Electronic supplementary information (ESI)

A computer-based prediction platform for the reaction of ozone with organic compounds in aqueous solution: Kinetics and mechanisms

Minju Lee,[a] *Lorenz C. Blum,*[b] *Emanuel Schmid,*[b] *Kathrin Fenner,*[c,d] *Urs von Gunten**[a,c,e]

[a]School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

[b]Scientific IT services (SIS), ETH Zurich, Zurich and Basel, Switzerland

[c]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, 8600 Duebendorf, Switzerland

[d]Department of Chemistry, University of Zurich, 8057 Zurich, Switzerland

[e]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zurich, Switzerland

*Corresponding author

phone: +41 58 765 5270, fax:+41 58 765 5802, email: vongunten@eawag.ch

This ESI includes 10 texts, 8 tables, 15 figures, and 17 schemes.

Text S1. Definition of reactive sites and their connections to k_{O3} prediction and pathway prediction

A reactive site is defined as a chemical moiety potentially reacting with ozone in a query compound and is therefore subject to a prediction in the current prediction platform. As shown in Table S1 and Fig. S1 (chemical structures), various reactive sites for five differing functional groups such as aromatic compounds, olefin/ethynyl groups, amines, heteroaromatic compounds, and organosulfur compounds were defined. The nomenclature of the reactive sites was chosen to not only give chemically meaningful names but also for convenient assignments of a reactive site to the subsequent k_{O3} prediction group (2nd and 3rd columns in Table S1) or reaction pathway group (4th column in Table S1).

Here, the defined reactive sites are individually discussed with respect to their assignment to a k_{O3} prediction group and a reaction pathway group. An *italic* font was adopted as a convention for individual reactive sites hereafter. More details for the respective predictions are provided in Texts S4 and S9.

It should be noted that the currently defined reactive sites and the corresponding predictions were based on a limited number of available reference studies and do not cover all the chemical structures of micropollutants potentially found in natural and technical aquatic systems. Therefore, the predictions contain uncertainty and further empirical investigations are recommended for the verification of the predictions as well as for updates of the prediction platform.

Aromatic compounds Seven reactive sites are defined for the aromatic compounds, namely *phenol*, *3alkoxybenzene*, *aniline*, *cyclic_aniline*, *12alkoxybenzene*, *benzazole_dervs*, and *benzene*. *3alkoxybenzene* and *12alkoxybenzene* are abbreviations for trialkoxybenzene and mono- and di-alkoxybenzene, respectively, and *dervs* in *benzazole_dervs stands* for derivatives. Five of those reactive sites, i.e., *phenol*, *aniline*, 3*alkoxybenzene*, *12alkoxybenzene*, and *benzene*, were defined based on the k_{O3} prediction models separately available for the respective compound groups (Text S4) as well as the distinctive reaction pathway groups (Text S9) (see below for *benzazloe_dervs*). For an ambiva-

lent case with multiple aromatic reactive sites in the same aromatic ring (e.g., phenol and aniline co-exist in the same benzene ring), the assignment is carried out based on the proposed hierarchy (see Text S6 for details).

As shown in Fig. S1 (below 'Subclassification for [*aniline*]'), different subgroups were defined for *aniline*. The *aniline* group is defined exclusively for primary and tertiary anilines, where the tertiary anlines with the cyclic amine nitrogen (e.g., *N*-phenylpiperidine) are excluded from this definition (see below). In contrast, secondary anilines are divided into two sub-groups, namely, secondary anilines with (i) an acyclic nitrogen and (ii) a cyclic nitrogen. As for (i), the benzene ring is to be defined as any corresponding aromatic group depending on the hierarchy of the substituents (Text S6) and the acyclic nitrogen is assigned as an *amine*, respectively. For example, the amine nitrogen and the benzene ring of 4-(methoxy)-*N*-methylaniline are assigned as an *amine* and a *12alkoxybenzene*, respectively. This definition is based on the observation of ozonation of diclofenac[1] which bears an aniline with an acyclic secondary nitrogen. The nitrogen of diclofenac was reported to be a dominant reactive site with ozone rather than the benzene ring, and this was also corroborated by the QSAR analysis for this compound using a Hammett substituent constant.[1] Moreover, diclofenac has been successfully included in the currently developed k_{O3} prediction model for amines using $E_{NBO,LP\text{-}N}$ rather than E_{HOMO} for the aniline group (data not shown). All this evidence suggests that the secondary aliphatic nitrogen seemingly behaves as an amine rather than activating a benzene ring. Therefore, the secondary aliphatic nitrogen of aniline was assigned as an *amine*, while its benzene ring was assigned to the corresponding aromatic group. Note that a prerequisite for these assignments as *aniline* and *amine* to be valid is that the nitrogen of the *aniline* and the *amine* should not be a part of any excluded moieties (e.g., amide) described in the 'Amine' section in Fig. S1 (see 'Amines' section below for more details).

For (ii), secondary anilines with the cyclic nitrogen (e.g., tetrahydroquinoline) are defined as *cyclic_aniline* together with tertiary anilines with a cyclic nitrogen. As shown in Fig. S1, any atom except hydrogen (expressed as 'A') can be a member of an aliphatic cyclic system along with the cyclic

nitrogen. A k_{O3} estimate of 450 $M^{-1}s^{-1}$ is assigned to *cyclic_aniline* as the average of experimental k_{O3}-values for hydrochlorothiazide (600 $M^{-1}s^{-1}$),[2] enrofloxacin (330 $M^{-1}s^{-1}$ in diprotonated form),[3] and ciprofloxacin (400 $M^{-1}s^{-1}$ in diprotonated form).[3] We assume that the benzene ring of these compounds is responsible for the observed k_{O3}. The reason why the k_{O3} estimate was given was because the experimental k_{O3} of this class of compounds could not be predicted well with the currently available k_{O3} prediction models with the E_{HOMO} for the benzene ring and the E_{NBO} for the cyclic amine (data not shown). Note that when the benzene ring of *cyclic_aniline* has a substituent corresponding to any aromatic compound group except *benzazole_dervs* or *benzene,* such an aromatic compound group will be detected over *cyclic_aniline* according to the hierarchy described in Text S6. For example, when the benzene ring of tetrahydroquinoline is substituted by a hydroxyl substituent, *phenol* is detected as the reactive site instead of the *cyclic_aniline*. This is based on the assumption that in presence of such substituent, the k_{O3} will be higher than the k_{O3} estimate (450 $M^{-1}s^{-1}$) due to the activation of the benzene ring by the substituent. Therefore, it has to be predicted by the corresponding k_{O3} prediction model. There is no empirical information for the oxidation products from the reaction of ozone with a *cyclic aniline* and it is currently assumed to be the same as for *aniline*. However, this remains to be verified empirically in future.

In addition to the reactive sites described above for the aromatic compound class, *benzazole_dervs* was additionally defined based on the experimental data for benzotriazoles. Differing benzotriazoles such as unsubstituted, 5-chloro, 5,6-dimethyl, and 5-methylbenzotriazole and their conjugate bases were included in the k_{O3} prediction model for the benzene group (Text S9). However, an explicit definition of *benzazole_dervs* rather than including it in *benzene* was preferred for clarity but also to account for the additional reaction pathways associated with the heteroatom of a 5-membered ring (e.g., *N* or *S*-oxide formation in Scheme S10). It is noted that benzazoles feature a benzene ring fused with azoles at the 4- and 5-positions. As an azole is defined as a five-membered heteroaromatic ring containing a nitrogen and at least one other heteroatom (nitrogen, oxygen, or sulfur), it by definition does not include isoindole, isobenzofuran, and 2-benzothiophene shown in Fig. S1 (top right chemical

structure of *benzazole_dervs*). Since those were assumed to behave similar to benzazoles for ozone reactions, *benzazole_dervs* was defined instead of *benzazole* to include isoindole, isobenzofuran, and 2-benzothiophene.

Olefins/Ethynyl As shown in Fig. S1, any carbon-carbon double bond (C=C) in an aliphatic system is defined as *olefin* with the exception of *olefin_conj*, which contains a conjugation with another functional group. The first exception included in *olefin_conj* is an olefin in a cyclic ring conjugated with an α-keto group, and the second is an olefin conjugated with both α-keto and α-amino groups at both carbons. β-cyclocitral, progesterone, medroxyprogesterone, norethindrone, and levonorgestrel were the basis for the former case and indigotrisulfonic acid was the reference compound for the latter case. All of these compounds were included in the k_{O3} prediction model as miscellaneous olefins (denoted as miscolefin in Table S1) using E_{HOMO} (Text S4) rather than $E_{NBO,C=C}$, which is typically used for olefins.[4] The detected products for progesterone during ozonation were according to the Criegee mechanism.[5] Based on the reported pathways for progesterone, the olefin pathways are assumed to be generally applicable to *olefin_conj* members. '*Ethynyl*' refers to a carbon-carbon triple bond. As there is no corresponding k_{O3} prediction model developed, a k_{O3} estimate of 200 $M^{-1}s^{-1}$ was assigned based on the second order rate constant for the ethynyl group in 17α-ethinylestradiol (Tables S1 and S2). A specific pathway reaction group named ethynyl was assigned.

Amines *Amine*, *hydrazine*, *N_sulfenamide*, *amine_conj*, *phenylazo*, *sulfonamide(1st)_prot*, *sulfonamide(1°)_deprot*, *sulfonamide(2°)_prot*, *sulfonamide(2°)_deprot*, and *sulfonamide(3°)* were defined as reactive amine sites (Table S1). For sulfonamides, 1°, 2°, and 3° stand for primary, secondary, and tertiary and "prot" or "deprot" stands for protonated or deprotonated, respectively. For all reactive sites to be valid as defined, the nitrogen should not be assigned to any of the excluded moieties in Fig. S1, which lists amide, imine, nitrile, nitro, nitroso, nitroxide, aminoacrylaldehyde, and quaternary amines. The nitrogen lone-pair electrons in such excluded moieties are considered to be inactivated either by multiple bonds (e.g. imine), by conjugation (amide), by protonation (protonated amine), or by substitution (quaternary ammonium). As a consequence, their reactivities with ozone are very low;

e.g., *N*-methylacetamide (0.6 $M^{-1}s^{-1}$),[6] tetranitromethane (10 $M^{-1}s^{-1}$),[7] *N*-nitrosodimethylamine (<0.1 $M^{-1}s^{-1}$),[8] protonated amines (<0.1 $M^{-1}s^{-1}$).[9] Therefore, those moieties were neglected in this study.

Amine is any primary, secondary, and tertiary amine, which does not belong to any other reactive sites in the amine group (e.g., *hydrazine* and *N-sulfenamide*). For *amine*, a reaction pathway group with the same name, i.e., amine, was assigned (see Text S9 for details). *Amine_conj* is defined as a nitrogen conjugated by either a C=N (amidine) or a C=S (thioamide) bond and no k_{O3} prediction model is available for this group. The reaction pathway group for the amines was also assigned for *amine_conj* based on ozonation of cylindrospermopsin containing a guanidine moiety (Text S9 for details).[10] Despite the fact that an azo group is structurally categorized for amine, *phenylazo* containing an azo moiety was assigned to the hetero_ar group for both k_{O3} and pathway predictions. Azobenzene, which was the only reference compound for *phenylazo*,[11] was reported to be oxidized to azoxybenzene. This is consistent with other heteroaromatic compounds in which the nitrogen is a major reaction center with ozone to form an oxide product (Text-S9). Moreover, its $E_{NBO,LP-N}$ did not fit with other amines but its E_{HOMO} correlated fairly with the other hetero_ar group members (Text S4). Empirical k_{O3} estimates were suggested for *sulfonamide(1°)_prot*, *sulfonamide(1°)_deprot*, and *sulfonamide(2°)_deprot*, based on experimental k_{O3} for the reference compounds as there is no available prediction model (Tables S1 and S2). Sulfonamide(2°)_deprot was assigned to the reaction pathway group for the amine based on ozonation of hydrochlorothiazide (Text S9).[2]

It should be noted that there are several reactive sites for the amine group (e.g., *hydrazine* and *sulfonamide(3°)*) for which no empirical k_{O3} estimate, no k_{O3} prediction model group, and no reaction pathway group is assigned. In other words, for these classes of reactive sites no prediction is available in the current prediction platform. Nevertheless, these reactive sites are explicitly presented to provide information as to the limitations of the currently developed prediction system and also to highlight the needs for further investigations. This is the same for many reactive sites for the organosulfur compound group below.

Heteroaromatic compounds *Adenine* and *guanine* were treated separately because the corresponding k_{O3} prediction models were developed separately (Text S4). Note that *guanine* refers not only to guanine but also to other purine bases with a keto group in the 6 position (e.g., hypoxanthine, xanthine, caffeine, and uric acid). A pathway prediction for *guanine* was developed based on ozonation of guanine, guanosine, and its derivatives, e.g., acyclovir.[12] Despite analogous chemical structures, the pathways for *guanine* were not extended for *adenine* in the current study (Text S9). Separate k_{O3} prediction models were developed for *thymine*, *uracil*, and *cytosine* (Text S4) and the pathway group, uracil, is applied for *thymine*, *uracil*, and *cytosine* (Text S9).

Pyridine and *diazine* are defined as heteroaromatic compounds in which the aromatic nitrogen is not conjugated with the exocyclic nitrogen, oxygen (as a double bond), or sulfur (as a double bond). In contrast, the conjugated forms for pyridine and diazine are defined as *pyridine_conj* and *diazine_conj*, respectively. This classification was based on the observation that non-conjugated pyridine and diazine (2-isopropyl-3-methoxypyrazine) were successfully included in the k_{O3} prediction model for the heteroaromatic compound group (hetero_ar) using E_{HOMO}, while a conjugated pyrimidine (2,4-diamino-5-methylpyrimidine) was included in the miscellaneous olefin group using E_{HOMO} (Text S4). In contrast to *pyridine* and *diazine*, which were defined separately from *pyridine_conj* and *diazine_conj*, no distinction by conjugation was made for *triazine*. Empirical k_{O3}-values were only available for conjugated 1,3,5-triazines such as atrazine and simazine[11] and those were included in the k_{O3} prediction model for the heteroaromatic compound group (hetero_ar) together with unconjugated ones (e.g., pyridine) (Text S4). It was assumed that other triazines such as 1,2,3-triazine or 1,2,4-triazine would behave similarly. For pyridine, diazine, and triazine, the same reaction pathway group, hetero_ar, was assigned (Text S9).

5hetero_ring was defined for 5-membered heteroaromatic compounds containing at least two neighboring aromatic carbons. Such compounds as imidazoles, indoles, isoxazoles were included in the miscellaneous olefin group for the k_{O3} prediction (Text S4). The corresponding reaction pathways

were assigned to olefins based on the detection of the Criegee products from ozonation of several *5hetero_ring* members such as tryptophan[13,14] histidine,[14] and vinylene carbonate.[15]

Quinolone and *quinolone_carboxylate* were defined based on the empirical observation for flumequine, enrofloxacin, and ciprofloxacin,[3] and k_{O3} estimates of 1.0 $M^{-1}s^{-1}$ and 2×10^4 $M^{-1}s^{-1}$ were assigned to the respective reactive sites (see below). Both neutral and anionic flumequine, which bear the respective *quinolone* and *quinolone_carboxylate*, were included in the k_{O3} prediction model for the miscellaneous olefin group in the previous study.[4] However, the miscellaneous olefin group in the current prediction was not used to predict k_{O3} for *quinolone* and *quinolone_carboxylate* because of the ambiguity of selecting an appropriate orbital for the quinolone moieties of flumequine, enrofloxacin, and ciprofloxacin (not addressed in the present study as it is highly hypothetical).

Quinolone is detected when the 4-pyridone of the quinolone moiety is not substituted by carboxylate (COO^-). For example, the quinolone moiety with a carboxylic group (COOH) as in enrofloxacin, ciprofloxacin, or fluemquine is detected as *quinolone*. However, it is important to note that for enrofloxacin and ciprofloxacin *cyclic aniline* is also detected for the benzene ring of the quinolone because it has the cyclic nitrogen substituent in a piperazine group (see Text S1 for the definition of *cyclic aniline*). A k_{O3} estimate of 1.0 $M^{-1}s^{-1}$ was assigned based on the observed k_{O3} for the neutral flumequine with no nitrogen substituent.[3] As shown in Fig. S1, when the 4-pyrdione of *quinolone* is substituted by carboxylate (COO^-), *quinolone_carboxylate* is detected instead. A k_{O3} estimate of 2.0×10^4 $M^{-1}s^{-1}$ was the average of the k_{O3}-values for deprotonated flumequine (1.8×10^3 $M^{-1}s^{-1}$),[3] monoprotonated ciprofloxacin (7.5×10^3 $M^{-1}s^{-1}$),[3] and monoprotonated enrofloxacin (4.6×10^4 $M^{-1}s^{-1}$),[3] all these compounds have COO^-. Note that three reference compounds used herein have a fluorine substituent. Although fluorine may have contributed to some extent to the observed k_{O3}, *quinolone* and *quinolone_carboxylate* were defined without fluorine for a general application. For reaction pathways, *quinolone* and *quinolone_carboxylate* are assigned to olefins. Two unfused carbons of 4-pyridone undergo the Criegee mechanism (see Text S9 below) based on the observed oxidation products of ciprofloxacin[16].

Organosulfur compounds In contrast to all the reactive sites mentioned above, there is no k_{O3} prediction model available for oragnosulfur compounds. Therefore, k_{O3} estimates are suggested for some reactive sites. However, similar to the amines above, reactive sites (e.g., thioketone, thioamide, and thiophosphoramide) with neither k_{O3} estimates nor available pathways were defined in this study. These reactive sites are presented with the intention to stress the necessity for further investigations in future. As explained above for the amines, for a reactive site to be valid, sulfur should not be a part of sulfonamide and sulfonic acid, which are defined as the excluded moieties in Fig. S1, because those moieties are known to be ozone-resistant.[11]

Text S2. Speciation analysis

A speciation analysis is included for a *query compound* undergoing acid-base speciation, thus providing relevant acid-base species and their relative concentrations (p) as a function of pH between 0 and 14. This is a prerequisite for a pH-dependent k_{O3} prediction, i.e., predicting an apparent k_{O3} for a reactive site of a *query compound* at a specific pH. A workflow of the speciation analysis is shown in Fig. S4. A built-in pK_a calculator plugin of Marvin (Chemaxon)[17] generates all acid-base species and predicts associated pK_a-values. For a *query compound*, which has multiple species with the same net charge (i.e., tautomers), a tautomeric fraction (f), which represents the relative population of each tautomer species, is additionally derived. Both pK_a and f calculation results are presented to the user and remain subject to confirmation or manual modification. As predicted values can be uncertain, it is always recommended that the user seek for experimental or theoretical values obtained from high level quantum chemical computations. Based on the confirmed or updated speciation parameters, p_i of a species (i) is recalculated and the corresponding species distribution as a function of pH is derived accordingly (e.g., panel (H) in Fig. 4).

Text S3. Methodology for the development of k_{O3} prediction models for organic compounds using quantum chemical computations and statistical evaluation of the model performance.

Quantum chemical models for predicting k_{O3} for differing classes of organic compounds such as aromatic compounds, olefins, amines, etc. were previously developed.[4] This approach was chosen as a k_{O3} prediction method in this study not only because of its comprehensive applicability to various organic compounds but also because the k_{O3} prediction procedure can be implemented with the help of existing quantum chemical computation softwares and other chemoinformatics applications. However, the previously proposed models[4] were not directly used but modified and adapted for the following reasons. The previously proposed quantum chemical model followed a computational protocol using *ab initio* methods (Hartree Fock (HF) or Density Function Theory (DFT)) for all the computation steps such as geometry optimization, frequency analysis, and single point calculation. All the steps were carried out at the same level of theory.[4] The prediction platform to be developed in this study was intended to be responsive and interactive with a user. Although the prediction performance of the previous models was (highly) satisfactory and the protocol was straightforward, performing *ab initio* quantum chemical computations for all aforementioned steps would have hampered this goal due to extended computation times, especially for large molecules (e.g., molecular weight > 300) with many microspecies. This is especially important given that the development of this prediction platform was targeted to be implemented by an end-user on a personal computer or as a web-based tool with minimal resources and maintenance. Moreover, as the prediction platform was intended to be available to the public, k_{O3} predictions with license-free softwares were sought for. ORCA 3.0.3,[18] which is a quantum chemical computation software available free of charge for academic users, was used instead of the commercial software Gaussian,[19] which was applied previously.[4] However, the commercial natural bond orbital (NBO) software 6.0[20] could not be replaced as there is no free substitute currently available. However, the license for the NBO software will be supported in an on-line prediction platform that is planned to be developed in the future. The details on modifications/adaptations of the original computation protocol to fulfill these requirements/specifications are given in Text S4.

General information about the establishment of k_{O3} prediction models and statistical parameters used to evaluate its performance is provided here. A k_{O3} prediction for organic compounds using quantum chemical computations exploits a linear relationship between the logarithms of empirical k_{O3}-values for the organic compounds and the corresponding orbital energies. The corresponding linear correlation is established by a least-squares regression with $\log(k_{O3}) = y_0 + aE$, where y_0, a, and E are the y-intercept, the slope, and the corresponding orbital energy of a compound, respectively. It has previously been derived which type of orbital energy most strongly correlates with which class of organic compound (e.g., aromatic compounds correlate with the highest occupied molecular orbital energies (E_{HOMO})).[4] Here, the same general relationships were used (Table S1) and the same set of training compounds were used unless stated otherwise. Statistical evaluations of the developed k_{O3} prediction models were carried out based on R^2, the mean unsigned error (MUE) in log units, and the root-mean square error (RMSE) in log units. The previous models were established by excluding outliers from the training compounds for which the predicted k_{O3} differed by more than a factor of 10, and a, y_0, and R^2 were reported accordingly. In contrast, those outliers were included in the reported MUEs and RMSEs to better demonstrate the predictive power of the models. In this study, the same compounds and the same outliers were used for the model development and evaluation for a consistent comparison of the model performances unless stated otherwise. A complete list of training compounds is given in Tables S4, S5, and S6, but their experimental k_{O3}-values can be found elsewhere.[4]

Text S4. Evaluation of k_{O3} prediction models developed with differing computational methods for the selected organic compound groups

The k_{O3} prediction models were developed for the 14 model groups in the 3rd column in Table S1 with seven combinations of computational methods, HF/6-31G//PM3//MMFF94, HF/SV//PM3//MMFF94, HF/SV//PM3//Dreiding, HF/3-21G//PM3//MMFF94, HF/3-21G//PM3//Dreiding, HF/3-21G//MMFF94, and HF/3-21G//Dreiding. All the models were developed based on the same methodology described in Text S3.

The computational methods presented above, divided by '//', comprise of either two or three consecutive computations to be conducted in a reverse order from the right to the left. 'MMFF94' or 'Dreiding' indicate a molecular mechanics method to obtain an initial 3D geometry from 1D SMILES using either a Merck molecular force field (MMFF94) method[21–25] or the Dreiding force field[26], respectively. Both methods were used as implemented in Marvin (Chemaxon). The 3D initial geometry of a *query compound* is then subjected to either semi-empirical PM3[27,28] geometry optimization followed by *ab initio* Hatree-Fock (HF) single point calculations or directly to the HF single point calculations without geometry optimization with the semi-empirical PM3 method. The semi-empirical PM3 method was performed in gas phase. Three differing basis sets (Pople 6-31G and 3-21G[29–33] and Ahlrichs split valence[34] (SV)) were explored for single point calculations.

The single point calculations for all seven computational methods were performed by an *ab initio* HF method. A semi-empirical PM3 method showed no meaningful correlations (R^2<0.4) along with computation failures for many compounds (data not shown). The DFT was not considered as it was previously reported that for some compounds (e.g., benzoate, benzensulfonate, and benzaldehyde) a molecular orbital located on the aromatic ring obtained by the DFT-B3LYP was not actually the HOMO but the HOMO-n (n=1,2,3, ···), which is a molecular orbital n levels lower than the HOMO.[4] In contrast, with the HF method, the HOMO consistently appeared on the aromatic ring for all the aromatic model compounds investigated. Because of this consistency, the HF method was preferred. Moreover, it was reported that the prediction performance of the HF method was superior or comparable to the DFT-B3LYP method.[4] Therefore, only the HF method was investigated in this study.

Although not presented in a computational method convention for brevity, auxiliary methods such as COSMO (conductor-like screening model), ECP (effective core potential), and NBO (natural bond orbital) were implemented for all single point calculations. For all single point calculations, the COSMO set to water (dielectric constant (ε) = 80.4 and refractive index = 1.33) was implemented to take the solvation effect into account. Los Alamos effective core potential (ECP)[35–37] was applied for iodine for which the 3-21G basis set was unavailable. While E_{HOMO} is obtained from the implementa-

tion of the described method above, for compounds containing olefins or amines, the NBO analysis was additionally implemented by the NBO program 6.0[20] to derive the corresponding natural bond orbital (NBO) energies ($E_{NBO,C=C}$ or $E_{NBO,LP-N}$).

The performances of all the developed k_{O3} prediction models were evaluated with statistical parameters such as R^2, MUE, and RMSE (Text S3) and are presented in Figs. S2 and S3. The average CPU time, which is the average of the individual CPU times measured for training compounds for differing computation methods relative to the reference method (i.e., HF/6-31G//PM3//MMFF94), is additionally presented for individual k_{O3} prediction model groups. The CPU times for 3D geometry generation was not included as it was completed in only a fraction of the CPU time required for single point calculations. The measurements were conducted on the same hardware (Intel® Core™2 Duo CPU at 3.00 GHz with 4 GB RAM) operated on Ubuntu Linux 14.04.4 LTS for all the computational methods.

Due to an extensive amount of data represented in Figs. S2 and S3, all k_{O3} prediction model groups are not individually discussed in detail. Rather, general discussions are given for specific model groups with noteworthy behaviors such as mono- and di-alkoxybenzene, trialkoxybenzene, guanine, and aniline. Along with the discussions given in this text, a brief summary of the prediction model evaluations is additionally provided in Table S7.

For most model groups, the prediction performance evaluated by R^2, MUE, and RMSE is consistently stable for differing methods, whereas the shortest CPU time (30% to 60% of the reference method) was required for the HF/3-21G method without the PM3 geometry optimization (i.e., HF/3-21G//MMFF94 and HF/3-21//Dreiding). As the MMFF was superior to the Dreiding force field for the mono- and di-alkoxybenzene group (see below), the HF/3-21G//MMFF method was chosen to be the default in this study.

Mono- and di-alkoxybenzene. Comparing the two molecular mechanics methods (i.e., MMFF94 and Dreiding force field), a worse performance was consistently observed with the Dreiding force field compared to the MMFF94 for the mono- and di-alkoxybenzene groups (Fig. S2) regardless of the

basis sets and the geometry optimization. As previously reported,[4] rotamers can play an important role for alkoxybenzenes as E_{HOMO} significantly fluctuates upon rotation of an alkoxy group. A better correlation was found with alkoxybenzenes mainly consisting of planar alkoxybenzenes for which the alkyl group is aligned to be planar to the benzene ring.[4] It turned out that initial geometries of more alkoxybenzenes were planar or nearly planar for the MMFF94 than for the Dreiding force field. For example, a planar and perpendicular geometry was obtained for 1,4-dimethoxybenzene with the MMFF94 and the Dreiding force field, respectively. The exclusion of 1,4-dimethoxybenzene from the correlation improved the performance of all methods with the Dreiding force field (e.g., R^2 increased by more than 0.1) (data not shown). Moreover, further improvements could be achieved by replacing E_{HOMO} of non-planar alkoxybenzenes with E_{HOMO} of their planar conformations, e.g., from R^2=0.80 to R^2=0.84 for the HF/SV method by using planar alkoxybenzenes (acebutolol, atenolol, 2,4-dichlorophenoxyacetic acid, and propranolol) (data not shown). Although automated geometric modifications could be pursued by programming an algorithm, it was not considered in this study to avoid additional complications. Overall, it was concluded that the MMFF94 has advantages over the Dreiding force field for the mono- and di-alkoxybenzene groups.

Trialkoxybenzenes. For the trialkoxybenzene group, poor correlations (R^2<0.5) were obtained for all the computational methods except for the HF/3-21G//MMFF94 method for which a partly meaningful correlation (R^2=0.69) was obtained. As shown in Fig. S5a (green triangles), however, a negative slope of the regression line was obtained for HF/3-21G//MMFF94. Similar to the mono- and di-alkoxybenzene groups, this phenomenon is related to the rotamers of trialkoxybenzenes.[4] However, no further investigation was conducted as automatic geometry modification was not planned to be developed in this study. Moreover, there are only four training compounds available for this model group, which makes it difficult to perform an in-depth evaluation. Nonetheless, the prediction performances evaluated with the MUE (0.26-0.51) and the RMSE (0.31-0.55) were comparable to those of other compound groups with good correlations. This is attributed to the fact that the calibration ranges for k_{O3} between the minimum and the maximum is only a factor of 30 for trialkoxybenzenes, while

other model groups span over several orders of magnitudes (e.g., 7 orders of magnitude for phenol). Therefore, its prediction may be acceptable as long as the derived orbital energy lays in the calibration range (e.g., E_{HOMO} of -0.33649 ~ -0.32043 hartree for the HF/3-21G//MMFF94), while it may completely fail when extrapolated.

Guanine. For guanine, an aggravated performance was observed for the method involving geometry optimization with a Dreiding force field-derived geometry, i.e., HF/SV//PM3//Dreiding (R^2=0.62) and HF/3-21G//PM3//Dreiding (R^2=0.65) compared to the other methods (R^2=0.94-0.98). This was because 5′-deoxyguanylic acid was an outlier in the Dreiding force field-based method. During the geometry optimization using the initial geometry obtained by the Dreiding force field, it was observed that an intramolecular hydrogen transfer from an amino group to a phosphate group occurred, which led to a significantly higher E_{HOMO} (data not shown). Because of this phenomenon, the MMFF94 method is recommended for guanines for the geometry optimization.

Aniline. It should be noted that in contrast to the other model groups for which k_{O3} prediction models were developed with the same training compounds as used in the previous study,[4] a k_{O3} prediction model for anilines has been established in this study with different sets of anilines. This was because k_{O3}-values for several anilines used for the previous model were updated with recently reported empirical k_{O3}-values.[38] Compared to the previous study, moreover, k_{O3}-values for anilines such as *p*-methylaniline, *p*-methoxyaniline, *p*-phenylenediamine, *o*-hydroxyaniline, *p*-hydroxyaniline, and *p*-aminobenzoic acid[38] were additionally included in the model. k_{O3} for the unsubstituted aniline was recently reported to be 1.3×10^6 $M^{-1}s^{-1}$,[38] which is almost an order of magnitude lower than the k_{O3}-value used for the previous model, i.e., 1.4×10^7 $M^{-1}s^{-1}$.[39] We have adapted the currently reported k_{O3} in this study as the empirical determination has been appropriately conducted using a well-established competitor (3-buten-2-ol) and at neutral to alkaline pH in which the ozone decomposition from the reaction with aminyl radicals can be reduced. Moreover, the reported k_{O3} was theoretically corroborated by an excellent correlation between k_{O3} for differing aromatic compounds including the unsubstituted aniline and the corresponding Gibbs free energy (ΔG^0) for the formation of an ozone adduct in

the para position.[38] As the previous k_{O3}-values for many other anilines were reported relative to the k_{O3} for unsubstituted aniline,[39] k_{O3}-values for such anilines were corrected accordingly and used in the current model development. Moreover, a significant correction was made for *N*,*N*-dimethylaniline. While the previous k_{O3} was 2.0×10^9 $M^{-1}s^{-1}$,[40] the k_{O3} that has been recently reported is 1.6×10^6 $M^{-1}s^{-1}$.[38] The k_{O3}-values for 22 anilines used for the model development in this study are compiled in Table S8 where they are also compared to the previous values.

With E_{HOMO}-values for 22 anilines and their up-to-date k_{O3} values, overall inferior correlations (R^2=0.43-0.64) to the original model (R^2=0.85 for the HF/6-31G method)[4] were obtained for the current model. Nonetheless, the prediction performance was better than with the original model, MUEs=0.24-0.30 and RMSEs=0.30-0.37 throughout all seven computational methods of the current model and MUE=0.34, RMSE=0.43 for the previous model, respectively. This is related to the fact that the current model was developed with the corrected k_{O3}-values, which were in a narrower range than the original model. While the k_{O3} range of the 22 anilines used for the current model only spans a factor of 50, it was ~ 4 orders of magnitudes for the previous model (Table S8).

It is worthwhile to note that the previous model for anilines has been developed using planar anilines which have two substituents on the amino group aligned to be planar to the phenyl ring because its correlation was better than that with pyramidal anilines (see reference (*4*) for more details). However, a somewhat worse performance was obtained using planar anilines with the new set of 22 anilines (data not shown). Therefore, it is concluded that the previously observed better correlation with planar anilines was rather fortuitous and it is recommended to use the current model rather than the previous model.

Other heteroaromatic compounds (hetero_ar). Although correlation to the HOMO was shown to be somewhat successful, no prediction model was proposed in the previous study for other heteroaromatic compounds because the model seemed to be rather premature to be established.[4] In this study, nonetheless, a prediction model was developed to provide kinetic information for these chemical moieties.

Text S5. Computational protocol for k_{O3} predictions

Based on the evaluation of the seven different combinations of computational methods in Text S4 above, the HF/3-21G//MMFF94 method was chosen as default in this study because it performed similarly or better compared to the other computational methods tested in this study, while the least CPU time was required (for more details see Text S4 above). The individual models are presented in Fig. S5 and their prediction performances as well as the parameters for linear regression equations are summarized in Table S3 in comparison with the original models. Note that this computational protocol does not carry out geometry optimization and frequency analysis. Nevertheless, the models' performances turned out to be acceptable although it was somewhat inferior to the original models (Table S3). This indicates that orbital energies obtained from molecular mechanics-level geometries of training compounds are similar to those of the optimized local minimum geometries from the original computational protocol involving geometry optimization and frequency analysis. Moreover, fortuitous cancellation of errors may have played a role as only relative orbital energies are relevant in the models while absolute orbital energies may still be biased. However, one should keep in mind that a prediction may fail for some compounds due to the use of an approximate geometry and further investigations are needed to be able to justify the general application of molecular mechanics-based geometries.

The CPU times required to complete single point calculations for the 294 individual compounds (10 outliers included) used for model development were measured and are plotted as a function of the number of atoms (atom count) in Fig. S6. A good relationship between the CPU time and the atom count was obtained when fitted by a quadratic regression. While less than 700 seconds were required for most of the training compounds with less than about 60 atoms, CPU times in the order of several thousand seconds were needed to complete single point calculations for macromolecules such as azithromycin, roxithromycin, tylosin, and microcyctin-LR with 124 – 145 atoms.

Text S6. Hierarchy for the assignment of an ambivalent aromatic compound

Compound groups such as olefins, amines, guanine, etc. are individually unique, therefore, the assignment of reactive sites can be conducted independently. However, reactive sites for aromatic compounds may overlap, e.g., aminophenol contains phenol and aniline in the same benzene ring. As such ambivalent compounds were not included in the model development, a hierarchy between aromatic compound models was established (arrow in top-left part of Fig. S1), based on the relative positions of individual k_{O3} prediction models in Fig. S5a. As phenolate has the highest k_{O3}-values, it is assigned to have the highest priority over all the other model groups. As shown in Fig. S5a, a hierarchy between trialkoxybenzene, aniline, and phenol is ambiguous because most of the data points are in close proximity. It is assumed in this study that trialkoxybenzenes should be given a higher hierarchy than aniline, followed by phenol. Another approach to resolve this issue would be to choose the model group with the highest predicted k_{O3} among the model groups. However, this approach has not been investigated yet. Therefore, further improvements/modifications can be expected in the future. Mono- and dialkoxybenzene and benzene are the lowest in the hierarchy, with the alkoxybenzene given higher hierarchy than the benzene as they have a higher slope (86.74 in Table S3) of the correlation than that of benzene (46.68 in Table S3) (Fig. S5a).

Text S7. Assignment of E_{HOMO-n} (n≥0) to the corresponding reactive sites of a micropollutant

As shown in Table S1, there are many reactive sites for which k_{O3} values are to be predicted using E_{HOMO-n}. Due to an intrinsic characteristic of HOMO-n that its electron density spreads over not only the reactive site of interest but also other parts of the micropollutant, the assignment of E_{HOMO-n} (n≥0) to the corresponding reactive sites can be ambiguous for certain micropollutants. In this section, the proposed procedure to assign E_{HOMO-n}-values to the individual reactive sites are discussed.

In the current k_{O3} prediction platform, E_{HOMO-n} is proposed to be assigned considering two conditions: (i) the degree of the localization of HOMO-n on the reactive site of consideration, which is estimated

as the sum of squares of molecular orbital coefficients (SS_{MOcoef}) of HOMO-n for the reactive site's atoms considered responsible for the reaction with ozone and (ii) the condition that HOMO-n needs to have a conjugation system consisted of *p* atomic orbitals occupying the reactive site' atoms (see below for more details for the respective conditions).

(i) The workflow for the $E_{HOMO\text{-}n}$ assignment based on the first condition is shown in Fig. S7. Quantum chemical computation outputs, in which the molecular orbital coefficients for HOMO-n are printed out, are a prerequisite for this process (see below). The first step is to identify reactive sites of a query compound for which k_{O3} is to be predicted using $E_{HOMO\text{-}n}$. $SS_{MOcoef\text{-}n(i)}$ for the identified reactive sites are subsequently estimated for respective occupied molecular orbitals from the highest occupied molecular orbital (i.e., HOMO where n=0) to lower HOMO-n (n=1,2,3,⋯). $SS_{MOcoef\text{-}n(i)}$, which is an empirical index introduced in this study, is the sum of squares of the molecular orbital coefficients (c) of the basis functions (ϕ) of HOMO-n for the reactive site i's atoms responsible for the reaction with ozone. c is derived from the best wavefunction (ψ) of the molecular orbital (MO) obtained by the variational method where MOs are constructed by a linear combination of atomic orbitals (LCAO). Therefore, SS_{MOcoef} can be used as an approximate index for the localized electron density of the HOMO-n on the atoms of a reactive site. $E_{HOMO\text{-}n}$ is assigned to a reactive site provided that the SS_{MOcoef} derived for the reactive site is higher than the minimum threshold value (see below for the derivation of threshold values). As shown in Fig. S7, the $E_{HOMO\text{-}n}$ assignment continues from E_{HOMO} (n=0) to a lower $E_{HOMO\text{-}n}$ (n≥1) until all identifed reactive sites are assigned with the corresponding $E_{HOMO\text{-}n}$.

In the following, the minimum SS_{MOcoef} threshold values for the $E_{HOMO\text{-}n}$ assignment for different reactive sites are presented. For aromatic compounds, six aromatic carbons of a benzene ring were considered as the atoms for which $SS_{MOcoef\text{-}n(i)}$ are calculated based on the fact that ozone attacks aromatic carbons (Text S9). For example, three aromatic compounds, i.e., benzene, phenol, and aniline, are presented in Fig. S8a. SS_{MOcoef} calculated for the part of the HOMO localized on the carbons of benzene, phenol, and aniline was 0.53, 0.49, and 0.43, respectively. Note that, as defined above, for phenol and aniline the contribution of a molecular orbital to the exocyclic oxygen or nitrogen was not

accounted for in the estimated SS_{MOcoef}. In this manner, SS_{MOcoef}-values for all the aromatic compounds used to develop the k_{O3} prediction models were calculated and were found to range from 0.24 for bisphenol A dianion to 0.53 for methylbenzoate. As shown in Fig. S8a, the two phenolates of the bisphenol A dianion have almost identical SS_{MOcoef} and their summation gives ~0.5 which is similar to the SS_{MOcoef} of benzene and phenol. This indicates that the electron density of the HOMO for the bisphenol A dianion is delocalized equally over the two phenolates. As the bisphenol A dianion was successfully included in the k_{O3} prediction model for the phenol group, the HOMO-n with a SS_{MOcoef} of > 0.25 are considered acceptable for aromatic compounds. Therefore, the minimum SS_{MOcoef} threshold for an aromatic ring was set to ≥0.22 with about 10% of a safety margin applied to 0.25 for the current prediction platform. It turned out that all the heteroaromatic model compounds have SS_{MOcoef}-values higher than 0.22. Therefore, the same threshold value was set to reactive sites belonging to heteroaromatic compounds. Consequently, the reactive sites with a minimum SS_{MOcoef} threshold of 0.22 include *phenol*, *aniline*, *12alkoxybenzene*, *3alkoxybenzene*, *benzene*, *benzazole_dervs*, *adenine*, *guanine*, *cytosine*, *thymine*, *uracil*, *pyridine*, *diazine*, *triazine*, *pyridine_conj*, and *diazine_conj*. For *5hetero_ring*, two adjacent carbons (e.g., imidazole) or three consecutive carbons (e.g., isoxazole) were used to estimate SS_{MOcoef} (atoms with black circles in Fig. S8b are the carbons used to estimate SS_{MOcoef}). This is based on the observation that the ozone attack occurs not at the hetero atoms but at the two adjacent carbons (see *5hetero_ring* in Text S1). SS_{MOcoef}-values of 0.32-0.35 were observed for the model compounds. Therefore, the threshold value was set to ≥0.3 for *5hetero_ring*. For *olefin_conj* with two adjacent carbons being the estimation center atoms (black circles in Fig. S8c) for the same mechanistic reason as for *5hetero ring* above, a rather low threshold value of ≥0.15 was set based on the indigotrisulfonic acid (Fig. S8c).

However, it should be noted that the minimum SS_{MOcoef} threshold values need to be considerd as an approximate suggestion because these threshold values were proposed based on a limited number of the model compounds and there may be micropollutants to which E_{HOMO-n} is not assigned as desired. As shown in Fig. S9, for example, the SS_{MOcoef} for two benzene rings of carbamazepine was estimated

to be 0.21 and 0.2 for E_{HOMO} (n=0), respectively. According to the defualt threshold value (0.22), not E_{HOMO} but a lower E_{HOMO-n} (n≥1) satisfying the threshold condition will be assigned. However, it was judged by the authors that the estimated SS_{MOcoef} (i.e., 0.21 and 0.2) value is significant enough to assign its E_{HOMO} to the benzene rings. In an attempt to resolve this issue, therefore, the threshold values in a configuration text file external to the GUI was made modifiable manually by the user. Note that technical details for the manual modification will be communicated with individual users for the current prediction platform. By setting a lower threshold of 0.2 for the benzene ring, k_{O3} of 3.1×10^2 $M^{-1}s^{-1}$, which was predicted using E_{HOMO}, was obtained for the two benzene rings of carbamazepine (Fig. 3 in the main manuscript). Note that it is not possible to empirically verify the predicted k_{O3} for the benzene ring because the observed k_{O3} predominantly represents the reaction of ozone with the olefin.

(ii) As shown in Fig. S8 for a few examples, such conjugation systems consisted of *p* atomic orbitals occupying the reactive site' atoms were found to at least partially cover the reactive site. This was true for all the model compounds using E_{HOMO-n}. Therefore, the visual examination of the selected HOMO-n to check for the presence of the conjugation system on the reactive site is recommended since the model compounds used to derive this methodology would not represent the entire chemical space of micropollutants of concern. Note that as the visual examination of the HOMO-n is not implemented in the current prediction system, the user needs to seek for an appropriate visualization software such as Molden[41] used in this study and performs the visual inspection manually. Cyproconazole shown in Fig. S10 demonstrates this necessity for visual examination (see the figure caption for a detailed explanation).

In summary, E_{HOMO-n} to reactive sites are automatically assigned based on the workflow shown in Fig. S7 and the default threshold SS_{MOcoef} values suggested above. However, manual modifications of the threshold values as well as the visual inspection of the HOMO-n based on the user's judgment can be necessary for certain compounds, which needs to be further improved in the future.

Text S8. Interpretation of predicted k_{O3}-values for reactive sites of a query compound and their relationships with experimental k_{O3}-values

Based on the computational protocol proposed in Text S5, k_{O3} for identified reactive sites of a *query compound* will be obtained. For some reactive sites, for which a k_{O3} prediction model is unavailable, empirical k_{O3} estimates when available are assigned instead (Table S1). The obtained k_{O3}-values can provide various useful informations. Note that k_{O3} prediction in the current prediction platform is carried out for only one conformer, assuming that the investigated conformer is the dominant species for the reaction with ozone, thus representative of the overall reactivity. Therefore, the treatment of multiple conformers as shown in Text S10 for cetirizine needs to be conducted externally. For a non-ionizing *query compound* (*i*) with *j* reactive sites (j=1,2,3,⋯,n), the 'rate constant prediction' process provides the same number of predicted $k_{O3,j}$ as the number of reactive sites. Comparison of $k_{O3,j}$ between differing reactive sites allows to identify which reactive site dominantly reacts with ozone. The summation of $k_{O3,j}$ for all the reactive sites (*j*) gives rise to a species (*i*)-specific k_{O3}, i.e., $k_{O3,i}$ as follows:

$$k_{O3,i} = \sum k_{O3,j} \text{ over all reactive sites } j \approx k_{O3,exp} \quad \text{(S1)}$$

In principle, $k_{O3,i}$ corresponds to an experimental k_{O3} ($k_{O3,exp}$) for a non-ionizing compound.

For an ionizing *query compound* having *i* acid-base microspecies, k_{O3} predictions are performed for individual microspecies. Therefore, $i \times j$ k_{O3}-values ($k_{O3,i\text{-}j}$) are derived in total. Based on p_i obtained from the speciation analysis (see Text S2 above), an apparent reactive site(*j*)-specific k_{O3} ($k_{O3,site(j)\text{-}app(pH)}$) or an apparent species(*i*)-specific k_{O3} ($k_{O3,species(i)\text{-}app(pH)}$) at a specific pH can be derived by Eqs. S2 and S3 below.

$$k_{O3,site(j)\text{-}app(pH)} = \sum k_{O3,i\text{-}j} \times p_{i(pH)} \text{ over all microspecies } i \quad \text{(S2)}$$

$$k_{O3,species(i)\text{-}app(pH)} = \sum k_{O3,i\text{-}j} \times p_{i(pH)} \text{ over all reactive sites } j \quad \text{(S3)}$$

In the same manner as for a non-ionizing compound above, a comparison of $k_{O3,site(j)\text{-}app(pH)}$ between different reactive sites allows to identify which reactive site dominantly reacts with ozone at a specific pH and a comparison of $k_{O3,species(i)\text{-}app(pH)}$ between different species allows to identify which species dominantly reacts with ozone at a specific pH, respectively.

The summation of $k_{O3,site(j)\text{-}app(pH)}$ over all the reactive sites (j) or of $k_{O3,site(i)\text{-}app(pH)}$ over all the species (i) gives rise to an apparent k_{O3} ($k_{O3,app(pH)}$) for the query compound at a specific pH, which in principle corresponds to an experimental apparent k_{O3} ($k_{O3,exp\text{-}app(pH)}$) at a specific pH.

$$k_{O3,app(pH)} = \sum k_{O3,site(j)\text{-}app(pH)} \text{ over all reactive sites } j$$

$$= \sum k_{O3,species(i)\text{-}app(pH)} \text{ over all microspecies } i \approx k_{O3,exp\text{-}app(pH)} \quad \text{(S4)}$$

Text S9. Selected reaction pathways from literature for pathway predictions

As explained in the main manuscript, reaction rules for pathway predictions were defined based on published reaction pathways in literature. In this section, the selected reaction pathways are discussed in detail divided into different reaction pathway groups, which are shown in the 4th column of Table S1. In the given Schemes S1-S17 for reaction pathways, various logical operators such as == (is equal to), != (is not equal to), && (and), and || (or) are presented in 'if ()' statements when necessary to provide specificity for certain reaction pathways. Reaction pathways will be appropriately modified/updated in future when new information is available.

Phenol (Schemes S1-S3). The discussion on the mechanisms for the reaction of ozone with phenol is divided into three categories; *ortho*, *para*, and *radical*. Ozone reactions initiated by ozone attack at ortho and para positions are discussed in '*ortho*' and '*para*' groups, and those involving radicals generated by a one-electron transfer between phenol and ozone are dealt with in the '*radical*' group.

Phenol-ortho (Scheme S1): The formation of hydroxylated products is the most commonly reported pathway in many studies for aromatic compounds in general.[1,38,42–50] Hydroxylation of phenol at the

ortho position gives rise to the formation of catechol provided that the ortho position is unsubstituted, i.e., substituted by hydrogen. The reaction mechanism was proposed as follows.[49] Initially, ozone attacks at the ortho position to form an ozone adduct (1 in Scheme S1) followed by a singlet oxygen release (3). It undergoes a rearrangement giving rise to a hydroxylated benzene provided that a substituent of the carbon to the ozone adduct was hydrogen (4-5). A substituent release (6) is expected when the substituent is a halide or a nitro group. This is based on the study of ozonation of *p*-chloro- or *p*-nitro-aniline,[38] in which such mechanisms were proposed to explain appreciable detections of chloride or nitrite and nitrate, respecitvely. Although an aryl carbocation was proposed therein following the substituent release without a further proposed pathway, it may be that the aryl carbocation can be stabilized by forming a carbonyl bond between oxygen of a hydroxyl group and the ipso carbon, which gives rise to 1,2-benzoquinone. Due to its unstability, 1,2-benzoquinone was not detected.[49] Various pathways of 1,2-benzoquinone seem to be possible; reaction with phenol to form a biphenyl dimer and the ensuing polymerization,[51] reaction with primary amines and the ensuing polymerization,[52] and reaction with thiols.[53] In the context of water treatment, 1,2-benzoquinone would undergo reactions with various moieties including the aforementioned ones present in dissolved organic matters (DOM), thus its detection is highly unlikely. An ozone adduct can transform into a primary ozonide with either an ipso carbon (7) or a meta carbon (12), both of which follow a typical Criegee mechanism (7-14) leading to a ring cleavage (see the section for 'olefin' below for more details). Similar to the hydro-xylation, a ring cleavage has been observed for many aromatic compounds.[42,44,45,49,54–57] The resulting product undergoes a post-Criegee mechanism which gives rise to differing products depending on the substituents such as halide or carboxyl groups, amongst which hydrogen peroxide elimination is the most commonly observed (see below for details about the post-Criegee mechanism). Exceptionally, *cis*,*cis*-muconic acid formed from water elimination rather than hydrogen peroxide elimination was detected in ozonation of unsubstituted phenol (11).[49] An ozone adduct may undergo a homolytic cleavage of an oxygen-oxygen bond, giving rise to a superoxide ion and an oxyl radical species (15). Further reaction pathways of the oxyl radical are unknown and the superoxide ion is a precursor to hydroxyl radicals. An oxyl radical can in principle be formed from the

dissociation of any ozone adduct. However, this reaction pathway for the oxyl radical is not explicitly presented hereafter for brevity.

Phenol-para (Scheme S2): The analogues reaction mechanisms such as hydroxylation, substituent release, and ring cleavage were proposed for the para position, resulting in the formation of the corresponding products. The formation of 1,4-benzoquinone was additionally proposed via reactions (23-30)

Phenol-radical (Scheme S3): A direct electron transfer from phenol to ozone gives rise to a phenol radical cation and an ozonide radical.[49] As the phenol radical cation has a pK_a of -2,[58] it quickly deprotonates to a phenoxyl radical in aqueous solution (34). An ozonide radical is a precursor to the hydroxyl radical.[59] The formation of a phenol radical cation can be also achieved via a dissociation of an ozone adduct.[11] As there is no mechanistic difference between the two routes in terms of concomitant reaction pathways, no distinction was made herein. Phenoxyl radicals can concurrently undergo various reactions. Superoxide ion described above can combine with the phenoxyl radical either in the para or in the ortho position, leading to the formation of the corresponding quinone (35-37 or 38-40). Quinones can also be formed either from disproportionation of the respective semiquinone radicals (41 or 43) or from two consecutive one-electron oxidations of hydroquinone (42) or catechol (44) by ozone. Alternatively, phenoxyl radicals may combine either at the ortho or at the para position to form various dimers (45-54). This is only important for higher concentrations.

Aniline (Schemes S4-S7). Reaction mechanisms for the reactions of ozone with aniline are discussed by dividing them into four categories; *ortho*, *para*, *nitrogen*, and *radical*.

Aniline-ortho (Scheme S4): The same reaction mechanisms such as hydroxylation, substituent release, and ring cleavage, which were applied to the phenol, are also considered for the reaction of aniline after attack of ozone at the ortho position. It is assumed herein that a benzoquinone imine is formed following deprotonation (60) of the product resulting from the substituent release. The 1,2-benzoquinone imine is prone to hydrolysis to yield 1,2-benzoquione and the corresponding amine.

Similar to above, a ring opening takes place via the Criegee mechanism. In addition, picolinic acid was reported as one of the products resulting from an intramolecular condensation reaction (68-70) of the Criegee product.[11,38]

Aniline-para (Scheme S5): The same reaction mechanisms such as hydroxylation, substituent release, and ring cleavage, which were applied for the ortho position of aniline (Scheme S4), are also considered for the reaction of aniline attacked by ozone at the para position.

Aniline-nitrogen (Scheme S6): An exocyclic nitrogen of aniline is available for ozone attack. Several previous studies reported formation of products involving oxidation of the nitrogen such as nitrosobenzene, nitrobenzene, and azobenzene.[38,60,61] It is worthwhile to mention that it was reported that the yields of the products are too low to consider a exocyclic nitrogen as a major site of ozone attack.[38] An ozone adduct to the nitrogen is initially formed (81) which is followed by various pathways. In the case of a primary amino group, an ensuing hydrogen transfer gives rise to a hydrotrioxide which decays either via hydrogen peroxide elimination to form nitrosobenzene (83, 84) or via a peroxyl radical ($HO_2^{\bullet}$) to form a nitroxyl radical. Despite the fact that a second order rate constant for the reaction between nitrosobenzene and azobenzene is very slow (k=8.4×10^{-6} M^{-1} s^{-1})[62] nitrosobenzene was reported to degrade over time. This indicates that there are hitherto unknown reaction pathways leading to the disappearance of nitrosobenzene. The nitroxyl radical can either disproportionate into nitrobenzene and aniline (87) or react with superoxide in the solvent cage to form nitrobenzene (89, 90) for which superoxide stems from reaction (88).[38] An oxygen transfer (91) is well-established for aliphatic amines in terms of the singlet oxygen and the ensuing products, i.e., *N*-oxide.[2,63–65] For aromatic amines, however, 7% of the singlet oxygen yield was reported for *N*,*N*-dimethylaniline without the detection of a *N*-oxide product.[64] Therefore, this pathway as well as an ensuing hydrogen shift to form a hydroxyl amino group (92) remains to be tentative.

Aniline-radical (Scheme S7): An electron transfer between ozone and aniline (or the dissociation of an ozone adduct) leads to the formation of an aniline radical cation and an ozonide radical (93). Depending on substituents, primary aniline radical cations have a pK_a range between 4 – 12,[48] depending

on which it deprotonates to an anilino radical provided that the amino group is primary or secondary (94). The resulting anilino radical can be further oxidized by ozone to form an ozone adduct radical which eventually leads to the formation of a 1,2-benzoquinone imine (95-99). 2-Haloaniline and 2-aminophenol may give rise to the same product via a cage reaction between the anilino radical and the ozonide radical (102)[1] or via another one-electron oxidation by ozone (103). These reaction pathways (95-103) apply the same for the para position. Alternatively, an anilino radical can be further oxidized by ozone to form a nitroxyl radical (104), which in turn participates in a radical disproportionation (87) or a nitrosation (89) as described above. There are other pathways for the aniline radical cation, which are in competition with a deprotonation of an amino group. For secondary or tertiary amino groups containing an alkyl group, a deprotonation can occur at the α-carbon to form a carbon-centered radical. The ensuing reactions are the same as those for aliphatic amines, details of which are presented below. Similarily, for anilines with a substituent such as an alkyl or an alkoxy group, a deprotonation occurs at the α-carbon of the substituent instead of the amino group (107 and 111), which via superoxide elimination gives rise to a dealkylated aniline and an aminophenol as well as their corresponding aldehyde products (108-110 and 112-114), respectively.[38] This mechanism was proposed to explain the formation of formaldehyde from ozonation of different anilines.[38] Moreover, an aniline radical cation can dimerize into a benzidine (105,106).[66] However, the formation of a benzidine was reported in a pulse radiolysis study in absence of dissolved oxygen as it was saturated by N_2O and an aniline radical cation is generated by another oxidant, i.e., $N_3^{\bullet}$. Indeed, it was not detected from ozonation of different anilines.[38] Therefore, this pathway is likely to be of minor importance.

Alkoxybenzene (Schemes S8-S9). Reaction mechanisms presented herein for alkoxybenzene are relevant to the reactive sites named *12alkoxybenzene* and *3alkoxybenzene* in Table S1.

Alkoxybenzene-ortho (Scheme S8): The same reaction mechanisms such as hydroxylation (116-118), substituent release (119), and ring cleavage (120-125), which were discussed above for phenol and aniline, are also considered. In addition to those, an alcohol elimination may occur.[67] For an alkoxybenzene with a hydrogen in ortho position, a proton transfer from the ortho position to the ter-

minal oxygen of the ozone adduct rearomatizes while forming a very unstable hydrotrioxide (126), which subsequently decays into an oxyl radical and a peroxyl radical (127). The two intermediates may undergo a cage reaction (128) provided that a substituent in para position with respect to the oxyl radical is an alkoxy group. Subsequently, alcohol and hydrogen peroxide elimination occur, giving rise to a 1,4-benzoquinone (129). Alternatively, an alkoxy substituent at the ipso position can be eliminated as alcohol, yielding 2-hydroxyphenol (130-135). A carbocation from reaction (115) may react with water to an intermediate with a hydroxyl group at the ipso position (130). This intermediate subsequently undergoes either an alcohol elimination (131) followed by an ensuing singlet oxygen elimination (132,133) or a singlet oxygen elimination (134) followed by an alcohol elimination (135), both of which give rise to 2-hydroxyphenol as a final product (135).

Alkoxybenzene-para (Scheme S9): The same reaction mechanisms as for alkoxybenzene-ortho such as hydroxylation (136-139), substituent release (140), and ring cleavage (141-143) are considered.

Alkoxybenzene-radical (Scheme S9): It was assumed that a superoxide elimination mechanism may occur, which initiates upon deprotonation of a exocyclic α-carbon of a substituent such as an alkyl or alkoxy group as described above for aniline-radical (107-114). This would lead to a change of an alkoxy group into a hydroxyl group and the formation of the corresponding aldehyde product.

Benzazole_dervs (Scheme S10): 1*H*-1,2,3-triazole-4,5-dicarbaldehyde was reported as a major oxidation product during ozonation of 1*H*-benzotriazole.[55] This can be formed via typical reaction pathways for aromatic compounds; ring cleavage of a benzene ring followed by further oxidation of the product by ozone via the Criegee mechanism. Therefore, the same reaction mechanisms such as hydroxylation, substituent release, and ring cleavage, which were discussed above, are also considered for both positions (6 and 7) of a benzene ring, while preventing an ozonide formation with a fused carbon, i.e., ozonide formation between C7 and C7a or C4 and C3a. When a nitrogen or a sulfur is present in the position 1 in a 5-membered ring, an oxygen transfer may additionally occur to form the corresponding oxide product. A subsequent proton shift may be possible for the resulting *N*-oxide product to form a hydroxylated product (148). Quantum chemical computations suggest that a *N*-oxide formation is

thermodynamically less favored than formation of an ozone adduct on a benzene ring (Lutze *et al*., private communication). Radical reactions were not considered because only little information is available.

Benzene (No scheme provided): The same reaction mechanisms such as hydroxylation, substituent release, and ring cleavage, which were discussed above, are also considered for benzenes. It seems that general rules of regioselectivity by substituents (e.g., ortho-para directors or meta directors) in electrophilic aromatic substitutions can also be applicable to the reaction of ozone with substituted benzenes other than the aromatic compound classes previously described (i.e., phenol, aniline, and alkoxybenzene). However, it was assumed in the current prediction platform that ozone may attack any carbon of the benzene ring because their transformation products from the reaction with ozone would be of minor importance compared to other compound groups. This is because dominant transformation products will be formed from the reaction with hydroxyl radicals due to their low reactivity with ozone. Nonetheless, it is planned to update this aspect in future. Radical reactions were not considered.

Olefins (Scheme S11): As shown in Table S1, the reaction pathway group 'olefin' is applied to reactive sites such as *olefin*, *olefin_conj*, *quinolone*, and *quinolone_carboxylate*. The reaction pathways for olefins with ozone are well established by the so-called Criegee mechanism.[68] A carbon-carbon double bond (C=C) reacts with ozone to form a primary ozonide (151) via a primary zwitterion (150). It subsequently decays into a secondary zwitterion (152). The secondary zwitterion via hydro-lysis decomposes into a α-hydroxyhydroperoxide and a carbonyl compound (153). Depending on substituents of these Criegee products, various ensuing pathways, which are named as a post-Criegee mechanism, are possible.[69,70] A α-hydroxyhydroperoxide gives rise to a carbonyl compound and hydrogen peroxide provided that its two substituents are not halogen (154). When it is substituted by a halogen, a peracid and a halogen ion are generated (159). Depending on a remaining substituent of the peracid, various decays can follow (160-162). Moreover, decarboxylations can occur for the hydroxyhydroperoxide substituted by a carboxyl group (163, 164). The ensuing product is then subject to another

decomposition depending on the presence of substituents such as a halogen and a carboxyl group (155-158). The secondary zwitterion formed in reaction (152) may also undergo decarboxylation provided that the carbon with cationic oxygen is substituted by a carboxyl group, giving rise to a ketone group (172). The additional presence of a carboxyl group to the opposite side leads the resul-ting product to decarboxylation yielding another ketone group (173). An identical product can be also formed via hydrolysis (174) in the absence of a carboxyl group. The final product is subject to further reactions (155-158). Products, which would result from the partial cleavage of a C=C bond, have been reported for a few olefins such as tetramethylethene, 1,1-dichlroethene, and trichloroethene.[69] In competition with a ring-closing reaction (151), a primary zwitterion may react with water (165) or undergo a chlorine shift (169). An oxygen is subsequently released (166, 167) and the halogen substituent when present is also released (167, 168). However, products from the partial cleavage were not detected for most olefinic compounds.[69–74] Therefore, this mechanism is considered to be of minor importance.

Amine (Scheme S12): As shown in Table S1, the amine pathway group is assigned for four reactive sites; *amine, amine_conj*, *phenylazo* and *sulfonamide(2°)_deprot*. As reaction pathways for *amine_conj* and *sulfonamide(2°)_deprot* belong to the pathways for *amine* under specific conditions, the overall reaction pathways for *amine* are first described (Scheme S12), followed by *amine_conj* and *sulfonamide(2°)_deprot*. As described above, the *phenylazo* is exceptionally defined as an amine for a structurally consistent categorization, even though, its pathways are implemented in hetero_ar. The reaction of amine with ozone is initiated by an electrophilic attack of ozone at the nitrogen lone-pair electrons to form an ozone adduct (175). A subsequent loss of oxygen to *N*-oxide (176) is commonly observed as a stable product for tertiary amines.[2,16,63,65,75–77] The resulting *N*-oxide transforms into a hydroxyl group (177) for secondary amines[2,42,54,78] or potentially primary amines. A proton transfer can occur for primary or secondary amines to form a hydrotrioxide (178), which subsequently decomposes into a nitroxyl radical and a peroxyl radical (179). The nitroxyl radical can also be formed via a dissociation of the ozone adduct (180) or via oxygen transfer for an aminyl radical by

ozone (187) formed from an electron-transfer-involving reaction pathway (see below for details). The nitroxyl radical and the aminyl radical can play a role to consume additional ozone via chain reactions (187 and 188).[1] This indicates a high ozone consumption compared to the degradation of an amine, therefore, supports a poor mass balance between ozone and amine frequently reported in relevant studies.[1,2,78] The nitroxyl radical can be further oxidized by ozone via electron transfer leading to an oxoazanium formation (181)[78] or disproportionate into two different pairs of products (182, 183).[11,79] Similar to aromatic compounds such as phenol and aniline described above, electron transfer-initiated reactions are of consideration for amine, which results in the formation of an amine radical cation. As described above, a secondary aniline nitrogen was defined as an amine for its k_{O3} and pathway prediction. The resulting radical cation species can have common reaction pathways for the aniline radical cation starting from reaction (94) in Scheme S7 as it has the same reaction center. Depending on pH of an aqueous system and the pK_a value of an amine radical cation, it can deprotonate to an aminyl radical (186). A pK_a range of 5.3 – 6.8 was reported for the amine radical cations of dimethylamine, diethylamine, pyrrolidine, and piperidine.[80] As already described above, an oxygen transfer by ozone (187) transforms an aminyl radical to a nitroxy radical followed by various ensuing pathways (181-183). In competition with deprotonation on the nitrogen, deprotonation can occur at the α-carbon to the nitrogen (189), yielding a carbon-centered radical species. Provided that the α-carbon is substituted by a carboxyl substituent, decarboxylation occurs, giving rise to the same carbon-centered radical. Most carbon-centered radicals are known to react with oxygen with a close to diffusion-controlled rate (k in the order of 10^9 $M^{-1}s^{-1}$) to form an α-amino-peroxyl radical (191).[81] The α-amino-peroxyl radical undergoes an unimolecular superoxide ion ($O_2^{\bullet-}$)-elimination to yield an iminium ion. It subsequently hydrolyzes into a carbonyl compound and an amine provided that a substituent of the amine is not a hydroxyl group.[82,83] If a hydroxyl group is present, it is assumed to deprotonate to nitrone which then hydrolyzes into the same products as above.[79] Alternatively, peroxy radicals can undergo a bimolecular recombination to form a tetroxide intermediate for which there are several possible reaction pathways.[84] The tetroxide decomposes into a carbonyl product, a product with a hydroxyl group, and oxygen (197), which is also known as the Russell mechanism.[85] Another possible reaction gives rise

to two carbonyl products and hydrogen peroxide, which is also known as the Bennett mechanism (198).[86] These two mechanisms necessitate the presence of hydrogen for the carbon bound to a tetraoxide, i.e., not possible for a tertiary peroxyl radical. An appreciable amount of hydrogen peroxide was reported for ozonaiton of morphorine (28% per compound degraded) and piperazine (60%), but not for piperidine (4.6%).[78] This indicates that the Bennett mechanism may have played an important role depending on the target compound. Alternatively, two oxyl radicals can be generated with oxygen (199), which undergo subsequent reactions (200, 201) to form the same products as above.[84,87]

The pathways for *amine_conj* are only based on ozonation of cylindrospermopsin bearing a single-bonded tertiary nitrogen in a tricyclic guanidine,[10] in which hydroxylation via oxygen transfer (176, 177 in Scheme S12) and dealkylation (175, 185, and 189-195 in Scheme S12) via electron transfer were reported. While the *N*-oxide formation is generally known to be the final product for tertiary amines as described above, it was proposed therein that an intermolecular hydrogen transfer takes place from a protonated secondary amine of the guanidine moiety to the oxide to form a hydroxyl group. However, as both *N*-oxide and hydroxylamine products have the same mass and identifications were conducted only with mass spectrometry, only *N*-oxide formation was tentatively considered in the current prediction. Moreover, it is assumed that a secondary nitrogen has the same fate.

The reaction pathways for *sulfonamide(2°)_deprot* are based on ozonation of hydrochlorothiazide,[2] in which a cyclic imine product (i.e., chlorothaizide) was reported to be a stable product, which can be formed via a partial dealkylation (175, 185, and 189-192 in Scheme S12). Although an imine product typically undergoes hydrolysis, the formation of an amidine seems to guarantee its stability by tautomerism. Therefore, the reaction rule is defined to cease for the case of amidine formation. Moreover, a protonation of the radical intermediate to form a radical cation was additionally proposed herein for *sulfonamide(2°)_deprot*. pK_a-values of aniline radical cations are several orders of magnitude higher than those of aniline (e.g., 4.63[38] and 7.05[48] for unsubstituted aniline and its radical cation and 5.08[38] and 8.50[48] for 4-methylaniline and its radical cation). Assuming that the same account can be also applied to secondary sulfonamides, the secondary sulfonamide radical species formed from oxidation

of the sulfonamide nitrogen anion would readily be protonated due to an increased basicity, which then undergoes a radical shift to α-carbon (189). Another reaction mechanism leading to the same product, i.e., chlorothiazide, via a tautomeric rearrangement was suggested. However, this pathway was not defined herein because it is considered rather exceptional.

Guanine (Scheme S13) The reaction pathways in Scheme S13 are based on guanine without an amino group at C(2). It was assumed in this study that the pathways can be generally applied to various guanine derivatives. The reaction pathways for guanine were not extended for *adenine* despite the similar chemical structure because the assumption for 1,2-acyl shift seems to be premature (see below). The oxidation of guanine analogues by ozone was reported to be initiated by the formation of an ozonide at the 4,5-double bond (202).[12,88] This is reminiscent of the Criegee mechanism. However, the formation of an epoxide intermediate via a subsequent singlet oxygen release (203) was proposed to take place for carboxy-acyclovir bearing a guanine moiety,[12] which is not a typical Criegee reaction. The detection of the epoxide was reported from ozonation of uric acid in microdroplets using online mass spectrometry,[88] and the yields of singlet oxygen per ozone were 40%, 37%, 16%, and 10% for 2′-deoxyguanosine, guanosine, 2′-deoxyadenosine, and adenosine, respectively. Further a 1,2-acyl shift (204) and the subsequent hydrolysis of the imidazole (205) give rise to the final product. An additional hydrolysis (206, 207) of the imine is expected for a guanine analogue in which an amino group is absent at C(2). For guanine with an amino group at C(8), a hydrolytic product formation (207) is not expected.

However, it is unclear whether or not the final product is formed via an epoxide formation from ozonation of acyclovir.[12] The conventional ozonation pathways can also lead to a singlet oxygen release, the intermediate with the same mass as the epoxide, and the same final product. Applying the conventional ozone reaction pathways to guanine we proposed reactions (208)-(212) based on an ozone adduct formed at C(5) (208). Singlet oxygen release (209) subsequently follows, leaving an oxygen anion in the intermediate with the same mass as the epoxide. Via ensuing rearrangement (210) and hydrolysis (211) the same product is formed. Moreover, the conversion of the amino group into an

imino group seems to be better explained by the presence of a carbocation (C4) (208) which leads to a imine bond shift from C(2)=N(3) to N(3)=C(4) (209). In this study, therefore, we chose to perform a pathway prediction for purines with the conventional ozonation pathways. It is noted that in the current study the pathway prediction for guanine is limited to the neutral form as little information is available for the ionic form.

Uracil (Scheme S14) The reaction pathways for the reaction of ozone with uracil are presented in Scheme S14, which is based on the reaction pathways proposed for thymine and 5-chlorouracil during ozonation.[64,89] Those pathways are assumed to be also applicable to uracil and cytosine analogues as a ring opening via the Criegee mechanism is also likely to occur for cytosines (see below). The C4 or C5 of a pyrimidine is the initial site of attack by ozone to form an ozone adduct (213), and subsequently undergoes the typical Criegee mechanism (214-218), which was already presented for aromatic compounds and olefins above. Asymmetric cleavage of the ozonide (215 versus 217) takes places, giving rise to two differing zwitterions (A and B). The zwitterionic intermediate A undergoes a series of subsequent reactions (219)-(225), giving rise to a hydrantoin. Hydrantoin can also be formed via the reactions (226)-(231) from A. The same reactions (226)-(231) can be applied to B, giving rise to a hydrotoin. The intermediate (B) alternatively undergoes fragmentation reactions (232)-(237) initiated by the deprotonation of N(3) (233). A partial oxidation rather than a ring opening may occur for a pyrimidine with the anionic N3. The carbocation at C6 of the intermediate (C) is neutralized with the formation of an imine bond (N1=C6) (238). This intermediate undergoes a single oxygen release (239) followed by hydrolysis (241) to form a glycol product. Alternatively, a halogen ion can be released (242) when a halogen substituent is present at the carbon of the ozone adduct to form 6-hydroxylpyrimidine-2,4,5-trione via the subsequent hydrolysis (243).

Hetero_ar (no Scheme provided) There is only one reaction pathway defined for non-conjugated *pyridine*, *diazine*, *triazine,* and *phenylazo*. It is a *N*-oxide formation (175, 176 in Scheme S12). Pyridine during ozonation was reported to yield almost 80% *N*-oxide per pyridine degraded.[90] Moreover, azoxybenzene was reported to be the major product from ozonation of azobenzene.[11] Based on these

reports, it is assumed that *N*-oxide formation may be generally applicable for non-conjugated *pyridine*, *diazine,* and *triazine, and phenylazo*. For *triazine*, an aromatic nitrogen conjugated with an amino group was excluded from *N*-oxide formation. Dealkylation of the exocyclic secondary or tertiary amino group were reported for atrazine.[91] Therefore, a dealkylation mechanism was additionally defined for triazine with a secondary or a tertiary amino substituent, which is the same pathway as the dealkylation of amine (175, 185, and 189-193 in Scheme S12).

Diaminopyrimidine (Scheme S15) Despite the fact that this reaction pathway group is defined for *diazine_conj*, by which diverse chemical structures are represented (Fig. S1), only the reaction mechanism for 2,4-diaminopyrimidines was defined herein (Scheme S15) due to a limited information for other moieties.

Ozonation products of the 2,4-diaminopyrimidine were reported for trimethoprim,[92,93] However, the presented pathways in Scheme S15 are not the same as the ones proposed therein.[92,93] In the reference study,[93] carbonylation of the carbon in the position 5 of a diaminopyrimidine moiety was proposed, giving rise to a product with the molecular weight of 324. However, the proposed product seems to have a MW of 322 rather than 324. Assuming that the MW 324 of the detected product is correct, we propose pathways giving rise to a product with the same mass, which can be explained based on the known ozone chemistry. In Scheme S15, the C(5) of a diaminopyrimidine is available for the attack by ozone to form a carbocation at the C(6) (244). Note that this pathway is also applicable to the C(6) but not shown for simplicity. Assuming that the carbocation is stable enough, it may react with water (245)[67] rather than undergoing other decay routes. Singlet oxygen can then be released from the ozone adduct with a subsequent proton addition (246). The resulting diol has the same mass (MW=324) of the detected product for trimethoprim. We additionally propose a halogen elimination in its presence to form a product with a carbonyl group (247 and 248).

Ethynyl (Scheme S16) An ethynyl group was reported to undergo reactions analogous to the Criegee mechanism for olefins.[47] As shown in Scheme S16, a primary ozonide is formed resulting in a C-C double bond (249). Further hydrolysis breaks down the C=C bond to form a carbonyl and a hydroxy-

hydropeorxide group for the respective carbons (250, 215). Hydrogen peroxide elimination gives rise to a glyoxal (252). Alternatively, hydrolysis may lead to the cleavage of the C-C bond, yielding two terminal carboxyl groups (253).

Sulfur (no Scheme provided) The oxygen transfer from ozone to sulfur via an ozone adduct is known to be predominant for most organosulfur compounds such as sulfides,[64,94] sulfinic acids,[64,95] and thiols.[64,96,97] For sulfides, sulfoxide is the stable ozone oxidation product. From ozonation of glutathione, thiol was reported to be oxidized to sulfonic acids.[96] Although a stepwise oxidation from thiol to sulfonic acid via successive oxygen transfer is expected, detailed mechanisms have not been elucidated yet. Therefore, a direct oxidation pathway to sulfonate from thiol and thiolate was defined without intermediates.

Post_ozone reactions (Scheme S17) Oxidation products formed via various reaction pathways described above can undergo various post-ozone reactions such as intramolecular rearrangements, hydrolysis, and fragmentation. Such reaction pathways, which are generally applicable, are discussed herein. They are simultaneously applied to predicted final products of all the compound groups. First, post-Criegee reactions ((154)-(164) in Scheme S11), which mainly deal with hydrolysis and decarboxylation, can be generally applied. In addition, it was reported from ozonation of progesterone that an intramolecular Bayer-Villiger type reaction ((254)-(255)) may occur to a minor extent for an α-hydroxyhydroperoxide product,[5] which converts ketones to esters. The resulting product from the Bayer-Villiger type reaction for progesterone is an anhydride, which undergoes hydrolysis to the corresponding carboxylic acids (256). The hydrolysis of anhydrides was defined to be generally applicable, not only specific to the product from the Bayer-Villiger type reactions.

Imine formation between primary amine and ketone via reactions ((257)-(259)) was observed from ozonation of carbamazepine.[73] As described above in the amine section (reaction (193) in Scheme S12), imine is prone to hydrolysis, therefore, a reverse reaction is possible, which depends on the stability of the imine. The imine product of carbamazepine during ozonation, 1-(2-benzaldehyde)-(1*H*,3*H*)-quinazoline-2,4-dione (BQD), was stabilized by forming an aromatic system. Moreover,

amidine was discussed above to be stablized by tautomerism. It is generally known that hydrolytic stability of imines are attributed to delocalization of the imine double bond with electron-withdrawing substituents, e.g., oxime and hydrazone.[98] Although there will be various other environments where imines are resistant to hydrolysis, only the aforementioned conditions are implemented.

Text S10. Case study: comparison between experimental k_{O3} and predicted k_{O3} for cetirizine

Fig. S12a shows a comparison between the experimental k_{O3}[2] and the predicted k_{O3} for cetirizine. The predicted k_{O3} values were obtained from the default method of this study (i.e., HF/3-21G//MMFF in Text S5) using a species distribution as shown in Fig. S13c. The acid-base speciation scheme for cetirizine is shown in Fig. S13b where individual acid-base species are denoted as X_i (i=1 to 8). Note that the most basic p$K_{a,3}$-value (Fig. S13b) is reported to be ~8.0 in literature.[99,100] However, a p$K_{a,3}$ of 7.0, which was suggested from the determination of k_{O3} for cetirizine,[2] was instead chosen for a consistent comparison between the measured k_{O3} and the predicted k_{O3}. As shown in Fig. S12a, the predicted apparent k_{O3} (dash-double-dot cyan line) was much higher than the experimental k_{O3} (yellow circles) for the entire pH range. Noticeably, the error increases to more than an order of magnitude with decreasing pH below 7. The experimental k_{O3} manifests the effect of the protonation of the amine nitrogen 1 (p$K_{a,3}$=7.0), whereas the predicted apparent k_{O3} is uninfluenced by pH until pH 3. In Figs. S12c and S12d, individual nitrogen-specific k_{O3} values of the species contributing to the predicted apparent nitrogen-specific k_{O3} in Fig. S12a are presented (black lines). The species X_2, X_6, and X_8 and the species X_3, X_7, and X_8 contribute to the predicted k_{O3} of nitrogens 1 and 2, respectively. For nitrogen 2 in Fig. S12d, the dominant species at pH > 8 and acidic to neutral pH (2-7) are X_8 (dotted black line) and X_7 (long-short dashed black line), respectively. Therefore, the plateau for the pH range 3-6 in Fig. S12a is a result of the nitrogen 2 of X_7. Interestingly, as shown in Fig. S14 (top row) a moderate hydrogen bond was formed between the hydrogen of the protonated amine and the carboxylate oxygen with a distance of 2.67 Å for X_7. Moreover, other species with a protonated amine such as X_2, X_3, and X_6 show even stronger hydrogen bonding because of shorter distances (1.57-2.25 Å) (Fig.

S14). Ionic intramolecular interaction between the two atoms makes the overall geometry of the species more stable. However, as the presented conformations were obtained by the MMFF94 method in the gas phase, they may not be relevant for aqueous solution where ionic charges can be locally stabilized in the polar environment. Therefore, initial geometries of cetirizine in spread-out conformations were optimized by the B3LYP/6-31G* method employing the integral equation formalism polarizable continuum model[101] (IEFPCM) as a solvation method. Note that this computation was performed using Gaussian.[19] The optimized geometries are presented in Fig. S14 (bottom row) in comparison with the geometries obtained by MMFF94. It can be seen that straight conformations were retained following the geometry optimization. When the total electronic energies obtained with the default method (HF/3-21G) using ORCA are compared (Fig. S14, top row), the optimized geometries in aqueous medium were more stable than the gas phase geometries obtained by the MMFF94 method. This supports the notion that the MMFF94 conformations are not necessarily relevant to the experimental data obtained in aqueous solution. The predicted k_{O3} values using the geometries relevant to the aqueous solution are presented in Figs. S12b, S12c, and S12d. Note that the predicted k_{O3} values for the benzene rings are neglected because of the low reactivity ($< 10^2$ $M^{-1}s^{-1}$). With the spread-out conformers, a predicted apparent k_{O3} decrease for nitrogen 2 was observed (Fig. S12b), which is consistent with the amine protonation and the experimental data. The absolute level and the extent of the decrease was, however, still small compared to the experimental data.

In fact, all the prediction outputs presented in Fig. S12 are obtained with the acid-base species of dextrocetirizine (*S* enantiomer) because the default MMFF94 method for the 3D geometry generation consistently yielded dextrocetirizine (Fig. S14). However, the racemic mixture of levocetirizine (*R* enantiomer) and dextrocetirizine was used to determine experimental k_{O3}.[2] To be consistent with the experimental conditions, therefore, the presence of levocetirizine needs to be taken into account and a k_{O3} prediction for levocetirizine in aqueous medium was performed as well. In the same way as above for dextrocetirizine, initial geometries were spread-out conformations, which were then optimized by B3LYP/6-31G* with the solvation method employed by Gaussian.[19] The optimized geometries were

submitted to single point calculations with HF/3-21G using ORCA and the corresponding k_{O3} were predicted accordingly. As shown in Fig. S15a, the predicted k_{O3} for levocetirizine overall shows a similar pattern to the predicted k_{O3} for dextrocetirizine (Fig. S12b) except that a shift of the dominant nitrogen from 1 to 2 at pH below 7 was observed for levocetirizine. The difference in the predicted apparent nitrogen-specific k_{O3} between levo- and dextro-cetirizines was within a factor of 5 at maximum. The predicted apparent k_{O3} values for the two enantiomers were combined with an equal weighting factor (0.5) to derive the predicted apparent k_{O3} for the racemic mixture, which is a more appropriate representation of the experimental data (Fig. S15b). However, even with the correct consideration of the enantiomers, the prediction still fails with an error of more than 2 orders of magnitude.

The cetirizine case study presented above highlighted some weaknesses of the developed k_{O3} prediction platform. The MMFF94 method, a default method for 3D geometry generation in this study, may give rise to geometries that are irrelevant to the aqueous phase, potentially leading to an error in predicting k_{O3}. Therefore, the development of k_{O3} prediction models with various conformer search tools generating an aqueous-phase relevant geometry will be of interest. Although it was evaluated for cetirizine that the overall improvement by using geometries relevant to aqueous phase was rather small, this aspect needs to be carefully taken into account because the potential error induced by the use of gas-phase geometries to a prediction error is not well known for other chemical structures.

Moreover, the currently developed platform considers only one conformer obtained from the MMFF94 method for the k_{O3} prediction. As shown for cetirizine, experiments may be conducted for a compound with isomerism. However, this aspect is not planned to be implemented in the current prediction platform due to a number of associated aspects that are difficult to resolve, e.g., the prediction of weighting factors (relative populations) of individual isomers. Nevertheless, one has to keep in mind that k_{O3} predictions may be significantly influenced by the presence of multiple conformers for some compounds.

Lastly, as it turned out above that the consideration of the geometry and enantiomers did not lead to a significant improvement, the poor k_{O3} prediction model for cetirizine seems to stem from the failure of the prediction model itself. Further investigations on whether this failure occurs consistently for similar compounds as cetirizine (e.g., a compound containing a piperazine moiety) would be helpful for a better understanding of the problem.

Table S1. A list of all reactive sites defined for the prediction platform and the associated k_{O3} estimate, k_{O3} prediction model group, and reaction pathway group. For structures see Fig. S1.

Reactive site	k_{O3} estimate[a,b], $M^{-1}s^{-1}$	k_{O3} prediction model group (associated orbital energy)	reaction pathway group
Aromatic compounds			
phenol	*prediction*	phenol (E_{HOMO-n})	phenol
aniline	*prediction*	aniline (E_{HOMO-n})	aniline
cyclic_aniline	4.5×10^{2}	*n.a.*	aniline
12alkoxybenzene	*prediction*	mono- and di-alkoxybenzene (E_{HOMO-n})	alkoxybenzene
3alkoxybenzene	*prediction*	trialkoxybenzene (E_{HOMO-n})	alkoxybenzene
benzene	*prediction*	benzene (E_{HOMO-n})	benzene
benzazole_dervs	*prediction*	benzene (E_{HOMO-n})	benzazole_dervs
Olefins/Ethynyl group			
olefin	*prediction*	olefin ($E_{NBO,\ C=C}$)	olefin
olefin_conj	*prediction*	miscolefin (E_{HOMO-n})	olefin
ethynyl	2×10^{2}	*n.a.*	ethynyl
Amines			
amine	*prediction*	amine ($E_{NBO,\ LP-N}$)	amine
hydrazine	*n.a.*	*n.a.*	*n.a.*
N_sulfenamide	*n.a.*	*n.a.*	amine
amine_conj	*n.a.*	*n.a.*	*n.a.*
phenylazo	*prediction*	hetero_ar (E_{HOMO-n})	hetero_ar
sulfonamide(1°)_prot	1	*n.a.*	*n.a.*
sulfonamide(1°)_deprot	7×10^{3}	*n.a.*	*n.a.*
sulfonamide(2°)_prot	*n.a.*	*n.a.*	*n.a.*
sulfonamide(2°)_deprot	3×10^{5}	*n.a.*	amine
sulfonamide(3°)	*n.a.*	*n.a.*	*n.a.*
Heteroaromatic compounds			
adenine	*prediction*	adenine (E_{HOMO-n})	*n.a.*
guanine	*prediction*	guanine (E_{HOMO-n})	guanine
cytosine	*prediction*	cytosine (E_{HOMO-n})	uracil
thymine	*prediction*	thymine (E_{HOMO-n})	uracil
uracil	*prediction*	uracil (E_{HOMO-n})	uracil
pyridine	*prediction*	hetero_ar (E_{HOMO-n})	hetero_ar
diazine	*prediction*	hetero_ar (E_{HOMO-n})	hetero_ar
triazine	*prediction*	hetero_ar (E_{HOMO-n})	hetero_ar
pyridine_conj	*prediction*	miscolefin (E_{HOMO-n})	*n.a.*
diazine_conj	*prediction*	miscolefin (E_{HOMO-n})	diaminopyrimidine
quinolone	1.0	*n.a.*	olefin
quinolone_carboxylate	2×10^{4}	*n.a.*	olefin
5hetero_ring	*prediction*	miscolefin (E_{HOMO-n})	olefin
Organosulfur compounds			
sulfide	8×10^{5}	*n.a.*	sulfur
disulfide	2×10^{5}	*n.a.*	sulfur
sulfide_6cyc	5×10^{4}	*n.a.*	sulfur
sulfide_5cyc	5×10^{3}	*n.a.*	sulfur
thiol_prot	3×10^{4}	*n.a.*	sulfur
thiol_deprot	3×10^{6}	*n.a.*	sulfur
thiophenol	*n.a.*	*n.a.*	*n.a.*
thiophosphoramide	*n.a.*	*n.a.*	*n.a.*
S_sulfenamide	*n.a.*	*n.a.*	*n.a.*
thioamide	*n.a.*	*n.a.*	*n.a.*
thioketone	*n.a.*	*n.a.*	*n.a.*
sulfide_conj	6×10^{2}	*n.a.*	*n.a.*
sulfoxide	10	*n.a.*	*n.a.*
sulfinic_acid	2×10^{6}	*n.a.*	sulfur
organothiophosphate	*n.a.*	*n.a.*	*n.a.*

n.a.: not available. [a]reference compounds, from which k_{O3} estimates are derived, are presented in Table S2. [b]a reactive site, to which '*prediction*' is assigned, is to be predicted by the corresponding k_{O3} prediction model group in the 3rd column.

Table S2. k_{O3} estimates for reactive sites for which no k_{O3} prediction model is available and the corresponding reference compounds from which k_{O3} estimates are derived.

Reactive site[a]	k_{O3} estimate, $M^{-1}s^{-1}$	Reference compound (empirical k_{O3}, $M^{-1}s^{-1}$)
Aromatic compounds		
cyclic_aniline	4.5×10^2	hychlorochlorothiazide (6×10^2),[2] ciprofloxacin diprotonated (4×10^2),[3] enrofloxacin diprotonated (3.3×10^2)[3]
Olefin/Ethynyl group		
ethynyl	2×10^2	1-ethinyl-1-cyclohexanol (2×10^2)[47]
Amine		
sulfonamide(1°)_prot	1	*N*,*N*-dimethylsulfamide(1.24[c])[102]
sulfonamide(1°)_deprot	7×10^3	*N*,*N*-dimethylsulfamide(7×10^3[,d])[102,103]
sulfonamide(2°)_deprot	3×10^5	hydrochlorthiazide (3×10^5)[2]
Heteroaromatic compounds		
quinolone	1	flumequine neutral (1.2)[3]
quinolone_carboxylate	2×10^4	flumequine deprotonated (1.8×10^3),[3] ciprofloxacin monoprotonated (7.5×10^3),[3] enrofloxacin monoprotonated (4.6×10^4)[3]
Organosulfur compounds		
thiol-prot	3×10^4	cysteine (4.2×10^4),[1] glutathione (2.0×10^4)[1]
thiol-deprot	3×10^6	cysteine anion (2.4×10^6),[1] glutathione anion(4.0×10^6)[1]
sulfide	8×10^5	aldicarb (4.3×10^5),[2] methionine (2.4×10^6)[1]
sulfide_5cyc[b]	5×10^3	penicillin G (4.8×10^3)[3]
sulfide_6cyc[b]	5×10^4	cephalexin ($5 - 10\times10^4$)[3]
sulfide_conj	6×10^2	EPTC (500;estimated)[4], molinate (500;estimated),[4] vydate (620)[2]
disulfides[c]	2×10^5	*trans*-1,2-dithiane-4,5-diol(2.1×10^5),[5] bis(2-hydroxyethyl) disulfide (1.7×10^5)[5]
sulfoxide	10	dimethylsulfoxide (8)[1]
sulfinic acid	2×10^6	methanesulfinate (2.0×10^6)

[a]see Fig. S1 for their chemical structures. [b]The number indicates the number of the atoms of a cyclic ring. [c]k_{O3} of 550 and 1.0×10^3 $M^{-1}s^{-1}$ was reported for cystine at pH 1 and 3.1, respectively. These values are significantly different from *trans*-1,2-dithiane-4,5-diol and bis(2-hydroxyethyl) disulfide and the reason for this is yet unclear. In this study, the representative k_{O3} for disulfides was set excluding cystine. [c]Apparent k_{O3} at pH 5.[102] [d]Estimated based on the reported apparent k_{O3} measurements between pH 5 and 10[102] and the reported pK_a.[103]

Table S3. The currently developed k_{O3} prediction models for various organic compound groups in comparison with the previously developed models

k_{O3} predictiom model group	Orbital energy	Present model[a]			Reference model[b]	
		slope(a) /y-intercept(y_0)[c]	N_{model}[d] (N_{total}[e])	R^2 (MUE[f], RMSE[g])	N_{model}[d] (N_{total}[e])	R^2 (MUE[f], RMSE[g])
Aromatic compounds						
phenol	E_{HOMO-n}	55.43/22.19	35 (36)	0.94 (0.46, 0.67)	35 (36)	0.94 (0.42, 0.61)
aniline[h]	E_{HOMO-n}	21.30/12.31	22 (22)	0.43 (0.26, 0.30)	16 (16)	0.85 (0.34, 0.43)
mono- and dialkoxybenzenes	E_{HOMO-n}	86.74/30.55	17 (17)	0.74 (0.71, 0.96)	17 (17)	0.87 (0.57, 0.69)
trialkoxybenzenes	E_{HOMO-n}	-95.32/-25.80	4 (4)	0.69 (0.26, 0.31)	4 (4)	0.95 (0.11, 0.12)
benzenes	E_{HOMO-n}	46.68/16.71	39 (40)	0.75 (0.58, 0.78)	40 (40)	0.82 (0.47, 0.61)
Olefin	$E_{NBO,C=C}$	31.26/16.78	45 (53)	0.79 (0.57, 0.76)	45 (53)	0.84 (0.51, 0.69)
Amine	$E_{NBO,LP-N}$	29.83/19.25	59 (59)	0.75 (0.61, 0.84)	59 (62)	0.83 (0.59, 0.86)
Heteroaromatic compounds						
adenine	E_{HOMO-n}	49.79/18.12	8 (8)	0.76 (0.63, 0.70)	8 (8)	0.73 (0.66, 1.26)
cytosine	E_{HOMO-n}	41.19/17.35	8 (8)	0.91 (0.35, 0.45)	8 (8)	0.98 (0.17, 0.28)
guanine	E_{HOMO-n}	57.06/21.45	8 (8)	0.95 (0.31, 0.33)	8 (8)	0.99 (0.14, 0.23)
thymine	E_{HOMO-n}	35.50/16.40	5 (5)	0.92 (0.22, 0.26)	5 (5)	0.94 (0.17, 0.34)
uracil	E_{HOMO-n}	50.55/20.75	8 (8)	0.80 (0.52, 0.65)	8 (8)	0.84 (0.45, 0.76)
other heteroaromatic compounds (hetero_ar)	E_{HOMO-n}	52.78/19.51	9 (9)	0.80 (0.48, 0.57)	*n.a*	*n.a*
Miscellaneous olefins (miscolefin)	E_{HOMO-n}	75.10/29.60	17 (17)	0.95 (0.43, 0.48)	32 (32)	0.89 (0.51, 0.67)

[a] k_{O3} prediction models developed with the HF/3-21G//MMFF94 method. See Texts S2 and S3 for the details. [b]Previously developed models with HF/6-31G method used throughout all the computations for all k_{O3} prediction model groups except for the amines for which HF/6-3|11++G** was used. See the reference[4] for more details. [c]applied to a linear regression model to predict k_{O3} in log units: $\log k_{O3} = aE + y_0$. [d]the number of model compounds used to develop a linear regression model. E, [e]the number of the compounds including both model compounds and outliers used to calculate MUEs and RMSEs. [f]mean unsigned error. [g]root-mean-square error. [h]mean unsigned error. [g]root-mean-square error.

Table S4. The selected model compounds used for the development of a k_{O3} prediction model for differing aromatic compound groups

No.	Phenol	Aniline[a]	Mono- and di-alkoxybenzene	Trialkoxybenzene	Benzene
1	bisphenol-A	aniline	acebutolol protonated	trimethoprim monoprotontaed	benzaldehyde
2	bisphenol-A dianion	3-chloro aniline	anisole	trimethoprim diprotontaed	benzene
3	2-chloro phenol	4-chloro aniline	atenolol protonated	1,3,5-trimethoxybenzene	benzene sulfonate
4	2-chloro phenolate	2-fluoro aniline	bezafibrate	3,4,5-trimethoxytoluene	benzoate
5	4-chloro phenol	3-fluoro aniline	carbofuran		chlorobenzene
6	4-chloro phenolate	4-fluoro aniline	2,4-dichlorophenoxyacetic acid		diazepam
7	2,3-dimethyl phenol	3-iodo aniline	1,4-dimethoxy benzene		diehtyl-*o*-phthalate
8	2,4-dimethyl phenol	4-iodo aniline	methoxychlor		1,3-dichlorobenzene
9	2,6-dimethyl phenol	2-methoxy aniline	4-methoxy-1-naphthalenesulfonic acid		ethylbenzene
10	3,4-dimethyl phenol	3-methyl aniline	2-methyl-4-chlorophenoxyacetic acid		galaxolide
11	2,6-di-*t*-butyl-4-methylphenol	4-methylsulfone aniline	2-methyl-4-chlorophenoxypropionic acid		ibuprofen
12	17b-estradiol	3-nitro aniline	metoprolol protonated		isopropylbenzene
13	17b-estradiol anion	4-nitro aniline	naproxen anion		methylbenzoate
14	estriol	sulfamethoxazole	1-phenoxy-2-propanol		1-methylnaphthalene
15	estriol anion	sulfamethoxazole anion	propranolol protonated		naphthalene
16	estrone	4-methyl aniline	2,4,6-tribromoanisole		nitrobenzene
17	estrone anion	4-methoxy aniline	2,4,6-trichloroanisole		toluene
18	catechol	4-phenylenediamine			tonalide
19	resorcinol	2-hydroxy aniline			1,2,3-trimethylbenzene
20	hydroquinone	4-hydroxy aniline			1,3,5-trimethylbenzene
21	*o*-cresol	4-aminobenzoic acid			*m*-xylene
22	*m*-cresol	*N*,*N*-dimethylaniline			*o*-xylene
23	*p*-cresol				*p*-xylene
24	4-nitrophenolate				alachlor
25	4-*n*-nonylphenol				chlorotoluron
26	4-*n*-nonylphenolate				diuron
27	4-octylphenol				isoproturon
28	pentabromo phenolate				linuron
29	pentachloro phenolate				metolachlor
30	phenol				propachlor
31	phenolate				benzotriazole
32	2-carboxylate phenol anion				benzotriazole anion
33	triclosan				5-chlorobenzotriazole
34	triclosan anion				5-chlorobenzotriazole anion
35	2,4,6-triiodophenol anion				5,6-dimethylbenzotriazole
36	*paracetamol*				5,6-dimethylbenzotriazole anion
37					5-methylbenzotriazole
38					5-methylbenzotriazole anion
39					quinoline
40					*amidotrizoic acid*

Refers the SI in the original reference[4] for experimental k_{O3} and their references. Outliers, which are excluded from the model calibration but included in the prediction performance evaluation (i.e., MUE, RMSE), are given in *italic* [a]See Table S8 for a list of the selected anilines and up-to-date k_{O3}-values used for the model development.

Table S5. The selected model compounds used for the development of a k_{O3} prediction model for olefins, amines, and heteroaromatic compounds

No.	Olefin	Amine	Heteroaromatic compound (hetero ar)
1	*cis*-1,2-dibromoethene	*N*-*α*-acetyllysine	azobenzene
2	1,1-dichloroethene	*N*-*ε*-acetyllysine	2-isopropyl-3-methoxypyrazine
3	*cis*-1,2-dichloroethene	alanine	pyridine
4	1,1-dichloropropene	*β*-alanine	pyridine protonated
5	ethene	arginine	atrazine
6	hexachlorocyclopentadiene	asparagine	deethylatrazine
7	propene	aspartate	deethyldeisopropylatrazine
8	trichloroethene	benzylamine	deisopropylatrazine
9	vinyl chloride	bromamine,	simazine
10	vinyl bromide	*n*-butylamine	
11	acrylic acid	*sec*-butylamine	
12	acrylic acid anion	*tert*-butylamine	
13	cinnamic acid	cyclohexanemethylamine	
14	cinnamic acid anion	cyclohexylamine	
15	fumaric acid	glutamine	
16	2-hexenoic acid anion	glycine	
17	maleic acid	hydroxylamine	
18	methacrylic acid	isoleucine	
19	methacrylic acid anion	leucine	
20	4-methoxy cinnamic acid	lysine	
21	4-methoxy cinnamic acid anion	monochloramine	
22	*trans*-*trans*-muconic acid dianion	phenylalanine	
23	4-nitro-cinnamic acid anion	serine	
24	sorbic acid	threonine	
25	sorbic acid anion	valine	
26	diethyl vinylphosphonate	acebutolol	
27	vinyl phosphonic acid	anatoxin	
28	vinyl phosphonic acid monoanion	atenolol	
29	vinyl phosphonic acid dianion	dibromamine,	
30	vinylsulfonate	dichloramine	
31	2-acetamidoacrylic acid	diclofenac	
32	acrylamide	diethylamine	
33	1,4-benzoquinone	dimethylamine	
34	but-1-en-2-ol	ethyl *N*-piperazinecarboxylate	
35	vinyl acetate	iminodiacetic acid	
36	anatoxin	methylchloramine	
37	cephalexin	metoprolol	
38	*cis*-3-hexen-1-ol	morpholine	
39	*β*-ionone	proline	
40	*trans*-*cis*-2,6-nonadienal	spectinomycin,	
41	oseltamivir acid	spectinomycin protonated	
42	1-penten-3-one	ciprofloxacin	
43	carbamazepine	1,4-diazabicyclo[2.2.2]octane	
44	tylosin protonated	DABCO monoprotonated	
45	MCLR	dimethylchloramine	
46	*trans-1,2-dichloroethene*	*N*,*N*-dimethylcyclohexylamine	
47	*3,4-dihydroxycinnamic acid*	dimethylethanolamine	
48	*3,4-dihydroxycinnamic acid anion*	enrofloxacin	
49	*maleic acid anion*	ethylenediaminetetraacetic acid	
50	*3-methoxy-4-hydroxy cinnamic acid*	EDTA monoptoronated	
51	*3-methoxy-4-hydroxy cinnamic acid anion*	lincomycin	
52	*acrylonitrile*	1-methylpyrrolidine	
53	*vinylene carbonate*	nitrilotriacetic acid	
54		tramadol	
55		triethylamine	
56		trimethylamine	
57		azithromycin	
58		roxithromycin	
59		tylosine	
60		*ammonia*	

Refers the SI in the original reference[4] for experimental k_{O3} and their references. Outliers, which are excluded from the model calibration but included in the prediction performance evaluation (i.e., MUE, RMSE), are given in *italic*

Table S6. The selected model compounds used for the development of a k_{O3} prediction model for adenine, guanine, cytosine, thymine, uracil, and miscellaneous olefins

No.	Adenine	Guanine	Cytosine	Thymine	Uracil	Miscellaneous olefins
1	adenine (9H)	aciclovir (1H)	cytidine	thymidine	5-chlorouracil	*β*-cyclocitral
2	adenine protonated (1,9H)	aciclovir anion	cytidine protonated	thymidine anion	5-chlorouracil anion (1H)	levonorgestrel
3	adenine anion	aciclovir protonated (1,7H)	cytosine (1H)	5'-thymidylic acid	1,3-dimethyluracil	medroxyprogesterone
4	adenosine	2'-deoxyguanosine (1H)	cytosine anion	thymine	6-methyluracil (1,3H)	norethindrone
5	adenosine protonated (1H)	5'-deoxyguanylic acid (1H)	cytosine protonated	thymine anion (1H)	6-methyluracil anion (1H)	progesterone
6	5'-adenylic acid	guanosine (1H)	2'-deoxycytidine		orotic acid anion	indigotrisulfonic acid
7	2'-deoxyadenosine	guanosine anion	2'-deoxycytidine protonated		uracil	imidazole
8	2'-deoxyadenosine protonated (1H)	guanosine protonated (1,7H)	5'-deoxycytidylic acid		uracil anion (1H)	imidazole protonated
9						4-methylimidazole
10						3,5-dimethylisoxazole
11						2,4-diamino-5-methylpyrimidine
12						2,4-diamino-5-methypyrimidine monoprotonated
13						trimethoprim deprotonated
14						tryptophan
15						3-(dimethylaminomethyl) indole protonated
16						N-(4)-acetylsulfamethoxazole deprotonated
17						N-(4)-acetylsulfamethoxazole protonated

Refers the SI in the original reference[4] for experimental k_{O3} and their references.

Table S7. A brief summary for the performance evaluation of the k_{O3} prediction model developed with seven different computational methods.

No.	Computational method	3D geometry generation	Geometry Optimization	Single point calculation	Remark[a]
1	HF/6-31G//PM3//MMFF	MMFF	PM3	HF/6-31G	- Most expensive computational costs - Overall stable prediction performance
2	HF/SV//PM3//MMFF	MMFF	PM3	HF/SV	- 2nd most expensive computational costs - Overall stable prediction performance
3	HF/SV//PM3//Dreiding	Dreiding	PM3	HF/SV	- 2nd most expensive computational costs - Poor performance for mono- and di-alkoxybenzene, trialkoxybenzene, and guanine attributable to the Dreiding force field method
4	HF/3-21G//PM3//MMFF	MMFF	PM3	HF/3-21G	- Overall stable prediction performance
5	HF/3-21G//PM3//Dreiding	Dreiding	PM3	HF/3-21G	- Overall stable prediction performance - Poor performance for mono- and di-alkoxybenzene, trialkoxybenzene, and guanine attributable to the Dreiding force field-derived initial 3D geometry
6	HF/3-21G//MMFF[b]	MMFF	*n. i.*	HF/3-21G	- The least expensive computational costs - Overall stable prediction performance - **Selected method for this study**
7	HF/3-21G//Dreiding	Dreiding	*n. i.*	HF/3-21G	- The least expensive computational costs - Poor performance for mono- and di-alkoxybenzene and trialkoxybenzene attributable to the Dreiding force field-derived initial 3D geometry

n.i. not implemented. [a]remarks were based on the evaluation outputs presented in Figs. S2 and S3 and detailed descriptions are given in Text S4. [b]default method chosen in this study for a k_{O3} prediction.

Table S8. Second-order rate constants (k_{O3}) of the selected anilines used for development of the previous model[4] and the current model.

Compound	Previous k_{O3}[a], $M^{-1}s^{-1}$	Current k_{O3}[b], $M^{-1}s^{-1}$
Aniline	1.40×10^{7}([39])	1.30×10^{6}([38])
3-chloroaniline	7.84×10^{6}([39])	7.28×10^{5}[c]
4-chloroaniline	1.04×10^{7}([39])	1.40×10^{6}([38])
2-fluoroaniline	7.28×10^{6}([39])	6.76×10^{5}[c]
3-fluoroaniline	8.12×10^{6}([39])	7.54×10^{5}[c]
4-fluoroaniline	1.23×10^{7}([39])	1.14×10^{6}[c]
3-iodoaniline	7.42×10^{6}([39])	6.89×10^{5}[c]
4-iodoaniline	9.24×10^{6}([39])	8.58×10^{5}[c]
2-methoxyaniline	1.72×10^{7}([39])	1.60×10^{6}[c]
3-methylaniline	1.88×10^{7}([39])	1.75×10^{6}[c]
4-methylsulfone aniline	4.70×10^{4}([3])	4.70×10^{4}([3])
3-nitroaniline	2.38×10^{6}([39])	2.21×10^{5}[c]
4-nitroaniline	1.40×10^{5}([39])	1.20×10^{5}([38])
sulfamethoxazole	4.70×10^{4}([3])	4.70×10^{7}([3])
sulfamethoxazole anion	5.70×10^{5}([3])	5.70×10^{5}([3])
N,*N*-dimethylaniline	2.00×10^{9}([40])	1.60×10^{6}([38])
4-aminobenzoate	*n.a.*	8.50×10^{5}([38])
4-methylaniline	*n.a.*	2.40×10^{6}([38])
4-methoxyaniline	*n.a.*	1.10×10^{6}([38])
2-hydroxyaniline	*n.a.*	3.70×10^{5}([38])
4-hydroxyaniline	*n.a.*	7.80×10^{5}([38])
p-phenylenediamine	*n.a.*	1.30×10^{6}([38])

n.a. not available when the model was established. [a]k_{O3} for anilines used for the development of the previous model[4], [b]k_{O3} for anilines used for the development of the current model, [c]corrected k_{O3} for anilines based on the relative difference of k_{O3} for unsubstituted aniline between the previous (i.e., 1.40×10^{7} $M^{-1}s^{-1}$)[39] and the recent (i.e., 1.30×10^{6} $M^{-1}s^{-1}$)[38] measurement.

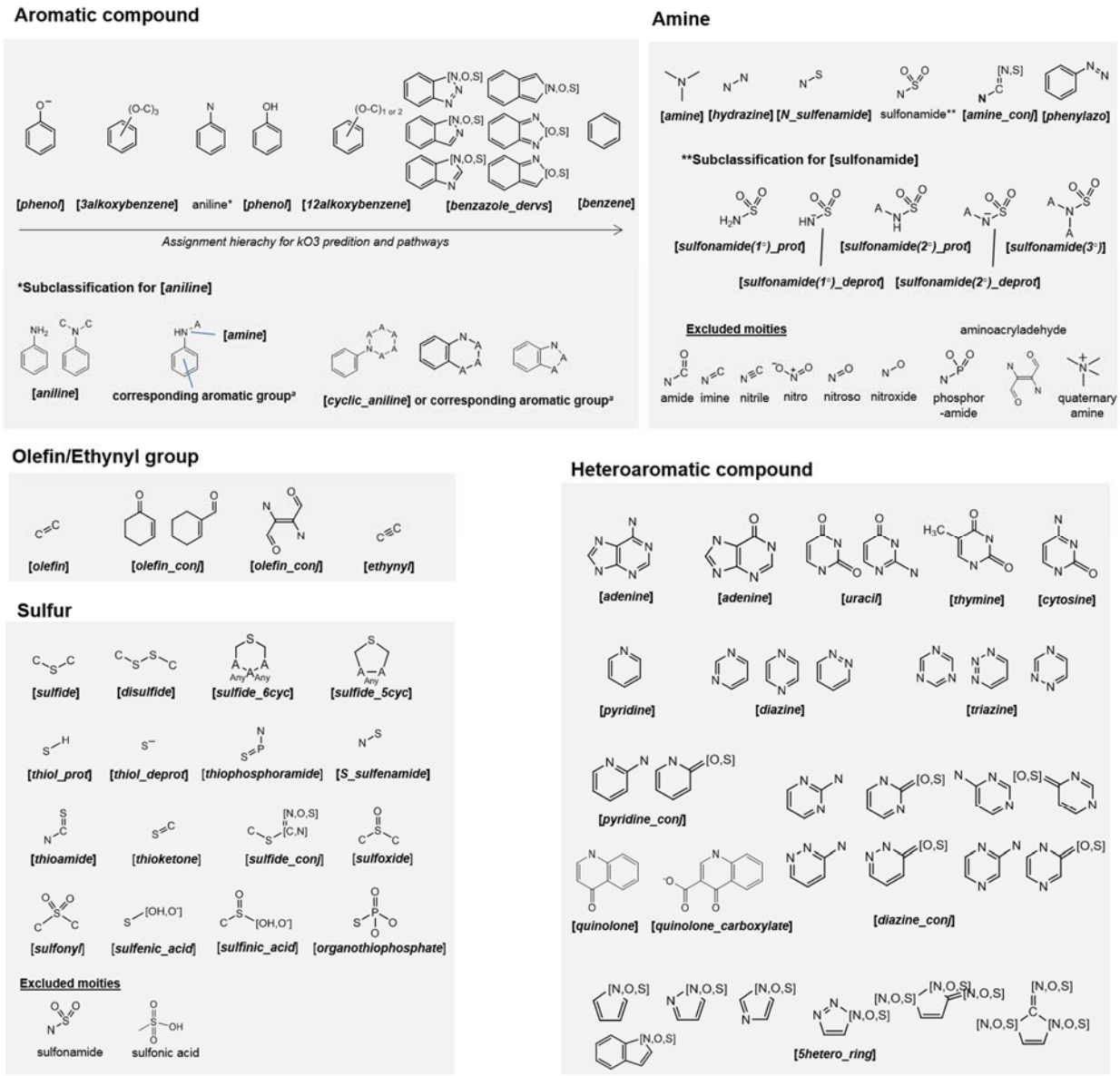


Fig. S1. Chemical structures for defined reactive sites. Names in **bold** and *italic* within brackets are the reactive sites in the first column in Table S1. 'A' stands for any aliphatic atom except for hydrogen. A bond denoted as 'any' shown in *sulfide_6cyc* and *sulfide_5cyc* indicates any bond between single, double, aromatic, and triple bonds. [a]A corresponding aromatic compound group to the substituent is to be assigned to the benzene ring (see Text S1 for more details).

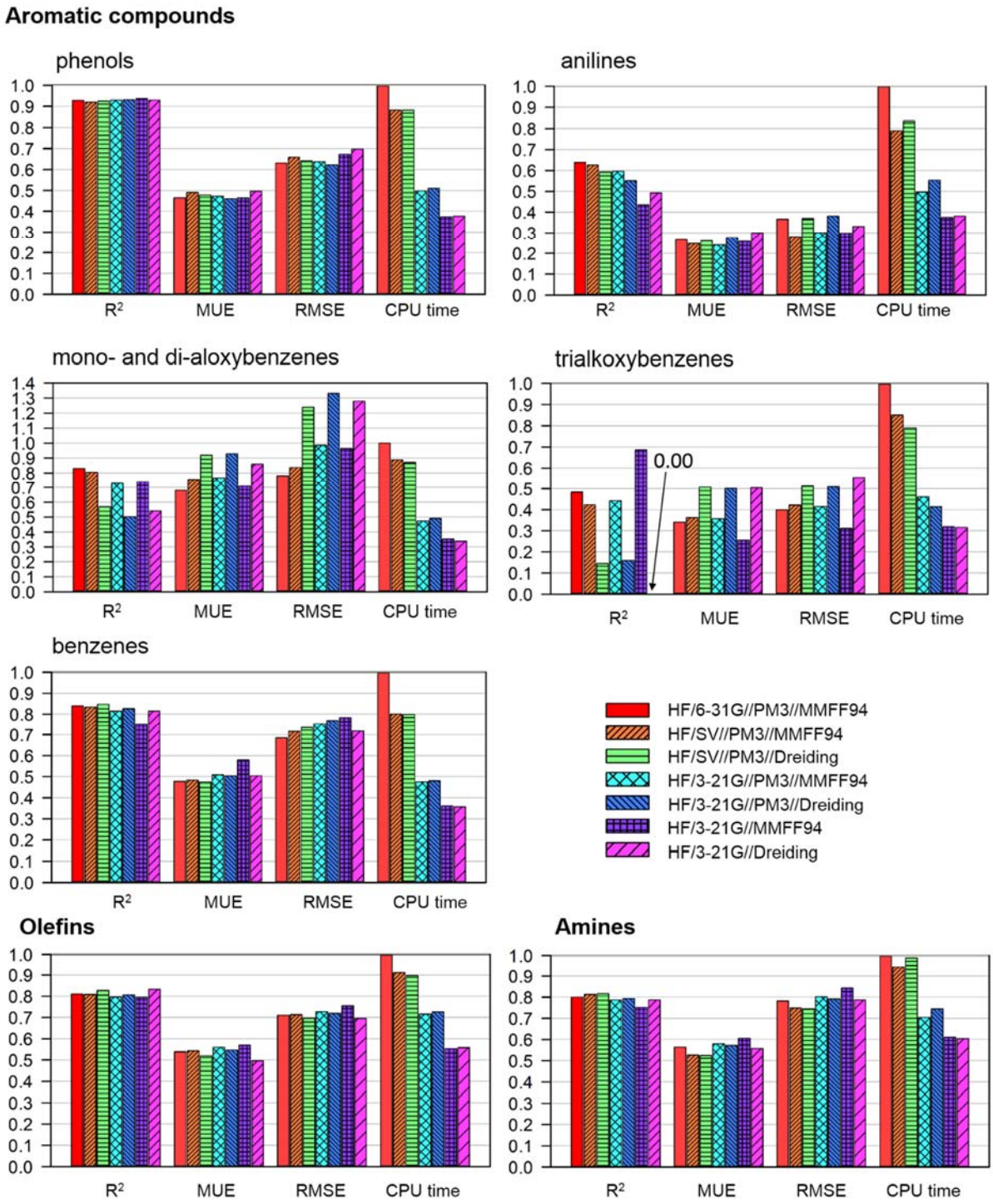


Fig. S2. Comparison of the performances of the k_{O3} prediction models for aromatic compound groups (phenols, anilines, mono- and di-alkoxybenzenes, trialkoxybenzenes, and benzenes), olefins, and amines developed with seven different computational methods. The relative CPU times were measured on the same hardware (Intel® Core™2 Duo CPU at 3.00 GHz with 4 GB RAM) operated on Ubuntu Linux 14.04.4 LTS. The numerical summary of R^2, MUE, and RMSE for HF/3-21G//MMFF94 selected in this study for all prediction models are given in Table S3. The absolute CPU times are given in Fig. S6 for the single point calculations of the selected method, i.e., HF/3-21G.

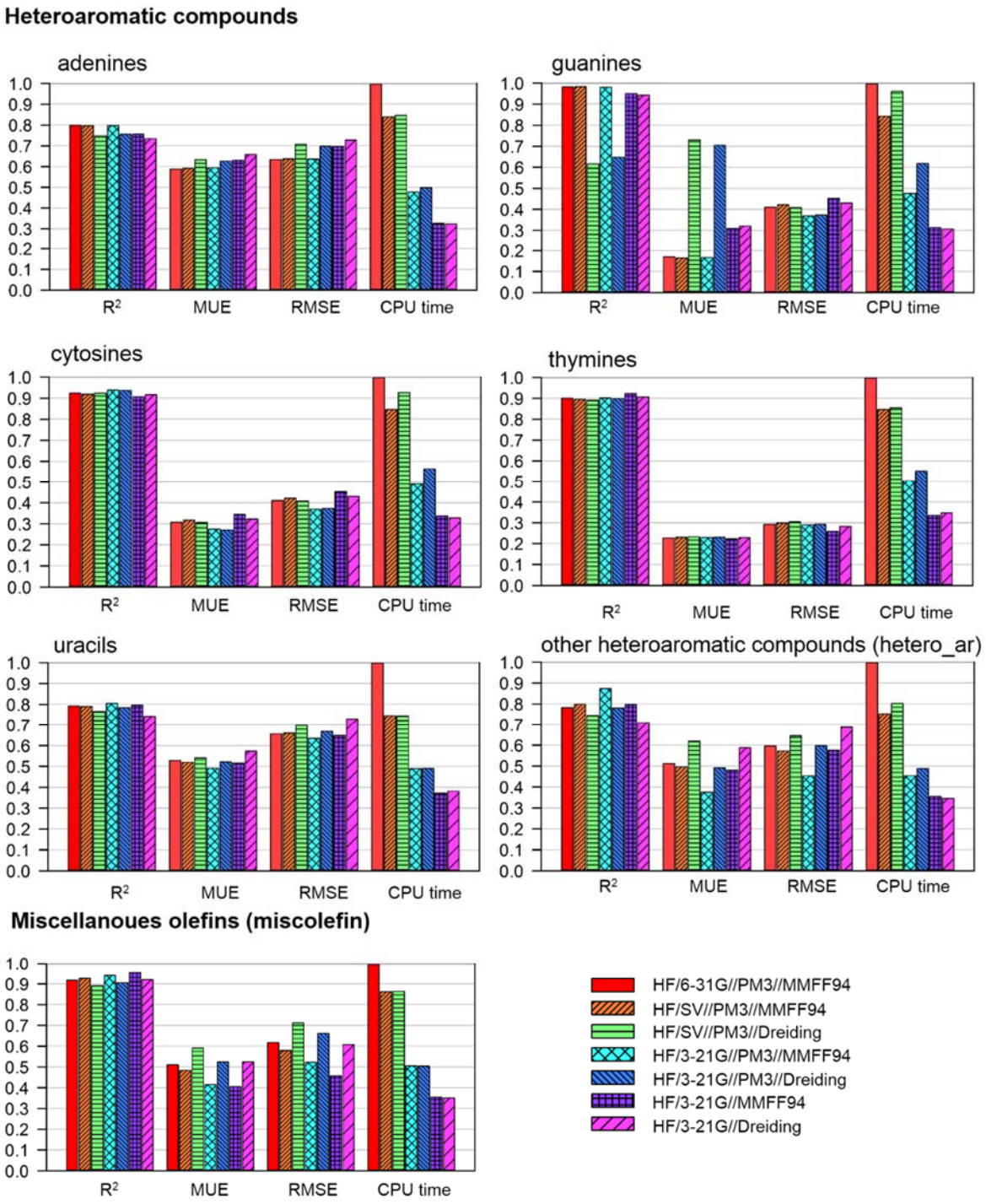


Fig. S3. Comparison of the performances of the k_{O3} prediction models for adenines, guanines, cytosines, thymines, uracils, heteroaromatic compounds (hetero_ar), and miscellaneous olefins (miscolefin) developed with seven different computational methods. The relative CPU times were measured on the same hardware (Intel® Core™2 Duo CPU at 3.00 GHz with 4 GB RAM) operated on Ubuntu Linux 14.04.4 LTS. The numerical summary of R^2, MUE, and RMSE for HF/3-21G//MMFF94 selected in this study for all prediction models are given in Table S3. The absolute CPU times are given in Fig. S6 for the single point calculations of the selected method, i.e., HF/3-21G.

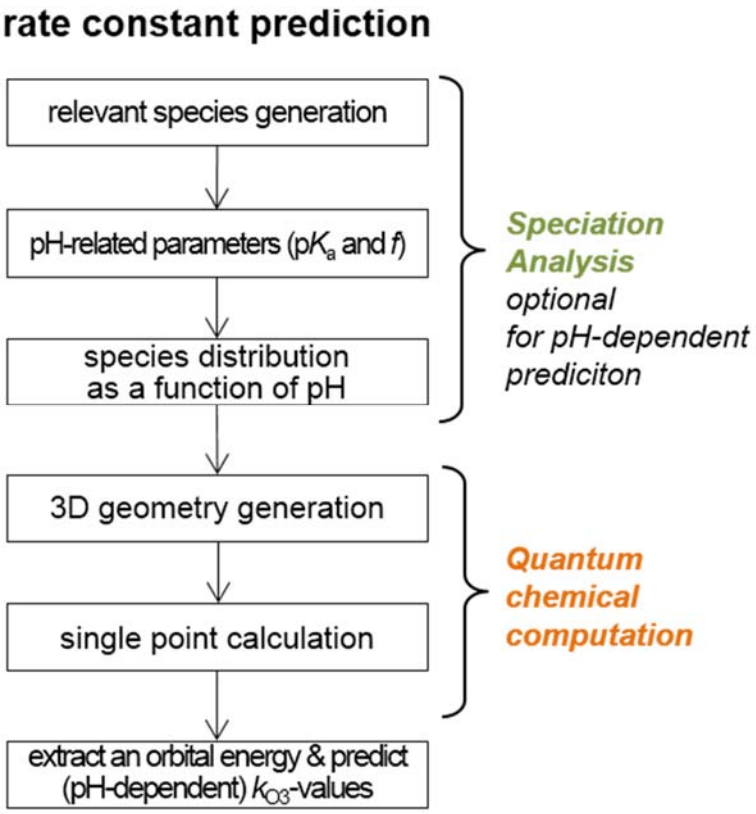


Fig. S4. Workflow of the pH-dependent rate constant prediction

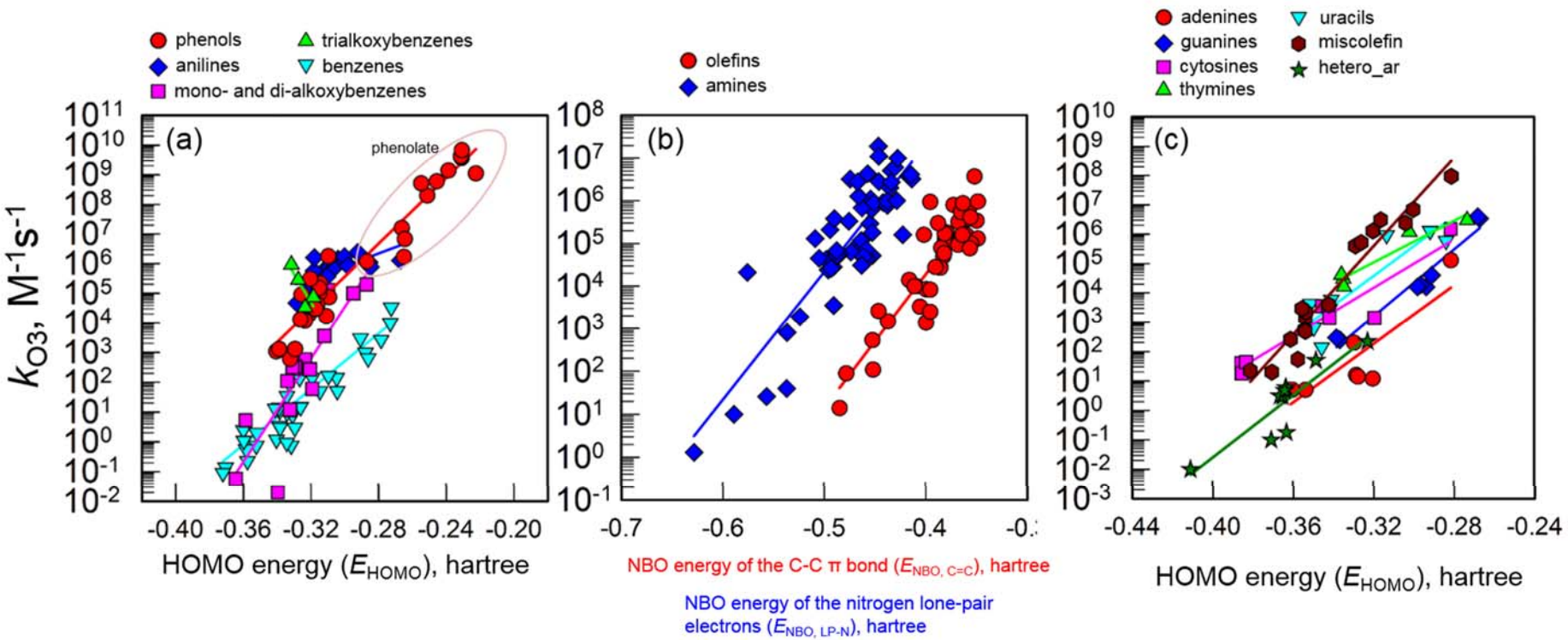


Fig. S5. Correlations between experimental k_{O3} for the reactions of ozone with (a) aromatic compound groups, (b) olefins, (b) amines, and (c) heteroaromatic compound groups with the corresponding orbital energies. All the orbital energies were obtained with the HF/3-21G//MMFF method. The statistical parameters for the prediction performance as well as slopes and y-intercept for the individual regression line equations are presented in Table S3.

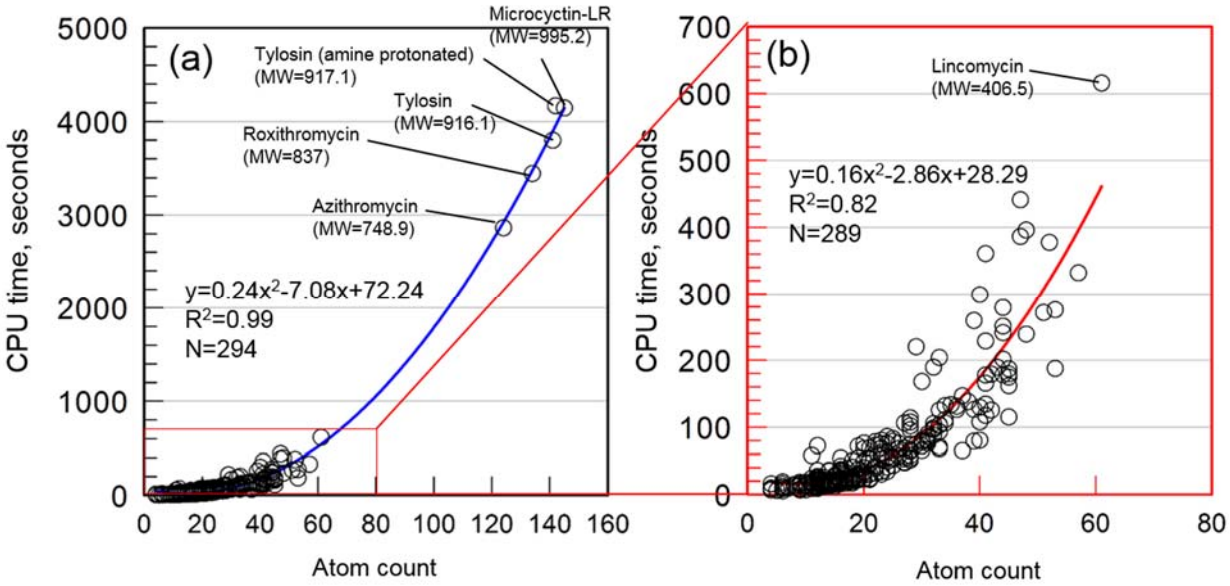


Fig. S6. The CPU times in seconds for (a) 294 training compounds (10 outliers included, N_{total} in Table S3) and (b) 289 training compounds excluding 5 macromolecules from (a) as a function of atom count of the training compound. The CPU time was measured for single point calculations with the HF/3-21G. The CPU times were measured on the same hardware (Intel® Core™2 Duo CPU at 3.00 GHz with 4 GB RAM) operated on Ubuntu Linux 14.04.4 LTS.

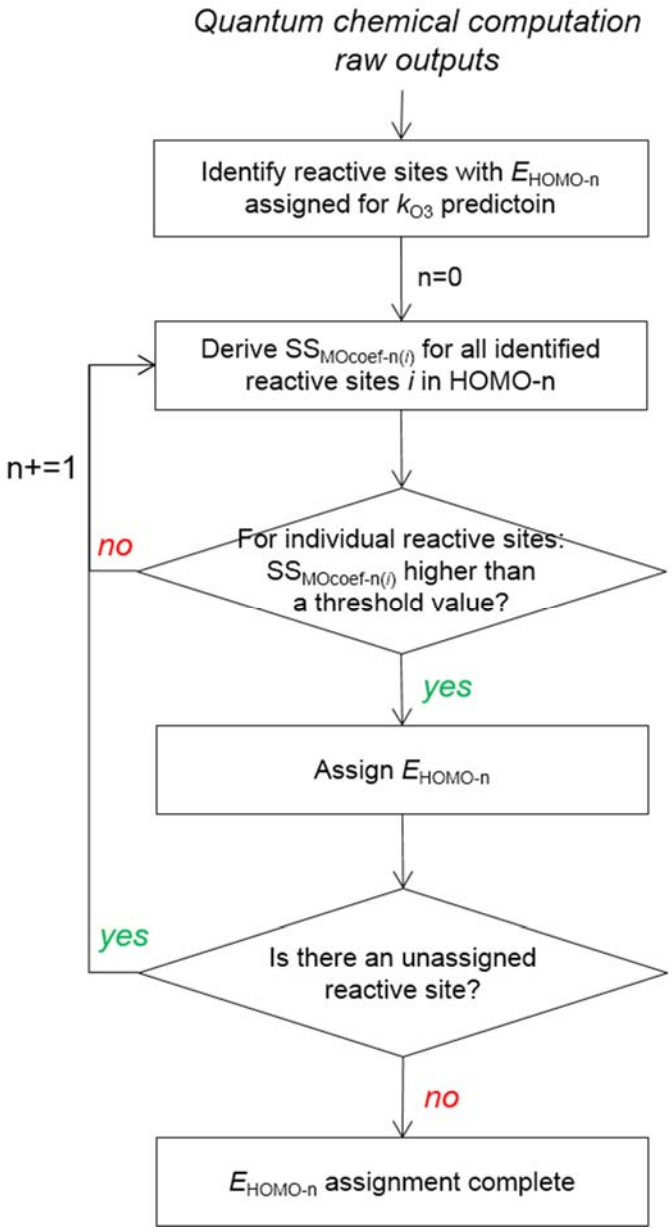


Fig. S7. Automated workflow in the k_{O3} prediction mode for the assignment of E_{HOMO-n} to the corresponding reactive sites. $SS_{MOcoef-n(i)}$ is the sum of squares of molecular orbital coefficients of atoms of a reactive site (*i*) responsible for the reaction with ozone in HOMO-n. Note that a user can manipulate the SS_{MOcoef} values based on user's judgment for which refer to Text S7 for more details.

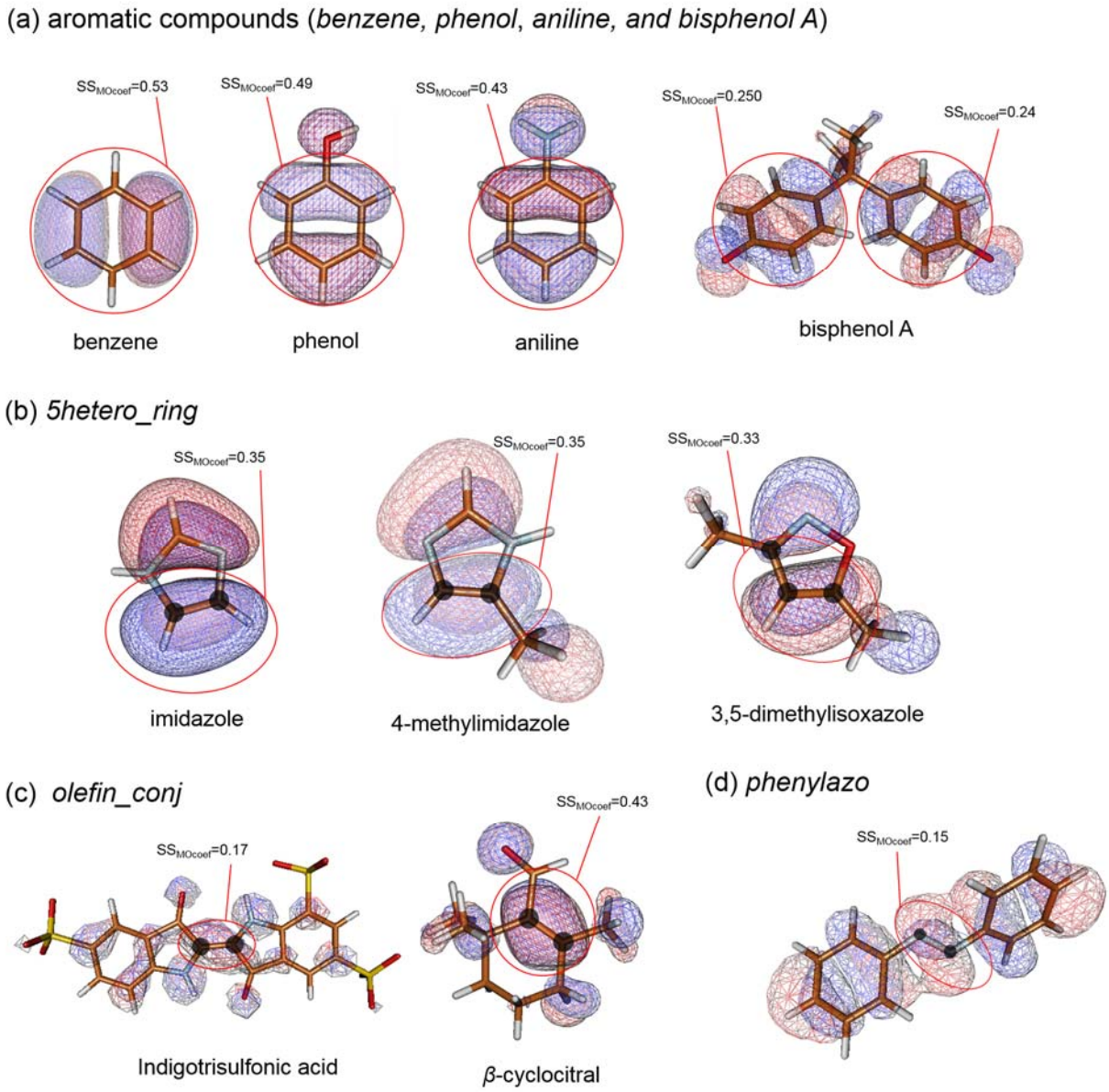


Fig. S8. HOMO and the estimated SS_{MOcoef} for the selected model compounds containing reactive sites such as (a) aromatic compounds (b) *5hetero_ring*, (c) *olefin_conj*, and (d) *phenylazo*. Computational method: HF/3-21G//MMFF94. The carbons with black circles are the atoms used to estimate SS_{MOcoef} for *5hetero_ring*, *olefin_conj*, and *phenylazo*.

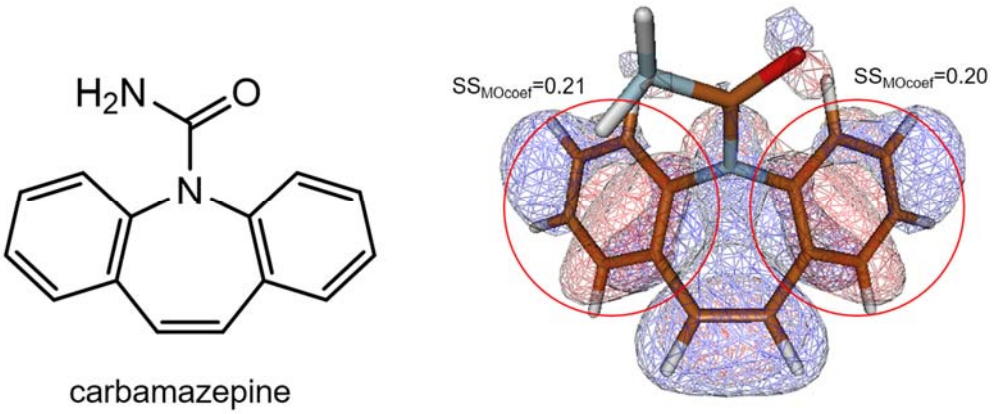


Fig. S9. The estimated SS_{MOcoef} for the two benzene rings of carbamazepine for HOMO-n obtained by the HF/3-21G//MMFF

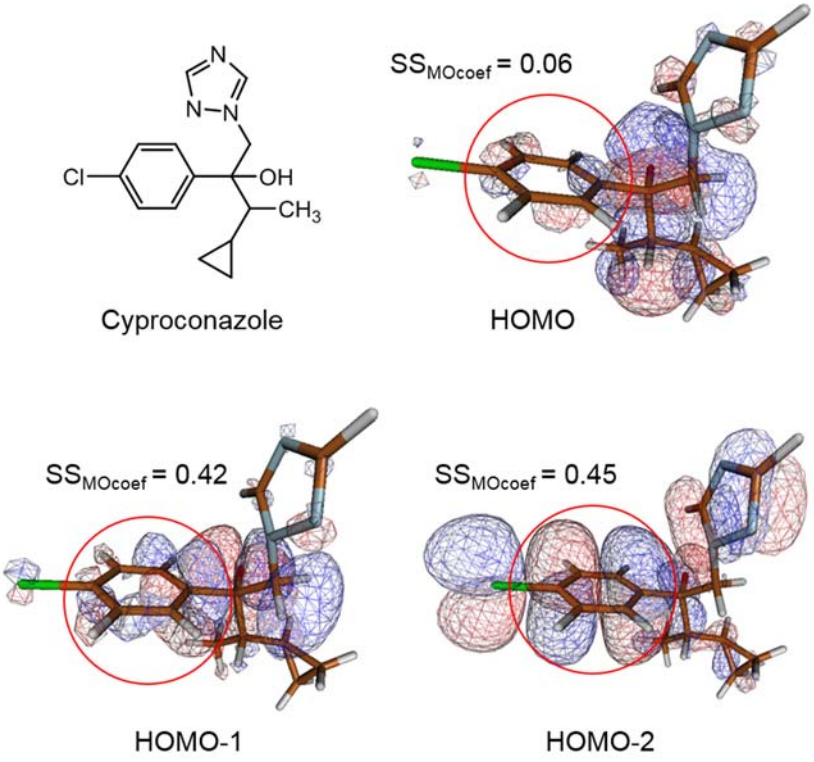


Fig. S10. HOMO, HOMO-1, and HOMO-2 of cyproconazole obtained by the HF/3-21G//MMFF and the estimated SS_{MOcoef} for the chlorobenzene ring for respective HOMO-n (n=0, 1, and 2). Based on the minimum SS_{MOcoef} threshold value for the benzene (i.e., 0.22) from Text S7, the E_{HOMO-1} is supposed to be assigned for its k_{O3} prediction because the estimated SS_{MOcoef} (0.42) is higher than 0.22. However, the visual examination manifests that the HOMO-1 has no conjugation system on the chlorobenzene ring. In contrast, the HOMO-2 has both SS_{MOcoef} of 0.45 and has the conjugation system for the chlorobenzene ring. Therefore, the E_{HOMO-2} should be selected for the k_{O3} prediction.

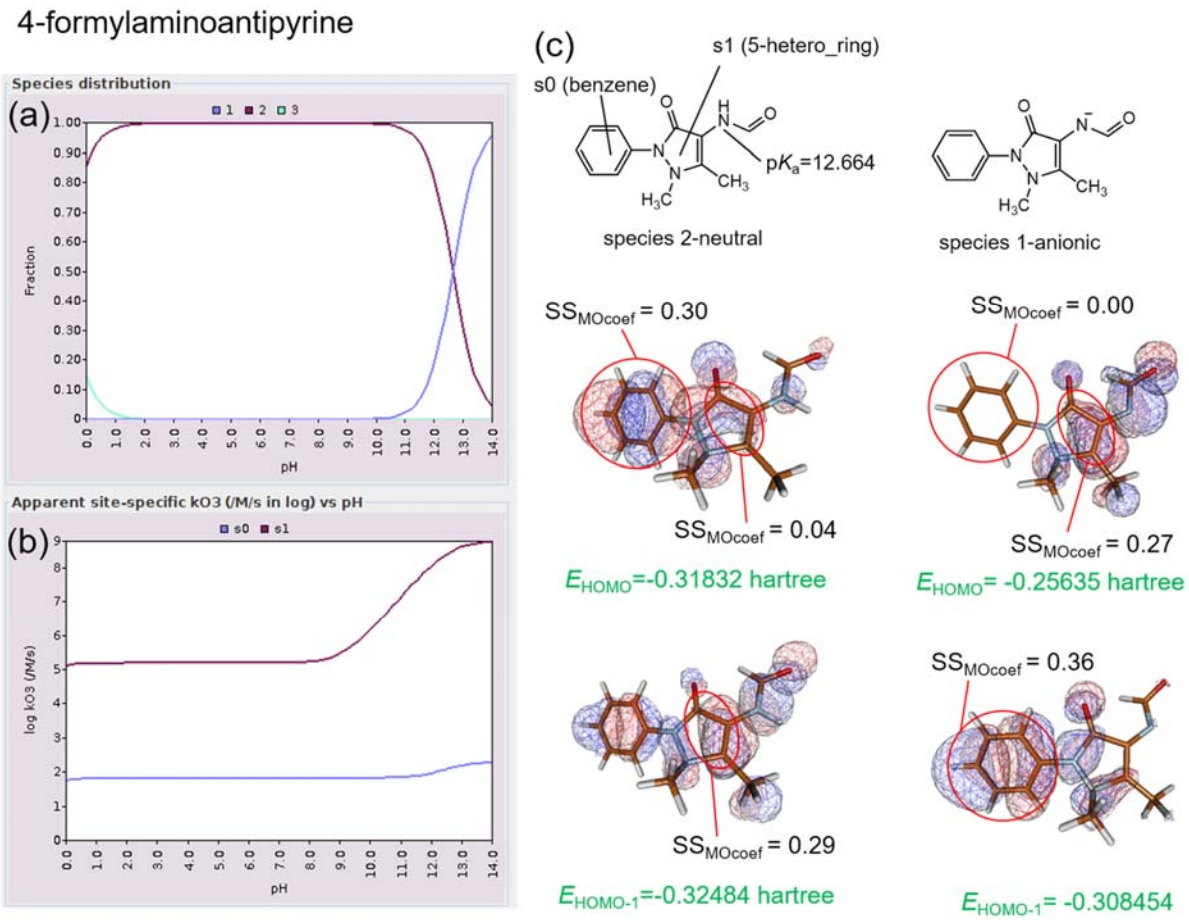


Fig. S11. Prediction of (a) species distribution and (b) apparent site-specific k_{O3} for 4-formylaminoantipyrine and (c) the E_{HOMO-n} assignment to the corresponding reactive sites. The predicted k_{O3} for the 5-membered heteroaromatic ring (*5hetero_ring*) of 4-formylaminoantipyrine increases by about four orders of magnitude with increasing pH from 9 to 14 (Fig S11b) because deprotonation of the amide nitrogen (pK_a=12.664 predicted by Marvin) leads to the increase in the assigned E_{HOMO}: E_{HOMO-1} (-0.32484 hartree) with the SS_{MOcoef} of 0.29 was assigned to the 5hetero_ring of the neutral 4-formylaminoantipyrine, whereas E_{HOMO} (-0.25635 hartree) with the SS_{MOcoef} of 0.27 was assigned to the anionic form (Fig. S11c).

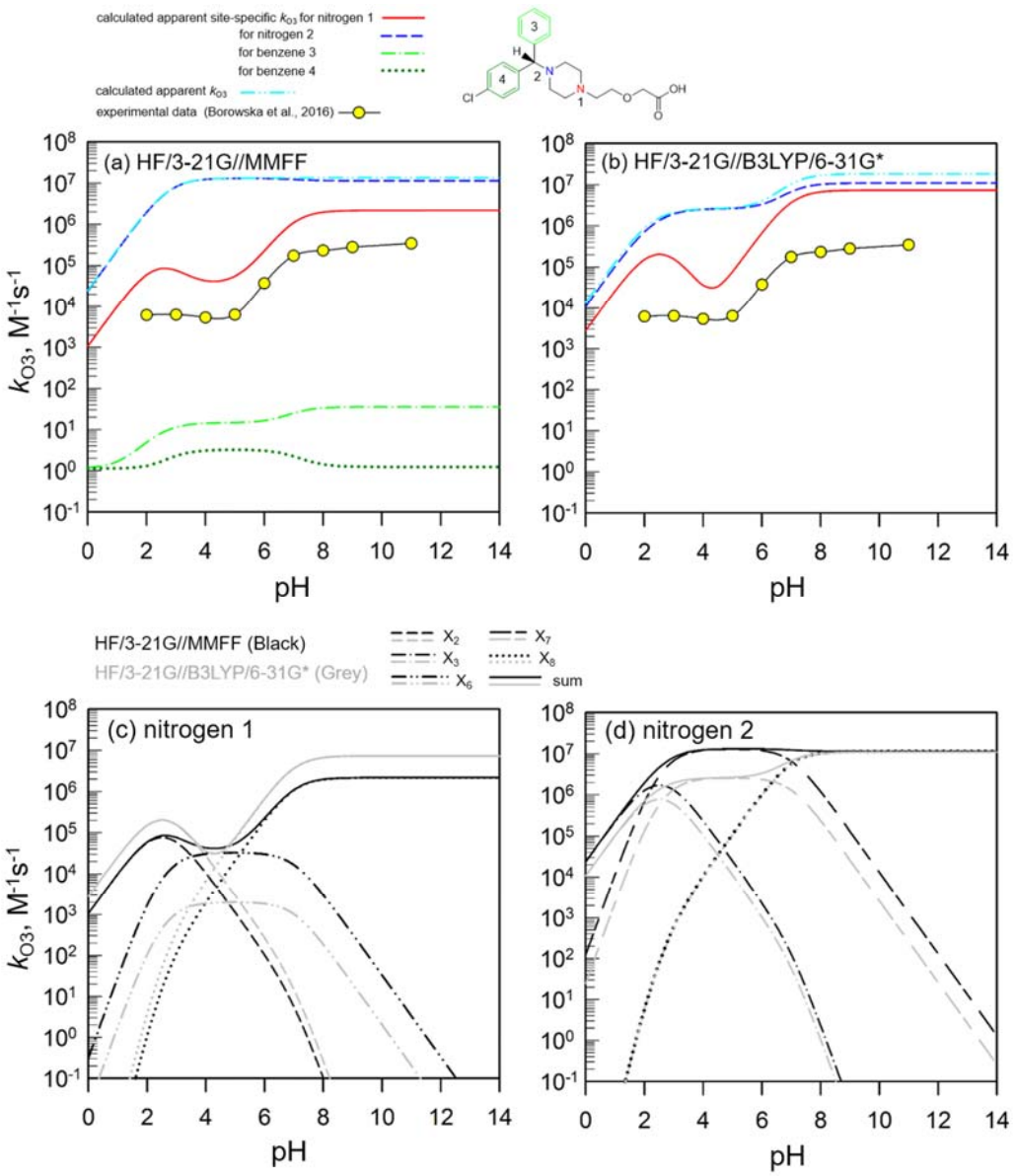


Fig. S12. Apparent second order rate constants for the reaction of ozone with cetirizine as a function of the pH. (a) Comparison between predicted and experimental k_{O3}-values, the prediction data were produced by the HF/3-21G//MMFF method. (b) Comparison between predicted and experimental k_{O3}-values, the prediction data were produced by the HF/3-21G//B3LYP/6-31G*. (c) Comparison of the predicted k_{O3} for nitrogen 1 of cetirizine for HF/3-21G//MMFF (black) and for HF/3-21G//B3LYP/6-31G* (grey). (d) Comparison of the predicted k_{O3} for the nitrogen 2 of cetirizine for HF/3-21G//MMFF (black) and for HF/3-21G//B3LYP/6-31G* (grey). The computations for HF method and B3LYP method were conducted with ORCA and Gaussian, respectively. All the predicted apparent site-specific k_{O3} were calculated based on Eq. S2 using the species distribution in Fig. S13. The species X_i corresponds to the species shown in Fig. S13(b). All the acid-base species of cetirizine used for the k_{O3} prediction were dextrocetirizine (*S* enantiomer) (see Text S10 for details). Note that for the data presented in Fig. S12a, the SS_{MOcoef} of 0.18 instead of the suggested value 0.22 (Text S7) was set as the threshold value for the benzene ring judged by the authors.

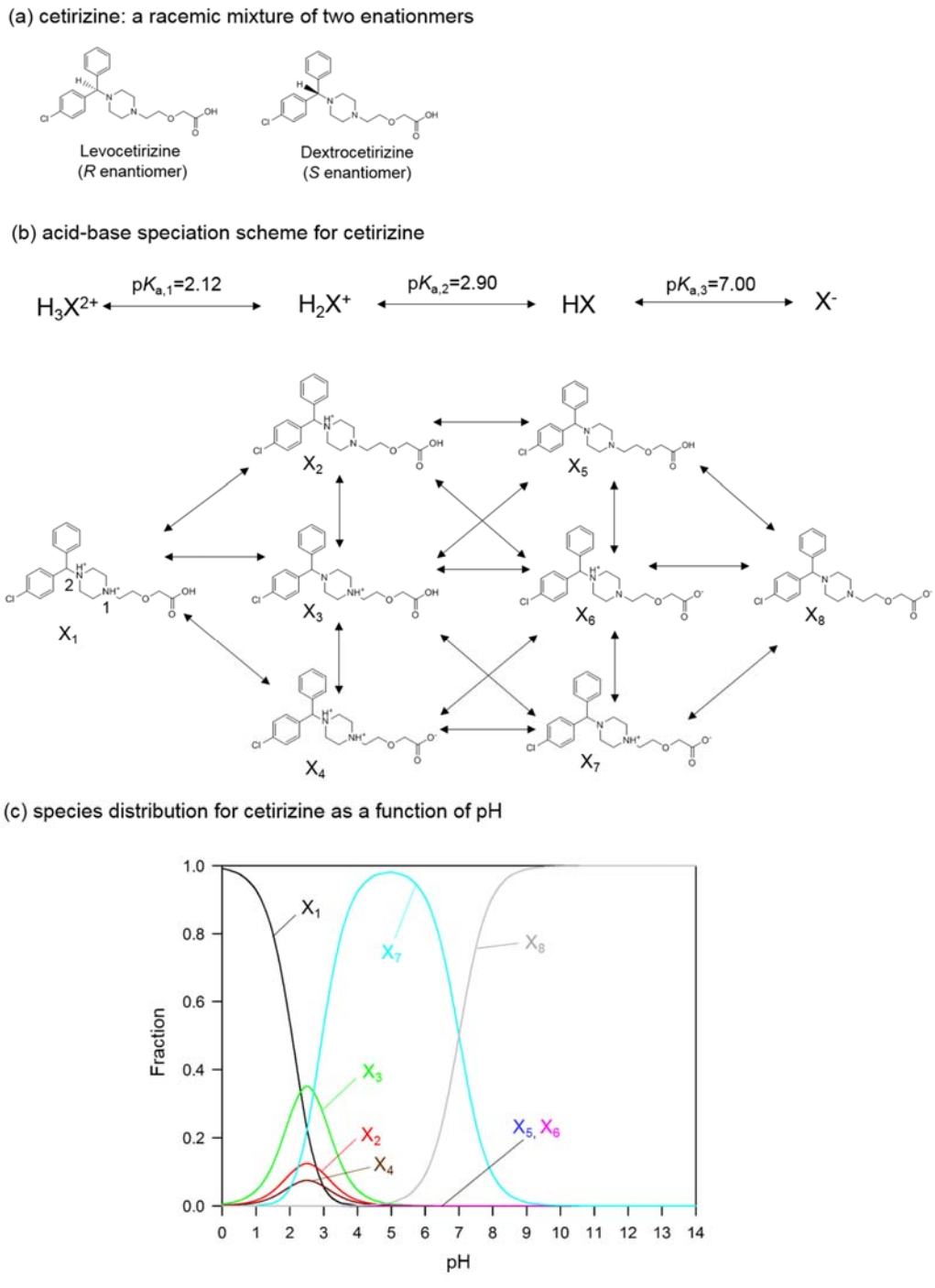


Fig. S13. Relevant information about cetirizine for k_{O3} prediction. (a) The chemical structures of a racemic mixture of cetirizine used for the determination of experimental k_{O3}.[2] (b) Acid-base speciation scheme of cetirizine. p$K_{a,1}$ and p$K_{a,2}$ values were from ref[99] and p$K_{a,1}$ were from ref.[2] (c) Species distribution for cetirizine as a function of pH between 0 and 14. Tautomeric fractions between X_2, X_3, and X_4 and X5, X6, and X7 were 0.226, 0.638, and 0.136 and 0.000, 0.001, and 0.999, respectively, from the literature.[99]

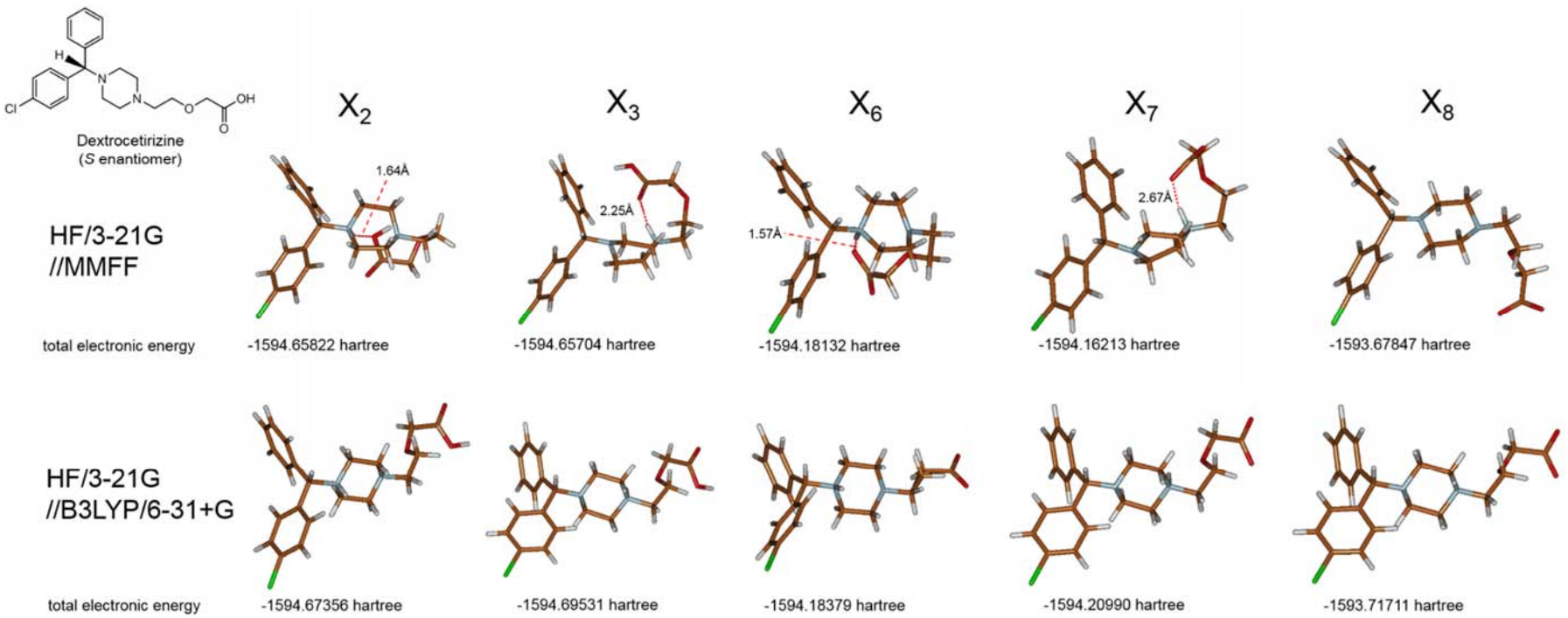


Fig. S14. Comparison of geometries and total electronic energies of acid-base species of dextrocetirizine used for the k_{O3} prediction presented in Figs. S13a and S13b for the HF/3-21G//MMFF (top row) and HF/3-21G//B3LYP/6-31G* (bottom row), respectively. The computations for HF method and B3LYP method were conducted with ORCA and Gaussian, respectively.

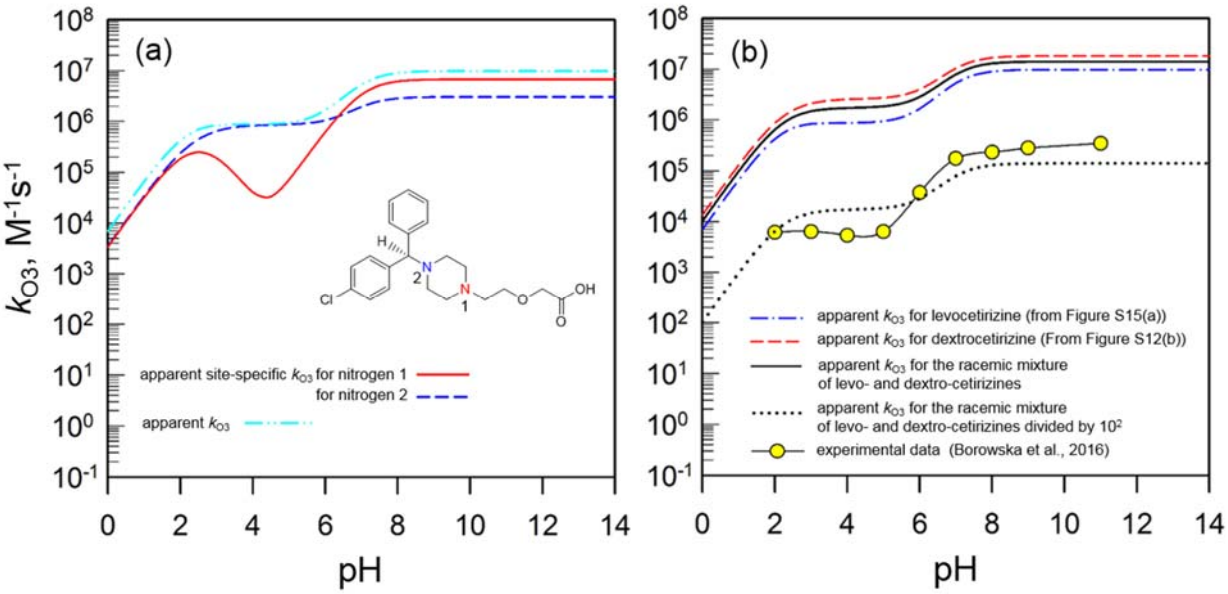
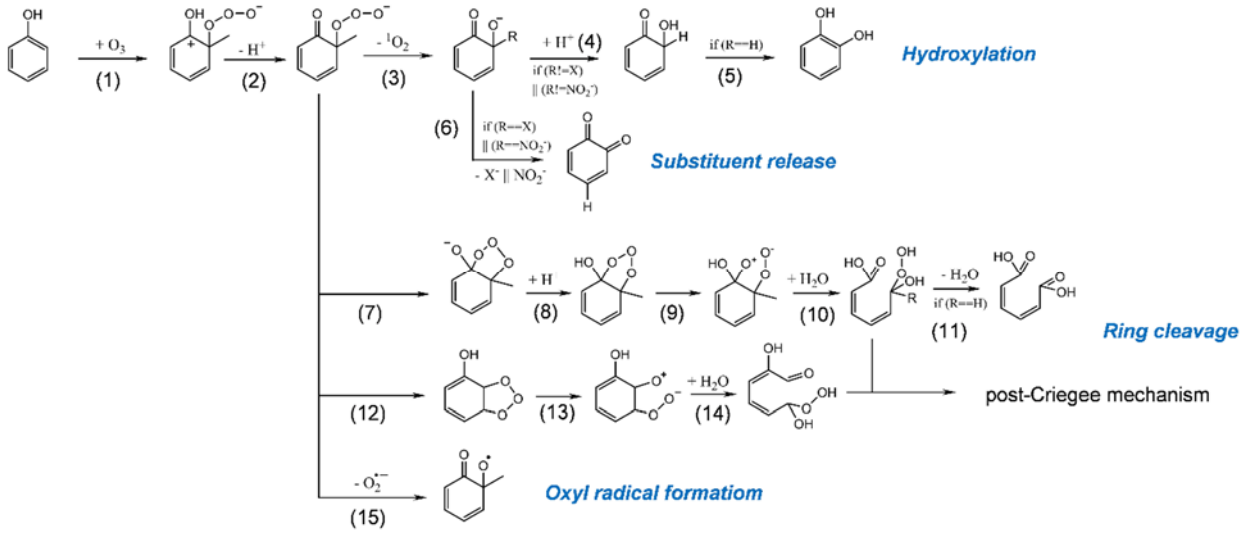
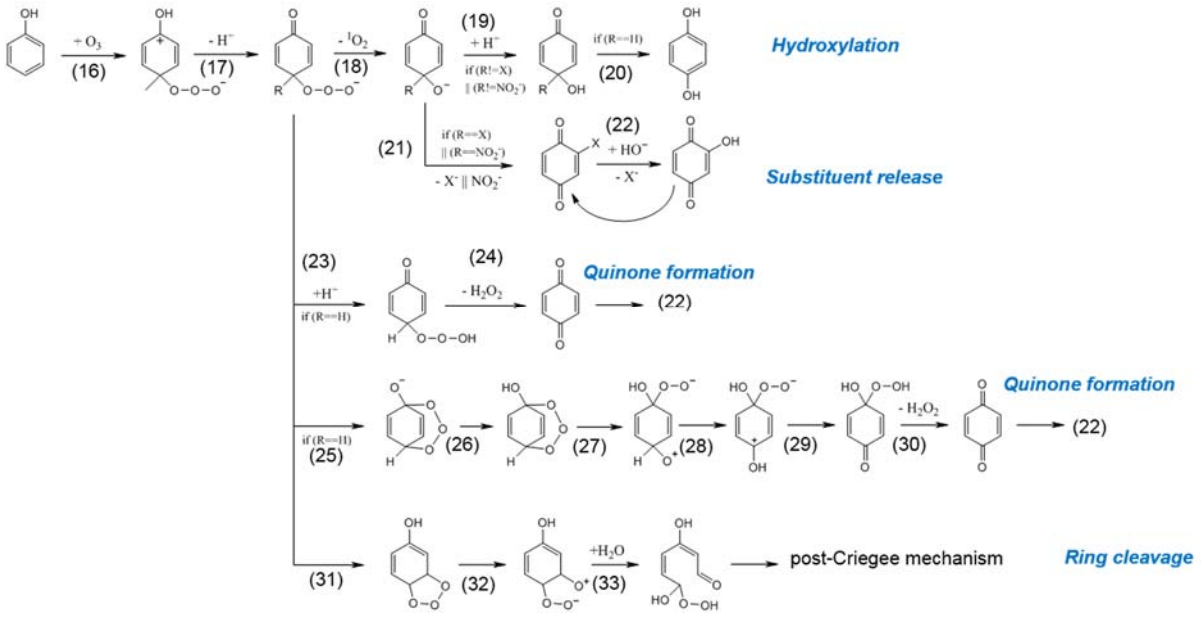


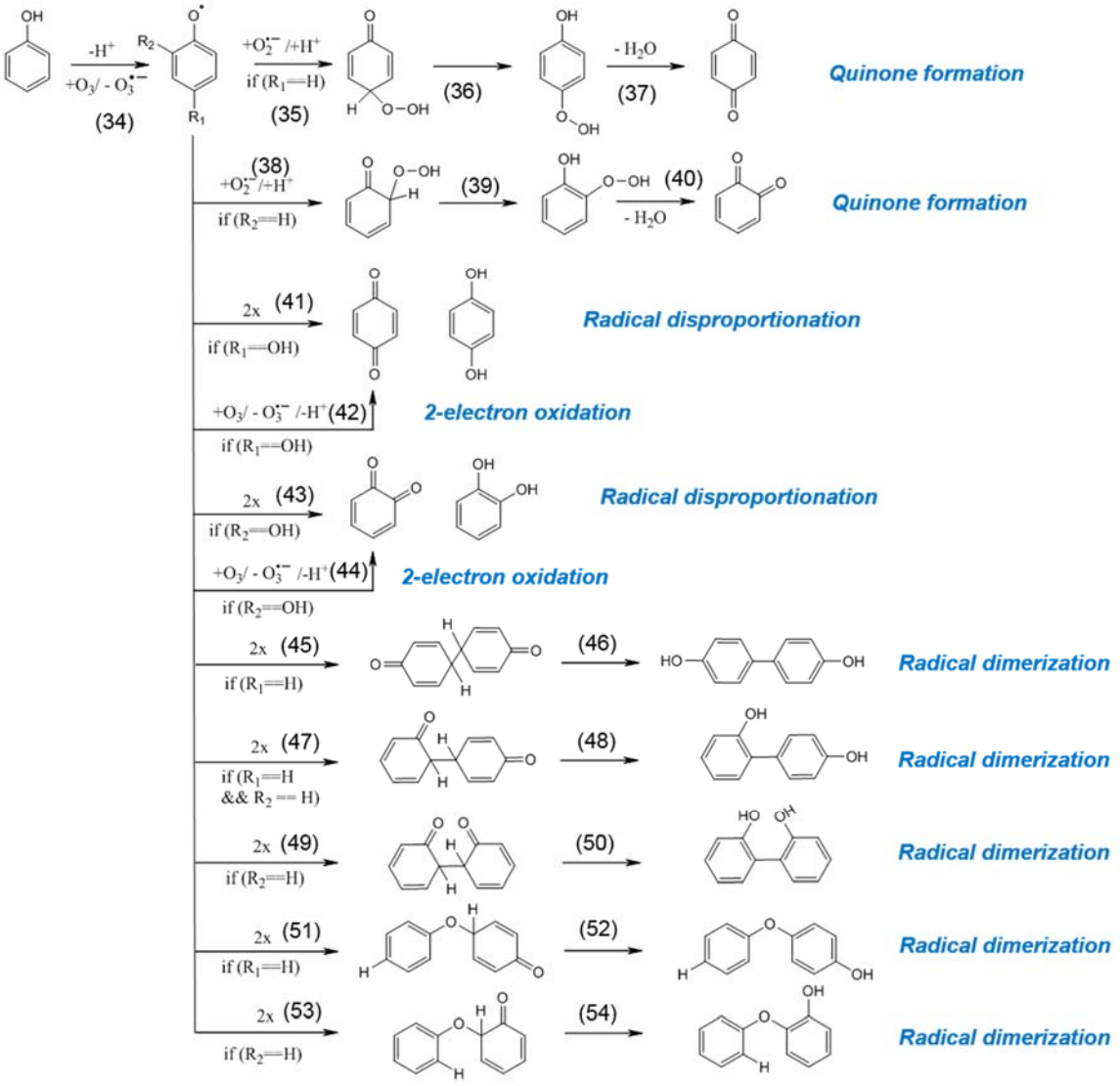
Fig. S15. Apparent second-order rate constants (k_{O3}) for the reaction of ozone with cetirizine as a function of pH. (a) predicted k_{O3} obtained by the HF/3-21G//B3LYP/6-31G* method for levocetirizine (*R* enantiomer) and (b) comparison between the experimental k_{O3}-values and the predicted k_{O3} obtained by the HF/3-21G//B3LYP/6-31G* method for the racemic mixture of levocetirizine and dextrocetirizine. The computations for HF method and B3LYP method were conducted with ORCA and Gaussian, respectively. All the predicted apparent site-specific k_{O3} were calculated based on Eq. S2 using the species distribution in Fig. S13.



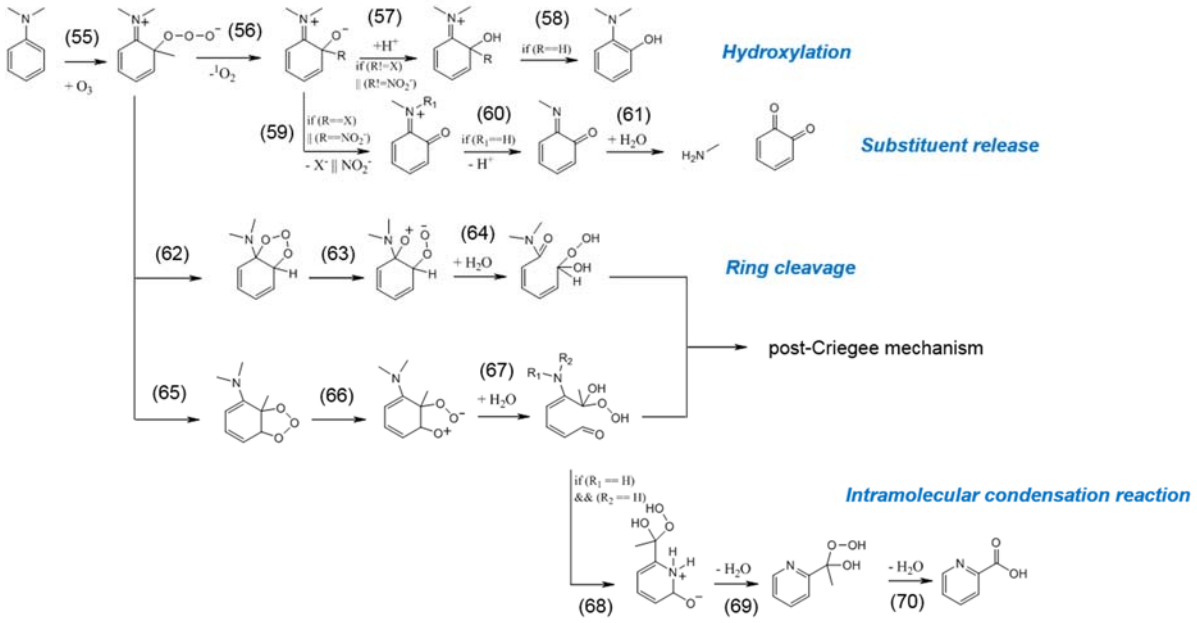
Scheme S1. Reaction pathways for the reactions of ozone with phenols at the ortho position adapted from various references (see Text S9) (logical operators such as '==' (is equal to), '!=' (is not equal to), and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



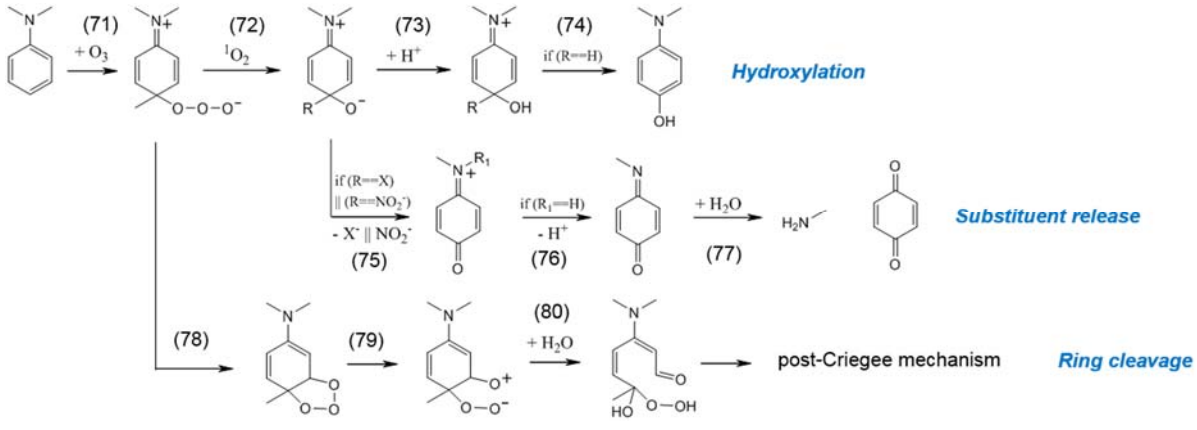
Scheme S2. Reaction pathways for the reactions of ozone with phenols at the para position adapted from various references (see Text S9) (logical operators such as '==' (is equal to), '!=' (is not equal to), and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)) . Implicit atoms can be hydrogen, any atom, or any organic moiety.



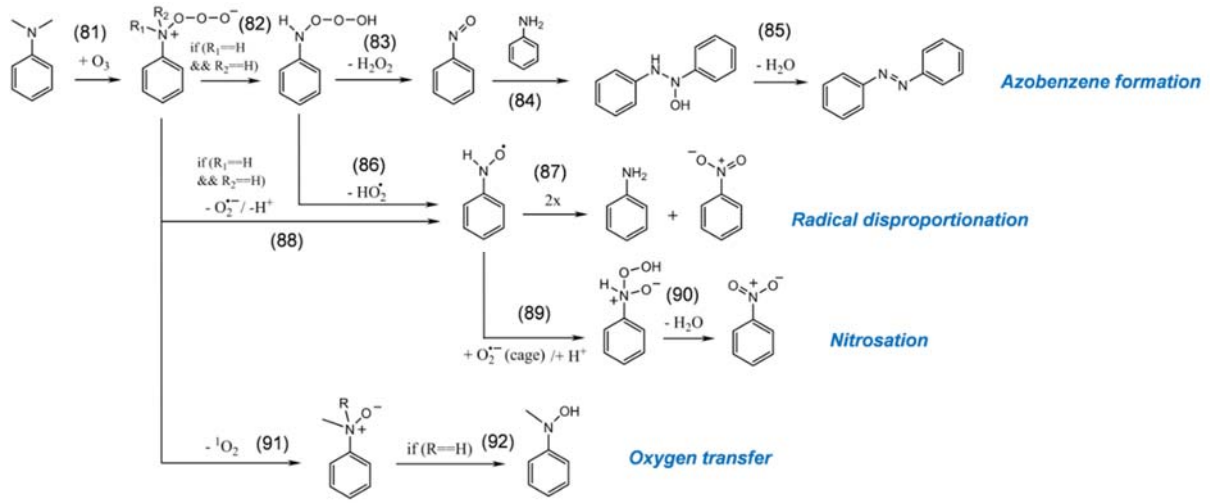
Scheme S3. Reaction pathways for the reactions of ozone with phenols via electron transfer adapted from various references (see Text S9) (logical operators such as '==' (is equal to) and '&&' (and) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



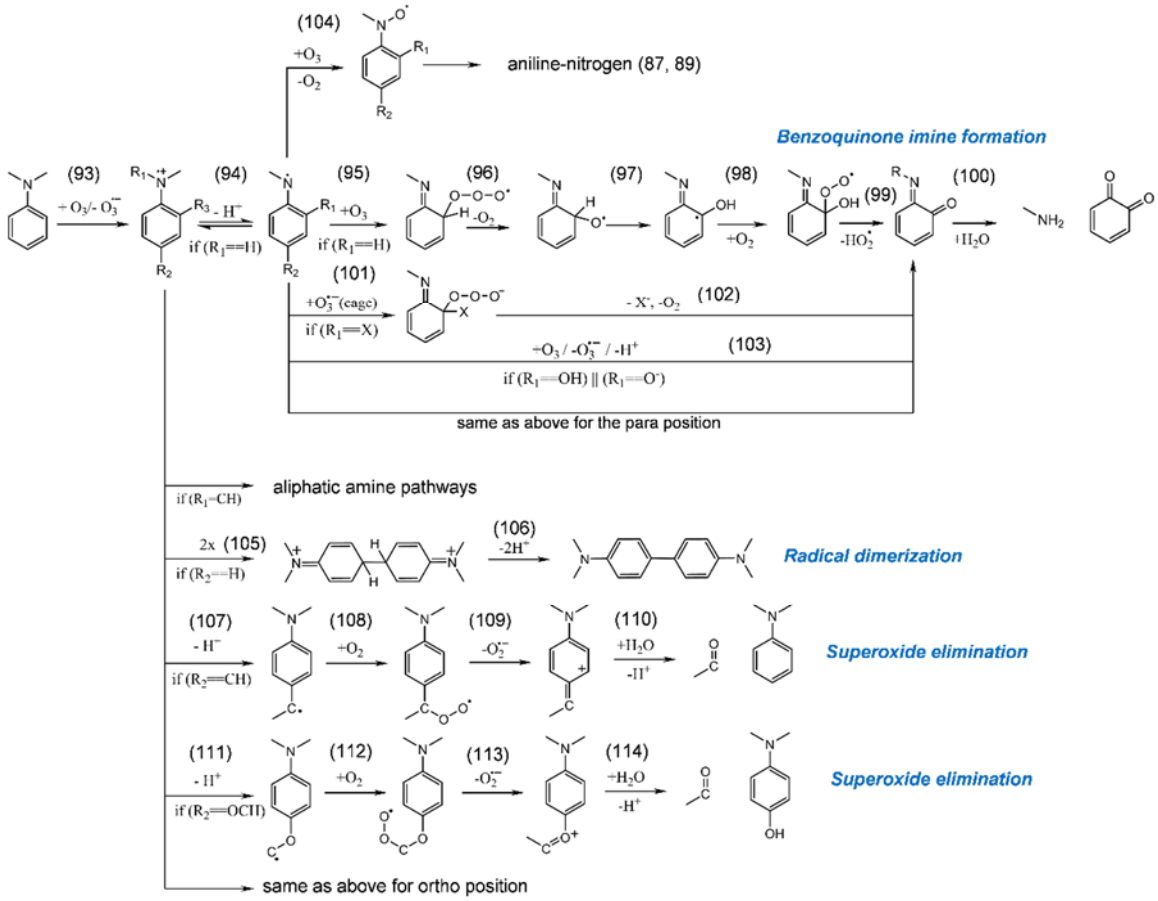
Scheme S4. Reaction pathways for the reactions of ozone with anilines at the ortho position adapted from various references (see Text S9) (logical operators such as '==' (is equal to), '&&' (and), and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



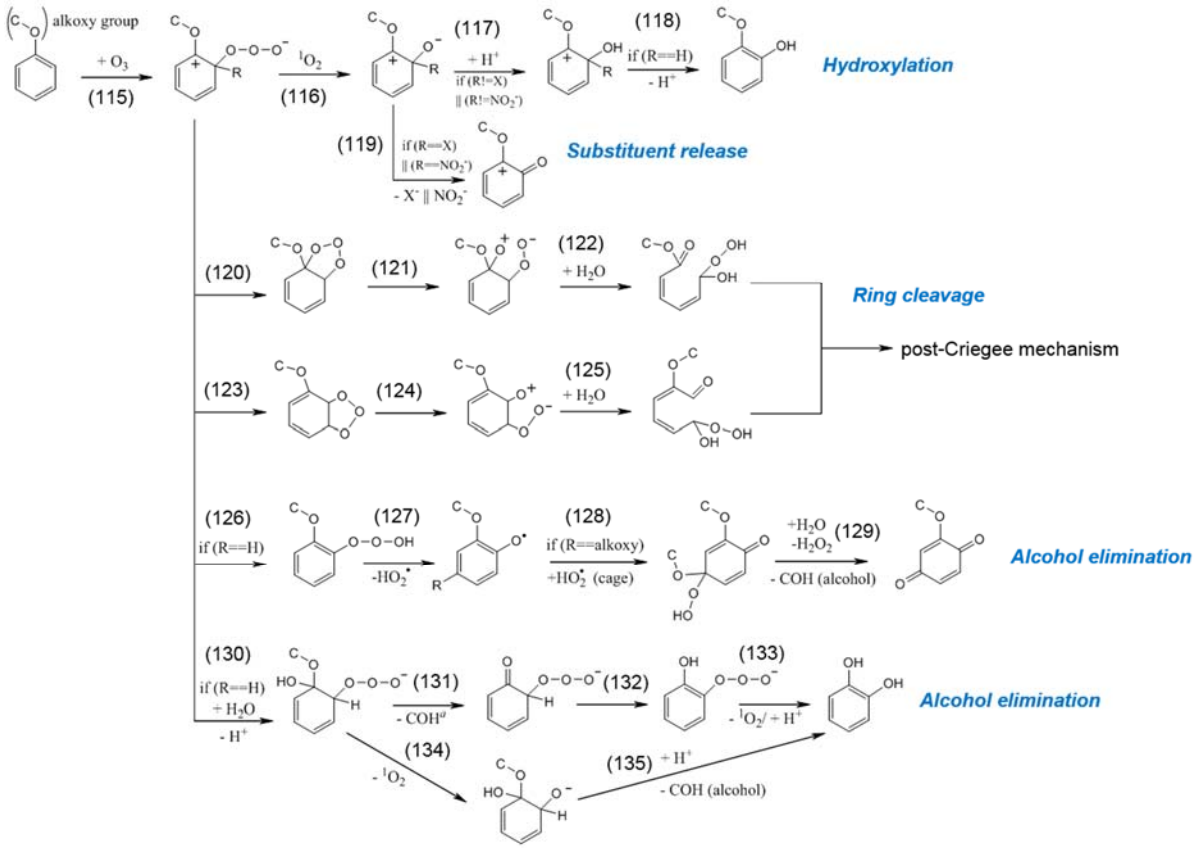
Scheme S5. Reaction pathways for the reactions of ozone with anilines at the para position adapted from various references (see Text S9) (logical operators such as '==' (is equal to) and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



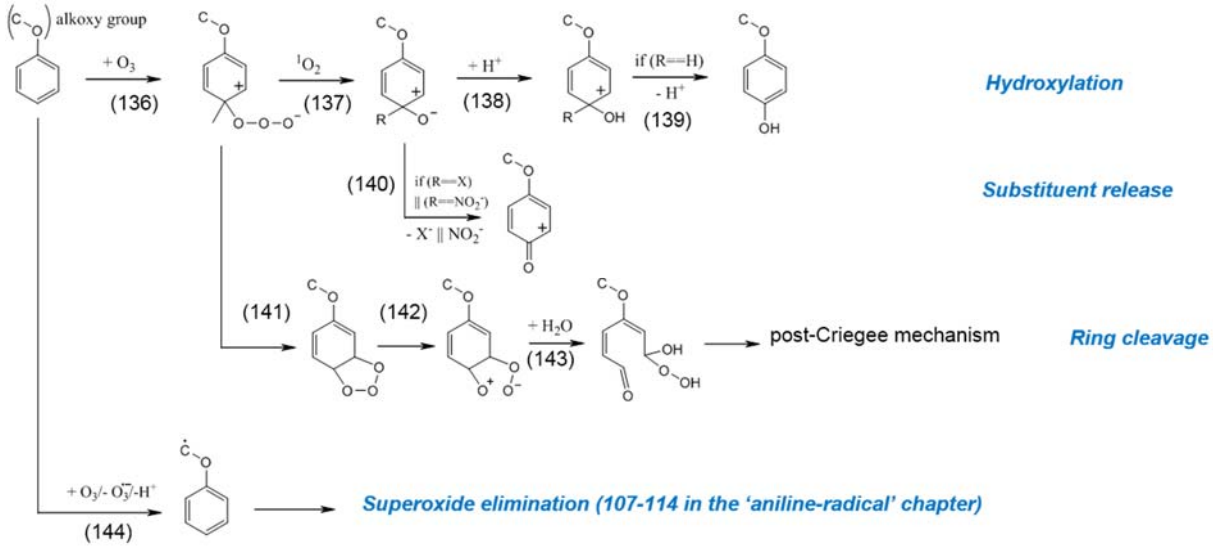
Scheme S6. Reaction pathways for the reactions of ozone with anilines at the exocyclic nitrogen adapted from various references (see Text S9) (logical operators such as '==' (is equal to) and '&&' (and) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



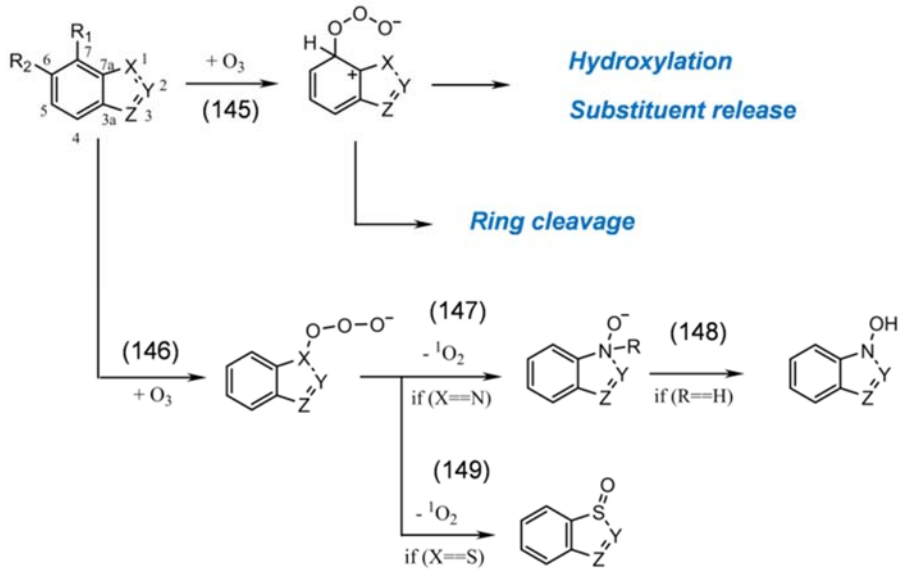
Scheme S7. Reaction pathways for the reactions of ozone with anilines via electron transfer adapted from various references (see Text S9) (logical operators such as '==' (is equal to) and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



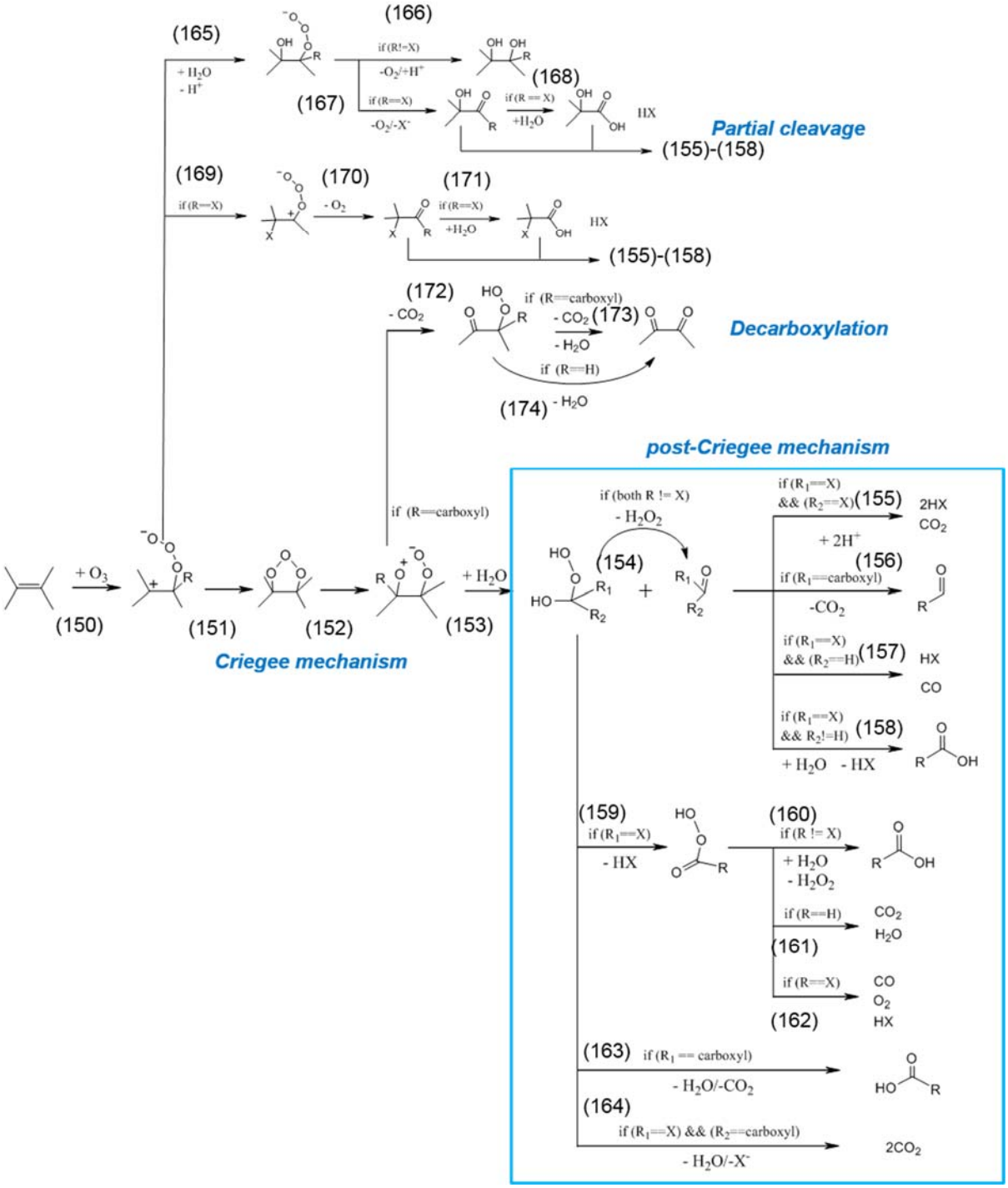
Scheme S8. Reaction pathways for the reactions of ozone with alkoxybenzenes at the ortho position adapted from various references (see Text S9) (logical operators such as '==' (is equal to) and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



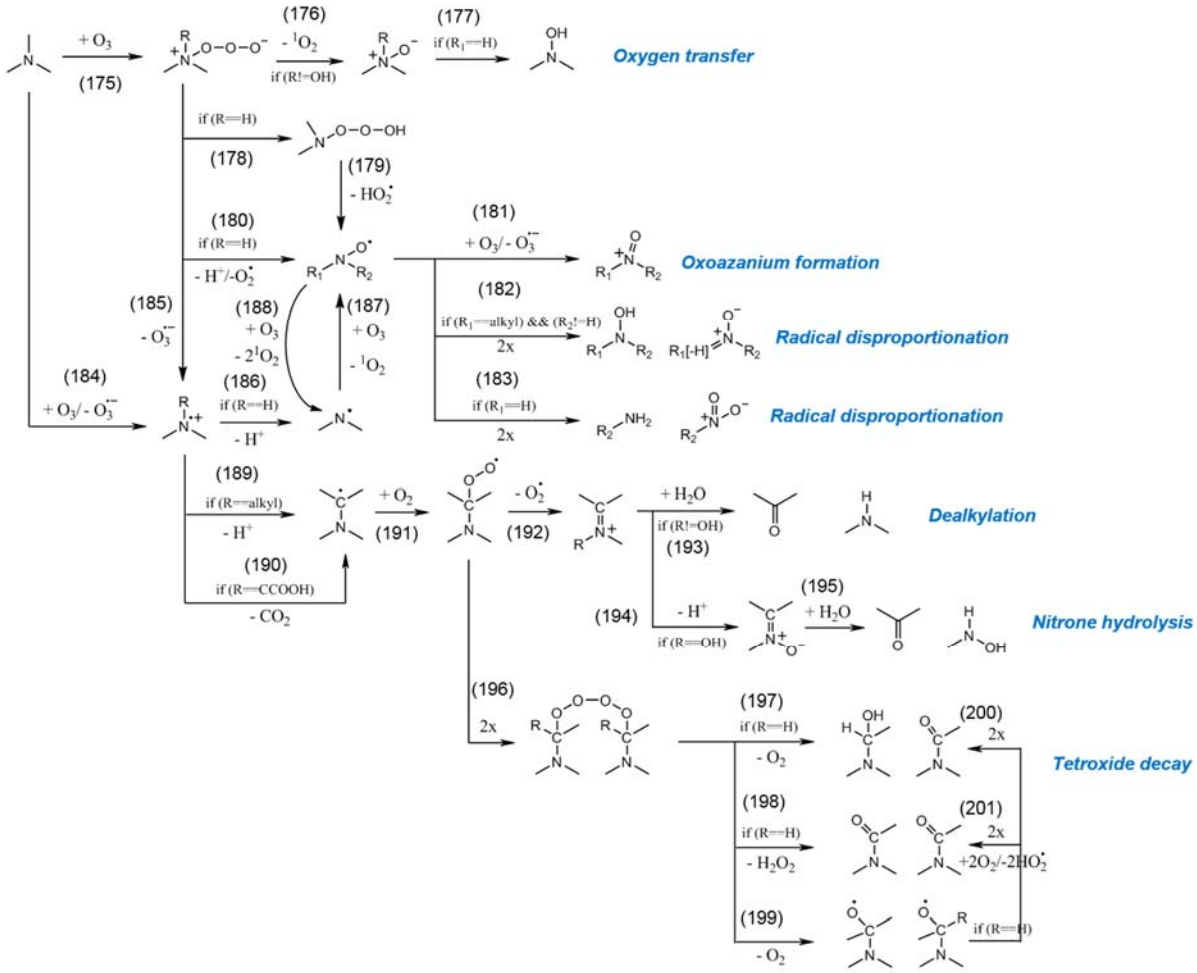
Scheme S9. Reaction pathways for the reactions of ozone with alkoxybenzenes at the para position or via electron transfer adapted from various references (see Text S9) (logical operators such as '==' (is equal to) and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



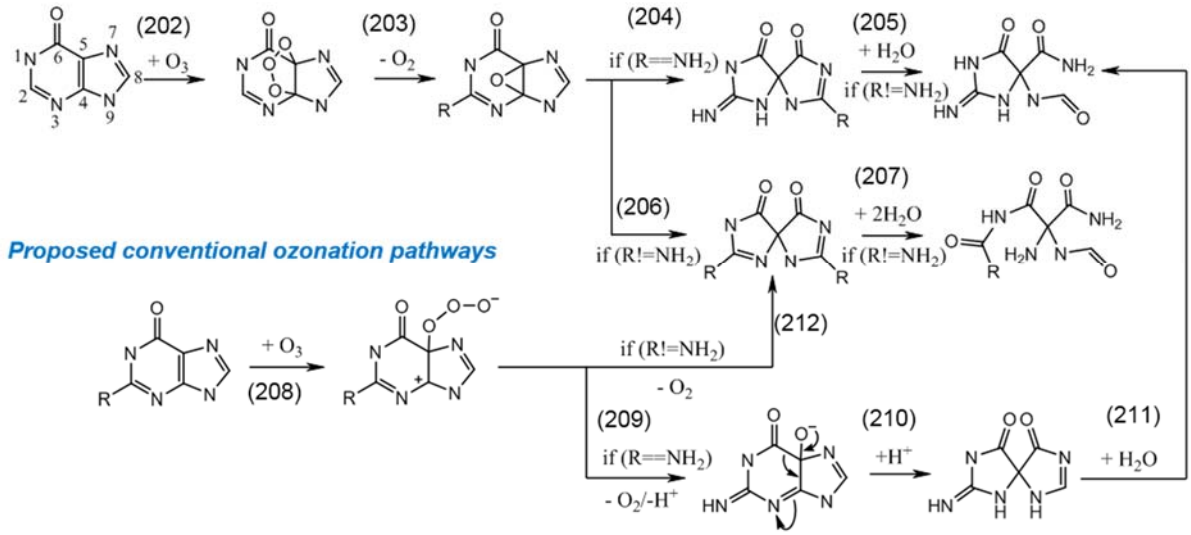
Scheme S10. Reaction pathways for the reactions of ozone with benzazoles adapted from various references (see Text S9) (a logical operator '==' (is equal to) is used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



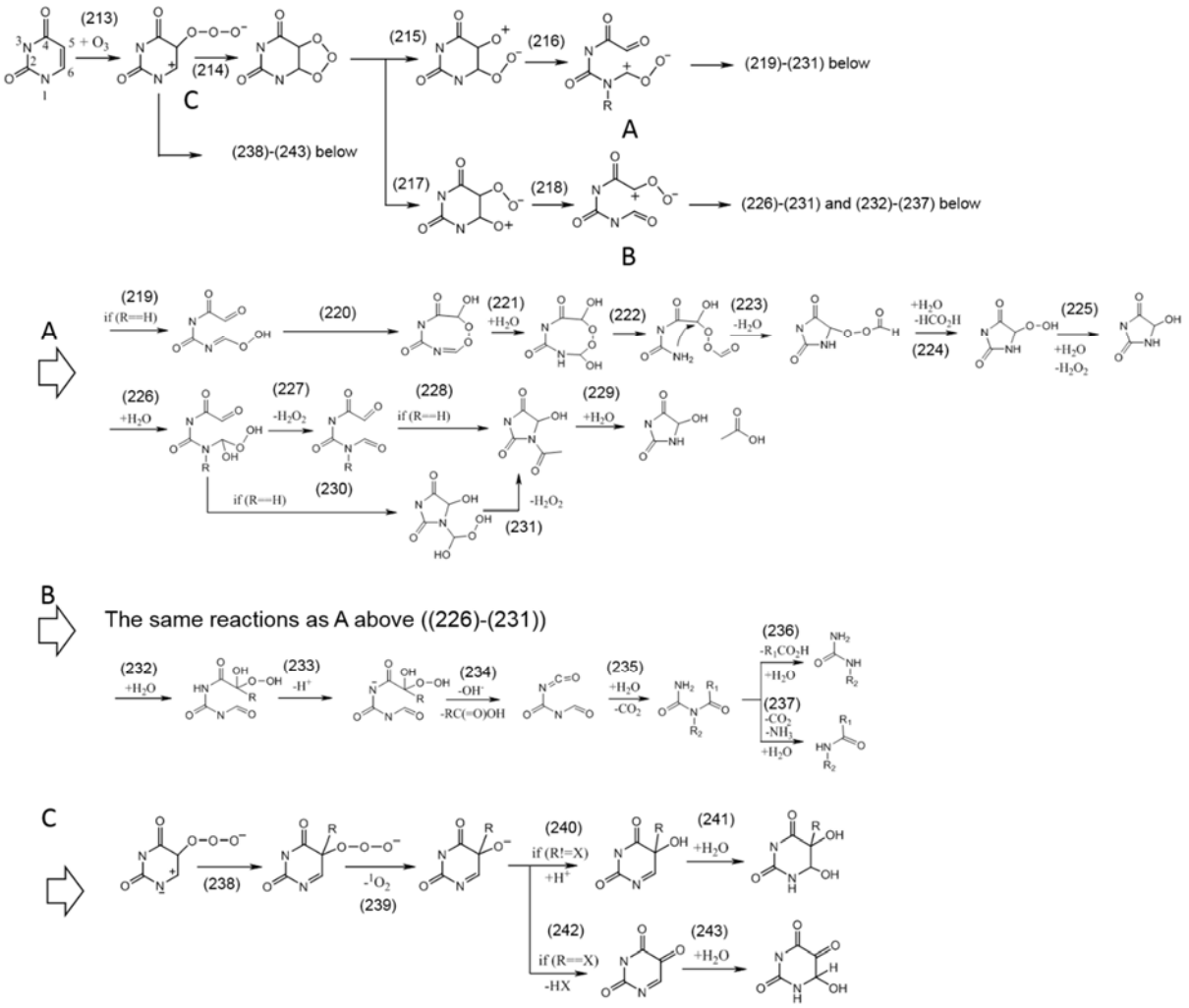
Scheme S11. Reaction pathways for the reactions of ozone with olefins adapted from various references (see Text S9) (logical operators such as '==' (is equal to), '!=' (is not equal to), '&&' (and), and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



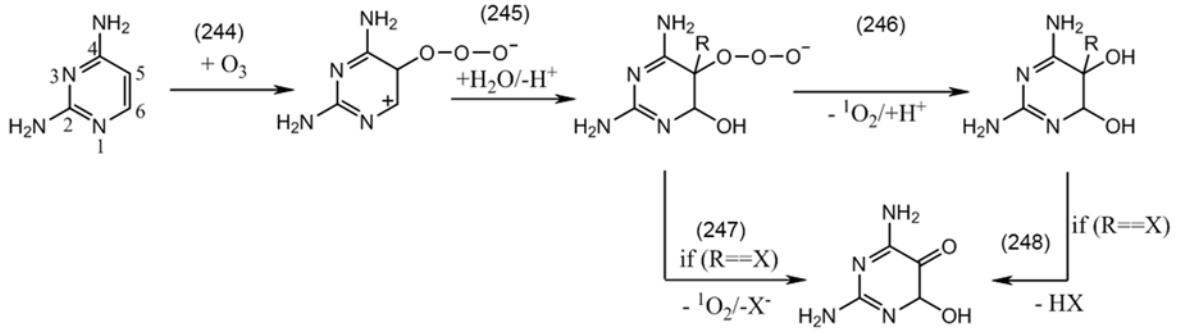
Scheme S12. Reaction pathways for the reactions of ozone with amines adapted from various references (see Text S9) (logical operators such as '==' (is equal to), '!=' (is not equal to), and '||' (or) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



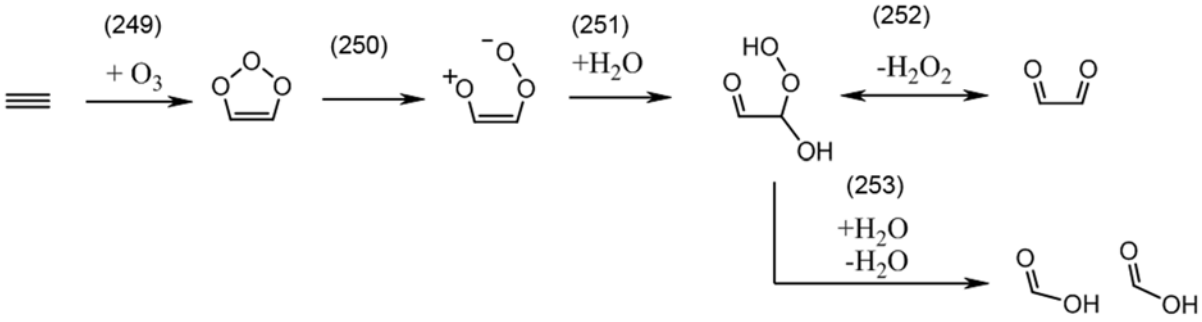
Scheme S13. Reaction pathways for the reactions of ozone with guanine analogues adapted from reference for reactions (202)-(207) and proposed herein for reactions (208)-(212) (logical operators such as '==' (is equal to) and '!=' (is not equal to) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



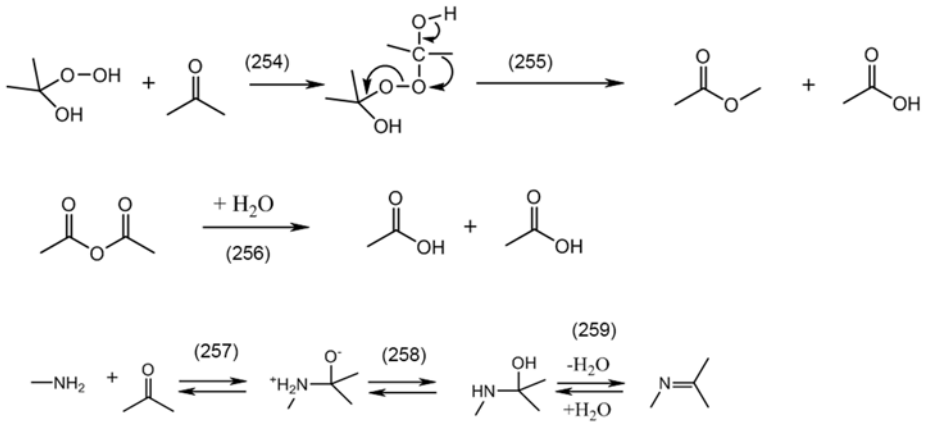
Scheme S14. Reaction pathways for the reaction of ozone with uracil adapted from the references[64,89] (logical operators such as '==' (is equal to) and '!=' (is not equal to) are used in 'if ()' statements to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



Scheme S15. Reaction pathways for the reactions of ozone with diaminopyrimidines proposed in the present study (a logical operator, '==' (is equal to), is used in a 'if ()' statement to present selectivity of reaction rules based on a substituent (R)). Implicit atoms can be hydrogen, any atom, or any organic moiety.



Scheme S16. Reaction pathways for the reactions of ozone with an ethynyl group adapted from reference[47]. Implicit atoms can be hydrogen, any atom, or any organic moiety.



Scheme S17. The selected post-ozonation reaction pathways adapted from various references (see Text S9). Implicit atoms can be hydrogen, any atom, or any organic moiety.

References

1 M. M. Sein, M. Zedda, J. Tuerk, T. C. Schmidt, A. Golloch and C. von Sonntag, *Environ. Sci. Technol.*, 2008, **42**, 6656–6662.
2 E. Borowska, M. Bourgin, J. Hollender, C. Kienle, C. S. McArdell and U. von Gunten, *Water Res.*, 2016.
3 M. C. Dodd, M.-O. Buffle and U. von Gunten, *Environ. Sci. Technol.*, 2006, **40**, 1969–1977.
4 M. Lee, S. G. Zimmermann-Steffens, J. S. Arey, K. Fenner and U. von Gunten, *Environ. Sci. Technol.*, 2015, **49**, 9925–9935.
5 E. Barron, M. Deborde, S. Rabouan, P. Mazellier and B. Legube, *Water Res.*, 2006, **40**, 2181–9.
6 W. A. Pryor, D. H. Giamalva and D. F. Church, *J. Am. Chem. Soc.*, 1984, **106**, 7094–7100.
7 R. Flyunt, A. Leitzke, G. Mark, E. Mvula, E. Reisz, R. Schick and C. von Sonntag, *J. Phys. Chem. B*, 2003, **107**, 7242–7253.
8 C. Lee, J. Yoon and U. von Gunten, *Water Res.*, 2007, **41**, 581–590.
9 J. Hoigné and H. Bader, *Water Res.*, 1983, **17**, 185–194.
10 S. Yan, A. Jia, S. Merel, S. A. Snyder, K. E. O'Shea, D. D. Dionysiou and W. Song, *Environ. Sci. Technol.*, 2016, acs.est.5b04540.
11 C. von Sonntag and U. von Gunten, *Chemistry of ozone in water and wastewater treatment: From basic principles to applications*, IWA publishing, 2012.
12 C. Prasse, M. Wagner, R. Schulz and T. A. Ternes, *Environ. Sci. Technol.*, 2012, **46**, 2169–78.
13 W. A. Pryor and R. M. Uppu, *J. Biol. Chem.*, 1993, **268**, 3120–3126.
14 T. Kotiaho, M. N. Eberlin, P. Vainiotalo and R. Kostiainen, *J. Am. Soc. Mass Spectrom.*, 2000, **11**, 526–535.
15 A. Leitzke, R. Flyunt, J. A. Theruvathu and C. von Sonntag, *Org. Biomol. Chem.*, 2003, **1**, 1012–1019.
16 B. DeWitte, J. Dewulf, K. Demeestere, V. Van De Vyvere, P. De Wispelaere and H. Van Langenhove, *Environ. Sci. Technol.*, 2008, **42**, 4889–4895.
17 *MarvinBeans/JChem suite (16.2.29.0)*, Chemaxon(http://www.chemaxon.com), 2016.
18 F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 73–78.
19 2009. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenb, .
20 E. D. Glendening, C. R. Landis and F. Weinhold, *J. Comput. Chem.*, 2013, **34**, 1429–1437.
21 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.
22 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 553–586.
23 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 616–641.
24 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 520–552.
25 T. A. Halgren and R. B. Nachbar, *J. Comput. Chem.*, 1996, **17**, 587–615.
26 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
27 J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 221–264.
28 J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209–220.
29 J. S. Binkley, J. A. Pople and W. J. Hehre, *J. Am. Chem. Soc.*, 1980, **102**, 939–947.
30 M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro and W. J. Hehre, *J. Am. Chem. Soc.*, 1982, **104**, 2797–2803.
31 K. D. Dobbs and W. J. Hehre, *J. Comput. Chem.*, 1986, **7**, 359–378.
32 K. D. Dobbs and W. J. Hehre, *J. Comput. Chem.*, 1987, **8**, 861–879.
33 K. D. Dobbs and W. J. Hehre, *J. Comput. Chem.*, 1987, **8**, 880–893.
34 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571.
35 P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 299.
36 P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 270.
37 W. R. Wadt and P. J. Hay, *J. Chem. Phys.*, 1985, **82**, 284.
38 A. Tekle-Röttering, C. von Sonntag, E. Reisz, C. vom Eyser, H. V. Lutze, J. Türk, S. Naumov, W. Schmidt and T. C. Schmidt, *Water Res.*, 2016, **98**, 147–159.
39 A. C. Pierpoint, C. J. Hapeman and A. Torrents, *J. Agric. Food Chem.*, 2001, **49**, 3827–3832.

40 C. Lee, C. Schmidt, J. Yoon and U. von Gunten, *Environ. Sci. Technol.*, 2007, **41**, 2056–2063.
41 G. Schaftenaar and J. H. Noordik, *J. Comput. Aided. Mol. Des.*, 2000, **14**, 123–134.
42 J. Benner and T. A. Ternes, *Environ. Sci. Technol.*, 2009, **43**, 5472–5480.
43 A. D. Coelho, C. Sans, A. Agüera, M. J. Gómez, S. Esplugas and M. Dezotti, *Sci. Total Environ.*, 2009, **407**, 3572–3578.
44 R. F. Dantas, M. Canterino, R. Marotta, C. Sans, S. Esplugas and R. Andreozzi, *Water Res.*, 2007, **41**, 2525–2532.
45 M. Deborde, S. Rabouan, P. Mazellier, J.-P. Duguet and B. Legube, *Water Res.*, 2008, **42**, 4299–4308.
46 C. Decoret, J. Royer, B. Legube and M. Dore, *Environ. Technol. Lett.*, 1984, **5**, 207–218.
47 M. M. Huber, T. A. Ternes and U. von Gunten, *Environ. Sci. Technol.*, 2004, **38**, 5177–5186.
48 M. Jonsson, J. Lind, T. E. Eriksen and G. Merenyi, *J. Am. Chem. Soc.*, 1994, **116**, 1423–1427.
49 E. Mvula and C. von Sonntag, *Org. Biomol. Chem.*, 2003, **1**, 1749–1756.
50 T. Nogami, T. Hishida, M. Yamada, H. Mikawa and Y. Shirota, *Bull. Chem. Soc. Jpn.*, 1975, **48**, 3709–3714.
51 H. Li, A. Guo and H. Wang, *Food Chem.*, 2008, **108**, 1–13.
52 R. Davies and J. L. Frahn, *J. Chem. Soc. Perkin Trans. 1*, 1977, 2295.
53 V. L. Singleton, M. Salgues, J. Zaya and E. Trousdale, *Am. J. Enol. Vitic.*, 1985, **36**, 50–56.
54 J. Benner and T. A. Ternes, *Environ. Sci. Technol.*, 2009, **43**, 5086–5093.
55 D. B. Mawhinney, B. J. Vanderford and S. A. Snyder, *Environ. Sci. Technol.*, 2012, **46**, 7102–7111.
56 M. K. Ramseier and U. von Gunten, *Ozone Sci. Eng.*, 2009, **31**, 201–215.
57 J. L. Sotelo, F. J. Beltran and M. Gonzalez, *Ind. Eng. Chem. Res.*, 1990, **29**, 2358–2367.
58 W. T. Dixon and D. Murphy, *J. Chem. Soc. Faraday Trans. 2*, 1976, **72**, 1221.
59 J. Hoigne and H. Bader, *Science (80-.).*, 1975, **190**, 782–784.
60 J. Sarasa, *Water Res.*, 2002, **36**, 3035–3044.
61 K. Turhan and S. Uzman, *Ann. Chim.*, 2007, **97**, 1129–1138.
62 J. Dalmagro, R. A. Yunes and E. L. Simionatto, *J. Phys. Org. Chem.*, 1994, **7**, 399–402.
63 S. G. Zimmermann, A. Schmukat, M. Schulz, J. Benner, U. von Gunten and T. A. Ternes, *Environ. Sci. Technol.*, 2012, **46**, 876–84.
64 F. Muñoz, E. Mvula, S. E. Braslavsky and C. von Sonntag, *J. Chem. Soc. Perkin Trans. 2*, 2001, 1109–1116.
65 F. Lange, S. Cornelissen, D. Kubac, M. M. Sein, J. von Sonntag, C. B. Hannich, A. Golloch, H. J. Heipieper, M. Möder and C. von Sonntag, *Chemosphere*, 2006, **65**, 17–23.
66 L. Qin, G. N. R. Tripathi and R. H. Schüler, *Zeitschrift für Naturforsch. A*, 1985, **40**, 1026–1039.
67 E. Mvula, S. Naumov and C. von Sonntag, *Environ. Sci. Technol.*, 2009, **43**, 6275–6282.
68 R. Criegee, *Angew. Chemie Int. Ed. English*, 1975, **14**, 745–752.
69 P. Dowideit and C. von Sonntag, *Environ. Sci. Technol.*, 1998, **32**, 1112–1119.
70 A. Leitzke and C. von Sonntag, *Ozone Sci. Eng.*, 2009, **31**, 301–308.
71 U. Jans, *ETH Zurich*, 1996.
72 A. Leitzke, E. Reisz, R. Flyunt and C. von Sonntag, *J. Chem. Soc. Perkin Trans. 2*, 2001, 793–797.
73 D. C. McDowell, M. M. Huber, M. Wagner, U. von Gunten and T. A. Ternes, *Environ. Sci. Technol.*, 2005, **39**, 8014–8022.
74 F. Muñoz and C. von Sonntag, *J. Chem. Soc. Perkin Trans. 2*, 2000, 661–664.
75 Y. Lester, H. Mamane, I. Zucker and D. Avisar, *Water Res.*, 2013, **47**, 4349–4356.
76 K. S. Tay and N. Madehi, *Sci. Total Environ.*, 2015, **520**, 23–31.
77 B. De Witte, H. Van Langenhove, K. Hemelsoet, K. Demeestere, P. De Wispelaere, V. Van Speybroeck and J. Dewulf, *Chemosphere*, 2009, **76**, 683–9.
78 A. Tekle-Röttering, K. S. Jewell, E. Reisz, H. V. Lutze, T. A. Ternes, W. Schmidt and T. C. Schmidt, *Water Res.*, 2016, **88**, 960–971.
79 U. von Gunten, *Water Res.*, 2003, **37**, 1443–67.
80 M. Jonsson, D. D. M. Wayner and J. Lusztyk, *J. Phys. Chem.*, 1996, **100**, 17539–17543.
81 P. Neta, R. E. Huie and A. B. Ross, *J. Phys. Chem. Ref. Data*, 1990, **19**, 413.

82 S. Das and C. von Sonntag, 1986.
83 S. Das, M. N. Schuchmann, H.-P. Schuchmann and C. Von Sonntag, *Chem. Ber.*, 1987, **120**, 319–323.
84 C. von Sonntag and H.-P. Schuchmann, *Angew. Chemie Int. Ed. English*, 1991, **30**, 1229–1253.
85 G. A. Russell, *J. Am. Chem. Soc.*, 1957, **79**, 3871–3877.
86 J. E. Bennett and R. Summers, *Can. J. Chem.*, 1974, **52**, 1377–1379.
87 C. von Sonntag, *Free-radical-induced DNA damage and its repair*, Springer, 2006.
88 S. Enami, M. R. Hoffmann and A. J. Colussi, *J. Phys. Chem. B*, 2008, **112**, 4153–4156.
89 R. Flyunt, J. A. Theruvathu, A. Leitzke and C. von Sonntag, *J. Chem. Soc. Perkin Trans. 2*, 2002, 1572–1582.
90 R. Andreozzi, A. Insola, V. Caprio and M. G. M. D'Amore, *Water Res.*, 1991, **25**, 655–659.
91 J. L. Acero, K. Stemmler and U. von Gunten, *Environ. Sci. Technol.*, 2000, **34**, 591–597.
92 J. Radjenović, M. Godehardt, A. Hein, M. Farré, M. Jekel and D. Barceló, *Environ. Sci. Technol.*, 2009, **43**, 6808–6815.
93 J. Kuang, J. Huang, B. Wang, Q. Cao, S. Deng and G. Yu, *Water Res.*, 2013, **47**, 2863–72.
94 M. C. Dodd, D. Rentsch, H. P. Singer, H.-P. E. Kohler and U. von Gunten, *Environ. Sci. Technol.*, 2010, **44**, 5940–5948.
95 R. Flyunt, O. Makogon, M. N. Schuchmann, K.-D. Asmus and C. von Sonntag, *J. Chem. Soc. Perkin Trans. 2*, 2001, 787–792.
96 S. Enami, M. R. Hoffmann and A. J. Colussi, *Chem. Res. Toxicol.*, 2009, **22**, 35–40.
97 S. Enami, M. R. Hoffmann and A. J. Colussi, *J. Phys. Chem. B*, 2009, **113**, 9356–9358.
98 J. Kalia and R. T. Raines, *Angew. Chemie Int. Ed.*, 2008, **47**, 7523–7526.
99 K. Y. TAM and L. QUÉRÉ, *Anal. Sci.*, 2001, **17**, 1203–1208.
100 A. Pagliara, B. Testa, P.-A. Carrupt, P. Jolliet, C. Morin, D. Morin, S. Urien, J.-P. Tillement and J.-P. Rihoux, *J. Med. Chem.*, 1998, **41**, 853–863.
101 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3093.
102 U. von Gunten, E. Salhi, C. K. Schmidt and W. A. Arnold, *Environ. Sci. Technol.*, 2010, **44**, 5762–5768.
103 D. Trogolo, B. K. Mishra, M. B. Heeb, U. von Gunten and J. S. Arey, *Environ. Sci. Technol.*, 2015, **49**, 4163–4175.