AEROSOL
R e s e a r c h
Open Access

Supplement of

A theory-informed, experiment-based constraint on the rate of autoxidation chemistry – an analytical approach

Lukas Pichelstorfer et al.

Correspondence to: Lukas Pichelstorfer (office@pi-numerics.com)

Supporting Information

autoCONSTRAINT input

There are three types of inputs read in: i) a chemical scheme (without rate coefficients) containing the expected reaction pathways. In the present work, this scheme is the αpin-test scheme, an analogy to the α-pinene autoxidation scheme developed by Roldin et al., (2019). To understand the resulting change in atomic composition between reactants and products, atomic compositions of all scheme-species need to be provided. The current version of autoCONSTRAINT requires the chemical scheme to be submitted in KPP notation (KineticPreProcessor, https://people.cs.vt.edu/asandu/Software/Kpp; Daescu et al., 2003; Damian et al., 2002; Sandu et al., 2003).

Input ii) covers the experimental data including the nitrate chemical ionization mass spectrometer (NO3- CIMS; Ehn et al., 2014; Jokinen et al., 2012) peak list. It covers a file containing the detected peaks concentration and, in a second column, the change in concentration per second (this part is 0 in case of fully developed steady state. The unit is cm^{-3}. Further, the experimental input contains the peaks' atomic composition which allows to relate them to model isomers from input i). The experimental input is completed by the environmental conditions relevant: concentrations of HO_2 and NO, relative humidity, temperature. If available, the summed concentration of all peroxy radicals ($\Sigma RO_2\bullet$) is also read in. Otherwise, it is estimated by means of modelling.

Input iii) consists of modelling parameters that supplement the experimental input: HO_2, NO and $\Sigma RO_2\bullet$ concentrations, if not determined experimentally, are read in. Simulation results containing species names and concentrations (in this work: MCM simulation; note that the underlying MCM chemical scheme does not cover autoxidation chemistry) are read in to estimate the distribution of RO_2 species between the MCM and the autoxidation scheme, where the schemes overlap.

autoCONSTRAINT output:

The model output contains: i) a summary of key inputs relevant for understanding the results. This covers a) a list of the closed shell species (CS_names.txt); b) a list of peroxy and alkoxy radical species names present in the autoxidation scheme (RO2_names.txt, RO_names.txt); c) sorted (with respect to descending mass) experimental composition, experimental concentrations and change in concentration per second (exp_comp.txt, exp_conc.txt and exp_dconc.txt).

ii) additional outputs depict the interpretation of the chemistry (chem_interpretation.txt; line by line interpretation of the autoxidation scheme). Further, all experimentally observed mass peaks are listed in "match_out.txt" together with the number of model-isomers, the species type (closed shell "CS isomers", autoxidation scheme peroxy radical "RO2 isomers" or MCM peroxy radical "RO2mcm isomers") and an isomer identifier (integer number; the integer number of CS isomers

and RO2 isomers refer to their line number in files: CS_names.txt and RO2_names.txt). Details on the solution of equ. (2) (see main text) for every experimentally determined mass peak are given by the file named "getrate_out.txt": it lists the mass peaks (sorted by descending mass) and their specifications (concentration, physical loss per second and composition). Further, the number of model isomers and specifications on their production pathways are provided. Finally, the determined rate coefficients and their relative contribution to the observed change in peak concentration ($\Delta C(CHON)_x$) are listed. Accordingly, the "getrate_out.txt" is meant to be a human-readable file providing insight to the process of constraining rate coefficients.

Output part iii) covers a summary of the dependencies "I" (i.e., $\sum_{m,n \to i} \square$ and $\sum_{u \to i} \square$ in equ. (2) in the main text), the rate coefficients "k" determined and their relative contribution to the observed change "x". There is an output file for each chemical reaction type (e.g., "k_RO2_NO.txt" specifying the reaction products of RO_2 and NO) containing a number of columns for each reaction channel (e.g., in case of $RO_2\bullet$ + NO there are five columns for the reaction pathways R2a-R2d: $RONO_2$, RC=O, $RO\bullet$ and fragmentation products; the first column contains the sum of all reaction channels). The lines of the arrays (I, k and x) refer to the $RO_2\bullet$ species listed in "RO2_names.txt". Except for alkoxy related outputs ("*_RO_autox.txt" files) which are related to "RO_names.txt". Table S1 provides an overview on the columns' meanings in the output files.

Table S1: list of reaction types reported in the output files [stands for a) "k": the rate coefficient; b) "x": contribution to the observed change in concentration via this reaction pathway; c) "I": the index number of species formed by this reaction pathway with regard to files CS_names.txt, RO2_names.txt and RO_names.txt]. In the rate coefficient file, the first column represents the sum of all pathways. In the other files, the first column is left empty.*

Reaction type	Product channels (2-5)			
*_RO2_sumRO2	ROH (R2a)	RC=O (R2b)	RO• (R2c)	Fragm. (R2d)
*_RO2_HO2	ROOH (R3a)	RO• (R3b)	Fragm. (R3c)	--
*_RO2_NO	RONO2 (R4a)	RC=O (R4b)	RO• (R4c)	Fragm. (R4d)
*_RO2_autox.	RO2 (R5a)	RC=O (R5b)	Fragm. (R5c)	--
*_RO_autox.	RO2 (R6a)	RC=O (R6b)	Fragm. (R6c)	--

Equ. (2) (see main text) variables and their dimensions

$\Delta C(CHON)_x$ represents the change rate in concentration of atomic mass peak x with composition $C_r H_s O_t N_u$ (r,s,t,u are integer numbers $\geq$ 0) at a certain point in time. For each of the mass peak concentrations, there exists a number of model-isomer species. This number is zero in case the mass peak's composition doesn't match a scheme's species composition. The sum of individual isomers' concentration changes is equal to the concetration changes of considered mass peak:

$$\Delta C(CHON)_x = \sum_{CHON_i = CHON_x} \Delta C_i$$

Model isomer species concentrations change according to the reaction equations $1-6$. Reaction equations are represented by terms depicting bimolecular reactions:

$$\sum_{m,n \to i} k_m * [R_{1,m}] * [R_{2,n}]$$

and unimolecular reactions:

$$\sum_{u \to i} k_u * [R_u]$$

Finally, non-chemical changes are accounted for:

$$\Delta C_{phys}$$

Dimensions:

$\Delta C(CHON)_x$ is an array of size n (n is the number of CIMS mass peaks observed). For each mass peak there is a number of model-isomers. The number of reaction terms forming the isomers in bimolecular (reaction types 1-4) and unimolecular (reaction types 5 and 6) reactions determine the content of $\sum_{m,n \to i} k_m * [R_{1,m}] * [R_{2,n}]$ and $\sum_{u \to i} k_u * [R_u]$

For each mass peak x, the physical loss rate at the considered point in time is a single, positive, real number. Accordingly, its dimension is same as the mass peak list.

Reagants ($R_{1,m}$ and $R_{2,m}$) in equ. (2) both represent peroxy radical concentrations for reaction type (R1). For types (R2a-c), (R3a-c) and (R4a-d), R_1 represents the peroxy radical concentration, while R_2 depicts the $\Sigma RO_2\bullet$, HO_2 and NO, respectively. R_u can represent $RO_2\bullet$ and $RO\bullet$ species, depending on the reaction forming the isomer (see (R5a-c) and (R6a-c)).

Exemplary solution of equ. (2)

To ease comprehension of the described approach, an example is given below:

In the simulation based NO_3- CIMS-like data (see "Reproduction of k-rates from known scheme" in the main text), mass peak number **214** with the atomic composition $C_{10}H_{16}O_6N_0$ is considered. According to "mach_out.txt", it has 3 isomers in the input chemical scheme (see S1). Four reaction pathways form the isomers. Accordingly, equ. 1 has 4 chemical production terms.

```
exp. species- index; # of isomers(RO2,RO2mcm,CS): 214   0        0       3
CS isomers: [ 3. 24. 27. 50.]
RO2 isomers: [0.]
RO2mcm isomers: [0.]
```

Figure S1: screenshot from match_out.txt output file showing the isomers found for the mass peak 214. The first line states the peak-list index, followed by the number of model isomers (autoxidation scheme: "RO2", MCM scheme: "RO2,mcm" and closed shell species: "CS"). In this case, all isomers are closed shell (CS). The first number in the brackets of second to 4th line indicate the quantity of isomers followed their identity via line indices in CS_names.txt.

An overview of the mass peak information, physical sink terms and chemical production terms is provided by "getrate_out.txt" (see Fig. S2). Information on the production channels can be read as follows:

"NO: 16172772.831692643 APh_RO2_O7 11723145665.52773 C(NO) -> APh_O5_O 24 0 16 1"

The term starts with production pathway type "NO" peroxy radical APh_RO2_O7 (with concentration 1.62e+7 cm^{-3}) reacts with NO (with concentration 1.17e+10 cm^{-3}) to form "Aph_O5_O", a closed shell species of index "24" (referring to "CS_names.txt"). The number of the reaction pathway to this mass peak is "0". The index of the peroxy radical is given as well ("16", referring to "RO2_names.txt" as well as to the line number in k_RO2_NO.txt and I_RO2_NO.txt). The last parameter "1" refers to the column in k_RO2_NO.txt, x_RO2_NO.txt and I_RO2_NO.txt, this is where the rate coefficient, the share in the calculated change and the dependencies are saved, respectively.

```
exp species i with concentration & CHON: 214     278320.32711749745     278320327.11749744
[10. 16.  6.  0.]
# of isomers (exp,exp_list,RO2,RO2mcm) 3.0      [24. 27. 50.]   0.0     0.0
  NO: 16172772.831692643 APh_RO2_O7 11723145665.52773 C(NO) -> APh_O5_O 24 0 16 1
  RO2pool: 16172772.831692643 APh_RO2_O7 4645490000.0 C(RO2pool) -> APh_O5_O 24 1 16 1
  HO2: 344928.70779206115 APo_RO2_O6 886398626.6830952 C(HO2) -> APo_O4_OOH 27 2 1 0
  RO2pool: 58101.693049280046 APo_RO2_O7 4645490000.0 C(RO2pool) -> APo_O4_2OH 50 3 7 0
all rates: [7.90266391e-13 1.59728474e-12 2.65162671e-11 1.39762415e-12
 0.00000000e+00 0.00000000e+00]
all contributions: [0.53834072 0.43117486 0.02912902 0.00135539 0.        0.        ]
```

Figure S2: screenshot from the getrate_out.txt output file. The first line lists the NO_3^--CIMS peak index (214), followed by the pyhsical losses in $s^{-1}cm^{-3}$ (2.78e+5), the peak hight in cm^{-3} (2.78e+8) and the atomic composition (C/H/O/N). The next line reports the number of isomers ("3") and the type CS, RO2, RO2mcm and indices. In the present case, all isomers are closed shell species. Production channels and respective information follow. Determined rates are reported in "all rates:…" and the relative contribution to the observed change is given in "all contributions: …".

By assigning relative values to the unknown rate coefficients, the number of unknowns reduces to one and equ. (2) can be solved. The relative values applied are not reported specifically in an output file. They can simply be calculated by dividing the rate coefficients.

Rate coefficients calculated are listed after "all rates: ". There are four values (in cm^3s^{-1}) for the four reaction pathways ("7.90266391e-13 1.59728474e-12 2.65162671e-11 1.39762415e-12"), followed by two zero rate coefficients which are reserved for potential contribution via the alkoxy pathway.

The relative contributions of the individual reaction pathways are listed after "all contributions: …". Similar to the rate results, the contributions are given for the reaction channels listed (unit is dimensionless). Again, some space for potential contribution via the alkoxy pathway is reserved.

Contribution of the individual In this example, the biggest contribution to the changes of mass peak 214 arise from the reaction 2b) APh_RO2_O7 + NO (~ 54%), followed by the reaction 4b) APh_RO2_O7 + $\Sigma RO_2\bullet$ (~ 43 %). Very minor contribution is expected via the pathways 1a) Apo_RO2_O6 + HO_2 (~ 3%) and reaction 4a) APo_RO2_O7 + $\Sigma RO_2\bullet$ (~ 0.1%).

Setting up equ (2) for mass peak 214:

The first term " $\Delta C\left(CHON_x\right)$ " is 0 for the steady state condition as the peak concentration is constant with respect to time. So is the second term of the equation " $\sum\limits_{CHON_i = CHON_x} \Delta C_i$ " .

The right hand side of the equation lists the chemical sources and sinks as well as the physical losses (i.e. wall loss and dilution):

$$\sum_{m,n \to i} k_m \ast \left[R_{1,m}\right] \ast \left[R_{2,n}\right] \ast I_{m,n \to i} + \sum_{u \to i} k_u \ast \left[R_u\right] \ast I_{u \to i} - \Delta C_{phys}$$

production terms of the closed shell species are:

k_1 * [APh_RO2_O7] * [NO]

k_2 * [APh_RO2_O7] * [$\Sigma RO_2\bullet$]

k_3 * [APo_RO2_O6] * [HO_2]

k_4 * [APo_RO2_O7] * [$\Sigma RO_2\bullet$]

loss terms are:

ΔC_{Phys} = [CS_{214}] * (k_{wall} + $k_{dilution}$)

The unknowns in the above terms are the rate coefficients $k_1 - k_4$. Making an estimate on the relative magnitude allows to substitute k_2-k_4 by terms of k_1 enabling equation 1) to be solved:

k_1 * ([APh_RO2_O7] * [NO] + k_2/k_1 * [APh_RO2_O7] * [$\Sigma RO_2\bullet$] + k_3/k_1 * [APo_RO2_O6] * [HO_2] + k_4/k_1 * [APo_RO2_O7] * [$\Sigma RO_2\bullet$]) − [CS214] * (k_{wall} + $k_{dilution}$) = 0

After rearranging:

k_1 = [CS214] * (k_{wall} + $k_{dilution}$) / ([APh_RO2_O7] * [NO] + k_2/k_1 * [APh_RO2_O7] * [$\Sigma RO_2\bullet$] + k_3/k_1 * [APo_RO2_O6] * [HO_2] + k_4/k_1 * [APo_RO2_O7] * [$\Sigma RO_2\bullet$])

values for relative rate coeffs are assigned (note that the relative rate coefficients have to be provided as an input):

k2/k1 ~ 2.021

k3/k1 ~ 33.55

k4/k1 ~ 1.768

Further, (k_{wall} + $k_{dilution}$) = 1.e-3 (in the present example)

This allows for solving the equation:

k_1 = 2.783e+8 * 1.e-3 / (1 *1.617e+7*1.172e+10 + 2.021 * 1.617e+7 * 4.645e+9 + 33.55 * 3.449e+5 * 8.864e+8 + 1.768 * 5.810e+4 * 4.645e+9) = 7.905e-13

By substituting k_1 in the relative k-rate expressions $k_2 - k_4$ can be calculated.

Contributions of the individual reaction pathways to the observed change can be derived by considering the production terms and applying the derived rate coefficients:

k_1 * [APh_RO2_O7] * [NO] = 7.905e-13 * 1.617e+7 * 1.172e+10 = 1.498e+5 cm^{-3}s^{-1}

k_2 * [APh_RO2_O7] * [$\Sigma RO_2\bullet$] = 1.597e-12 * 1.617e+7 * 4.645e+9 = 1.200e+5 cm^{-3}s^{-1}

k_3 * [APo_RO2_O6] * [HO_2] = 2.656e-11 * 3.449e+5 * 8.863e+8 = 8.119e+3 cm^{-3}s^{-1}

k_4 * [APo_RO2_O7] * [$\Sigma RO_2\bullet$] = 1.328e-12 * 5.810e+4 * 4.645e+9 = 3.584e+3 cm^{-3}s^{-1}

The sum of all channels is 2.782e+5 cm^{-3}s^{-1}. Division of the individual channels by the sum-flux results in the individual contributions x:

x_1 = 1.498e+5 / 2.782e+5 = 0.538

x_2 = 1.200e+5 / 2.782e+5 = 0.431

x_3 = 8.119e+3 / 2.782e+5 = 0.029

x_4 = 3.584e+3 / 2.782e+5 = 0.001

On sorting the input CIMS data

Solution of equ. (2) is approached in a specific order with regard to CIMS-peaks. First, the input atomic mass peak-list is sorted with regard to mass (descending). This is necessary as the determination of peroxy autoxidation reactions (see (R5a)) is derived as a balance: the autoxidation reactions has to balance the formation of closed shell species and branch towards the alkoxy path (see Fig 1 and equ. (2)). To derive the H-shift + O_2 addition reaction rate coefficient of a RO_2 with

O_x, the loss terms (CS including ROOR, fragmentation products and alkoxy pathways) need to be summed up for the RO_2 with O_{x+2}. As a consequence, the analysis is started at the upper end of the mass distribution.

Description PyCHAM

The CHemistry with Aerosol Microphysics in Python (PyCHAM) box model simulated a continuous flow aerosol chamber using a fixed and continuous influx of α-pinene and O_3 under illumination of simulated artificial lamps of actinic flux set equal to that of the Julich Plant Aerosol Chamber. Continuous removal of chamber air was set to an air change rate of 10^{-3} s^{-1}. Gas-wall partitioning was switched off. Particles were not present in simulations. A mass spectrum of results was generated by summation of components with identical molar masses, and reporting the summed molar concentrations (the mass spectrum was extracted at 5h of simulated time; see supporting data for details on the simulation). Therefore, by analogy to chemical ionisation mass spectrometry, the ratio of abundance to mass spectrometer signal was treated as 1:1.

Inputs PyCHAM

The full set of input files to repeat the PyCHAM simulations analysed in the this work can be found here in the supporting data (see supporting data: PyCHAM/input). The simulation time was 6 hours with constant influx of O_3, α-pinene, H_2O and NO. The chemical reaction is started by considering a UV light source that is on from 2h to 6h simulation time. Losses are considered by activating a dilution factor of 10^{-3} s^{-1}. The runs were conducted under normal pressure (101300 Pa), at 288 K and with a relative humidity of 50%. Results were saved to file every 10 minutes.

Effect of the fragmentation pathway on the rate coefficients:

As discussed in the main text, section "2.1 Key aspects of the approach", subsection "Fragmentation products", any fluxes through the various fragmentation channels (i.e., R2d, R3c, R4d, R5c and R6c) have to be considered when determining R5a and R6a. This is described in the subsection "H-shift rates" of section 2.1. Briefly, any fragmentation reaction resulting in a loss of precursor molecules (RO_2^*) per second is compensated by the same flux [molecules*s^{-1}] by the combined terms of R5a and R6a to keep the mass balance (see Fig S3).
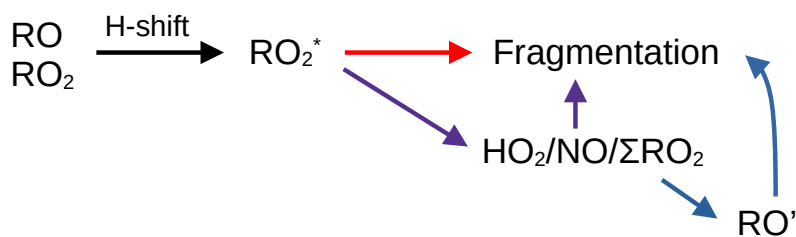
Fig S3: Fragmentation pathways and relation to formation of RO_2^* via alkoxy and peroxy radical autoxidation.

The effect of changing the rate coefficient for fragmentation k_{frag} is illustrated in Fig S3:

a) for peroxy radicals (red pathway) the change effects the sum of H-shift of RO and RO_2.

b) for fragmentation reactions via R2d, R3c and R4d (purple pathway), the change of k_{frag} has similar effects.

c) changes of fragmentation via RO' (blue pathway) additionally require changed formation of RO' via R2c, R3b and R4c.

For the present approach, effects of fragmentation do not affect the rate coefficients downstream the reaction scheme, but upstream. This results from the empirical quantification of rate coefficients and clearly distinguishes the autoCONSTRAINT method from structure-based predictions of rate coefficients.
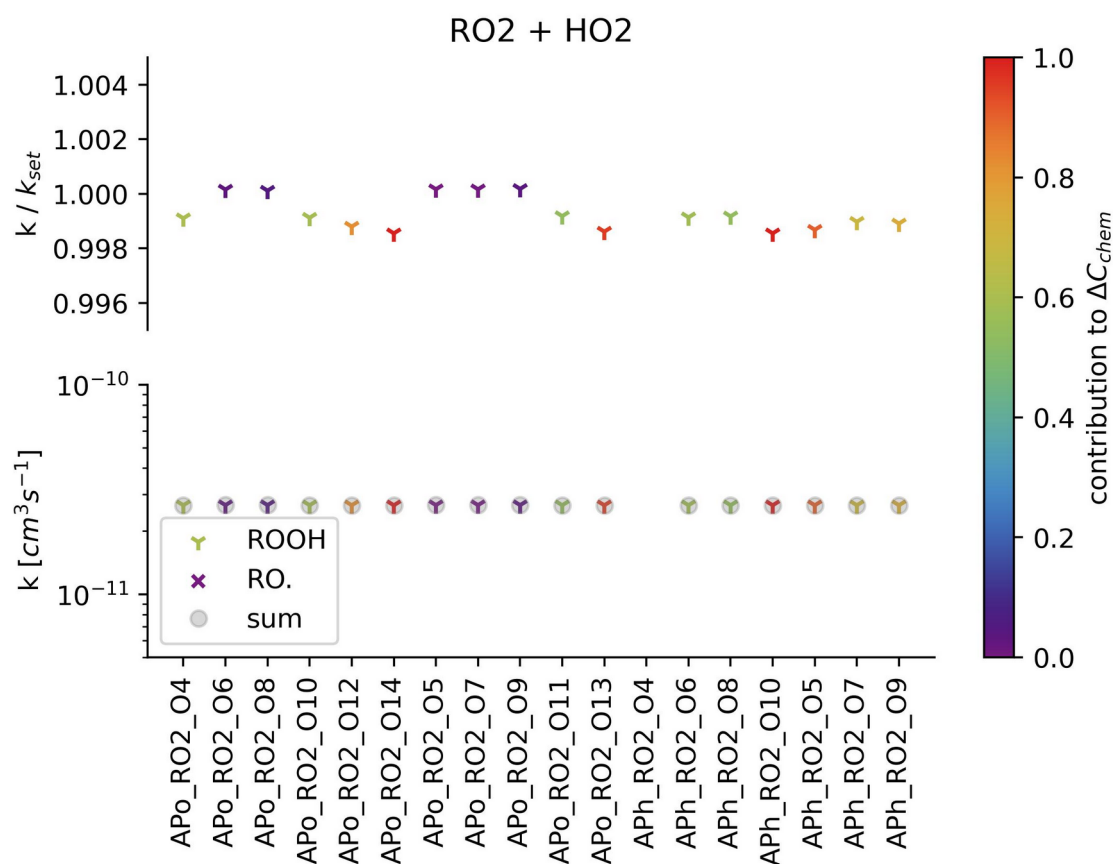
Figure S4: peroxy radicals (names are given by the abscissa) react at a rate (denoted on the ordinate lower) to form ROOH (downward facing triangle) and alkoxy radicals (crosses). The sum of the individual reaction channel's rates is shown by a circle. *The precision of rate coefficient recovery from the (simulated) mass spectrum is shown on the upper ordinate (recovered over set rate coefficient).* The color code denotes the calculated contribution of the reaction to the change (by means of chemical reactions) in CIMS mass peak concentration.
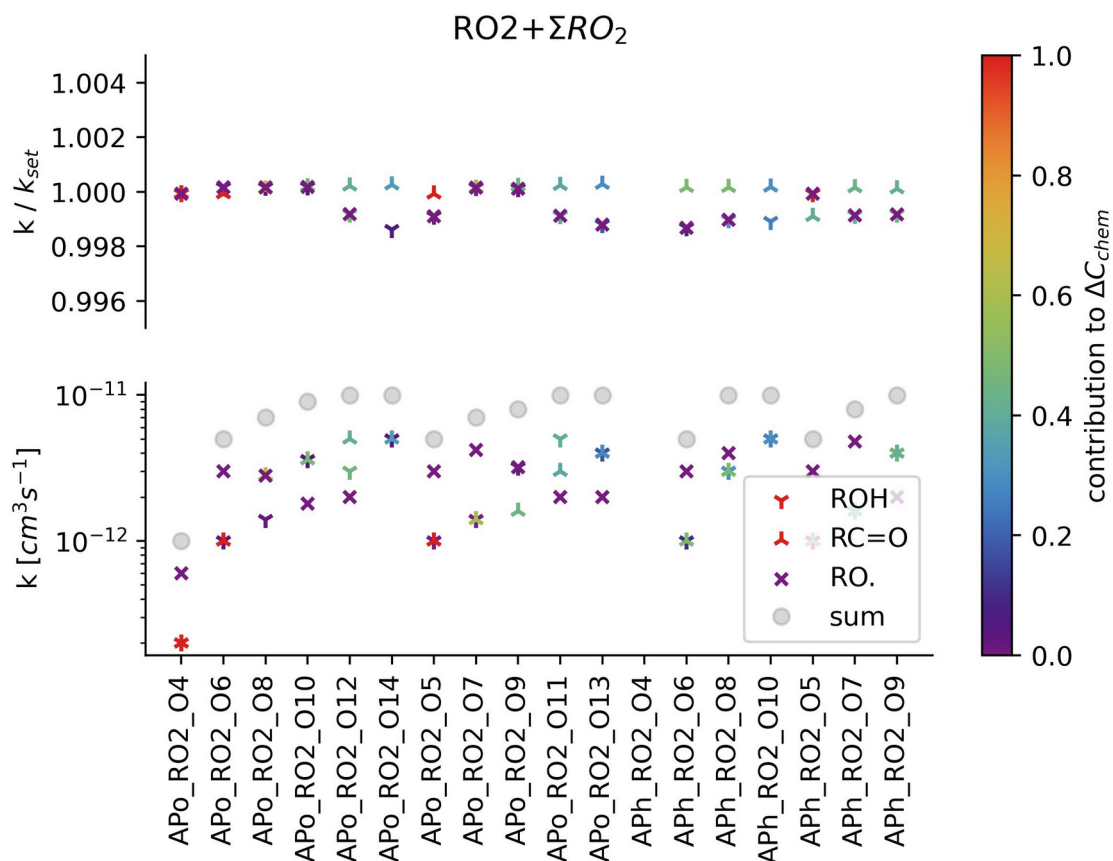
Figure S5: peroxy radicals (names are given by the abscissa) react at a rate (denoted on the lower ordinate) to form ROH (downward facing triangle), alkyl species (upward facing triangle) and alkoxy radicals (crosses). The sum of the individual reaction channel's rates is shown by a circle. *The precision of rate coefficient recovery from the (simulated) mass spectrum is shown on the upper ordinate (recovered over set rate coefficient).* The color code denotes the calculated contribution of the reaction to the change (by means of chemical reactions) in CIMS mass peak concentration.
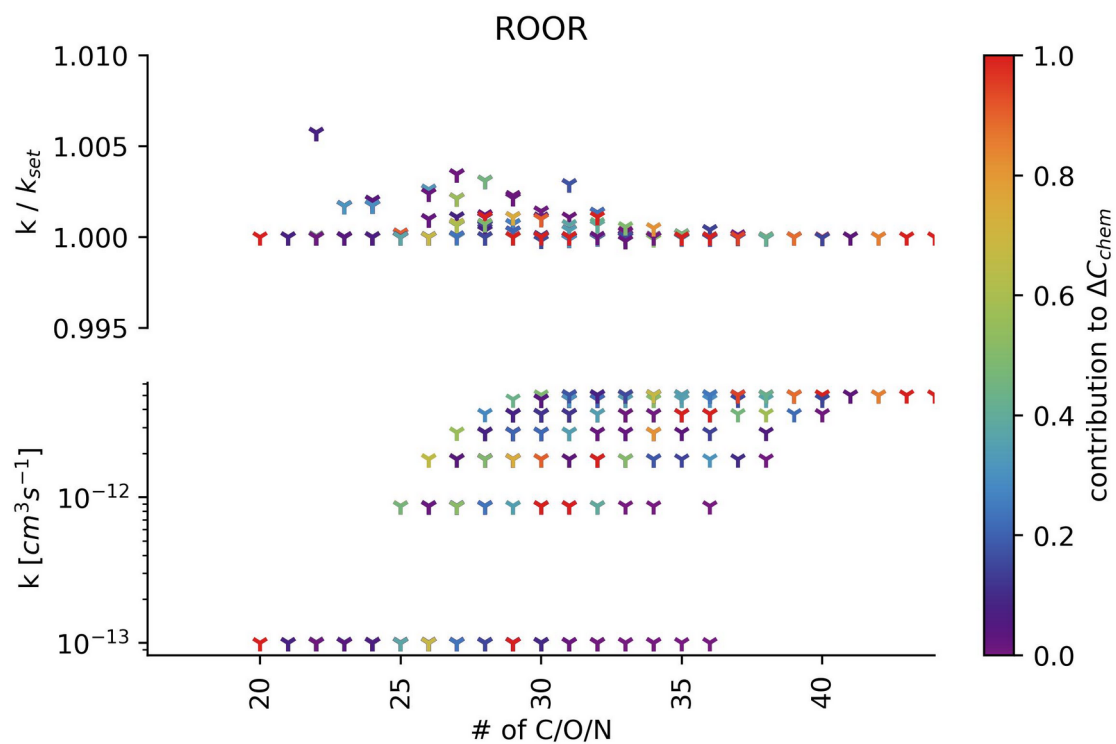
Figure S6: the abszissa shows peroxy radicals (denoted by their number of carbon + oxygen + nitrogen atoms) reacting at the rate coefficient k (lower ordinate) to form ROOR species. *The precision of rate coefficient recovery from the (simulated) mass spectrum is shown on the upper ordinate (recovered over set rate coefficient).* The color code denotes the calculated contribution of the reaction to the change (by means of chemical reactions) in CIMS mass peak concentration.
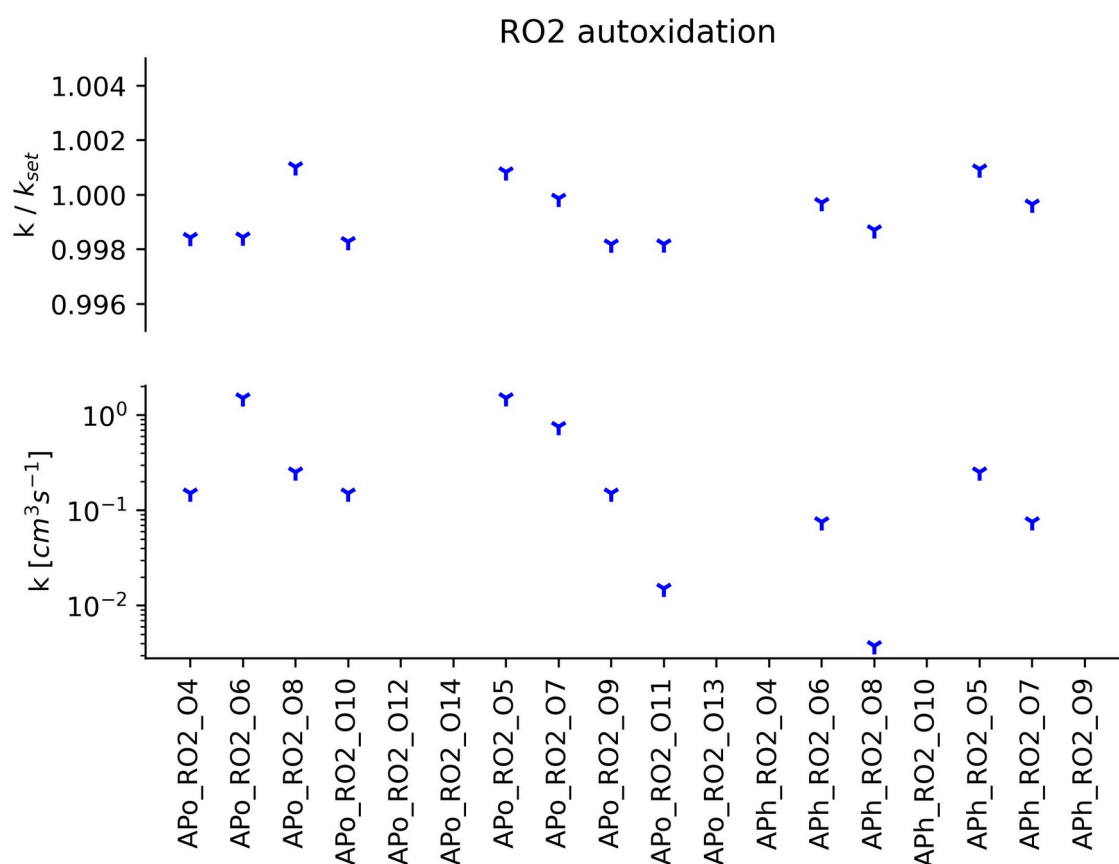
Figure S7: peroxy radicals (names are given by the abscissa) react at a rate (denoted on the lower ordinate) to form RO_2. *The precision of rate coefficient recovery from the (simulated) mass spectrum is shown on the upper ordinate (recovered over set rate coefficient)..*

SI-tables:

Table S2: list of influx potentials via the alkoxy- (RO•) and peroxy radical (RO2•) pathway to compensate for the consumption of the peroxy radicals listed in the first column. More detailed information on the initiation potentials is saved to the output file "getrate_out.txt" (look for "--- get the H-shift rates ---").

	P(RO•)	P(RO2•)
APo_RO2_O14	0	1
Apo_RO2_O13	1.11E-01	-1.32E+02
Apo_RO2_O12	0.62	-12.27
Apo_RO2_O11	0.33	-12.27
Apo_RO2_O10	0.74	-6.96
Aph_RO2_O10	2.70E-02	-5.30E+02
Apo_RO2_O9	0.14	-1.65
Aph_RO2_O9	0.34	-25.54
Apo_RO2_O8	0.97	-0.33
Aph_RO2_O8	0.13	-25.53
Apo_RO2_O7	0.11	-0.32
Aph_RO2_O7	0.68	-6.96

Apo_RO2_O6	1	-12.27
Aph_RO2_O6	0	1
Apo_RO2_O5	0.06	1
Aph_RO2_O5	1	1
Apo_RO2_O4	1	1