

Correlation coefficients between experimental and calculated $-\log(\text{OH})$ values for external test set. Best models are indicated by bold.

Split	Version*	Threshold	Run 1	Run 2	Run 3	Average	
A	1	0	0.6805	0.6873	0.6876	0.6851	
		1	0.6942	0.6889	0.6896	0.6909	
		2	0.6791	0.6845	0.6838	0.6825	
		3	0.6628	0.6663	0.6663	0.6651	
		2	0	0.7226	0.7195	0.7222	0.7214
			1	0.7169	0.7112	0.7194	0.7158
	3	2	0.6996	0.6866	0.6934	0.6932	
		3	0.6845	0.6875	0.6887	0.6869	
		0	0.7874	0.7900	0.7862	0.7879	
		1	0.7880	0.7804	0.7922	0.7868	
		2	0.7769	0.7759	0.7857	0.7795	
		3	0.7698	0.7639	0.7608	0.7648	
B	1	0	0.6654	0.6621	0.6704	0.6660	
		1	0.6713	0.6611	0.6645	0.6656	
		2	0.6561	0.6589	0.6523	0.6558	
		3	0.6186	0.6147	0.6218	0.6184	
		2	0	0.6938	0.6809	0.6756	0.6834
			1	0.6781	0.6947	0.6862	0.6864
	2		0.6611	0.6581	0.6587	0.6593	
	3	3	0.5538	0.5954	0.5727	0.5739	
		0	0.7822	0.7774	0.7756	0.7784	
		1	0.7855	0.7898	0.7813	0.7855	
		2	0.7414	0.7343	0.7434	0.7397	
		3	0.7046	0.7070	0.6940	0.7019	
3		0.7161	0.7129	0.7225	0.7171		
C	1	1	0.7151	0.7224	0.7179	0.7184	
		2	0.7120	0.7155	0.7132	0.7136	
		3	0.7108	0.7092	0.7061	0.7087	
		2	0	0.7562	0.7501	0.7482	0.7515
			1	0.7507	0.7378	0.7419	0.7435
			2	0.7541	0.7505	0.7521	0.7522
	3	3	0.6918	0.6966	0.6968	0.6951	
		0	0.7860	0.7943	0.7860	0.7888	
		1	0.7855	0.7975	0.7942	0.7924	
		2	0.7922	0.7943	0.7968	0.7944	
		3	0.7906	0.7922	0.7940	0.7923	

Version 1: SMILES-based model;

Version 2: HFG-based model;

Version 3: model is calculated with the representation of the molecular structure by SMILES combined with HFG.

Table 2

Statistical quality and criteria of predictability for three models (Split A, B, and C) calculated with N^* and T^* (Fig. 2).

Split A	$-\log(\text{OH}) = 10.7474(\pm 0.0036) + 0.08580(\pm 0.0002) \cdot \text{HybridDCW}(0,13)$ $n = 119, R^2 = 0.8849, q^2 = 0.8814, s = 0.398, F = 899$ (sub-training set); $n = 110, R^2 = 0.8821, r_{\text{pred}}^2 = 0.8814, s = 0.416$ (calibration set); $n = 231, R^2 = 0.7782, r_{\text{pred}}^2 = 0.7743, s = 0.415$ (test set); Estimation of the predictability: $n = 231$ $R^2 = 0.7782$ $r_0^2 = 0.7780$ $r_0^2 = 0.7079$ $(r^2 - r_0^2) / r^2 = 0.0002$ $(r^2 - r_0^2) / R^2 = 0.0903$ $k = 1.0031$ $k' = 0.9955$ $R_{\text{m}}^2(\text{test}) = 0.7695$
Split B	$-\log(\text{OH}) = 10.9428(\pm 0.0037) + 0.0819(\pm 0.0003) \cdot \text{HybridDCW}(1,12)$ $n = 119, R^2 = 0.8622, q^2 = 0.8561, s = 0.434, F = 732$ (sub-training set); $n = 106, R^2 = 0.9145, r_{\text{pred}}^2 = 0.9112, s = 0.354$ (calibration set); $n = 235, R^2 = 0.7820, r_{\text{pred}}^2 = 0.7776, s = 0.432$ (test set); Estimation of the predictability: $n = 235$ $R^2 = 0.7820$ $r_0^2 = 0.7769$ $r_0^2 = 0.6631$ $(R^2 - r_0^2) / R^2 = 0.0065$ $(R^2 - r_0^2) / R^2 = 0.1521$ $k = 1.0074$ $k' = 0.9912$ $R_{\text{m}}^2(\text{test}) = 0.7262$
Split C	$-\log(\text{OH}) = 11.2276(\pm 0.0035) + 0.0949(\pm 0.0004) \cdot \text{HybridDCW}(2,11)$ $n = 129, R^2 = 0.8458, q^2 = 0.8393, s = 0.445, F = 697$ (sub-training set); $n = 100, R^2 = 0.8947, r_{\text{pred}}^2 = 0.8888, s = 0.351$ (calibration set); $n = 231, R^2 = 0.7875, r_{\text{pred}}^2 = 0.7840, s = 0.446$ (test set); Estimation of the predictability: $n = 231$ $R^2 = 0.7875$ $r_0^2 = 0.7870$ $r_0^2 = 0.7428$ $(R^2 - r_0^2) / R^2 = 0.0006$ $(R^2 - r_0^2) / R^2 = 0.0568$ $k = 0.9960$ $k' = 1.0025$ $R_{\text{m}}^2(\text{test}) = 0.7702$

Criteria of predictability: $(R^2 - r_0^2) / R^2 < 0.1$ or $(R^2 - r_0^2) / R^2 < 0.1$; $0.85 < k < 1.15$; and $0.85 < k' < 1.15$ [6]; $R_{\text{m}}^2(\text{test}) > 0.5$ [7].

noise to the model, hence rare component is not involved in the building up of the model. N_{epoch} is the number of epochs of the Monte Carlo optimization.

Fig. 2 shows the theoretical influence of the threshold and of the number of epochs of the Monte Carlo method optimization for the correlation coefficient between the experimental and calculated values of an endpoint.

Three versions of the calculation of optimal descriptors were examined: Version 1 with only SMILES; Version 2 with only HFG; and Version 3 using both SMILES and graph. The parametrization for SMILES-based descriptor was as follows: $\alpha=1$, $\beta=1$, $\gamma=0$, $x=1$, $y=1$, $z=1$, and $\lambda=1$. The parametrization for HFG-based descriptor included: $\alpha=0$, $\beta=1$, $\gamma=0$, and $\delta=0$.

3. Results and discussion

Our study provides the opportunity to test performance of the standard SMILES and graph approaches together with the methods

that combined them both. This is done for a large set of data. Table 1 shows the statistical characteristics of the models obtained with the versions 1, 2 and 3. One can see that version 3 gives best prediction (the statistical characteristics for the test set) for three random splits.

Table 2 contains statistical characteristics for the sub-training, calibration, and test sets for descriptors calculated with version 3 for three random splits. There are few differences that one can find for examined splits in Table 2. However, it is evident that the models for three random splits are quite good. They satisfy the criteria of Golbraikh and Tropsha [6] as well as the criterion of Roy and Roy [7]. Fig. 3 contains the representation of these models graphically.

The statistical characteristics of models for the endpoint $-\log(\text{OH})$ described in Ref [26] are the following: the range of r^2 is 0.82–0.87; the range of s is 0.43–0.49; and the range of $R_m^2(\text{test})$ is 0.75–0.80. Thus, the models which are calculated with Eq. (3) (Table 2) can be considered as a valid alternative for the above-mentioned models of $-\log(\text{OH})$ [26] since their statistical quality is very similar [26]. However the models based on CORAL-descriptor (calculated with Eq. (3)) are based on two-dimensional descriptors, while the aforementioned approach [26] requires three-dimensional optimization and the quantum-mechanics data on chemical structure, i.e., they are much more complex and time-consuming. Hence CORAL-models may be more comfortable for the praxis.

Our approach gives the possibility to detect stable promoters of increase or decrease of the $-\log(\text{OH})$. Table 3 contains examples of the molecular attributes which are stable promoters of increase or decrease for $-\log(\text{OH})$. Thus, the models based on the CORAL-descriptors can give a list of structural alerts related to an endpoint. A compound can contain both promoters of increase and promoters

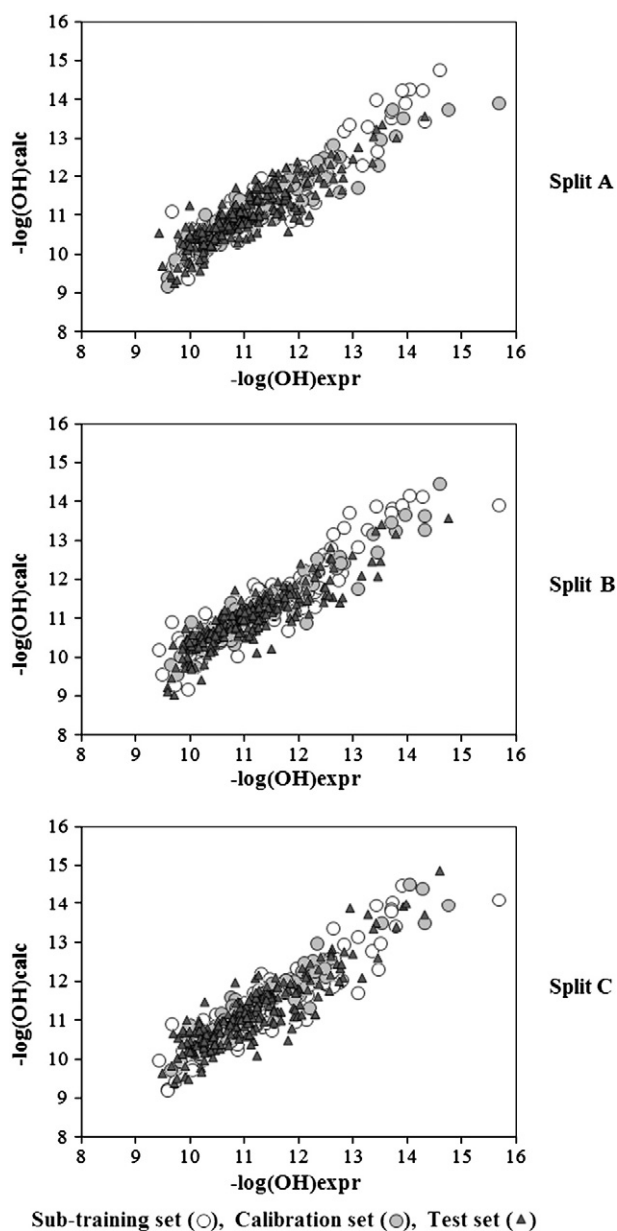


Fig. 3. Graphical representation of the performance of CORAL models for splits A,B, and C (see Table 2).

Table 3

Example of interpretation of the influence of attributes of the molecular structure (taken from SMILES or HFG) which have considerable prevalence in sub-training, calibration, and test sets.

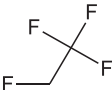
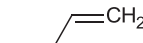
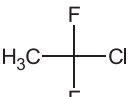
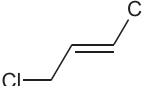
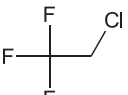
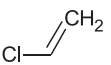
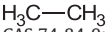
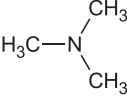
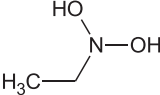
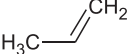
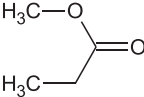
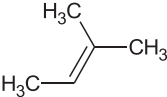
Structural attributes	Description	Promoter ^a of increase or decrease in $-\log(\text{OH})$	Distribution in sub-training set/ calibration set/ test set
EC1-C...7... ^b	Value of the Morgan's extended connectivity of first order (for carbon) that is equal to 7	Increase	81/76/169
EC1-C...10...	Value of the Morgan's extended connectivity of first order (for carbon) that is equal to 10	Increase	58/51/111
EC1-O...3...	Value of the Morgan's extended connectivity of first order (for oxygen) that is equal to 3	Increase	17/17/35
NOSP0000	The absence of nitrogen, oxygen, sulphur, and phosphorus.	Increase	57/58/111
EC1-C...6...	Value of The Morgan's extended connectivity of first order (for carbon) that is equal to 6	Decrease	37/33/72
EC1-C...8...	Value of The Morgan's extended connectivity of first order (for carbon) that is equal to 8	Decrease	35/45/82
HALO0000	The absence of fluorine, chlorine, and bromine.	Decrease	94/87/205
=xxx1xxxxxxxxx	Presence of double bond and cycles	Increase	22/16/47

^a The increase means that three correlation weights which are obtained in three runs of the Monte Carlo optimization are larger than zero; The decrease means that three correlation weights which are obtained in three runs of the Monte Carlo optimization are smaller than zero.

^b Fig. 1 contains clarification of this code.

Table 4

Examples of compounds which contain carbon atoms characterized by Morgan degrees of 6, 7, 8, and 10. The values 7 and 10 are promoters of increase in $-\log(\text{OH})$; values 6 and 8 are promoters of decrease in $-\log(\text{OH})$.

	Presence of EC1-C...7... and EC1-C...10...	Presence of EC1-C...6... and EC-C...8...
Presence of halogens	 <p>CAS 811-97-2; FCC(F)(F)F; $-\log(\text{OH}) = 14.05$</p>	 <p>CAS 107-05-1; CICC=C; $-\log(\text{OH}) = 10.77$</p>
	 <p>CAS 75-68-3; CC(F)(F)Cl; $-\log(\text{OH}) = 14.33$</p>	 <p>CAS 10061-02-6; ClC=C(Cl); $-\log(\text{OH}) = 10.85$</p>
	 <p>CAS 75-88-7; FC(F)(F)CCl; $-\log(\text{OH}) = 13.98$</p>	 <p>CAS 75-01-4; ClC=C; $-\log(\text{OH}) = 11.18$</p>
Absence of halogens	 <p>CAS 74-84-0; CC; $-\log(\text{OH}) = 12.59$</p>	 <p>CAS 75-50-3; CN(C)C; $-\log(\text{OH}) = 10.22$</p>
	 <p>CAS 79-24-3; CCN(O)O; $-\log(\text{OH}) = 12.82$</p>	 <p>CAS 115-07-1; CC=C; $-\log(\text{OH}) = 10.56$</p>
	 <p>CAS 554-12-1; CCC(=O)OC; $-\log(\text{OH}) = 12.57$</p>	 <p>CAS 513-35-9; C(C)=C(C)C; $-\log(\text{OH}) = 10.06$</p>

of decrease of an endpoint. Consequently, there is a combination of different factors which play a role in a given chemical. Table 4 contains examples of compounds which contain some structural attributes from Table 3. One can see from the Table 4, that the presence of halogens together with the EC1-C...7... and EC1-C...10... is an indicator of large values for $-\log(\text{OH})$, i.e. this combination is an indicator of low hydroxyl radical reactivity, and it is in agreement with the

Table 5

The checking of the models with randomization. The \bar{R}_r^2 is average for 10 probes of the Y-scrambling with 100 random permutations. The criterion ${}^cR_p^2 = R\sqrt{R^2 - \bar{R}_r^2}$ should be larger than 0.5 [29].

Probe	Split A	SplitB	Split C
	R_r^2	R_r^2	R_r^2
1	0.1326	0.0326	0.1160
2	0.0839	0.0937	0.0932
3	0.1492	0.1177	0.0330
4	0.0753	0.1175	0.0649
5	0.0471	0.0653	0.1862
6	0.0889	0.1441	0.0963
7	0.2659	0.1220	0.0876
8	0.0522	0.0380	0.1016
9	0.1305	0.1336	0.0303
10	0.1233	0.1039	0.0807
\bar{R}_r^2	0.1149	0.0968	0.0890
${}^cR_p^2$	0.7184	0.7320	0.7417

literature [28]. Table 5 shows that Y-randomization [29] also confirms the robustness of suggested models. The details of described computational experiments are available on the Internet (<http://www.insilico.eu/coral>, folder "RateConstants").

4. Conclusions

This study evaluates novel applications of the CORAL software. The hybrid models taking into account the representation of molecular structure by both SMILES and graph (hydrogen filled) were tested on the predictions of rate constants of hydroxyl radical reactions of poly-brominated diphenyl ethers and (benzo-)triazoles. We concluded that the hybrid models are more accurate than models calculated using solely SMILES or hydrogen-filled graph approaches. Finally, the statistical quality of 2D CORAL models is approximately the same as the statistical quality of models [26] based on the 3D optimization and quantum mechanics data.

Acknowledgements

The authors express their gratitude to ANTARES (project number LIFE08-ENV/IT/00435), National Science Foundation (NSF) Interdisciplinary Nanotoxicity Center (Grant # HRD-0833178) and NSF-Experimental Program to Stimulate Competitive Research Award #: 362492-190200-01\NSFEPS0903787 for financial support.

References

- [1] B. Furtula, I. Gutman, *Journal of Chemometrics* 25 (2011) 87–91.
- [2] B. Hollas, I. Gutman, N. Trinajstić, *Croatica Chemica Acta* 78 (2005) 489–492.
- [3] G. Melagraki, A. Afantitis, *Current Medicinal Chemistry* 18 (2011) 2612–2619.
- [4] A. Afantitis, G. Melagraki, P.A. Koutentis, H. Sarimveis, G. Kollias, *European Journal of Medicinal Chemistry* 46 (2011) 497–508.
- [5] T. Ivanciuc, O. Ivanciuc, D.J. Klein, *Molecular Diversity* 10 (2006) 133–145.
- [6] A. Golbraikh, A. Tropsha, *Journal of Molecular Graphics and Modelling* 20 (2002) 269–276.
- [7] P.K. Ojha, I. Mitra, R.N. Das, K. Roy, *Chemometrics and Intelligent Laboratory Systems* 107 (2011) 194–205.
- [8] I. Mitra, A. Saha, K. Roy, *European Journal of Medicinal Chemistry* 45 (2010) 5071–5079.
- [9] A.G. Mercader, P.R. Duchowicz, F.M. Fernández, E.A. Castro, *Journal of Chemical Information and Modeling* 50 (2010) 1542–1548.
- [10] L.M.A. Mullen, P.R. Duchowicz, E.A. Castro, *Chemometrics and Intelligent Laboratory Systems* 107 (2011) 269–275.
- [11] A.P. Toropova, A.A. Toropov, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, *Journal of Computational Chemistry* 32 (2011) 2727–2733.
- [12] E. Benfenati, A.A. Toropov, A.P. Toropova, A. Manganaro, R. Gonella Diaza, *Chemical Biology & Drug Design* 77 (2011) 471–476.
- [13] A.A. Toropov, A.P. Toropova, A. Lombardo, A. Roncaglioni, E. Benfenati, G. Gini, *European Journal of Medicinal Chemistry* 46 (2011) 1400–1403.
- [14] A.P. Toropova, A.A. Toropov, R.G. Diaza, E. Benfenati, G. Gini, *Central European Journal of Chemistry* 9 (2011) 165–174.
- [15] A.P. Toropova, A.A. Toropov, E. Benfenati, G. Gini, *Chemometrics and Intelligent Laboratory Systems* 105 (2011) 215–219.
- [16] A.P. Toropova, A.A. Toropov, E. Benfenati, G. Gini, *Central European Journal of Chemistry* 9 (2011) 75–85.
- [17] A.P. Toropova, A.A. Toropov, E. Benfenati, D. Leszczynska, J. Leszczynski, *Journal of Mathematical Chemistry* 48 (2010) 959–987.
- [18] D. Weininger, *Journal of Chemical Information and Computer Sciences* 28 (1988) 31–36.
- [19] D. Weininger, A. Weininger, J.L. Weininger, *Journal of Chemical Information and Computer Sciences* 29 (1989) 97–101.
- [20] D. Weininger, *Journal of Chemical Information and Computer Sciences* 30 (1990) 237–243.
- [21] ACD/ChemSketch Freeware, version 11.00, Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2011 www.acdlabs.com.
- [22] B.F. Rasulev, A.A. Toropov, A.T. Hamme II, J. Leszczynski, *QSAR and Combinatorial Science* 27 (2008) 595–606.
- [23] A.A. Toropov, B.F. Rasulev, J. Leszczynski, *QSAR and Combinatorial Science* 26 (2007) 686–693.
- [24] A.A. Toropov, E. Benfenati, *Bioorganic & Medicinal Chemistry* 14 (2006) 2779–2788.
- [25] A.A. Toropov, A.P. Toropova, *Journal of Molecular Structure (THEOCHEM)* 578 (2002) 129–134.
- [26] P.P. Roy, S. Kovarich, P. Gramatica, *J. Comput. Chem.* 32 (2011) 2386–2396.
- [27] CORAL, <http://www.insilico.eu/CORAL> June 15 2011 Accessed.
- [28] K.-H. Kim, Z.-H. Shon, H.T. Nguyen, E.-C. Jeon, *Atmospheric Environment* 45 (2011) 1369–1382.
- [29] I. Mitra, A. Saha, K. Roy, *Molecular Simulation* 36 (2010) 1067–1079.