Theor Chem Acc (2010) 127:355–367
DOI 10.1007/s00214-009-0724-8

REGULAR ARTICLE

# Predicting rate constants of OH radical reactions with organic substances: advances for oxygenated organics through a molecular orbital HF/6-31G** approach

**Anna Böhnhardt · Ralph Kühne · Ralf-Uwe Ebert · Gerrit Schüürmann**

Received: 23 October 2009 / Accepted: 29 December 2009 / Published online: 16 January 2010


**Abstract** The molecular orbital OH (MOOH) approach is a perturbational quantum chemical method to predict rate constants of OH radical reactions with organic compounds. Going beyond previous AM1 parameterizations, a first ab initio implementation employing the HF/6-31G** level of calculation has been developed. For a set of 799 organic compounds with experimental rate constants, $k_{OH}$, varying over more than six orders of magnitude, the new MOOH-HF method is superior to both MOOH-AM1 and Atkinson's increment scheme, yielding a predictive squared correlation coefficient ($q^2$) of 0.95 and a root-mean-square error of 0.29 log units. For oxygenated compounds, MOOH-HF shows significant improvements over MOOH-AM1, which holds in particular for aldehydes and ketones. The discussion includes detailed comparative analyses of the model performances for individual compound classes.

**Keywords** Atmospheric degradation · Hydroxyl radical · Rate constant · Molecular orbital model · MOOH method · Atkinson increment scheme · Consensus modelling

A. Böhnhardt · R. Kühne · R.-U. Ebert · G. Schüürmann (✉)
UFZ Department of Ecological Chemistry,
Helmholtz Centre for Environmental Research,
Permoserstr. 15, 04318 Leipzig, Germany
e-mail: gerrit.schuurmann@ufz.de

A. Böhnhardt · G. Schüürmann
Institute for Organic Chemistry,
Technical University Bergakademie Freiberg,
Leipziger Str. 29, 09596 Freiberg, Germany

## 1 Introduction

The reaction with OH radicals is the major tropospheric sink for most volatile organic compounds. Accordingly, their residence time in the air is often largely determined by the respective rate constant, $k_{OH}$, and thus can be approximately quantified through pseudo first-order kinetics, making use of an average 24-h tropospheric OH radical concentration of $9 \times 10^5$ molecule/cm$^3$ [1, 2]. For organic compounds, $k_{OH}$ varies by more than six orders of magnitude [3], corresponding to atmospheric lifetimes from minutes to hundreds of years.

As the experimental determination of OH radical rate constants is expensive and time-consuming, prediction methods to provide at least approximate $k_{OH}$ values are welcome. The latter are of particular interest for screening compounds with respect to persistent organic pollutant (POP) criteria as laid down in the Stockholm Convention [4] and to persistent, bioaccumulative and toxic (PBT) properties according to the new European chemical legislation REACH [5]. Moreover, respective calculation schemes allow one to predictively assess molecular structures of compounds that are not yet synthesized, thus providing guidance in the design of industrial chemicals with respect to environmental persistence criteria.

Currently, an increment scheme developed by Atkinson [6–8] is the most widely applied approach for predicting $k_{OH}$. The Atkinson method requires two-dimensional (2D) molecular structures as the only input information. Basic rate constants are used to describe the reactivity of specific molecular sites. Assuming additivity of intramolecular effects, the influence of neighbouring groups is quantified through fixed group factors.

An independent approach based on local electronic properties of 3D molecular structures was introduced by

Klamt [9, 10]. This approach, named molecular orbital OH (MOOH) method, employs local molecular parameters that are derived from quantum chemical calculations. Thus, also nonlinear intramolecular interactions are taken into account as far as these are captured in the chosen level of quantum chemical calculation.

Modified and recalibrated in a recent study, the MOOH approach has been shown to surpass the Atkinson method for a set of 805 experimental rate constants [3]. Furthermore, it became apparent that the performance of both methods strongly depends on the compound class, with both methods showing different strengths and shortcomings [3]. For example, MOOH-AM1 yields significantly better results for alkenes, whereas the Atkinson method was still superior for predicting most classes of oxygenated compounds.

Establishing a new, HF/6-31G** based MOOH model, the aim of this study is to examine whether the performance of the MOOH approach can be improved by the application of a higher level quantum chemical model. HF/6-31G** was chosen as standard model chemistry, keeping in mind that the MOOH approach as outlined below focuses on the LCAO–MO (linear combination of atomic orbitals–molecular orbital) coefficients and MO energies of organic molecules in their ground state (and that HF yields the best one-determinant wavefunction for the basis set selected). Statistical performances of the different MOOH models and the Atkinson method are evaluated in detail, using a large set of organic compounds with up-to-date experimental $k_{OH}$ values.

## 2 Materials and methods

### 2.1 Data set

Starting point for the model calibration and validation (see below) was a set of 805 compounds with experimental, room-temperature second-order rate constants $k_{OH}$ as collected previously [3]. All $k_{OH}$ values used in this study refer to gas-phase reactions. As the 6-31G** basis set used in this work is not parameterized for iodine, the following six compounds were excluded from the data set: methyl iodide, iodoethane, 1-iodopropane, 2-iodopropane, 3-iodo-1-propene and iodobenzene. For the resultant data set with 799 organic compounds, the experimental rate constant covers more than six orders of magnitude, ranging from $1.91 \times 10^{-16}$ (pentafluoromethyl ether) to $5.5 \times 10^{-10}$ (1-naphthol) $cm^3$ $molecule^{-1}$ $s^{-1}$. Therefore, log $k_{OH}$ is the target property for regression model calibration and prediction.

### 2.2 Mechanistic domain and calibration

The MOOH approach covers the following reaction mechanisms of organic compounds with the OH radical: Abstraction of hydrogen (H) atoms from $sp^3$ carbon atoms ($k^{C}_{H-abs}$), H abstraction from aliphatic carbon enhanced through lone pairs of neighbour oxygen ($k^{C(O)}_{H-abs}$), abstraction of aldehydic H ($k^{CHO}_{H-abs}$), OH addition to aromatic atoms ($k^{aro}_{add}$), OH addition to double and triple bonds ($k^{C=C}_{add}$ and $k^{C\equiv C}_{add}$, respectively) and H abstraction from the hydroxyl groups of alcohols and carboxylic acids ($k^{OH}_{H-abs}$).

Consequently, the mechanistic domain of the MOOH models is defined through compounds where the overall $k_{OH}$ may cover the following individual contributions:

$$k_{OH} = \sum k^{C}_{H-abs} + \sum k^{C(O)}_{H-abs} + \sum k^{CHO}_{H-abs} + \sum k^{aro}_{add} + \sum k^{C=C}_{add} + \sum k^{C\equiv C}_{add} + \sum k^{OH}_{H-abs} \quad (1)$$

For a given molecule, the second-order rate constant $k_{OH}$ is calculated as the sum of individual rate constants, taking into account each intramolecular reaction site with its site-specific reaction mechanism.

The model equation for H abstraction was calibrated first using a subset of 230 saturated hydrocarbons and ethers. For these compounds, OH radical reaction occurs solely via H abstraction from saturated carbon ($sp^3$ hybridization). This reaction path is usually much slower than other reaction mechanisms involving OH radicals.

Subsequently, the other reaction mechanisms were calibrated separately, using subsets of compounds that offered only H abstraction from $sp^3$ carbon (without activating oxygen at $\alpha$ carbon) and the respective additional mechanism. In this way, monofunctional subsets were generated for OH addition to aromatic atoms ($k^{aro}_{add}$) and OH addition to double and triple bonds ($k^{C=C}_{add}$ and $k^{C\equiv C}_{add}$, respectively).

As different functional groups are reaction sites for the enhancement of H abstraction through lone pairs of neighbour oxygen ($k^{C(O)}_{H-abs}$), the subset to calibrate the respective subset also contains bi- and multifunctional oxygenated compounds possessing the following functional groups: ketone group, hydroxyl group, carboxylic acid group and carboxylic acid ester group. Although aldehydes also show enhanced H abstraction from $sp^3$ carbon, they were assembled in a separate, monofunctional subset, as they dominantly react via abstraction of aldehydic H ($k^{CHO}_{H-abs}$).

Multifunctional compounds offering more than one degradation pathway with similar degradation kinetics were excluded from calibration. This strategy was applied in order to avoid the possibility of contaminating the model equation for one fast reaction mechanism through the presence of a second fast reaction pathway. Consequently, calibration was carried out using only 670 compounds of the total set of 799 compounds, leaving the remainder of 129 compounds with at least two different fast reaction pathways for the model validation.

### 2.3 Computational details

The program AOPWIN [11, 12] was used to calculate $k_{OH}$ according to the Atkinson method. Gaussian 03 program package [13] was used to carry out geometry optimizations at the HF/6-31G** level for the total set of 799 compounds. The same program was applied to conduct UB3LYP/6-31G** calculations on pre-reactive complexes between the OH radical and the selected organic compounds. All minima were confirmed by frequency calculations. The resultant Gaussian 03 output files and files containing structural information from our in-house software ChemProp [14] were used as molecule-specific input files, providing information on eigenvalues, eigenvectors and geometry for subsequent MOOH calculations.

The molecular parameters feeding the MOOH model equations as well as the resultant $k_{OH}$ values were calculated using a FORTRAN program that included a modified version of the original MOOH FORTRAN subroutine. For calibrating the nonlinear MOOH model equations, the Marquardt method [15] was implemented based on a FORTRAN algorithm from literature [16].

### 2.4 Statistical performance

The calibration quality was quantified through the squared correlation coefficient,

$$r^2 = 1 - \frac{\sum_i \left(y_i^{\text{fit}} - y_i^{\text{obs}}\right)^2}{\sum_i \left(y_i^{\text{obs}} - y^{\text{mean}}\right)^2} \tag{2}$$

In Eq. 2, $y_i^{\text{fit}}$ and $y_i^{\text{obs}}$ denote the fitted (calibrated) and observed target value (in our case log $k_{OH}$) of compound $i$, and $y^{\text{mean}}$ the experimental mean value of the data set. The squared correlation coefficient, $r^2$, has a value ranging between 0 (no correlation) and 1 (perfect correlation), and focuses on the trend rather than on absolute values. It automatically corrects for systematic errors also if applied to existing regression models [17].

Additionally, the predictive squared correlation coefficient,

$$q^2 = 1 - \frac{\sum_i \left(y_i^{\text{pred}} - y_i^{\text{obs}}\right)^2}{\sum_i \left(y_i^{\text{obs}} - y^{\text{mean}}\right)^2} \tag{3}$$

was used to evaluate the prediction performance [17]. Here, $y_i^{\text{fit}}$ has been replaced by $y_i^{\text{pred}}$, the value predicted by (not fitted through) the regression model under investigation. In contrast to $r^2$, $q^2$ ranges from 1 (perfect agreement) to $-\infty$, with $q^2 = 0$ representing the case where the model prediction is as good as taking the experimental mean as predictor for all values. When deriving a regression model, $q^2 = r^2$ for the training set in case of parallel calibration of all multilinear parameters (because in this case $y_i^{\text{pred}} = y_i^{\text{fit}}$).

In this study, however, $q^2 \leq r^2$ for the total compound set because of two reasons: first, only 670 of the 799 compounds were used for model calibration, and second, a step-wise parameter calibration has been employed as outlined above (and specified below when presenting the results). In addition to $r^2$ and $q^2$, the following parameters were used to obtain further information on the statistical performance: root-mean-square error (rms), systematic error (bias), mean error (me), maximum negative error (mne, largest underestimation) and maximum positive error (mpe, largest overestimation).

### 2.5 Molecular parameters

The MOOH approach is based on local molecular parameters that are derived from quantum chemical calculations. These descriptors were introduced by Klamt [9] and contain energy and charge information that reflects the local characteristics of the respective reaction site.

The charge-limited donor energy at atomic site $r$, $EQ_{\text{occ}}(q, r)$, is calculated as mean of the topmost doubly occupied molecular orbital energies $E_i$, weighted by the charge-factor $2q^{-1}\sum_{\mu(r)} c_{\mu,i}^2$:

$$EQ_{\text{occ}}(q,r) = \sum_{i=\text{HOMO}}^{1} E_i \cdot w_q(i,r)$$

$$w_q(i,r) = \begin{cases} 2q^{-1}\sum_{\mu(r)} c_{\mu,i}^2 \text{ if } 2\sum_{j=\text{HOMO}}^{\text{i}}\sum_{\mu(r)} c_{\mu,j}^2 < q \\ 1 - 2q^{-1}\sum_{m=\text{HOMO}}^{\text{i+1}}\sum_{\mu(r)} c_{\mu,m}^2 \\ \quad \text{if } 2\sum_{m=\text{HOMO}}^{\text{i+1}}\sum_{\mu(r)} c_{\mu,m}^2 \leq q \leq 2\sum_{j=\text{HOMO}}^{\text{i}}\sum_{\mu(r)} c_{\mu,j}^2 \\ 0 \text{ if } 2\sum_{m=\text{HOMO}}^{\text{i+1}}\sum_{\mu(r)} c_{\mu,m}^2 > q \end{cases} \tag{4}$$

$EQ_{\text{occ}}(q, r)$ quantifies the energy required to donate charge $q$ from atomic site $r$ of the compound under investigation, evaluating for each molecular orbital (MO) the associated site-specific weights provided by the linear combination of atomic orbitals (LCAO). In Eq. 4, the LCAO–MO coefficient $c_{\mu i}$ refers to the contribution of the $\mu$th AO to the $i$th MO, and the sum index $\mu(r)$ indicates that only AOs $\mu$ located at atomic site $r$ are taken into account. Moreover, the term $2\sum_{\mu(r)} c_{\mu,i}^2$ provides an approximate quantification of the electron population of the $i$th occupied MO at site $r$, ignoring off-diagonal contributions to the electron density (that would be zero in case of semiempirical schemes neglecting diatomic differential overlap). Note that the

same approximation has already been successfully applied when employing this kind of local parameters for predicting the H-bond donor and acceptor strengths of organic compounds [18–20]. The pre-defined charge limit $q$ is a cut-off value determining the number of occupied orbitals (starting from the highest occupied molecular orbital, HOMO) that are actually taken into account. Note that the formula symbol $q$ denotes a net atomic charge and should not be confused with the symbol often used for the predictive squared correlation coefficient, $q^2$ (Eq. 3 and [17]). As both $q$ and $q^2$ are widely used as symbols for the respective properties we adhere to this nomenclature. At the same time, the sufficiently different contexts when referring to one or other property should avoid respective misunderstanding.

The effective donor energy at atomic site $r$, $EE_{\mathrm{occ}}(E_{\mathrm{ref}}, r)$, is defined as weighted mean of the energies $E_{\mathrm{i}}$ of all occupied MOs, and can be understood as a generalized site-specific HOMO energy:

$$EE_{\mathrm{occ}}(E_{\mathrm{ref}}, r) = \sum_{i=\mathrm{HOMO}}^{1} E_i \cdot w_{\mathrm{ref}}(i, r)$$
$$w_{\mathrm{ref}}(i, r) = \frac{\sum_{\mu(r)} c_{\mu i}^2 \cdot \exp\left\{\frac{E_i}{E_{\mathrm{ref}}}\right\}}{\sum_{j=\mathrm{HOMO}}^{1} \left(\sum_{\nu(r)} c_{\nu j}^2 \cdot \exp\left\{\frac{E_j}{E_{\mathrm{ref}}}\right\}\right)} \quad (5)$$

The weighting factor of the $i$th MO with respect to site $r$, $w_{\mathrm{ref}}(i, r)$, is defined as a Boltzmann-type term, taking into account both the local electron density in terms of the respective squared LCAO–MO coefficients, $c_{\mu i}^2$ (considering all AOs $\mu$ located at site $r$), and the MO energy $E_{\mathrm{i}}$ relative to a pre-defined reference energy $E_{\mathrm{ref}}$. In this way, the actual contribution of a given MO energy $E_{\mathrm{i}}$ to $EE_{\mathrm{occ}}(E_{\mathrm{ref}}, r)$ is triggered by both its electron density at site $r$ and the (exponentially weighted) ratio of $E_{\mathrm{i}}$ to $E_{\mathrm{ref}}$.

The energy-limited acceptor charge $QE_{\mathrm{vac}}(\varepsilon, r)$ quantifies the local charge acceptor capacity at atomic site $r$:

$$QE_{\mathrm{vac}}(\varepsilon, r) = 2 \cdot \sum_{k=\mathrm{LUMO}}^{n} \sum_{\mu(r)} c_{\mu k}^2 \cdot w_k(\varepsilon)$$
$$w_k(\varepsilon) = \begin{cases} 1 \text{ if } E_k < \varepsilon - 0.5 \\ (E_k - \varepsilon) + 0.5 \text{ if } \varepsilon - 0.5 \le E_k \le \varepsilon + 0.5 \\ 0 \text{ if } E_k > \varepsilon + 0.5 \end{cases} \quad (6)$$

The pre-defined penetration energy $\varepsilon$ is a cut-off value limiting the summation over the charge contributions $c_{\mu k}^2$ of the lowest unoccupied MOs $k$ of AOs $\mu$ located at site $r$. Thus, $QE_{\mathrm{vac}}(\varepsilon, r)$ characterizes the readiness of the molecule to populate unoccupied MOs $k$ by additional electron charge, constrained by an upper limit of the accordingly active unoccupied MOs through $\varepsilon$ via the weighting factor $w_{\mathrm{k}}(\varepsilon)$. Note further that as with $EQ_{\mathrm{occ}}(q, r)$ (Eq. 4), overlap contributions to the electron density at site $r$ are not taken into account for $EE_{\mathrm{occ}}(E_{\mathrm{ref}}, r)$ (Eq. 5) and $QE_{\mathrm{vac}}(\varepsilon, r)$ (Eq. 6). The term $2\sum_{\mu(r)} c_{\mu,i}^2$ therefore provides only an approximation of the electron population of the $i$th occupied MO at atomic site $r$ when employing the HF/6-31G** level of theory.

For $\pi$ systems such as alkenes, atomic $\pi$ orbitals are of major interest for characterizing reactivity. This aspect is taken into account by $\pi$-specific local molecular parameters that are calculated for the respective $\pi$ orbitals alone instead of summing over all atomic orbitals. In the following, these $\pi$-specific parameters are denoted by the index $\pi$ in the respective formula.

Addition to aromatic rings is expected to occur preferably at unsubstituted carbon atoms. The deformation energy $\Delta_{\mathrm{def}}^{\mathrm{aro}}$ is a parameter introduced by Klamt [9] to account for this effect, quantifying the additional energy demand in case of substituted aromatic carbon. The $\Delta_{\mathrm{def}}^{\mathrm{aro}}$ values were derived from AM1 calculations on the energy required to bend the substituent of the respective aromatic carbon by 45° from planarity. As $\Delta_{\mathrm{def}}^{\mathrm{aro}}$ has been found to be principally determined by the atom type of the first substituent atom, Klamt [9] generated a list of $\Delta_{\mathrm{def}}^{\mathrm{aro}}$ values for the most common substituent atom types (see Table 1) for both calibration and prediction purposes. Reflecting the fact that addition to the inner atoms of condensed aromatics is very unlikely, $\Delta_{\mathrm{def}}^{\mathrm{aro}}$ has been set to 100 kcal/mol for these atoms [9].

As outlined above, the local parameters charge-limited donor energy $EQ_{\mathrm{occ}}(q, r)$, effective donor energy at atomic site $r$, $EE_{\mathrm{occ}}(E_{\mathrm{ref}}, r)$, and energy-limited acceptor charge $QE_{\mathrm{vac}}(\varepsilon, r)$ depend on the values selected for $q$, $E_{\mathrm{ref}}$ and $\varepsilon$. Adequate values for these parameters were determined through successive variation.

We examined the differences between AM1 based and HF/6-31G** molecular descriptors. AM1 and HF/6-31G** calculations yield similar, but not completely corresponding values for $EQ_{\mathrm{occ}}(q, r_{\mathrm{H}})$, $EQ_{\mathrm{occ}}^{\pi}(q, r_{\mathrm{C}})$ and $EE_{\mathrm{occ}}(E_{\mathrm{ref}}, r_{\mathrm{C}})$. Significant differences were observed for the AM1 and HF/6-31G** values for $QE_{\mathrm{vac}}(\varepsilon, r_{\mathrm{C}})$ values. This parameter is used together with the charge-limited $\pi$-electron donor charge $EQ_{\mathrm{occ}}^{\pi}(q, r_{\mathrm{C}})$ to describe the reactivity of alkenes. Notwithstanding a poor correlation between the corresponding AM1 and HF/6-31G** results for most carbon atoms, the largest $QE_{\mathrm{vac}}(\varepsilon, r_{\mathrm{C}})$ values with both AM1 and HF/6-31G** are observed for $sp^2$ carbons neighboured by

**Table 1** Deformation energies $\Delta_{\mathrm{def}}^{\mathrm{aro}}$ relative to hydrogen (in kcal/mol) [4]

| Atom type | –S– | –H | –I | –Br | $-C_{\mathrm{any}}$ | –Cl | –O– | $-NO_2$ | –F | –N< |
|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta_{\mathrm{def}}^{\mathrm{aro}}$ | −3 | 0 | 0 | 3 | 5 | 6 | 8 | 8 | 11 | 19 |

chlorine atoms. This finding probably reflects the overlap of chlorine and $sp^2$ carbon $\pi$-orbitals and the resulting mesomeric effect. $QE_{vac}(\varepsilon, r_C)$ is rather small for most other carbon atoms and appears to function as a trigger for resonance.

## 3 Results and discussion

In the following discussion, the units are $10^{-12}$ $cm^3$ $molecule^{-1}$ $s^{-1}$ for rate constants, eV for MO energies and electron charges $e$ for charges.

### 3.1 Hydrogen abstraction from aliphatic carbon

Hydrogen abstraction from aliphatic carbon atoms is slow compared with most other OH radical reaction mechanisms, such as addition to unsaturated bonds [21, 22]. Consistently, calibration of the model equation for hydrogen abstraction requires a fitting set of compounds with this mechanism being the only OH degradation pathway. As the hydrogen abstraction reactions of ketones, esters, alcohols and carboxylic acids are accelerated by the formation of hydrogen-bonded pre-reactive complexes [10, 23–30], this additional lone pair-enhanced reactivity is accounted for separately (see below).

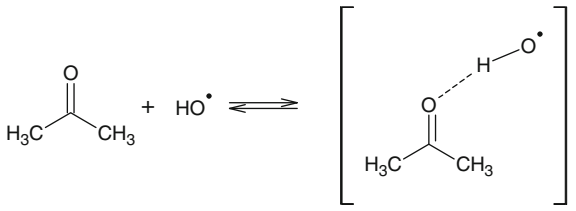
Taking the 230 aliphatic compounds from the present data set, nonlinear calibration resulted in the following MOOH-HF model equation for the H abstraction from aliphatic carbon:

$$k^{C}_{H-abs}(r_H) = \exp\left\{ \frac{12.13}{\left(1+\exp\{\,-0.87\cdot(EQ_{occ}(0.12,r_H)\,+12.73)\,\}\right)^{0.86}} - 8.33 \right\} \quad (7)$$

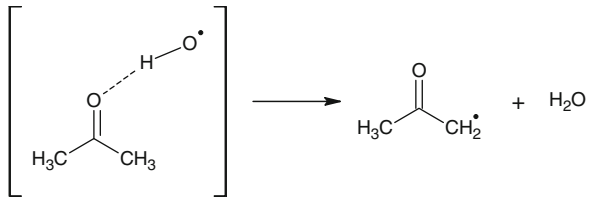
As can be seen from Eq. 7, the parameter triggering the susceptibility of reaction site H for OH attack is the local charge-limited donor energy at that site, $EQ_{occ}(q, r_H)$ (Eq. 4), donating charge $q = 0.12$ in the initial course of the (indirectly modelled) reaction in response to the oxidizing power of OH. Summation over all respective sites with H attached to aliphatic carbon yields the first term of Eq. 1. The agreement between calculated and experimental values is very good ($q^2 = 0.97$, rms = 0.26), slightly better than for MOOH-AM1 ($q^2 = 0.96$, rms = 0.28) [3], and outperforms the Atkinson method ($q^2 = 0.93$, rms = 0.39).

### 3.2 Enhanced reactivity for H abstraction from aliphatic carbon

Hydrogen abstraction reactions of ketones, esters, alcohols and carboxylic acids proceed via the formation of H-bonded pre-reactive complexes between the OH radical and the oxygen lone pairs of the respective functional group [10, 23–25, 27–30]. Taking acetone as an example, the following two-step mechanism is assumed.

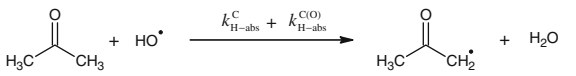



(8)




(9)

In this way, initial stabilization of the OH radical through the oxygen lone pair has an effect on the overall H abstraction rate from aliphatic carbon:




(10)

The MOOH approach applies the following equations to describe the lone pair effect on H abstraction from aliphatic carbon for a given reaction site $r_H$:

$$\Delta k^{C(O)}_{H-abs}(r_H) = k^{C}_{H-abs}(r_H)\cdot\sum_i w_i^{lp}\cdot f_i^{steric}(r_H) \quad (11)$$

$$k^{C(O)}_{H-abs}(r_H) = k^{C}_{H-abs}(r_H) + \Delta k^{C(O)}_{H-abs}(r_H) = k^{C}_{H-abs}(r_H)\cdot\left(1+\sum_i w_i^{lp}\cdot f_i^{steric}(r_H)\right) \quad (12)$$

Following the original approach of Klamt, the reactivity enhancement at a given reaction site $r_H$ is taken into account through a lone pair factor $w_i^{lp}$ and a steric factor $f_i^{steric}(r_H)$. For each $r_H$, the summation is carried out over all relevant oxygen lone pairs. The sum of the terms $k^{C}_{H-abs}(r_H)$ (Eq. 7) and $\Delta k^{C(O)}_{H-abs}(r_H)$ (Eq. 11) describes the local reactivity at reaction site $r_H$; the overall reaction rate for hydrogen abstraction from aliphatic carbons is obtained by summation over all reaction sites $r_H$. In Eq. 11, $f_i^{steric}(r_H)$ describes the spatial availability of the H atom for the attacking OH as a function of the distance between oxygen and the reaction site H of interest, $r_H$.

$$f_i^{\text{steric}}(r_H) = \exp\left[-\frac{\left(d_i^{lp}(r_H) - d_{opt}^{lp}\right)^2}{2\,t^2}\right] \quad (13)$$

In Eq. 13, $d_i^{lp}(r_H)$ denotes the distance between H (attached at aliphatic carbon) and oxygen $i$ carrying the lone pair, $d_{opt}^{lp}$ is the respective optimal distance and $t$ indicates a tolerance in the distance that was set to 0.7 Å. The original method derived $d_{opt}^{lp}$ from AM1 calculations, using a standard hydrogen bond length of 2.16 Å [9]. For the updated AM1-MOOH [3] and the present MOOH-HF calibration, the $d_{opt}^{lp}$ values were derived from UB3LYP/6-31G** geometry optimizations of the H-bonded complexes between the OH radical and the organic compound (see Eq. 8) and employed in Eq. 13 to calculate $f_i^{\text{steric}}(r_H)$.

Acetone, acetic acid, acetic acid methyl ester and alcohol were used as reference substances representing their respective compound class. Geometry optimizations and frequency calculations were carried out for the respective H-bonded complexes with OH. The resulting hydrogen bond distances are shown in Table 2, and range from 1.83 Å for the alcohol pre-reactive complex to 1.89 Å for the acetone complex. These hydrogen bond distances and the corresponding O–H···O angles were used to determine the $d_{opt}^{lp}$ values. The impact of $d_{opt}^{lp}$ on the overall performance of the model turned out to be rather small. Therefore, no additional further efforts were undertaken to derive further optimized $d_{opt}^{lp}$ values. A more detailed computational analysis of this H abstraction mechanism via a pre-activated complex is a subject of an ongoing study.

The lone pair factor $w_i^{lp}$ is a constant parameter specific for the respective functional group, and describes the susceptibility of the oxygen lone pair to interact with OH. As in our previous study [3], three factors were used to describe the different oxygen lone pairs in oxygenated organic compounds. The first value, $w_{C=O}^{lp}$, is used to characterize the carbonyl oxygen of ketones and carboxylic acids. As ester carbonyl oxygen shows a distinctly less enhancing effect on hydrogen atom reactivity, a second value $w_{OOR}^{lp}$ is introduced to account for this finding. The third factor $w_{OH}^{lp}$ refers to the oxygen lone pair effect of alcohols. In accord with experimental $k_{OH}$ data, the $sp^3$ oxygen of ester functions as well as of ether oxygen are not considered as enhancing the reactivity for H abstraction. The abstraction of hydroxylic hydrogen atoms in alcohols and acidic hydrogen atoms in carboxylic acids is taken into account by the constant contributions of $k_{H-abs}^{alc}(r_H)$ and $k_{H-abs}^{acid}(r_H)$.

**Table 2** Hydrogen bond distances between the OH radical hydrogen and the oxygens of acetone, acetic acid, acetic acid methyl ester and alcohol (see H-bonded complex formation according to Eq. 8)

| Compound | O···H·[Å] |
|---|---|
| [$CH_3COCH_3$···HO·] | 1.87 |
| (E)-[$CH_3COOH$···HO·] | 1.89 |
| (E)-[$CH_3COOCH_3$···HO·] | 1.86 |
| (Z)-[$CH_3COOCH_3$···HO·] | 1.88 |
| [$CH_3COH$···HO·] | 1.83 |

Geometry optimizations were carried out at the UB3LYP/6-31G** level of theory using the Gaussian 03 program package [13]. Frequency calculations were conducted to confirm the minima. Note that for (Z)-[$CH_3COOCH_3$···HO·], one spurious imaginary frequency of less than 5i $cm^{-1}$ was found. However, based on examination of the frequency vectors, we assume that this imaginary frequency is caused by numerical imprecision

Among the formyl esters, the formyl H attached to carbonyl carbon, HC(=O)OR, offers an additional site of H abstraction that must not be neglected [31]. This reaction path mechanistically differs from the oxygen-enhanced H abstraction from $sp^3$ carbon, but shows analogy to the abstraction of aldehydic H atoms. Therefore and in accord with our previous MOOH-AM1 calibration, formyl esters were excluded from calibration and only used for validation, treating the reaction as aldehydic H abstraction (see below) [3]. The same approach was used for the HF/6-31G**-level model.

For the 148 oxygenated compounds with a lone pair effect as described, calibration at the quantum chemical HF/6-31G** level yielded the following lone pair factors:

$$w_{C=O}^{lp} = 13.9$$
$$w_{OOR}^{lp} = 4.72 \quad (14)$$
$$w_{OH}^{lp} = 1.52$$

Moreover, the MOOH constants to account for H abstraction from hydroxyl groups ($k_{H-abs}^{alc}$) as well as from carboxylic acids ($k_{H-abs}^{acid}$) turned out to be

$$k_{H-abs}^{alc}(r_H) \approx 0.025$$
$$k_{H-abs}^{acid}(r_H) \approx 0.26 \quad (15)$$

Note that the data set used for calibration also contains compounds showing two or more of the respective functional groups, following the calibration strategy specified above. For a given alcohol or carboxylic acid, addition of the $k_{H-abs}^{alc}(r_H)$ or $k_{H-abs}^{acid}(r_H)$ values per OH or OOH functionality yields the term $\sum k_{H-abs}^{OH}$, describing the contribution of H abstraction from hydroxyl groups in Eq. 1. Further, note that because the term $k_{H-abs}^{C}$ was calculated according to Eq. 7, the overall rate constant $k_{OH}$ consists of a predicted and a fitted part, thus leading to a small difference between calibration ($r^2$) and prediction ($q^2$) statistics.

With regard to this subset of 148 oxygenated compounds, MOOH performs significantly better than MOOH-AM1 ($q^2$ 0.84 vs. 0.78), with an overall prediction quality comparable to the Atkinson method ($q^2 = 0.84$).

The MOOH performance increased for the monofunctional subsets of esters, ketones, alcohols and acids, and here the subset of ketones showed the largest improvement (HF/6-31G**: $q^2 = 0.85$, AM1: $q^2 = 0.80$). For the group of bi- and multifunctional oxygenated compounds, MOOH-HF yields the overall largest increase in prediction performance as compared to the AM1-based model ($q^2$ 0.76 vs. 0.49), suggesting that within the MOOH perturbational approach, the interactions between the respective functional groups cannot be covered adequately at the AM1 level.

Inspection of $k_{OH}$ for selected ketones as compared to related alkanes provides mechanistic insight into its dependence on the electronic structure at the relevant C–H sites as related to the oxygen lone pair effect. As can be seen from Table 3, $k_{OH}$ is slightly smaller for acetone (as simplest ketone) than for ethane (as corresponding alkane) with regard to both experimental and MOOH-calculated values. However, $k^{C}_{H-abs}$ (Eq. 7) yields a significantly smaller value for acetone (0.04) than for ethane (0.24), which results from the impact of the oxygen lone pair on the relevant H sites in terms of the associated charge-limited donor energies $EQ_{occ}$ (see Eqs. 4 and 7). In this case, the additional lone pair effect in terms of $\Delta k^{C(O)}_{H-abs}$ (Eq. 11) almost (but not fully) compensates for this decrease in intrinsic reactivity, raising $k_{OH}$ by 0.27 to the final predicted value of 0.31 for acetone. The situation is similar for 2-butanone with a larger remaining difference in the overall $k_{OH}$ as compared to *n*-butane. For 2-pentanone, the direct lone pair contribution to $k_{OH}$ in terms of $\Delta k^{C(O)}_{H-abs}$ is already larger than the $k^{C}_{H-abs}$ result for *n*-pentane (3.85 vs. 3.02), the latter of which is ca. fourfold larger than the corresponding alkane contribution to $k_{OH}$ of 2-pentanone (3.02 vs. 0.82). This analysis shows that in case of ketones, the intrinsically reduced reactivity for the OH-mediated H abstraction as compared to corresponding alkanes is at least partly (and sometimes more than fully) compensated for by the oxygen lone pair effect as modelled through Eq. 11. The final balance of the two opposing contributions of the oxygen lone pair to $k_{OH}$ thus depends in a subtle way on the details of its electronic structure impact on the relevant reaction sites.

### 3.3 Aldehydic reactivity

For most aldehydes, abstraction of the aldehydic H is the dominating reaction pathway [21, 22, 31]. In contrast to the H abstraction reactions described above, this reaction does not take place at aliphatic carbon, but at $sp^2$ carbon. Therefore, a new equation is required to parameterize the respective aldehydic reactivity.

Additionally, hydrogen abstraction from $sp^3$ carbon contributes to the overall $k_{OH}$ for aldehydes, including the above-described reactivity enhancement through the carbonyl oxygen lone pairs. In analogy to the AM1-based models, we use the ketone lone pair factor $w^{lp}_{C=O}$ and the respective optimal distance $d^{lp}_{opt}$ also for aldehydes. Accounting for H abstraction from aliphatic $sp^3$ carbon through Eqs. 7 and 11–13, calibration for the subset of 29 aldehydes with experimental $k_{OH}$ values resulted in the following MOOH-HF equation for aldehydic H:

$$k^{CHO}_{H-abs}(r_H) = 10 * *\{1.07 \cdot EQ_{occ}(0.12, r_H) + 13.3\} \quad (16)$$

According to Eq. 16, the abstraction of aldehydic H from carbonyl carbon is triggered by the charge-limited donor energy at reaction site $r_H$, $EQ_{occ}(q, r_H)$ with $q = 0.12$, and thus by the energy associated with donating the charge of 0.12 electrons during the rate-determining step of the (not explicitly modelled) reaction. This model gives very good results ($q^2 = 0.95$) and outperforms the MOOH-AM1 method ($q^2 = 0.89$) as well as the Atkinson method ($q^2 = 0.57$).

### 3.4 Increased performance for oxygenated compounds

Generally, oxygenated compounds show the largest difference in performance between MOOH-HF and MOOH-AM1.

**Table 3** Experimental and predicted $k_{OH}$ values for selected ketones and related alkanes

| Ketone | Exp. $k_{OH}$ | Predicted $k_{OH}$ | | | Alkane | Exp. $k_{OH}$ | Predicted $k_{OH}$ |
|---|---|---|---|---|---|---|---|
| | | $k^{C}_{H-abs}$ (Eq. 7) | $\Delta k^{C(O)}_{H-abs}$ (Eq. 11) | $k^{C(O)}_{H-abs}$ (Eq. 12) | | | $k^{C}_{H-abs}$ (Eq. 7) |
| Acetone | 0.18 | 0.04 | 0.27 | 0.31 | Ethane | 0.24 | 0.33 |
| n.a.[a] | | | | | Propane | 1.10 | 1.26 |
| 2-Butanone | 1.10 | 0.24 | 1.25 | 1.49 | *n*-Butane | 2.30 | 2.14 |
| 2-Pentanone | 3.14–4.66 | 0.82 | 3.85 | 4.67 | *n*-Pentane | 3.80 | 3.02 |
| 2-Hexanone | 6.37 | 1.42 | 5.63 | 7.05 | n.a.[a] | | |

For the ketones, both the alkane contribution ($k^{C}_{H-abs}$, Eq. 7) and the oxygen lone pair contribution ($\Delta k^{C(O)}_{H-abs}$, Eq. 11) to $k_{OH}$ are specified in addition to the predicted overall rate constant ($k^{C(O)}_{H-abs}$, Eq. 12). The rate constant $k_{OH}$ is given in $10^{-12}$ $cm^3$ $molecule^{-1}$ $s^{-1}$

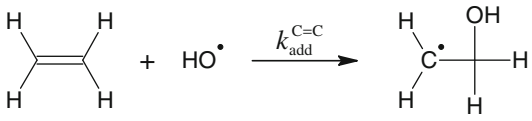
[a] No suitable compound with experimental $k_{OH}$ available

To illustrate this finding, Table 4 shows experimental and calculated log $k_{OH}$ for selected ketones, esters, alcohols, multifunctional oxygenated compounds and aldehydes (three compounds per compound class). Taking 3,3,3-trifluoro-propanal as an example, the MOOH prediction error reduces from 0.48 to 0.08 when going from AM1 to HF-6-31G**. Among the subset of aldehydes, HF-MOOH particularly seems to improve results for halogenated species.

The largest difference in log $k_{OH}$ prediction is found for 3-hydroxy-2-butanone (acetoin, exp. log $k_{OH}$:–11.0; MOOH-HF, log $k_{OH}$:–10.9; MOOH-AM1, log $k_{OH}$: –12.0), and 3,3,4,4,5,5,6,6,6-nonafluoro-1-hexanol shows the largest prediction error with MOOH-HF (exp. log $k_{OH}$:–13.0, MOOH-HF log $k_{OH}$:–11.8, MOOH-AM1 log $k_{OH}$:–11.9). For the subset of 148 ketones, alcohols, carboxylic acids and esters, the MOOH-HF rms is smaller by 0.05 log units (rms 0.32 vs. 0.37), with an intercorrelation of the prediction errors of $r^2 = 0.66$. For the subset of 29 aldehydes, the MOOH-HF rms is smaller by 0.08 log units (rms 0.19 vs. 0.27) and the intercorrelation of the prediction errors $r^2 = 0.54$.

### 3.5 OH addition to C=C bonds

The addition of OH radicals to $sp^2$ carbon involved in C=C bonds is the major degradation mechanism for olefinic compounds [21, 22]:




(17)

As an additional but minor reaction pathway, H abstraction from $sp^3$ carbon atoms was taken into account using Eq. 7.

The site-specific reactivity of the $sp^2$ carbon atoms is characterized by the charge-limited $\pi$-electron donor charge $EQ^{\pi}_{occ}(0.32, r_C)$ (Eq. 4 confined to $\pi$ electron MOs, donation charge $q = 0.32$ at relevant $sp^2$ carbon site $r_C$ determined trough successive variation during the nonlinear calibration procedure) and by the energy-limited acceptor charge $QE_{vac}(4.2, r_C)$ (Eq. 6, penetration energy $\varepsilon = 4.2$ eV at relevant $sp^2$ carbon site $r_C$). For the 123 alkenes with experimental $k_{OH}$ values, nonlinear regression yielded the following model equation:

$$k_{add}^{C=C}(r_C) = \exp\left\{1.42 \cdot EQ^{\pi}_{occ}(0.32,\ r_C) - 0.79 \cdot QE_{vac}(4.2,\ r_C) + 16.41\right\} \quad (18)$$

Summation over all olefinic reaction sites, $r_C$ of a molecule yields the term $\sum k_{add}^{C=C}$ and describes the contribution of OH addition to double bonds to the overall OH reactivity of the compound. For this reaction pathway, again MOOH-HF performs better than MOOH-AM1, and both MOOH models yield much better results than the Atkinson method ($q^2$: 0.90 vs. 0.88 vs. 0.51, rms: 0.21 vs. 0.24 vs. 0.48; see Table 5).

**Table 4** Experimental and calculated log $k_{OH}$ for selected ketones, esters, alcohols, multifunctional oxygenated compounds and aldehydes

| Subset | Chemical name | CAS | Experimental | MOOH-AM1 | MOOH-HF |
|---|---|---|---|---|---|
| Aldehydes | Formaldehyde | 000050-00-0 | −11.07 | −10.75 | −11.05 |
| | Glyoxal | 000107-22-2 | −10.96 | −10.56 | −11.11 |
| | 3,3,3-Trifluoro-propanal | 000460-40-2 | −11.52 | −11.99 | −11.60 |
| Alcohols | 1-Butanol | 000071-36-3 | −11.08 | −10.58 | −10.95 |
| | 1-Pentanol | 000071-41-0 | −10.95 | −10.60 | −10.91 |
| | 2,3-Butanediol | 000513-85-9 | −10.63 | −11.61 | −11.21 |
| Esters | Methyl acetate | 000079-20-9 | −12.48 | −11.91 | −12.21 |
| | Methyl difluoroacetate | 000433-53-4 | −12.83 | −12.61 | −12.81 |
| | Acetic acid, methoxy-, ethyl ester | 003938-96-3 | −11.08 | −11.30 | −11.07 |
| Ketones | 2-Pentanone | 000107-87-9 | −11.37 | −11.12 | −11.33 |
| | 2-Hexanone | 000591-78-6 | −11.11 | −10.83 | −11.15 |
| | Nopinone | 024903-95-5 | −10.80 | −10.45 | −10.70 |
| Bi- and multifunctional oxygenated compounds | Acetoin | 000513-86-0 | −10.99 | −12.00 | −10.93 |
| | 2-Hydroxypropanoic acid, methyl ester | 000547-64-8 | −11.55 | −12.05 | −11.51 |
| | 4-Hydroxy-3-hexanone | 004984-85-4 | −10.82 | −11.46 | −11.05 |

Log $k_{OH}$ refers to $k_{OH}$ in $10^{-12}$ $cm^3$ $molecule^{-1}$ $s^{-1}$

**Table 5** Statistical performances of the MOOH-HF method, the re-calibrated MOOH-AM1 method, the original Klamt parameterization and the Atkinson increment scheme

| Data set | Method | $n$ | $r^2$ | $q^2$ | rms | bias | me | mne | mpe |
|---|---|---|---|---|---|---|---|---|---|
| Mechanistic domain | MOOH-HF | 799 | 0.95 | 0.95 | 0.29 | −0.02 | 0.22 | −1.12 | 1.21 |
| | MOOH-AM1 | 799 | 0.94 | 0.94 | 0.32 | −0.01 | 0.23 | −1.28 | 1.34 |
| | Klamt | 799 | 0.93 | 0.93 | 0.35 | −0.06 | 0.25 | −1.28 | 1.34 |
| | Atkinson | 799 | 0.93 | 0.92 | 0.35 | 0 | 0.20 | −2.56 | 2.20 |
| Training set | MOOH-HF | 670 | 0.96 | 0.96 | 0.27 | 0 | 0.20 | −0.96 | 1.21 |
| | MOOH-AM1 | 670 | 0.95 | 0.95 | 0.29 | 0 | 0.21 | −1.03 | 1.08 |
| | Klamt | 670 | 0.94 | 0.93 | 0.33 | −0.05 | 0.24 | −1.28 | 1.11 |
| | Atkinson | 670 | 0.92 | 0.92 | 0.37 | 0 | 0.21 | −2.56 | 2.20 |
| Test set (extrapolation) | MOOH-HF | 129 | 0.91 | 0.88 | 0.40 | −0.13 | 0.31 | −1.12 | 1.14 |
| | MOOH-AM1 | 129 | 0.87 | 0.86 | 0.43 | −0.08 | 0.32 | −1.28 | 1.34 |
| | Klamt | 129 | 0.87 | 0.85 | 0.45 | −0.10 | 0.34 | −1.24 | 1.34 |
| | Atkinson | 129 | 0.95 | 0.95 | 0.26 | −0.02 | 0.19 | −0.49 | 1.02 |
| Saturated hydrocarbons and ethers | MOOH-HF | 230 | 0.97 | 0.97 | 0.26 | 0 | 0.20 | −0.86 | 0.77 |
| | MOOH-AM1 | 230 | 0.96 | 0.96 | 0.28 | 0 | 0.21 | −1.02 | 1.02 |
| | Klamt | 230 | 0.96 | 0.96 | 0.30 | −0.01 | 0.23 | −1.09 | 1.07 |
| | Atkinson | 230 | 0.93 | 0.93 | 0.39 | 0.01 | 0.24 | −1.72 | 1.27 |
| Hydrogen bond acceptors (no aldehydes) | MOOH-HF | 148 | 0.84 | 0.84 | 0.32 | −0.01 | 0.24 | −0.89 | 1.21 |
| | MOOH-AM1 | 148 | 0.79 | 0.78 | 0.37 | −0.01 | 0.29 | −1.03 | 1.08 |
| | Klamt | 148 | 0.76 | 0.70 | 0.43 | −0.19 | 0.33 | −1.22 | 0.91 |
| | Atkinson | 148 | 0.85 | 0.84 | 0.31 | 0.03 | 0.19 | −0.87 | 1.61 |
| Ketones | MOOH-HF | 28 | 0.86 | 0.85 | 0.27 | 0.02 | 0.20 | −0.89 | 0.42 |
| | MOOH-AM1 | 28 | 0.81 | 0.80 | 0.32 | 0.07 | 0.25 | −1.03 | 0.36 |
| | Klamt | 28 | 0.81 | 0.80 | 0.31 | −0.07 | 0.20 | −1.16 | 0.24 |
| | Atkinson | 28 | 0.89 | 0.89 | 0.23 | −0.01 | 0.15 | −0.80 | 0.38 |
| Esters (no formyl esters) | MOOH-HF | 38 | 0.90 | 0.90 | 0.24 | −0.02 | 0.18 | −0.70 | 0.51 |
| | MOOH-AM1 | 38 | 0.87 | 0.87 | 0.27 | −0.01 | 0.19 | −0.77 | 0.59 |
| | Klamt | 38 | 0.89 | 0.79 | 0.34 | −0.23 | 0.28 | −0.94 | 0.41 |
| | Atkinson | 38 | 0.88 | 0.88 | 0.26 | −0.03 | 0.18 | −0.64 | 0.93 |
| Alcohols | MOOH-HF | 55 | 0.80 | 0.79 | 0.39 | 0 | 0.30 | −0.74 | 1.21 |
| | MOOH-AM1 | 55 | 0.77 | 0.77 | 0.41 | 0.02 | 0.33 | −0.98 | 1.08 |
| | Klamt | 55 | 0.76 | 0.67 | 0.49 | −0.20 | 0.39 | −1.22 | 0.91 |
| | Atkinson | 55 | 0.86 | 0.79 | 0.39 | 0.11 | 0.22 | −0.37 | 1.61 |
| Acids | MOOH-HF | 6 | 0.73 | 0.69 | 0.26 | −0.06 | 0.19 | −0.42 | 0.30 |
| | MOOH-AM1 | 6 | 0.71 | 0.66 | 0.27 | −0.07 | 0.22 | −0.42 | 0.30 |
| | Klamt | 6 | 0.66 | 0.46 | 0.35 | 0.16 | 0.20 | −0.11 | 0.73 |
| | Atkinson | 6 | 0.76 | 0.65 | 0.28 | 0.12 | 0.16 | −0.10 | 0.60 |
| Bi- and multifunctional oxygenated compounds | MOOH-HF | 21 | 0.76 | 0.76 | 0.35 | −0.03 | 0.29 | −0.58 | 0.57 |
| | MOOH-AM1 | 21 | 0.62 | 0.49 | 0.50 | −0.16 | 0.43 | −1.01 | 0.76 |
| | Klamt | 21 | 0.57 | 0.30 | 0.59 | −0.33 | 0.49 | −1.19 | 0.82 |
| | Atkinson | 21 | 0.86 | 0.83 | 0.29 | −0.06 | 0.19 | −0.87 | 0.71 |
| Aldehydes | MOOH-HF | 29 | 0.95 | 0.95 | 0.19 | 0.02 | 0.13 | −0.31 | 0.63 |
| | MOOH-AM1 | 29 | 0.90 | 0.89 | 0.27 | 0.02 | 0.21 | −0.48 | 0.79 |
| | Klamt | 29 | 0.90 | 0.88 | 0.28 | 0.06 | 0.19 | −0.35 | 1.11 |
| | Atkinson | 29 | 0.65 | 0.57 | 0.55 | 0.22 | 0.32 | −0.32 | 2.20 |
| Alkenes | MOOH-HF | 123 | 0.90 | 0.90 | 0.21 | 0.01 | 0.15 | −0.76 | 0.76 |
| | MOOH-AM1 | 123 | 0.88 | 0.88 | 0.24 | 0.01 | 0.16 | −0.67 | 0.99 |
| | Klamt | 123 | 0.85 | 0.81 | 0.30 | −0.01 | 0.18 | −1.28 | 0.86 |
| | Atkinson | 123 | 0.62 | 0.51 | 0.48 | −0.05 | 0.23 | −2.56 | 1.29 |

**Table 5** continued

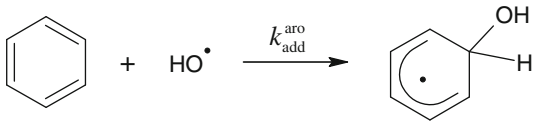
| Data set | Method | $n$ | $r^2$ | $q^2$ | rms | bias | me | mne | mpe |
|---|---|---|---|---|---|---|---|---|---|
| Aromatics | MOOH-HF | 133 | 0.89 | 0.89 | 0.28 | 0 | 0.20 | −0.96 | 0.87 |
| | MOOH-AM1 | 133 | 0.90 | 0.90 | 0.27 | 0.01 | 0.19 | −0.89 | 0.90 |
| | Klamt | 133 | 0.89 | 0.88 | 0.29 | −0.03 | 0.22 | −1.06 | 0.95 |
| | Atkinson | 133 | 0.94 | 0.94 | 0.21 | −0.02 | 0.14 | −0.65 | 0.83 |
| Alkynes | MOOH-HF | 7 | 0.96 | 0.96 | 0.10 | −0.01 | 0.07 | −0.12 | 0.18 |
| | MOOH-AM1 | 7 | 0.95 | 0.94 | 0.12 | 0.03 | 0.09 | −0.16 | 0.18 |
| | Klamt | 7 | 0.97 | 0.96 | 0.10 | 0.04 | 0.08 | −0.14 | 0.13 |
| | Atkinson | 7 | 0.99 | 0.99 | 0.06 | −0.01 | 0.04 | −0.08 | 0.08 |

The total set contains 799 compounds, of which 670 compounds were used as training set for MOOH-HF

$n$ number of compounds, $q^2$ predictive squared correlation coefficient (Eq. 2; see [17]), *rms* root-mean-square error, bias systematic error, *me* mean error, *mne* maximum negative error (largest underestimation), *mpe* maximum positive error (largest overestimation)

### 3.6 OH addition to aromatic rings

Addition of OH radicals to aromatic carbon is the dominating degradation pathway for most aromatic compounds [21, 22]:




(19)

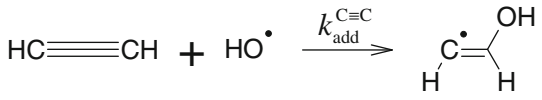
As a minor reaction mechanism, H abstraction from alkyl substituent groups is taken into consideration (Eq. 7). Calibrating the MOOH model with the subset of 133 aromatic compounds, we derived the following equation:

$$k_{add}^{aro}(r_C) = \exp\left\{\frac{8.76}{(1+\exp\{-1.92\cdot(EE_{occ}(2.96, r_C)+10.65)\})} - 0.082\cdot\Delta_{def}^{aro} - 3.48\right\} \quad (20)$$

Now, the main electronic factor governing $k_{OH}$ at aromatic carbon is the effective donor energy $EE_{occ}(E_{ref}, r_C)$ with $E_{ref} = 2.96$ eV, where the actual contribution of occupied MO energies $E_i$ is triggered by both $\exp(E_i/E_{ref})$ and by the local electron density of these MOs at reaction site $r_C$ (see Eq. 5). Moreover, the substituent deformation energy associated with transition state formation is taken into account through parameter $\Delta_{def}^{aro}$ (see above and Table 1). Surprisingly, the MOOH-HF results ($q^2 = 0.89$) are slightly inferior to the MOOH-AM1 results ($q^2 = 0.90$). However, the difference observed is only marginal and may also be caused by experimental and numerical imprecision. The widely applied Atkinson method still performs best for this subset of compounds (see Table 5).

### 3.7 OH addition to C≡C bonds

The set of available experimental data contains only seven alkynes [3]. This subset was used for calibration of the current MOOH models. The major reaction mechanism is the addition of the OH radical to the sp carbon:




(21)

Again, H abstraction from aliphatic carbon was taken into account as a minor reaction path (Eq. 7). Nonlinear optimization yielded the following regression equation:

$$k_{add}^{C\equiv C}(r_C) = \exp\{2.74\cdot EQ_{occ}(0.32, r_C) + 29.4\} \quad (22)$$

It follows that according to the MOOH model, OH addition to C≡C bonds is governed by the charge-limited donor energy at the sp carbon, $EQ_{occ}(q, r_C)$ with $q = 0.32$, and depends on the energy required to donate 0.32 electrons from this site. The statistics are slightly better than with MOOH-AM1, but still inferior to the results of the Atkinson method (see Table 5). Due to the small set of data, reliability of both the Atkinson approach and the MOOH models is limited concerning this compound class.

### 3.8 Multimechanistic compounds

The compound classes addressed so far consist of substances that react with OH radicals via a dominating reaction mechanism, including only H abstraction from aliphatic carbon and, in the case of aldehydes, enhanced reactivity due to the oxygen lone pair of the –CHO group as a minor contribution. However, the degradation of multifunctional compounds is determined by the

concurrent presence of several fast degradation pathways in addition to H abstraction from $sp^3$ carbon (if applicable). With regard to our data set of 799 organic compounds with experimental reaction rate constants $k_{OH}$, this concerns 12 formyl esters, 6 hydroxylated aldehydes, 3 keto-aldehydes, 2 hydroxylated alkynes, 29 aromatics with oxidized and/or olefinic side chains and 64 oxidized alkenes that have been collected in a former study [3]. Accordingly and as mentioned above, these compounds were excluded from calibration to avoid cross-contamination of pathway-specific model parameters. Because former studies revealed a poor model performance for halomethanes [3, 9], this subset of 14 compounds was also removed from the training set. The resultant 129 mostly bi- and multifunctional compounds were used as an external test set to study the prediction performance of MOOH-HF.

Note, however, that this validation set would usually be considered outside the chemical domain of the MOOH model, because the latter had been calibrated predominantly with hydrocarbons and monofunctional compounds. As aforementioned, the model equation for the enhancement of H abstraction through lone pairs of neighbour oxygen ($k_{H-abs}^{C(O)}$) was calibrated using a subset that also contained bi- and multifunctional oxygenated compounds. However, these oxygenated compounds only contain functional groups that are reaction sites for one uniform mechanism, whereas the multifunctional compounds in this test set offer reaction sites for different fast reaction mechanisms. Thus, application of MOOH-HF to this external test set examines whether and to what degree the model is robust with respect to extrapolation from mono- to multisubstitution and from one dominating reaction mechanism to the summation of multiple significant reaction pathways.

Figure 1 shows the distribution of externally predicted versus experimental log $k_{OH}$ for the validation set of 129 compounds. The resultant statistics ($q^2 = 0.88$, rms = 0.40) is superior to the results achieved with MOOH-AM1 ($q^2 = 0.86$, ms = 0.43) [3], and shows that multifunctional compounds can already be treated with reasonable statistics without prior calibration. The Atkinson method calibration had included compounds with respective substitution patterns, and still yields best results for this subset ($q^2 = 0.95$, rms = 0.26).

### 3.9 Overall statistics

Figure 2 shows the data distribution of predicted versus experimental log $k_{OH}$ for the HF/6-31G**-level MOOH model. For the total set of 799 organic compounds, MOOH-HF ($q^2 = 0.95$, rms = 0.29) is slightly superior to MOOH-AM1 ($q^2 = 0.94$, rms = 0.32) and to the Atkinson increment method ($q^2 = 0.92$, rms = 0.35). For

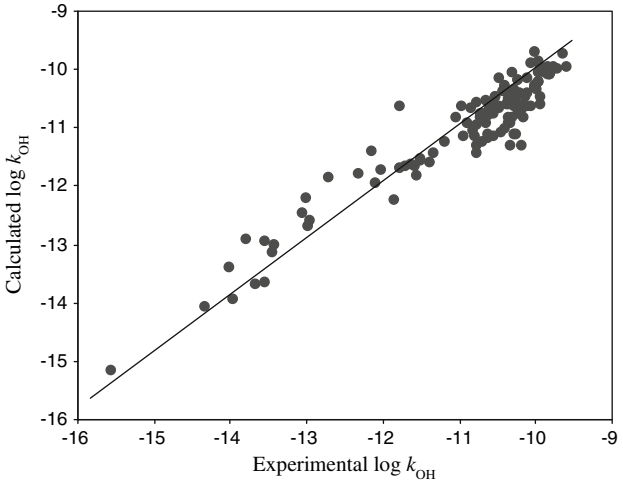



**Fig. 1** Calculated versus experimental log $k_{OH}$ (with $k_{OH}$ in $10^{-12}$ cm$^3$ molecule$^{-1}$ s$^{-1}$) for the test set of 129 compounds, yielding $r^2 = 0.91$, $q^2 = 0.88$ and rms = 0.40 (see Table 5 for further statistical parameters)

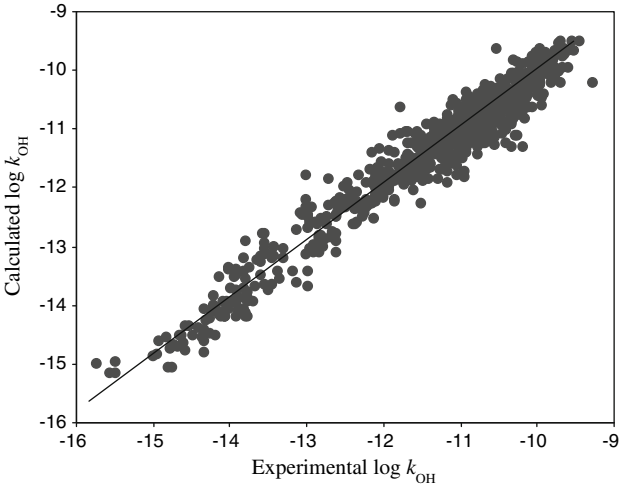



**Fig. 2** Calculated versus experimental log $k_{OH}$ (with $k_{OH}$ in $10^{-12}$ cm$^3$ molecule$^{-1}$ s$^{-1}$) for the total data set of 799 compounds, yielding the statistics $r^2 = 0.95$, $q^2 = 0.95$ and rms = 0.29 (see Table 5 for further statistical parameters)

oxygenated compounds such as aldehydes and ketones, however, MOOH-HF yields a significant improvement over the AM1-level MOOH model ($q^2$: 0.95 vs. 0.90 and 0.86 vs. 0.81; rms: 0.19 vs. 0.27 and 0.27 vs. 0.32). For most other compound classes, the difference in performance is small with regard to global statistics.

However, MOOH-HF yields less outliers than MOOH-AM1. With the HF/6-31G**-level approach, there are three compounds exceeding one order of magnitude as prediction error, compared to eight respective outliers observed when applying MOOH-AM1. The numbers of compounds exceeding 2 · rms are 46 for MOOH-HF (5.8%, 2 · rms = 0.58) and 49 for the MOOH-AM1

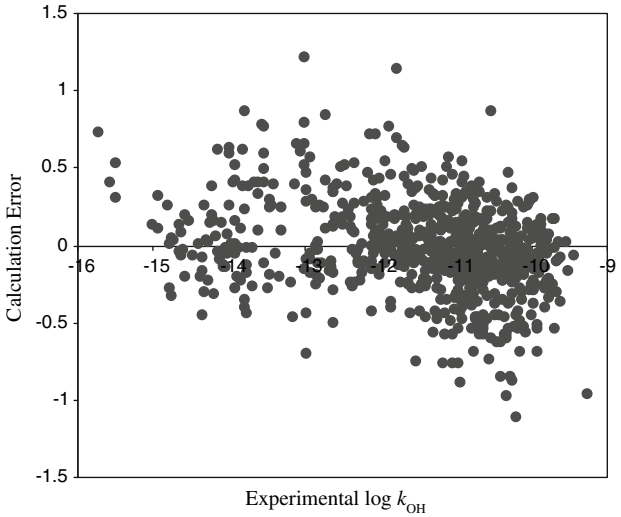



**Fig. 3** Logarithmic calculation error versus experimental log $k_{OH}$ (with $k_{OH}$ in $10^{-12}$ cm$^3$ molecule$^{-1}$ s$^{-1}$) for the total data set of 799 compounds (see Table 5 for information about the overall statistical performance)

method (6.1%, 2 · rms = 0.64). For the Atkinson method, 52 compounds exceed 2 · rms (6.5%, 2 · rms = 0.70), 24 of which exceed one order of magnitude as prediction error. The error distribution in terms of prediction error versus experimental log $k_{OH}$ is depicted in Fig. 3 and reveals that overestimations (positive prediction errors) of log $k_{OH}$ are observed more frequently with smaller log $k_{OH}$, whereas underestimations (negative prediction errors) occur more often with larger values for log $k_{OH}$, a trend that has also been observed for MOOH-AM1 and the Atkinson method [3]. Attention should be paid to this finding for future work on the MOOH approach.

With regard to the overall prediction errors, MOOH-HF shows a significant correlation with MOOH-AM1 ($r^2 = 0.62$), but appears to be essentially independent from the Atkinson method ($r^2 = 0.10$). It follows that for screening untested or even unsynthesized compounds as well as for predictive applications in the regulatory context, a consensus modelling approach is recommended, using approaches of both the Atkinson and the MOOH models, thus increasing confidence in the prediction significantly in those cases where these approaches yield results of sufficient agreement.

The MOOH method can be easily applied by the combined use of the Gaussian 03 package and the MOOH software. A licensed copy of an executable program to automatically predict OH radical rate constants, $k_{OH}$, from Gaussian 03 output files can be obtained upon request from the corresponding author. Currently, both the Atkinson method and the MOOH method are confined to predicting reaction rate constants at 298 K. To explore the scope of modelling, the temperature dependence of $k_{OH}$, the starting point would be a collection and systematic analysis of a respective set of temperature-dependent $k_{OH}$ data. Note that with regard to the air–water partition coefficient (Henry's law constant in dimensionless form) as a free-energy molecular parameter, it had been possible to extend standard prediction schemes referring to 298 K by a group contribution method that allows one to predict the respective temperature variation, taking the (predicted or experimental) value at 298 K as reference for the substance of interest, for a wide range of organic compounds [32].

## 4 Conclusions

The MOOH method has been implemented on the HF/6-31G** level using accordingly calibrated local molecular reactivity parameters and an up-to-date set of experimental rate constants for the gas-phase reaction of organic compounds with OH radicals. For the overall set of 799 compounds, MOOH-HF is superior to the previously introduced AM1-level approach as well as to Atkinson's increment scheme. Significant improvements concern oxygenated compounds in general, and aldehydes and ketones in particular. The overall MOOH-HF prediction errors show almost no correlation with the ones of the Atkinson scheme, suggesting to combine both in a consensus modelling approach in order to still increase the confidence in prediction for compounds where both methods yield similar results. To extend the applicability of the MOOH method to substrates containing heteroatoms, such as nitrogen, sulphur and phosphorous, future studies may derive regression equations on further reaction mechanisms. The results of this study also encourage adapting the approach to further electronic structure methods such as the density functional theory level of calculation.

**Acknowledgments** The authors thank Dr. Andreas Klamt for making available the original MOOH source code. This research was financially supported by the European Commission through the projects NOMIRACLE (FP6 contract no. 003956) and OSIRIS (contract no. 037017).

## References

1. Prinn RG, Huang J, Weiss RF, Cunnold DM, Fraser PJ, Simmonds PG, McCulloch A, Harth C, Salameh P, O'Doherty S, Wang RHJ, Porter L, Miller BR (2001) Science 292:1882–1888
2. Kurylo MJ, Orkin VL (2003) Chem Rev 103:5049–5076
3. Böhnhardt A, Kühne R, Ebert R-U, Schüürmann G (2008) J Phys Chem A 112:11391–11399
4. Stockholm convention on persistent organic pollutants—declaration. OJ L 209, 31.7.2006, pp 3–29
5. Guidance on information requirements and chemical safety assessment. http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_en.htm?time=1246100282#A. Accessed 27 June 2009

6. Atkinson R (1985) Chem Rev 85:69–201
7. Atkinson R (1987) Int J Chem Kinetics 19:799–828
8. Kwok ESC, Atkinson R (1995) Atmos Environ 29:1685–1695
9. Klamt A (1993) Chemosphere 26:1273–1289
10. Klamt A (1996) Chemosphere 32:717–726
11. AOPWIN 1.91 (2000) U. S. Environmental Protection Agency, Washington, DC, USA
12. Meylan WM, Howard PH (1993) Chemosphere 26:2293–2299
13. Gaussian 03, Revision C.02, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pompelli C, Ochterski JW, Ayala PY, Morokuma Keiji, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian, Inc., Pittsburgh, PA, USA
14. Schüürmann G, Kühne R, Kleint F, Ebert R-U, Rothenbacher C, Herth PA (1997) In: Chen F, Schüürmann G (eds) QSAR in environmental sciences. SETAC Press, Pensacola, pp 93–114
15. Marquardt DW (1963) SIAM J Appl Math 11:431–441
16. Brandt S (1999) Datenanalyse, 4th edn. Spektrum Akademischer Verlag, Heidelberg
17. Schüürmann G, Ebert R-U, Chen J, Wang B, Kühne R (2008) J Chem Inf Model 48:2140–2145
18. Schwöbel J, Ebert R-U, Kühne R, Schüürmann G (2009) J Comput Chem 30:1454–1464
19. Schwöbel J, Ebert R-U, Kühne R, Schüürmann G (2009) J Chem Inf Model 49:956–962
20. Schwöbel J, Ebert R-U, Kühne R, Schüürmann G (2009) J Phys Chem A 113:10104–10112
21. Seinfeld JH, Pandis SN (1998) Atmospheric chemistry and physics: from air pollution to climate change. Wiley, New York
22. Atkinson R, Arey J (2003) Chem Rev 103:4605–4638
23. Vandenberk S, Vereecken L, Peeters J (2002) Phys Chem Chem Phys 4:461–466
24. Rosado-Reyes CM, Francisco JS (2006) J Phys Chem A 110:4419–4433
25. Aloisio S, Francisco JS (2000) J Phys Chem A 104:3211–3224
26. D'Anna B, Bakken V, Beukes JA, Nielsen CJ, Brudnik K, Jodkowski JT (2003) Phys Chem Chem Phys 5:1790–1805
27. Galano A, Alvarez-Idaboy JR, Bravo-Perez G, Ruiz-Santoyo ME (2002) Phys Chem Chem Phys 4:4648–4662
28. Galano A, Alvarez-Idaboy JR, Ruiz-Santoyo ME, Vivier-Bunge A (2002) J Phys Chem A 106:9520–9528
29. Vasvári G, Szilágyi I, Bencsura Á, Dóbé S, Bérces T, Henon E, Canneaux S, Bohr F (2001) Phys Chem Chem Phys 3:551–555
30. Smith I, Ravishankara AR (2002) J Phys Chem A 106:4798–4807
31. Mellouki A, Le Bras G, Sidebottom H (2003) Chem Rev 103:5077–5096
32. Kühne R, Ebert R-U, Schüürmann G (2005) Environ Sci Technol 39:6705–6711