http://dx.doi.org/10.5935/0103-5053.20130223

J. Braz. Chem. Soc., Vol. 24, No. 11, 1781-1788, 2013.
Printed in Brazil - ©2013 Sociedade Brasileira de Química
0103 - 5053 $6.00+0.00

Article

QSAR Models of Reaction Rate Constants of Alkenes with Ozone and Hydroxyl Radical

Yueyu Xu,[a] Xinliang Yu,[a] and Shihua Zhang*,[a,b]*

[a]*College of Chemistry and Chemical Engineering,* [b]*Network Information Center, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China*

As constantes de velocidade da reação do ozônio com 95 alcenos (-logk_{O3}) e do radical hidroxila (•OH) com 98 alcenos (-logk_{OH}) na atmosfera foram previstas por modelos de relações quantitativas entre estrutura e atividade (QSAR). Cálculos usando a teoria do funcional da densidade (DFT) foram realizados para os respectivos alcenos no estado fundamental e para as estruturas do estado de transição para o processo de degradação na atmosfera. Técnicas de regressão linear múltipla (MLR) e de redes neurais de regressão generalizada (GRNN) foram utilizadas para desenvolver os modelos. O modelo GRNN de -logk_{O3} com base em três descritores e propagação ideal σ de 0,09 tem erro quadrático médio (*rms*) de 0,344; o modelo GRNN de -log$k$OH com quatro descritores e propagação ideal σ de 0,14 produz um erro *rms* de 0,097. Comparado com os modelos da literatura, os modelos GRNN neste artigo mostram estatísticas melhores. A importância dos descritores associados aos estados de transição na previsão de k_{O3} e k_{OH} nos processos de degradação atmosférica foi demonstrada.

The reaction rate constants of ozone with 95 alkenes (–logk_{O3}) and the hydroxyl radical (•OH) with 98 alkenes (–logk_{OH}) in the atmosphere were predicted by quantitative structure-activity relationship (QSAR) models. Density functional theory (DFT) calculations were carried out on respective ground-state alkenes and transition-state structures of degradation processes in the atmosphere. Stepwise multiple linear regression (MLR) and general regression neural network (GRNN) techniques were used to develop the models. The GRNN model of –logk_{O3} based on three descriptors and the optimal spread σ of 0.09 has the mean root mean square (*rms*) error of 0.344; the GRNN model of –logk_{OH} having four descriptors and the optimal spread σ of 0.14 produces the mean *rms* error of 0.097. Compared with literature models, the GRNN models in this article show better statistical characteristics. The importance of transition state descriptors in predicting k_{O3} and k_{OH} of atmospheric degradation processes has been demonstrated.

Keywords: atmospheric degradation, general regression neural network, quantitative structure-activity relationship, reaction rate constant, transition states

Introduction

Organic compounds emitted into the atmosphere can result in many adverse effects, such as photochemical air pollution, acid deposition, long-range transport of chemicals, changes of the stratospheric ozone layer and global weather modification, through a complex array of chemical and physical transformations.[1] The reactions of chemicals with OH radicals and ozone (O_3) during the daytime and NO_3 radicals at night are the most important degradation processes in the troposphere, so the lifetime and the upper concentration limit of the individual chemicals is assessed by determining their reaction rate constants with •OH, •NO_3 and O_3. These reaction rate constants can be obtained from experiments that may be quite costly, time-consuming, and laborious. But the experimental rate constants are available for only a limited number of organic compounds. Thus, it is useful to develop theoretical models predicting these reaction rate constants. Quantitative structure–activity relationship (QSAR) models are regression models describing the relationships between chemical structures and activities in a data-set of chemicals.[2] Once a QSAR model is developed successfully, it can be used to predict the activity of new chemicals.

In recent years, several QSAR models predicting reaction rate constants k_{O3} of O_3 have been reported. Pompe and

*e-mail: yxliang5602@sina.com.cn, shihua_zh@sina.cn

Veber described a 6-parameter model for $\log k_{O3}$ of O_3 and 117 organic compounds using multiple linear regression (MLR) analysis. The prediction capabilities of selected MLR model were evaluated by performing 10-fold cross-validation procedure. The average root-mean squared (*rms*) error in prediction of logarithm of reaction rate constants ($\log k_{O3}$) was 0.99.[3] Gramatica *et al.* produced models for the estimation of $-\log k_{O3}$ of O_3 with 125 heterogeneous chemicals. The optimum MLR model contained six parameters and had a *rms* error of 0.73.[4] Fatemi introduced a 6-parameter model for $-\log k_{O3}$ of 137 organic compounds, by using artificial neural networks (ANN). The *rms* errors for the training, prediction and validation sets were 0.357, 0.460 and 0.481, respectively.[5] Ren *et al.* developed models of $-\log k_{O3}$ for 116 organic compounds with projection pursuit regression (PPR) and support vector regression (SVR). The PPR model based on 7-descriptor had a *rms* error of 1.041 for the test set, which are smaller than the results obtained by the two SVR models (1.339 and 1.165, respectively).[6] Recently, Yu *et al.* used the radicals from organic compounds to calculate quantum chemical descriptors and developed 3-parameter SVR model for $-\log k_{O3}$. The *rms* errors for the training, validation and test sets were 0.680, 0.777 and 0.709, respectively.[7] In addition, two MLR model of $-\log k_{O3}$ in aqueous solution were, respectively, built for 39 aromatic pollutants organic compounds and 26 substituted phenols. The square regression coefficients R^2 were 0.791 and 0.826, respectively.[8,9]

Besides the most widely referenced AOPWIN model in EPI Suite that can be used for the calculations of reaction rate constants k_{OH} and the accuracy of k_{OH} was approximately 90% at 25 °C,[10,11] numerous QSAR studies have also been reported for predicting the rate constants k_{OH} of organic compounds. Gramatica *et al.* reported three QSAR models for k_{OH} with the prediction *rms* errors above 0.400.[12,13] Öberg constructed a model with the prediction standard error of 0.501 log units, through selecting 333 descriptors and compressing to 7 latent variables.[14] Böhnhardt *et al.* predicted the k_{OH} with the semiempirical AM1 method. The AM1-MOOH model yielded an overall *rms* error of 0.32 log units.[15] Wang et al. used 22 molecular structural descriptors to obtain a QSAR model with a *rms* error of 0.430 for the external validation set.[16] Fatemi and Baher developed six QSAR models of k_{OH} for 98 alkenes with linear and nonlinear techniques. These models based on five molecular descriptors were evaluated by a leave-24-out cross-validation test and had errors of $rms \geq 0.16$.[17] Toropov *et al.* examined the k_{OH} of 78 organic aromatic pollutants and developed a model with correlation coefficients (r^2) for the test sets of the four random splits were 0.75, 0.91, 0.84, and 0.80.[18]

All these models stated above are based on descriptors from the ground states of molecules (or radicals). According to classical chemical theory, reaction rate constants correlate with the ground-state reactants and the structures of transition states (or energy-rich intermediates). These models above should be defective when only the ground-state descriptors were used. What is certainly true is that techniques for finding a transition state are more difficult than finding a ground-state structure.[19] First, relatively little is known about transition-state geometries, at least by comparison with our extensive knowledge about ground-state molecules. Second, finding a saddle point is probably (but not necessarily) more difficult than finding a minimum due to theories and techniques. Third, the energy surface in the vicinity of a transition state is likely to be more "shallow" than that of a minimum. This "shallowness" suggests that the former is likely to be less well described in terms of a simple quadratic function than the latter. Last, the transition-state calculation, in the case of radical-molecule reactions, must be carried out on open-shell system with an odd number of electrons and correlation energy must be taken into account in those calculations since it plays a vital role for the transition-state properties. Therefore, transition-state calculation is extremely hardware intensive and time-consuming. For example, transition-state optimization typically requires two to three times the number of steps as geometry optimization.[19]

The gas-phase reactions of the alkenes with O_3 and hydroxyl radicals are of importance as atmospheric loss processes, since the available data shows that the k_{O3} and k_{OH} values at room temperature are in the ranges of 10^{-15} to 10^{-20} and 10^{-10} to 10^{-11} cm^3 molecule^{-1} s^{-1}, respectively, which are larger than the values of other compounds, such as alkanes. The purpose of this work is to produce QSAR models for k_{O3} of 95 alkenes and for k_{OH} of 98 alkenes. Quantum chemical descriptors used are calculated from ground-states of reactants and energy-rich transition states of degradation processes in the atmosphere.

Methods

Statistical methods

Stepwise multiple linear regression (MLR) is used widely in seeking an optimum linear combination of variables from the subsets of the N variables.[20,21] This technique only adds one parameter to a model at a time and always in the order from most to least important. Some important statistical parameters, the correlation coefficient R, standard error *se*, *t*-test, *Sig*.-test (or *p*-value), and variance inflation factor (*VIF*), were used to evaluate the statistical quality. A good QSAR model will have a low

value for *se* and a high *R* close to 1. The *t*-test measures the statistical significance of variables. The larger (in absolute terms) a test statistic value is, the more significant the associated variable will be. All variables with the *p*-values below a specified α cutoff values (the default level, 0.05) indicate that statistical significance is kept in the model. *VIF* can be used to identify whether excessively high multicollinearity coefficients exist among the descriptors. Generally, descriptors with *VIF* < 10 show multicollinearity coefficients for descriptors do not exceed 0.90, which indicates these descriptors may be acceptable.

General regression neural network (GRNN), proposed by Specht as the category of probabilistic neural networks, is a very useful tool to perform predictions and comparisons of system performance in practice.[22] Compared with other neural networks, such as back propagation, the GRNN paradigm has the advantages that it requires no iterative training and that it is unnecessary to define the number of hidden layers or the number of neurons per layer in advance.

The basic idea of GRNN is that each (*x*, *y*) data point for an input vector to be evaluated is computed as a mean value weighted by the influence which each Parzen window has on the input vector.[22] Suppose that *f*(*x*, *y*) represents the known joint continuous probability density function of a vector random variable, *x*, and a scalar random variable, *y*, and let *X* be a particular measured value of the random variable *x*, the regression of y on *X*, i.e., the conditional mean, is given by:

$$\hat{Y} = E[y \mid X] = \frac{\int_{-\infty}^{\infty} yf(X,y)dy}{\int_{-\infty}^{\infty} f(X,y)dy} \quad (1)$$

where $\hat{Y}$ is the estimate output of *Y*, by considering *X* as the system input.

The sample values X^i and Y^i of the random variables *x* and *y* can be used to estimate the density *f*(*x*, *y*), by introducing a nonparametric strategy based on Parzen's window.

$$\hat{f}(X,Y) = \frac{1}{(2\pi)^{(p+1)/2}\sigma^{(p+1)}} \cdot \frac{1}{n}\sum_{i}^{n} \exp\left[-\frac{(X-X^i)^T(X-X^i)}{2\sigma^2}\right] \cdot \exp\left[-\frac{(Y-Y^i)^2}{2\sigma^2}\right] \quad (2)$$

where *n* is the number of samples, *p* is the dimension of the vector variable *x*, and σ is the spreading factor (smoothing parameter or width coefficient) of Gaussian function.

By using Parzen windows estimation, the GRNN estimator can be expressed as follows:

$$\hat{Y}(X) = \frac{\sum_{i}^{n} \exp\left[-\frac{D_i^2}{2\sigma^2}\right] Y^i}{\sum_{i}^{n} \exp\left[-\frac{D_i^2}{2\sigma^2}\right]} \quad (3)$$

where D_i^2 is a scalar function.

$$D_i^2 = (X - X^i)^T (X - X^i) \quad (4)$$

As the spreading factor σ becomes very large, $\hat{Y}$ assumes the mean value of the observed, Y^i, and as σ goes to 0, $\hat{Y}$ assumes the value of the Y^i associated with the observation closest to *X*. For intermediate values of σ, all values of Y^i are taken into account, but those corresponding to points closer to *X* are given larger weight values.[22] For GRNN, only the spreading factor σ needs to be tuned. For a bigger σ value, the possible representation of the point of sample evaluated is possible for a wider range of *X*. For a small σ value the representation is limited to a narrow range of *X*.

Figure 1 shows the basic structure of a GRNN including the input, pattern, summation and output layers.[22] The input layer has a full interconnection to the pattern layer and brings all of the (scaled) measurement variables *X* into the network. The input neurons are merely distribution units, which are equal to the dimension of the vector variable *x*. The pattern layer contains the Parzen windows (Gaussian activation function, $\exp(-D_i^2/2\sigma^2)$, which approximates a density function by constructing it out of many simple parametric probability density functions. The width of these Parzen windows is specified by the spreading factor σ. The number of units equals to the number of sample observations. The summation layer consists of two types of nodes. One belongs to the denominator nodes and the other belongs to the numerator nodes. The output unit yields the desired estimate of $\hat{Y}$ values.

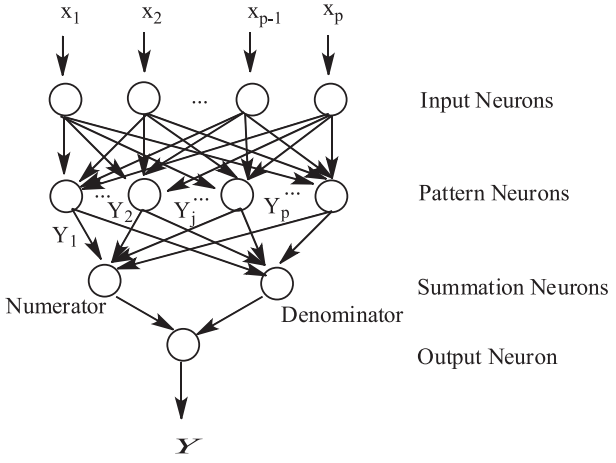


Figure 1. Structure model of a GRNN model.

Data set

Supplementary Information Table S1 listed the rate constants (k_{O3}) for the reaction of ozone with 95 alkenes,[7] which were measured at 25 °C and 101.3 kPa. The experimental data, reported in $cm^3\ s^{-1}$ $molecule^{-1}$, were transformed to logarithmic units and multiplied by −1 to

obtain positive values. The minimum and maximum values of $-\log k_{O3}$ were 13.1 for α-Terpinene (No. 39) and 20.4 for 1,1-dichloroethene (No.7), respectively. The former has 10 carbon atoms and the later has 2 carbon atoms. The experimental $-\log k_{O3}$ values in Table S1 were randomly split into a training set (60 alkenes) and a prediction set (35 alkenes).

Supplementary Information Table S2 listed the experimental rate constants ($\log k_{OH}$) for the reactions of the OH radical with 98 alkenes at 25 °C and 101.3 kPa. These experimental $\log k_{OH}$ data have been studied by Fatemi and Baher.[17] Both the training and test sets consist of 49 alkenes. The training sets in Tables S1 and S2 were used to develop models, which were tested with respective prediction set.

Quantum chemical descriptors

The reactions between O_3 and alkenes proceed by initial O_3 addition to the bond to yield an energy-rich ozonide which rapidly decomposes to a carbonyl and an initially energy-rich biradical.[1] The degradation reaction process of alkenes with O_3 can be expressed with Scheme 1.

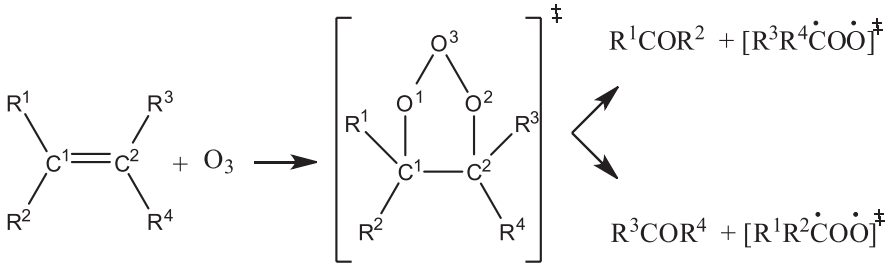
Thus the structures of the ground-states ($R^1R^2C^1{=}C^2R^3R^4$) and energy-rich transition states ($R^1R^2C^1C^2(O_3)R^3R^4$) should correlate with the reaction rate constants (k_{O3}). The transition-state complexes, characterized by a single imaginary vibrational frequency, were fully optimized and followed by frequency calculations. Nine quantum chemical descriptors were calculated for each transition state. These descriptors include the molecular average polarizability (α_I), the molecular dipole moment (μ_I), the energy of the lowest unoccupied molecular orbital (E_{ILUMO}), the energy of the highest occupied molecular orbital (E_{IHOMO}), the most positive net atomic charge on hydrogen atoms in a molecule (q_{IH}), the net charge of the most negative atom (q_I^-), the total energy (E_{IT}), the sum of the Mulliken charges of O^1 and O^2 (Q_{IO12}), the sum of the atomic polar tensor (APT) charges on C^1 and C^2 (q_{IC12}). Seven quantum chemical descriptors were derived for each ground state, which are the molecular average polarizability (α_G), the molecular dipole moment (μ_G), the energy of the lowest unoccupied molecular orbital (E_{GLUMO}), the energy of the highest occupied molecular orbital (E_{GHOMO}), the most positive net atomic charge on hydrogen atoms in a molecule (q_{GH}), the net charge of the most negative atom (q_G^-), and the total energy (E_{GT}). All these calculations were performed using density functional theory (DFT) in Gaussian 09 program (Revision A.02), at the B3LYP level of theory with 6-31G(d) basis set.

To fit $\log k_{OH}$, we calculated 12 quantum chemical descriptors from the energy-rich transition states, $R^1R^2{\bullet}C^1C^2(OH)R^3R^4$, formed from the reactions of the OH radical with alkenes ($R^1R^2C^1{=}C^2R^3R^4$). These descriptors are α_I, μ_I, $E_{I\alpha HOMO}$ and $E_{I\alpha LUMO}$ (for alpha spin states), $E_{I\beta HOMO}$ and $E_{I\beta LUMO}$ (for beta spin states), q_{IH}, q_I^-, E_{IT}, Q_{IO12}, q_{IC12}, and Q_{IC12} (the sum of the APT charges on C^1 and C^2 with hydrogens summed into heavy atoms). For each ground-state alkene, the same seven descriptors (α_G, μ_G, E_{GLUMO}, E_{GHOMO}, q_{GH}, q_G^-, and E_{GT}) were calculated. The DFT/UB3LYP/6-31G(d) and DFT/B3LYP/6-31G(d) methods in Gaussian 09 program were respectively adopted to optimize and calculate radical transition states and ground-state alkenes.

Results and discussion

Models for reaction rate constants $-\log k_{O3}$

By correlating the rate constants $-\log k_{O3}$ of 95 organic compounds in Table S1 to the 16 descriptors calculated in this article with stepwise regression analysis,[20,21] five MLR models (in Table 1) were obtained. In order to make a comparison with previous studies, the three parameters in Model 3 were taken as the optimal subset of descriptors, since the previous SVR model has minimum descriptors ($n = 3$).[7] The calculation values of three parameters E_{GHOMO} (the energy of the highest occupied molecular orbital of ground-state alkenes), Q_{IO12} (the sum of the Mulliken charges on O^1 and O^2 of intermediates), and q_{IC12} (the sum of the atomic polar tensor (APT) charges on C^1 and C^2 of intermediates) are listed in Table S1. The statistical parameters corresponding to the standardized and non-



Scheme 1. Degradation reaction process of alkenes with O_3.

standardized regression equations based on the training set in Table S1 were summarized below:

$$-\log k_{O3} = -0.562E_{GHOMO} + 0.207\ q_{IC12} + 0.319\ Q_{IO12} \quad (5)$$

$$-\log k_{O3} = 12.344 - 43.576E_{GHOMO} + 0.809\ q_{IC12} + 12.188\ Q_{IO12} \quad (6)$$

$R = 0.926$, $R^2 = 0.857$, $se = 0.558$, $F = 120.064$, $N = 60$, where R is the correlation coefficient, se is the standard error of estimation, F is the Fischer ratio, N is the number of compounds used.

The standardized coefficients in Equation 5 measure the relative importance of different variables, i.e., the larger the standardized coefficient (in absolute value) is, the more significant the variable will be. Usually, the non-standardized regression equations are used to predict the values of the dependent variable. The rate constants $-\log k_{O3}$ calculated with Equation 6 are listed in Table S1 and depicted in Figure 2, whose error bars give a good representation of the typical 5% error in the measurement of the rate constants. The statistical results of Equation 6 are listed in Table 2. *Sig.*-test suggests that the three descriptors E_{GHOMO}, Q_{IO12}, and q_{IC12} are significant descriptors and the *VIF*-test shows that the descriptors are not strongly correlated with each other.

As can be seen from standardized coefficients in Equation 5 or *t*-test values in Table 2, the most significant descriptor appearing in Equation 5 is the descriptor E_{GHOMO}, i.e., the energy of the highest occupied molecular orbital of ground-state molecules. This descriptor denotes the energetics of the reactant molecular orbitals involved in the reaction: the more reactive molecules possess a high E_{HOMO}. Therefore, the alkene molecule with a larger E_{HOMO} tends to lose electrons and leads to increased susceptibility of O_3 attacking which result in a larger k_{O3} value. The next two significant descriptor are Q_{IO12} (the sum of the Mulliken charges of O^1 and O^2 of energy-rich intermediates) and q_{IC12} (the sum of the atomic polar tensor charges of C^1 and C^2 of energy-rich intermediates), respectively. The smaller Q_{IO12} and q_{IC12} are, the larger k_{O3} will be. An energy-rich intermediate with the more positive net charges on C^1 and C^2 or on O^1 and O^2 indicates that the intermediate lies in higher energy-rich state. Thus the reaction will be relatively slow, and its k_{O3} will decrease.

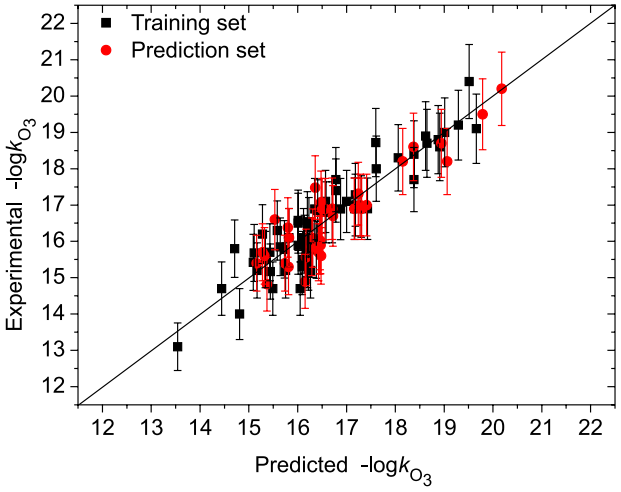


Figure 2. Predicted *vs.* experimental $-\log k_{O3}$ values from the MLR model for $-\log k_{O3}$. Error bars represent the typical 5% error in the measurement of the rate constants.

Table 1. Model summary for $-\log k_{O3}$

Model	*R*	*R* square	Adjusted *R* square	Std. error of the estimate
1[a]	0.873	0.763	0.760	0.672
2[b]	0.903	0.815	0.811	0.596
3[c]	0.920	0.847	0.842	0.545
4[d]	0.929	0.864	0.858	0.518
5[e]	0.933	0.871	0.864	0.506

[a]Predictors: (Constant), E_{GHOMO}; [b]predictors: (Constant), E_{GHOMO}, Q_{IO12}; [c]predictors: (Constant), E_{GHOMO}, Q_{IO12}, q_{IC12}; [d]predictors: (Constant), E_{GHOMO}, Q_{IO12}, q_{IC12}, E_{GLUMO}; [e]predictors: (Constant), E_{GHOMO}, Q_{IO12}, q_{IC12}, E_{GLUMO}, α_1.

We used the function newgrnn in MATLAB (R2012a for Windows) to build general regression neural networks (GRNN).[22] The best subset of descriptors (E_{GHOMO}, Q_{TO12} and q_{TC12}) selected for the MLR models were fed to GRNN as input vectors, and the reaction rate constants $-\log k_{O3}$ were taken as the output. The 30-fold (or leave-two-out) cross-validation strategy was used to train GRNNs and the circulation method was used to find the optimal spread

Table 2. Descriptor coefficients in MLR models of $-\log k_{O3}$

Descriptor	Unstandardized coefficients	Std. error	Standardized coefficients	*t*	Sig.	VIF
Constant	12.344	2.581	/	4.783	1.297×10^{-5}	/
E_{GHOMO}	–43.576	5.751	–0.562	–7.577	3.858×10^{-10}	2.155
Q_{IC12}	0.809	0.220	0.207	3.686	5.164×10^{-4}	1.236
Q_{IO12}	12.188	2.712	0.319	4.495	3.545×10^{-5}	1.975

parameter σ, which varied from 0.01 to 2 with the step being 0.01. The mean square error (MSE) was used to evaluate the accuracy of GRNN models. In the end, the optimal spread σ is determined as 0.09 and the minimum MSE value of two validation samples is 0.0041. The results from the optimal GRNN method are listed in Table S1 and depicted in Figure 3, which indicate that the predicted $-\log k_{O3}$ values are close to the experimental values. The *rms* errors of the training and test sets are 0.265 and 0.448, respectively, and the mean *rms* error of 95 chemicals is 0.344. These results are smaller than the corresponding values (0.540, 0.540 and 0.540, respectively) obtained from the MLR model, i.e., Equation 6. Thus the GRNN model has better prediction accuracy than the MLR model, although the latter is accurate and acceptable when compared to the previous models.[3-7]

We further predicted rate constants $-\log k_{O3}$ for the test set in Table S1 with the approaches reported by Yu *et al.*[7] The MLR and SVR models from the training set in Table S1 produced *rms* errors of 0.681 and 0.663, respectively, which are larger than that (*rms* = 0.448) of the present GRNN model (σ = 0.09). Therefore, combining the quantum chemical descriptors from the ground-states and the energy-rich intermediates to predict $-\log k_{O3}$ of alkenes is feasible.

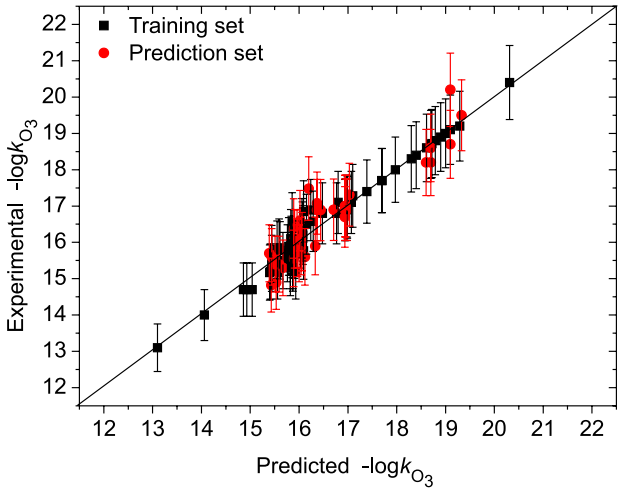


Figure 3. Predicted *vs.* experimental $-\log k_{O3}$ values from the GRNN model for $-\log k_{O3}$. Error bars represent the typical 5% error in the measurement of the rate constants.

Models for reaction rate constants $-\log k_{OH}$

Similar analysis methods were used to develop QSAR models for $-\log k_{OH}$ of 98 alkenes in Table S2. The optimal standardized and non-standardized regression equations were, respectively,

$$-\log k_{OH} = -0.254\mu_I - 0.183\, E_{I\beta HOMO} - 0.551\, E_{GHOMO} + 0.556\, Q_{IC12} \quad (7)$$

$$-\log k_{OH} = 5.219 - 0.215\mu_I - 4.284\, E_{I\beta HOMO} - 16.187\, E_{GHOMO} + 1.253\, Q_{IC12} \quad (8)$$

$R = 0.940$, $R^2 = 0.884$, $se = 0.153$, $F = 83.692$, $N = 49$.

The MLR model (i.e., Equation 8) of $-\log k_{OH}$ includes a subset of descriptors: the molecular dipole moment of energy-rich intermediates (μ_I), the energy of the lowest unoccupied molecular orbital for alpha spin states of intermediates ($E_{I\alpha LUMO}$), the sum of the APT charges on C^1 and C^2 with hydrogens summed into heavy atoms of intermediates (Q_{IC12}), and the energy of the highest occupied molecular orbital of ground-state alkenes (E_{GHOMO}). Table 3 shows the descriptors in the MRL model all are significant descriptors and do not contaminate each other. As stated above, the descriptors $E_{I\beta HOMO}$, E_{GHOMO}, and Q_{IC12} are correlated with the reaction rate constants. In addition, the dipole moment descriptor μ can reflect the polarity of a molecule. A larger descriptor μ_I indicates a higher reactivity, which leads to a high k_{OH} value.

The values of four descriptors and predicted $-\log k_{OH}$ were listed in Table S2 and depicted in Figure 4 (for the MLR model) and Figure 5 (for the GRNN model). For the GRNN model with the optimal spread σ of 0.14, the *rms* errors of the training and test sets are 0.069 and 0.119, respectively, which are smaller than the corresponding *rms* values (0.144 and 0.134, respectively) of the MLR model (i.e., Equation 8) in this article. The mean *rms* errors of the MLR and GRNN models are 0.140 and 0.097, respectively. These *rms* errors are lower than the results (0.16-0.28)

Table 3. Descriptor coefficients in MLR model of $-\log k_{OH}$

Descriptor	Unstandardized coefficients	Std. error	Standardized coefficients	*t*	Sig.	VIF
Constant	5.219	0.394	/	13.257	5.718×10^{-17}	/
μ_I	–0.215	0.057	–0.254	–3.762	4.955×10^{-4}	1.728
$E_{I\beta HOMO}$	–4.284	2.057	–0.183	–2.083	4.311×10^{-2}	2.928
E_{GHOMO}	–16.187	2.116	–0.551	–7.650	1.290×10^{-9}	1.968
Q_{IC12}	1.253	0.188	0.556	6.676	3.420×10^{-8}	2.628

of the six QSAR models in the literature.[17] Moreover, compared to the literature models,[17] our models have fewer descriptors (4:5) and use more samples for the test set.

We also predicted rate constants $-\log k_{OH}$ for the test set in Table S2 with the Atkinson scheme.[11,15,23,24] The *rms* error of the test set is 0.210, which are larger than the results of 0.134 from the present MLR model and 0.119 from the present GRNN model ($\sigma = 0.14$). Therefore, both the MLR and GRNN models of $-\log k_{OH}$ based on quantum chemical descriptors from ground states of reactants and radical transition states are successful in predicting $-\log k_{OH}$ values of alkenes.

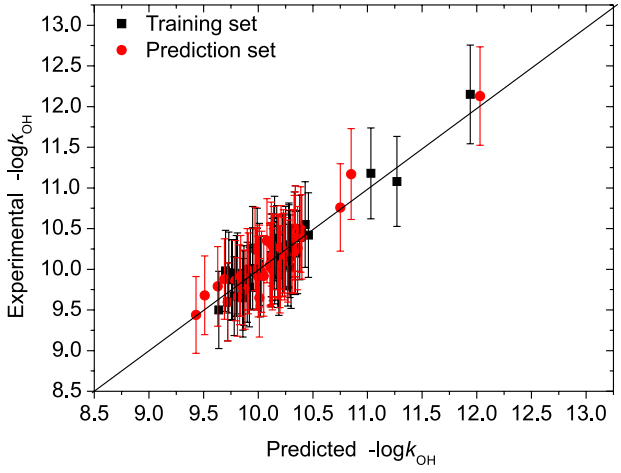


Figure 4. Predicted *vs.* experimental $-\log k_{OH}$ values from the MLR model for $-\log k_{OH}$. Error bars represent the typical 5% error in the measurement of the rate constants.

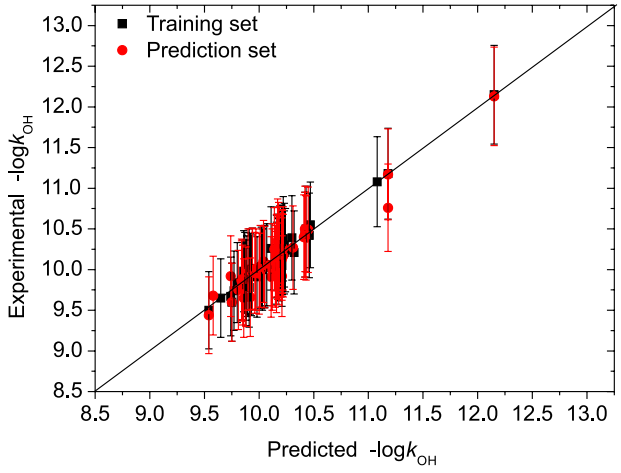


Figure 5. Predicted *vs.* experimental $-\log k_{OH}$ values from the GRNN model for $-\log k_{OH}$. Error bars represent the typical 5% error in the measurement of the rate constants.

Conclusions

QSAR models based on the MLR and GRNN approaches were successfully developed for reaction rate constants $-\log k_{O3}$ of 95 alkenes and $-\log k_{OH}$ of 98 alkenes. Quantum chemical descriptors used were obtained from the ground-states of alkenes and energy-rich transition states of degradation processes in the atmosphere. The present work tests that transition states have important effects on k_{O3} and k_{OH} of alkenes in degradation processes. Our models overcome the defects of the existing models only based on ground-state descriptors. The optimal GRNN models in this investigation are expected to have good predictive performance.

Supplementary Information

Tables S1 and S2 showing experimental rate constants and descriptors used are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgements

The study was supported by the Natural Science Research Foundation of Hunan Province (No. 12JJ6011) and the National Natural Science Foundation of China (Grant No. 20972045).

References

1. Atkinson, R.; *Atmos. Environ.* **2007**, *41*, 200.
2. Karelson, M.; Lobanov, V. S.; Katritzky, A. R.; *Chem. Rev.* **1996**, *96*, 1027.
3. Pompe, M.; Veber, M.; *Atmos. Environ.* **2001**, *35*, 3781.
4. Gramatica, P.; Pilutti, P.; Papa, E.; *QSAR Comb. Sci.* **2003**, *22*, 364.
5. Fatemi, M. H.; *Anal. Chim. Acta* **2006**, *556*, 355.
6. Ren, Y.; Liu, H.; Yao, X.; Liu, M.; *Anal. Chim. Acta* **2007**, *589*, 150.
7. Yu, X.; Yi, B.; Wang, X.; Chen, J.; *Atmos. Environ.* **2012**, *51*, 124.
8. Jiang, J.; Yue, X.; Chen, Q.; Gao, Z.; *Bull. Environ. Contam. Toxicol.* **2010,** *85*, 568.
9. Liu, H.; Tan, J.; Yu, H.; Liu, H.; Wang, L.; Wang, Z.; *Int. J. Environ. Res.* **2010**, *4*, 507.
10. Atkinson, R.; *Chem. Rev.* **1986**, *86*, 69.
11. Kwok, E. S. C.; Atkinson, R.; *Atmos. Environ.* **1995**, *29*, 1685.
12. Gramatica, P.; Pilutti, P.; Papa, E.; *Atmos. Environ.* **2004**, *38*, 6167.
13. Gramatica, P.; Pilutti, P.; Papa, E.; *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1794.
14. Öberg, T.; *Atmos. Environ.* **2005**, *39*, 2189.
15. Böhnhardt, A.; Kühne, R.; Ebert, R.-U.; Schüürmann, G.; *J. Phys. Chem. A* **2008**, *112*, 11391.
16. Wang, Y.; Chen, J.; Li, X.; Wang, B.; Cai, X.; Huang, L.; *Atmos. Environ*, **2009**, *43*, 1131.

17. Fatemi, M. H.; Baher, E.; *SAR QSAR Environ. Res*. **2009**, *20*, 77.
18. Toropov, A. A.; Toropova, A. P.; Rasulev, B. F.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J.; *J. Comput. Chem*. **2012**, *33*, 1902.
19. Hehre, W. J.; *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction Inc.: California, 2003.
20. Fatemi, M. H.; Baher, E.; Ghorbanzade'h, M.; *J. Sep. Sci*. **2009**, *32*, 4133.
21. Yu, X.; Yi, B.; Wang, X.; *J. Comput. Chem*. **2007**, *28*, 2336.
22. Specht, D. F.; *IEEE Trans. Neural Networks* **1991**, *2*, 568.
23. Atkinson, R; *Chem. Rev.* **1985**, *85*, 69.
24. Atkinson, R.; *Int. J. Chem. Kinet.* **1987**, *19*, 799.

Submitted: June 20, 2013
Published online: September 13, 2013