This document is the accepted manuscript version of the following article:
Lee, M., Blum, L. C., Schmid, E., Fenner, K., & von Gunten, U. (2017). A computer-based prediction platform for the reaction of ozone with organic compounds in aqueous solution: kinetics and mechanisms. Environmental Science: Processes and Impacts, 19(3), 465-476. https://doi.org/10.1039/C6EM00584E

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]

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

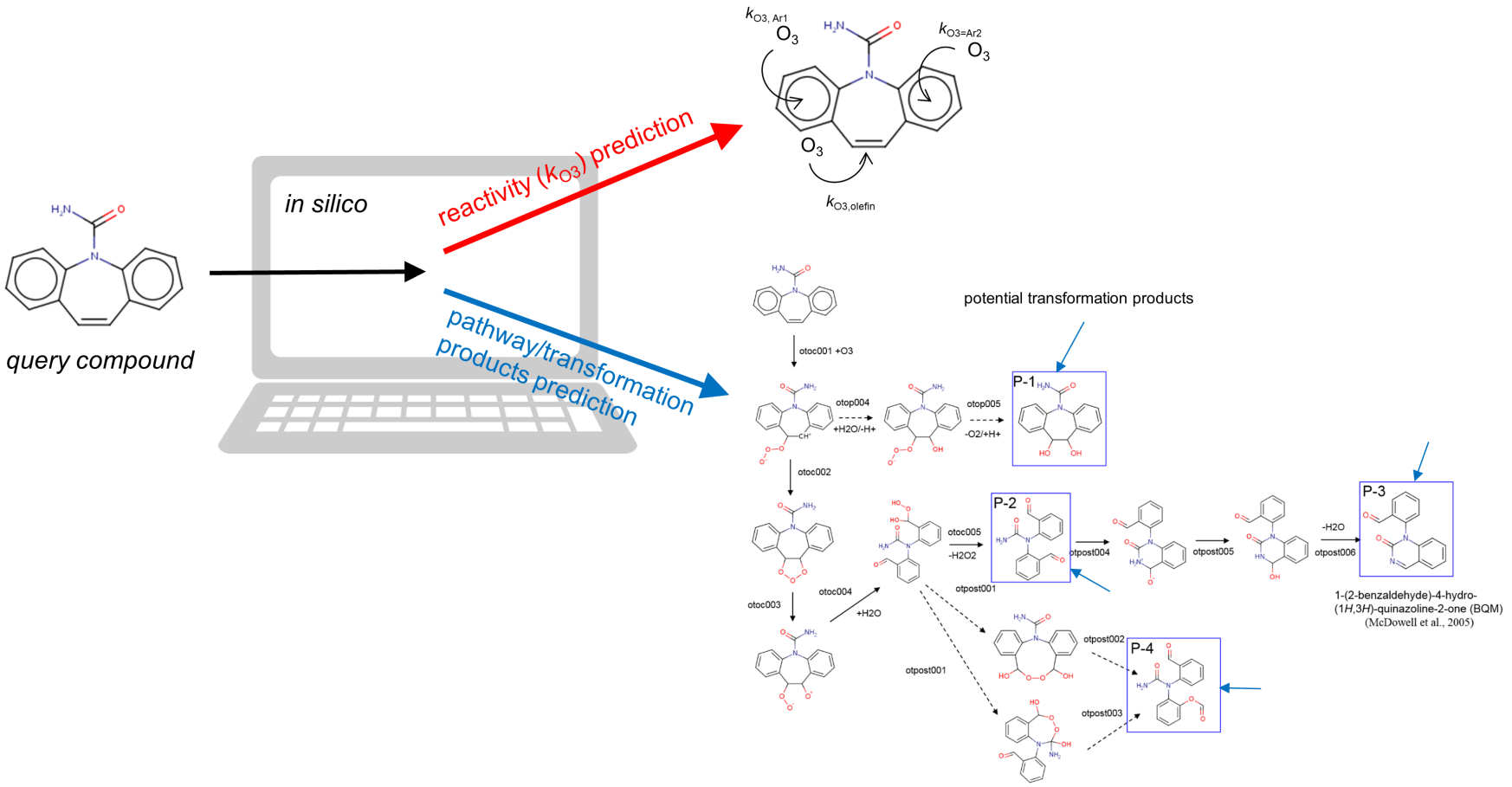
Submitted to Environmental Science: Processes & Impacts

*Corresponding author

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

†Electronic supplementary information (ESI) available

22 Table of Contents (TOC) Art



23

24

Abstract

Ozonation of secondary wastewater effluents can reduce the discharge of micropollutants by transforming their chemical structures. Therefore, a better understanding of the formation of transformation products during ozonation is important. In this study, a computer-based prediction platform for kinetics and mechanisms of the reactions of ozone with organic compounds was developed to enable *in silico* predictions of transformation products. With the developed prediction platform, reaction kinetics expressed as second-order rate constants for the reactions of ozone with selected organic compounds (k_{O3}, $M^{-1}s^{-1}$) can be predicted based on an adapted k_{O3} prediction model from a previous study (Lee et al., 2015[1]) (average model error of about a factor of 6 for 14 compound classes with 284 model compounds). Ozone reaction mechanisms reported in literature have been reviewed and, using chemoinformatics tools, encoded into about 340 individual reaction rules that can be generally applied to predict transformation products of micropollutants. Predictions for k_{O3} and/or transformation products were overall consistent with experimental data for three micropollutants used as validation compounds (e.g., carbamazepine, tramamol, and triclosan). However, limitations of the current k_{O3} prediction platform were also identified: ambiguous assignment of the n-th highest occupied molecular orbital energy (E_{HOMO-n}) to the reactive sites, potential errors associated with the use of a gas-phase geometry, poor k_{O3} prediction for certain compounds (cetirizine). Therefore, the current prediction tool should not be considered as a substitute for experimental studies and experimental data are still required in the future to obtain a more robust prediction model. Nonetheless, the developed prediction platform, made available as a stand-alone graphical user interface (GUI) application, will to provide useful information about aqueous ozone chemistry to various groups of end users such as environmental chemists, engineers, or toxicologists.

Introduction

Ozonation, which has been used for drinking water disinfection since the early 20th century,[2] has recently been considered as a viable tertiary treatment process for mitigating the release of various micropollutants (e.g., pharmaceuticals, personal care products, pesticides, etc.) from secondary wastewater effluents.[2–6] However, as ozone mainly transforms the chemical structure of a micropollutant rather than mineralizing it, the formation of transformation products and their potential biological effects has remained a concern.[2,7–11] Therefore, numerous studies employing various *in vitro* and *in vivo* bioassays to assess the water quality after ozonation have been carried out.[11–20] The findings generally support the beneficial effects of ozonation, which was demonstrated both for individual compounds with exposure-based assays[11–15] and for ozonated wastewaters through effect-based assays.[16–20] It is noteworthy, however, that in certain wastewaters elevated biological effects were observed after ozonation, which were later reduced in a subsequent biological treatment step.[21–23] The identification of ozone transformation products is important for the empirical or theoretical assessment of their toxicity and biodegradability. However, considering the ever increasing number of chemicals detected in natural and technical aquatic systems, an empirical elucidation of the transformation products for all relevant micropollutants is a formidable task to be achieved.

Over the last decades, knowledge on aqueous ozone chemistry regarding reaction kinetics and mechanisms for organic compounds has greatly advanced.[2,24,25] Several hundred second-order rate constants (k_{O3}) for the reactions of ozone with organic compounds were experimentally determined. Various classes of compounds such as aromatic compounds, olefins, amines, and organosulfur compounds were investigated. The range of k_{O3} spans over >11 orders of magnitude from $<10^{-2}$ $M^{-1}s^{-1}$ – 10^{9} $M^{-1}s^{-1}$.[2] With these experimental k_{O3} values as a basis, several k_{O3} prediction tools such as a quantitative structure-activity relationship (QSAR) model[26] and a molecular orbital energy-based quantum chemical model[1] have been developed. Moreover, for numerous organic substances, reaction mechanisms were proposed based on the identified intermediates and final products.[2] Common intermediates and transformation products were often reported for differing organic compounds with similar ozone-reacting moieties: Criegee products for olefins (e.g., cephalexin,[11] progesterone[27]), hydroxylation and ring cleavage for

aromatic ring-containing compounds (e.g., phenol,[28,29] bisphenol A,[30] and methoxylated benzenes[31]) and *N*-oxide formation and *N*-dealkylation for tertiary amino group-containing compounds (e.g., tramadol[32] and clarithromycin[33]), or hydroxylamine formation for secondary amines (e.g., propranolol,[34] piperidine,[35] and morpholine[35]), respectively. Therefore, it seems possible to deduce chemical structures of ozone transformation products of hitherto uninvestigated compounds if a reference study is available.

In the present study, a computer-based prediction platform for kinetics and pathways for reactions of organic compounds with ozone in aqueous solution was developed based on available kinetic and mechanistic information for aqueous ozone reactions. It largely consists of two prediction modes: (i) reactivity (k_{O3}) prediction and (ii) pathway prediction. In (i), k_{O3}-values for chemical moieties of a query compound potentially reacting with ozone are provided. Either k_{O3} is predicted by a prediction protocol that has been adapted from a previously developed quantum chemical model[1] or an experimental k_{O3} of a reference compound that is structurally similar to the chemical moiety of interest is suggested as an estimate for the k_{O3} value. In (ii), ozone reaction pathways proposed in the relevant peer-reviewed literature have been generalized into transformation rules to predict potential reaction pathways and transformation products. A detailed description of the development of the prediction platform is given and its application for some examples is demonstrated.

Materials and methods

Chemoinformatics and quantum chemical computation tools

Various applications of the MarvinBeans/JChem package (Linux version 16.2.29.0, ChemAxon)[36] in Java have been used as follows: Marvin for drawing input molecules, converting them into SMILES (simplified molecular-input line-entry system)[37] strings, and visualizing molecular structures, the pK_a calculator plugin for generating acid-base species, predicting acid dissociation constants, and producing a species distribution as a function of pH, and Conformer plugin and Reactor for generating the 3D structure of a molecular structure from a SMILES string for further quantum chemical computations and for enumerating reaction pathways, respectively. For quantum chemical computations, ORCA[38] was used for semi-empirical and *ab initio* quantum chemical computations. Natural bond orbital (NBO) analysis was

performed by the NBO program 6.0.[39] Molden 5.4[40] was used for a graphical representation of chemical structures and molecular orbitals.

Development of a computer-based prediction platform

A computer-based prediction platform for ozone reactions was developed based on the workflow given in Fig. 1. To initiate a prediction, a *query compound* has to be submitted. A *query compound* is limited to organic compounds comprising of carbon and atoms such as hydrogen, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine. With the submission of a *query compound*, a user can choose between two prediction modes: (i) *direct prediction* and (ii) *pH-dependent prediction*. In (i) the prediction is performed exclusively for the species specified as a chemical structure by the user, while in (ii) a comprehensive pH-dependent prediction for a query compound undergoing acid-base speciation is performed.

For a *direct prediction*, a *query compound* is subject to (a) a *reactive site search* (Fig. 1), which detects chemical moieties in the *query compound* potentially reacting with ozone and presents them to the user. In a further step, the user can choose between two prediction modes, i.e., either (b) *rate constant (k_{O3}) prediction* or (c) *pathway enumeration*. In (b) k_{O3} is predicted for the identified reactive sites, while in (c) the reaction pathways for the reactive site selected by the user is enumerated, respectively. More information on the *reactive site search* and the two prediction modes is given below. Both (b) *rate constant (k_{O3}) prediction* and (c) *pathway enumeration* can be operated independently or in combination. In the latter case, (b) precedes (c) because *rate constant (k_{O3}) prediction* provides information as to which reactive site dominantly reacts with ozone based on the predicted k_{O3} values for the individual sites, and this is decisive for (c) *pathway enumeration*. For a *pH-dependent prediction*, (a) *reactive site search* and (d) *pH-dependent rate constant prediction* are comprehensively carried out for all the relevant acid-base species of the *query compound* for a specific pH or a pH range (see below for details). The pH-dependent prediction outputs provide information on the extent of the contribution of individual acid-base species to the overall reactivity at a certain pH. Based on the respective contributions of the individual species, the user can export the species of interest ((e) *species selection*) to the direct prediction train for further pre-

dictions (arrow going from (e) to (a) in Fig. 1). Note that (a) and (b) had already been executed for the exported species from the *pH-dependent prediction*. Therefore, no additional calculations for (a) and (b) are necessary. Later generation oxidation products can be predicted by feeding back the predicted transformation products as *query compounds*.

Reactive site search. A reactive site is a chemical moiety potentially reactive with ozone, for which a prediction can be made. The entire list of reactive sites currently defined (48 sites in total for five compound groups) is presented in the first column in Table S1† (see Fig. S1† for their chemical structures). The naming convention for the reactive sites in Table S1† has been proposed not only as an identification of the chemical moiety transformed, but also considering their appropriate/convenient assignments to the subsequent prediction models. As shown in Table S1†, individual reactive sites are coupled with the corresponding k_{O3} *prediction model group* and *reaction pathway group* in the 3rd column and the 4th column, respectively. Details of definitions for individual reactive sites and their assignments are presented in Text S1†. A SMARTS (SMiles ARbitrary Target Specification)[41] string is used for identifying substructural patterns (i.e., reactive sites) in a *query compound*. Two differing sets of chemical compounds, ~300 compounds used to develop k_{O3} prediction models in this study and ~ 500 environmentally relevant micropollutants (374 from Schymanski et al., 2014[42]) used as in-house target analytes in the Department of Environmental Chemistry, Eawag (Switzerland), have been selected to train the *reactive site search* module. It has been manually checked that reactive sites for all the compounds are assigned as intended (data not shown).

Rate constant prediction. k_{O3} for a reactive site denoted as '*prediction*' in the 2nd column in Table S1† is to be predicted by the corresponding quantum chemical models (3rd column in Table S1†) adapted from the originally proposed model.[1] For a reactive site with no k_{O3} prediction model assigned, either an empirical k_{O3} estimate (2nd column in Table S1† and Table S2† for a detailed derivation) is used when there is a reference value available (e.g., 200 $M^{-1}s^{-1}$ for ethynyl group determined for 1-ethinyl-1-cyclohexanol[10]), or k_{O3} is considered unknown (denoted as *n.a.* in 2nd column in Table S1†) when no empirical information is available either (e.g., thiophenol). As shown in Table S2†, an experimental k_{O3} for a reference compound (or an average k_{O3} when multiple reference compounds were used) was used as

a k_{O3} estimate for such reactive sites without k_{O3} prediction model. Therefore, these k_{O3} estimates may bear a large uncertainty because they do not take into account the substituent effects for derivatives of the reference structure. As mentioned above, '*prediction*' in the 2nd column in Table S1† indicates that quantum chemical computations will be conducted to predict k_{O3} for the reactive site. This is achieved by applying the corresponding prediction model with the associated orbital energy (3rd column in Table S1†) following the workflow shown in Fig. S4†. A k_{O3} prediction consists of a '*speciation analysis*' followed by a '*quantum chemical computation*'. Briefly, a speciation analysis is for a query compound undergoing acid-base speciation. The user can decide to conduct it for a specific pH or the pH range 0-14. Compound-specific speciation parameters such as dissociation constants (pK_a) and a tautomeric fraction (*f*), if relevant, can be either predicted on the fly or provided by the user. Based on this, a species distribution between pH 0 and 14 for the query compound is derived. *f* indicates the fraction of individual tautomers (i.e., species with the same net charge). A more detailed explanation of the species analysis is given in Text S2†.

Adopting to a previously developed k_{O3} prediction model based on molecular orbitals,[1] k_{O3} predictions for reactive sites were performed by quantum chemical computations deriving orbital energies (e.g., n-th highest occupied molecular orbital energy (E_{HOMO-n}) or natural bond orbital energy (E_{NBO})) corresponding to the reactive site. The obtained orbital energies (E_{HOMO-n} and/or E_{NBO}) are then used in a linear prediction model to predict k_{O3} for a reactive site. The computation protocol denoted as 'HF/3-21G//MMFF94', which was adapted from the original method in a previous study,[1] has been proposed to be universally applicable to all the compound groups in this study. HF/3-21G//MMFF94 implies that an orbital energy is derived from a single point calculation with the *ab initio* HF method using the Pople 3-21G basis set[43–47] for a geometry obtained using the molecular mechanics MMFF94 method.[48–52] The HF method was chosen over the DFT-B3LYP method because it has previously been shown to perform comparable to the B3LYP method. Furthermore, the HOMO (n=0) rather than the lower HOMO-n was consistently found to be most appropriate for the k_{O3} predictions with the HF method, rendering it a more straightforward application.[1] This protocol provides a similar prediction performance as the original model[1] while being much less computationally expensive. To this end, more information is provided in the ESI for the meth-

odology for development of the k_{O3} prediction model (Text S3†), the evaluation of the k_{O3} prediction models with different computational methods (Text S4†), the detailed description and the computation costs of the computational methods chosen in this study for k_{O3} predictions (Text S5†), the hierarchy for the assignment of ambivalent aromatic compounds (Text S6†), and the assignment of E_{HOMO-n}-values (n = 1,2,3, ··· indicating the occupied molecular orbital, n levels lower than E_{HOMO} where n=0) to the corresponding reactive sites (Text S7†). Combining both the predicted k_{O3} and the speciation information for a *query compound*, k_{O3}-values corresponding to a reactive site at a specific pH (or as a function of pH) can be predicted. A derivation of different types of predicted k_{O3} values such as (apparent) site-specific or species-specific k_{O3} values and their correspondence to experimentally measured k_{O3}-values is described in detail in Text S8†.

Pathway enumeration. Ozone reaction pathways proposed in peer-reviewed research articles have been comprehensively reviewed and compiled. The selected reaction pathways are discussed in detail in Text S9†. Using Reactor of the JChem package (Chemaxon), all the selected reaction pathways (Schemes S1-S17†) were segmented into individual unit reactions encoded as reaction rule files. Series of unit reaction rules have then been organized into pre-defined tree-like patterns to reproduce the structure of the reference pathways from literature. For instance, the reaction rule tree for phenols (Schemes S1-S3†) is shown in Fig. 2. It consists of three major branches, namely, "ortho", "para", and "radical". In Fig. 2, the sequence of reactions for the ortho position of phenols is shown, based on which the reaction rules were defined. The ortho branch is initiated with a reaction rule '*otapho01*'. When a *query compound* contains a phenol group, the rule '*otapho01*' is triggered to produce an ozone adduct zwitterion at the ortho position. Although only the neutral form of phenol is shown in Fig. 2 for simplicity, the reaction rules were encoded to be triggered for all phenol- and phenolate-containing compounds, respectively. The system continues to trigger subsequent reaction rules with the previous product until there is no match found. The 'post-Criegee mechanism' mentioned in Fig. 2 is the pathways defined for the olefin group in Scheme S11† and is also applied to the resulting products formed from phenolic compounds via reactions (7)-(10) or (12)-(14) (Fig. 2). Criegee mechanism is discussed in more detail in Text S9†. To date, about 340 unit reaction rules have been defined, based on the reaction pathways in Schemes S1-S17†.

Note that due to the lack of information, regioselectivity and stereoselectivity were generally not considered in this study except for a few prominent cases (e.g., para- and ortho-positions of phenol or aniline are only considered for the reaction with ozone in Schemes S1, S2, S4, and S5†).

Results and discussion

A standalone graphical user interface (GUI) has been developed for the prediction platform. Demonstrations for different prediction modes, namely, direct k_{O3} prediction, pH-dependent k_{O3} prediction, and pathway enumeration, are presented below for selected micropollutants.

Rate constant (k_{O3}) prediction

Direct k_{O3} prediction. The main window of the developed prediction platform is shown in Fig. 3. In the left panel named 'Input', the chemical structure for a *query compound* can be manually drawn on a canvas or be imported as a 1D SMILES string. Carbamazepine is displayed as an example. The user can then choose between direct prediction (A) and pH-dependent prediction (B) (Fig. 3). Direct prediction performs a reactive site search and displays the results on the right panel. For carbamazepine, two benzene rings and one olefin group were detected as reactive sites. Note that none of the two nitrogens in carbamazepine has been detected since amide nitrogens have a very low ozone reactivity[2] (see Text S1† for details). In the right panel, the user can choose to predict k_{O3} (C), reaction pathways (D), or transformation products (E). For k_{O3} prediction, no selection of the reactive site is necessary because relevant orbital energies are obtained all together from a single quantum chemical computation for the entire structure. In contrast, one needs to select a reactive site for pathway or product prediction as the reaction rules are applied to a specific reactive site (see below for more details).

Upon choosing '*Predict kO3*', the quantum chemical computations are carried out following the workflow shown in Fig. S4† without speciation analysis and the predicted k_{O3}-values are displayed in the third column ([kO3 /M/s]) after the computation has been finished. Together with the predicted k_{O3}, k_{O3} estimates are also presented for the assigned reactive sites. While a k_{O3} of 4.5×10^4 $M^{-1}s^{-1}$ was predicted for

the olefin, a k_{O3} of 3.1×10^2 $M^{-1}s^{-1}$ was reported for both benzene rings (see below for further discussion about the $E_{HOMO\text{-}n}$ assignment for predicting k_{O3} for the benzene rings). The predicted k_{O3}-values indicate that the olefin moiety is the main site for ozone attack. Indeed, it was reported that the main oxidation products of carbamazepine resulted from an ozone attack at the olefin.[8] Based on Eq. S1† in Text S8†, the species-specific k_{O3} is calculated as 4.6×10^4 $M^{-1}s^{-1}$ (i.e., $4.5\times10^4+3.1\times10^2+3.1\times10^2$), which is lower than the experimental k_{O3} (~ 3×10^5 $M^{-1}s^{-1}$,[53]) by a factor of 6.5. Compared to a mean unsigned error (MUE) of 0.57 for log k_{O3} (corresponding to a factor of $10^{0.57}$=3.7) for the general k_{O3} prediction model for olefins (Table S3†), the prediction for carbamazepine is rather at the bad end of the model predictions.

pH-dependent k_{O3} prediction. Carbamazepine has no functional group undergoing acid-base speciation in the environmentally relevant pH-range. Thus, its reactivity with ozone is barely influenced by pH. However, there are numerous organic compounds with changing acid-base speciations in the environmentally relevant pH range. This may significantly modify their reactivity, eventually yielding differing transformation products. Therefore, a pH-dependent k_{O3} prediction has been implemented in the prediction platform. Upon selecting *pH-dependent prediction* (B) for a *query compound* (e.g., tramadol) in the main window, a new window pops up (Fig. 4). This window contains four panels, and to the right of them control tables are shown. The upper-left pK_a panel (F) shows the moieties undergoing acid-base speciation for which the predicted pK_a-values are presented in the pK_a table (J). The bottom-left panel (G) shows the acid-base microspecies for the *query compound* and relevant information is shown in the microspecies table (K) with the tautomeric fraction (*f*) in the last column. Both pK_a and *f* can be manually replaced if empirical or theoretical values with a better accuracy are available. By selecting '*Plot species distribution*', a species distribution (H) for the query compound is generated based on the given pK_a- and *f*-values. The middle-bottom panel (I) created by selecting '*Calculate and Plot*' is a plot for apparent site-specific k_{O3} ($k_{O3,site(j)\text{-}app(pH)}$ in Eq. S2†) for individual reactive sites for a user-defined pH range in Table (L). Apparent species-specific k_{O3} ($k_{O3,species(i)\text{-}app(pH)}$ in Eq. S3†) for all microspecies can be calculated at a user-defined pH in panel (M) by selecting '*Calculate*'. In the last column, the % contribution of $k_{O3,species(i)\text{-}app(pH)}$ for individual microspecies (*i*) to the overall apparent k_{O3} at the chosen pH is given. Upon selecting a microspecies and clicking '*Selection to Main Window*', the selected microspecies is ex-

ported to the main window where further k_{O3} predictions and pathway predictions for the forwarded microspecies can be implemented.

As shown in Figs. 5a and b, good agreements were obtained between the predicted overall apparent k_{O3} ($k_{O3,app}$) and the experimental $k_{O3,app}$ for tramadol within a factor of 1.2 – 5.8 and for triclosan within a factor of 1.3– 4.0, respectively. For tramadol, the predicted dominant sites for the reactions with ozone are the tertiary amine above and the benzene ring below pH 4.4, respectively. This is consistent with the experimental data, which shows the appearance of a plateau at pH < 5 (Fig. 5a). For triclosan, it was predicted that the ozone-phenol reaction would dominate over the whole pH range (Fig 5b).

Interestingly, the predicted k_{O3} (red dashed lines in Figs 5a and b) for the aromatic ring of tramadol and the dichlorobenzene ring of triclosan decreased with decreasing pH between pH 8 – 10. This effect is more significant in the case of tramadol. This is attributed to the fact that E_{HOMO} for the respective aromatic rings is lowered upon the protonation of the amino group and the hydroxyl group, respectively. For tramadol, one more factor seems to be involved in lowering k_{O3}: a hydrogen bond between the alcohol oxygen and the protonated amine hydrogen is predicted. The potential role of hydrogen bonds is discussed in detail for cetirizine in Text S10†. In brief, as the current prediction platform uses a gas-phase geometry for quantum chemical computations, a hydrogen bond present in the gas-phase geometry, which may be irrelevant in the aqueous medium under consideration here, may lead to an error in predicted k_{O3} values for certain micropollutants. However, a geometry relevant for aqueous phase has not been implemented in the current prediction system because it involves comprehensive evaluations of different approaches (e.g., the use of an aqueous-relevant conformer generator using molecular mechanics theory, *ab initio* geometry optimization in aqueous phase, etc), which will be implemented in the future.

In contrast to the reasonable predictions above, a poor prediction with more than two orders of magnitude difference compared to the measurements was observed for cetirizine (see Text S10† for more details). Despite various attempts such as the use of aqueous-phase relevant geometries and a consideration of enantiomers were made to improve the prediction, only minor improvements could be achieved. Over-

all, it seems that the current prediction model performs poorly for the piperazinyl moiety of cetirizine, supported by unsubstituted piperazine (experimental k_{O3} = 2.6×10^4 M^{-1}s^{-1} and predicted k_{O3} = 1.2×10^6 M^{-1}s^{-1}). However, it is not known whether it generally fails for (un)substituted piperazinyl groups because empirical k_{O3} data for this class of compounds is limited.

In addition, it is worthwhile mentioning two limitations of the current k_{O3} prediction model identified during development and testing with micropollutants, which need to be carefully taken into account if relevant cases appear. Firstly, the assignment of $E_{HOMO\text{-}n}$ to the corresponding reactive site can be ambiguous for certain micropollutants (e.g., carbamazepine, see Text S7† for more details). This is due to the fact that HOMO-n delocalizes not only over the site known to react with ozone but also other sites that are unreactive with ozone (Text S7†). In contrast, E_{NBO} is straightforward to assign to the corresponding reactive sites (i.e, olefins or amines) because the NBO is predominantly localized on the reactive site of interest. To avoid undesired assignement of $E_{HOMO\text{-}n}$, careful examinations of the calculated SS_{MOcoef} (sum of squares of molecular orbital coefficients) as well as visual inspection of HOMO-n assisted by a visualization software such as Molden[54] are proposed (for details refer to Text S7†). The examination needs to be conducted manually in the current platform. Secondly, the increase of k_{O3} with increasing pH, which is predicted for micropollutants with reactive sites being in the vicinity of functional groups such as amines, amides, or alcohols deprotonating to an anionic form, may be inaccurate. It is because the effect of their deprotonation on the predicted k_{O3} has not been calibrated in the developed k_{O3} prediction models due to a lack of empirical k_{O3} values for the deprotonated species. 4-Formylaminoantipyrine is given as an example in Fig. S11b† with more details in the caption.

Pathway enumeration

By selecting a reactive site for the pathway prediction and choosing '*Predict reaction pathways*' (D) in Fig. 3, a pathway enumeration is carried out. Based on the k_{O3} prediction described above, it was found that the olefin and the amine moieties are the dominant reactive sites with ozone at pH 7 for carbamazepine or tramadol, respectively. The corresponding predicted pathways for carbamazepine and tramadol are presented in Figs. 6 or 7, respectively. Note that raw pathway outputs were re-arranged for an appro-

priate representation in Figs. 6 and 7 with no changes in the contents. As shown in Fig. 6, carbamazepine underwent a typical Criegee mechanism (otoc001-otoc005) (Scheme S11†), giving rise to a Criegee product (P-2) following H_2O_2 elimination. Additionally, a partial oxidation (otop004 and otop005) giving rise to the product (P-1) has also been proposed (see Text S9†). Pathways with dashed arrows are considered less likely, based on the reference information and authors' judgment. Prior to H_2O_2 elimination, a Bayer-Villiger type reaction (otpost001-otpost002) may occur to form an ester (P-4) (Text S9† and Scheme S17† for details), which is of minor importance (dashed arrows). Upon an intramolecular reaction (otpost004-otpost006), the Criegee product (P-2) transforms into P-3, 1-(2-benzaldehyde)-4-hydro-(1*H*,3*H*)-quinazoline-2-one (BQM), which was experimentally found to be the major oxidation product of carbamazepine during ozonation.[8] Note that the predicted BQM from ozonation of carbamazepine is not an external validation but only a confirmation for the correct definition of the reaction rules because the intramolecular reactions (otpost004-otpost006) were defined based on the reference study for carbamazepine.[8] To our knowledge, there is no other empirical data available to demonstrate these intramolecular pathways. Therefore, their general applicability remains to be shown. For ozonation of tramadol (Fig. 7), tramadol *N*-oxide (P-5), *N*-desmethyl-tramadol (P-6), and the *N*-deaminated-tramadol (P-10) were predicted to be the major products. This is consistent with the experimental observation with the exception that deamination was of minor importance in the experiments.[32] Minor products (P-8, P-9, P-12, and P-13) formed via the decay of tetroxide intermediates were also considered (dashed arrows). In contrast to carbamazepine, the reaction rules applied to tramadol were defined based on data for other reference compounds (Text S9†) and found to be successfully applicable to tramadol. In that sense, the predicted tramadol pathway can be considered as a validation by external data. Overall, predicted transformation products undetected in experimental studies can be false positives or result from analytical limitation. The mass balance is often incomplete in experimental studies because of the limited analytical capability, indicating an only partial elucidation of transformation products. Transformation products that are considered to be stable, and thus potentially relevant for empirical detection, can be highlighted by selecting '*Predict transformation products*' (E) (Fig. 3). If doing so, only the products in blue boxes in Figs. 6 and 7 will be presented for carbamazepine or tramadol, respectively.

Conclusions and perspectives

A computer-based system for predicting kinetics and pathways for the reaction of ozone with micropollutants was developed in this study. Kinetic information (k_{O3}) for a micropollutant in question can be predicted with molecular orbital energy-based k_{O3} prediction models using quantum chemical computations. Reaction pathways and the concomitant transformation products can be predicted by applying generalized reaction pathway rules defined based on empirical reaction mechanisms reported in the scientific literature. The developed prediction platform can serve various purposes. As demonstrated above, the predicted k_{O3} values can be used to identify the predominant site for ozone attack, which enables targeted pathway predictions. Moreover, k_{O3} values derived based on Eqs. S1† or S4† can be used for estimating the elimination efficiency of micropollutants during ozonation with the theoretical/methodological details given elsewhere.[1,2,4,25] Environmental engineers or treatment plant operators can utilize the predicted k_{O3} for the optimization/design of ozone treatment processes for abating micropollutants. Pathway prediction can further be used as a tool to support screening for suspect ozone transformation products using high-resolution mass spectrometry. Moreover, the predicted transformation products can be screened for their toxicity using available *in silico* toxicity prediction tools. Thus, environmental chemists and toxicologists are considered another group of end-users who would benefit from the predictions of the system presented. However, it should be highlighted that the current prediction system has been established based on the knowledge and information available to date. Therefore, the prediction should be considered as a complimentary tool rather than a substitute for experimental studies, which are still required to verify prediction outputs and improve the performance of the *in-silico* tool. Some further improvements to be pursued with priority in the future are highlighted below. Firstly, the scope of the current prediction platform is exclusive for ozone. Pathway predictions are possible for any micropollutant regardless of their k_{O3} provided that the corresponding reaction pathway rules are available. However, as k_{O3} for a micropollutant decreases, the contribution of hydroxyl radicals to its abatement increases in the absence of a radical scavenger, leading to transformation products formed from reactions with hydroxyl radicals rather than with ozone. Therefore, further efforts will be made to include reactions with hydroxyl radicals in the prediction system. Secondly, for a few aforementioned limitations in predicting k_{O3} (e.g., ambiguous

E_{HOMO-n} assignment, the use of a gas phase geometry, and a poor k_{O3} prediction for a certain chemical class (e.g., piperazinyl group)), appropriate improvements are necessary which may require different approaches (e.g., exploring other quantum molecular descriptors). Moreover, further experimental investigations with more micropollutants to generate a database for external validation and with micropollutants containing yet unexplored chemical structures (e.g., the ones with k_{O3} estimates in Table S1†) are recommended. The prediction platform is currently available as a standalone GUI application that can be installed on a personal computer. The development of a web-based prediction platform adapted from the developed GUI application, to which anyone has free access without the need for installing any software/application, will be pursued in future.

Acknowlegements

This study was financially supported by the Swiss Federal Office for the Environment (FOEN). The authors would like to thank Holger Lutze (Universität Duisburg-Essen) and Peter R. Tentscher (Eawag) for fruitful discussions. We thank Chemaxon Ltd. and the ORCA development team at the Max-Planck-Institute for Chemical Energy Conversion for providing academic licenses for MarvinBeans/JChem package and ORCA, respectively.

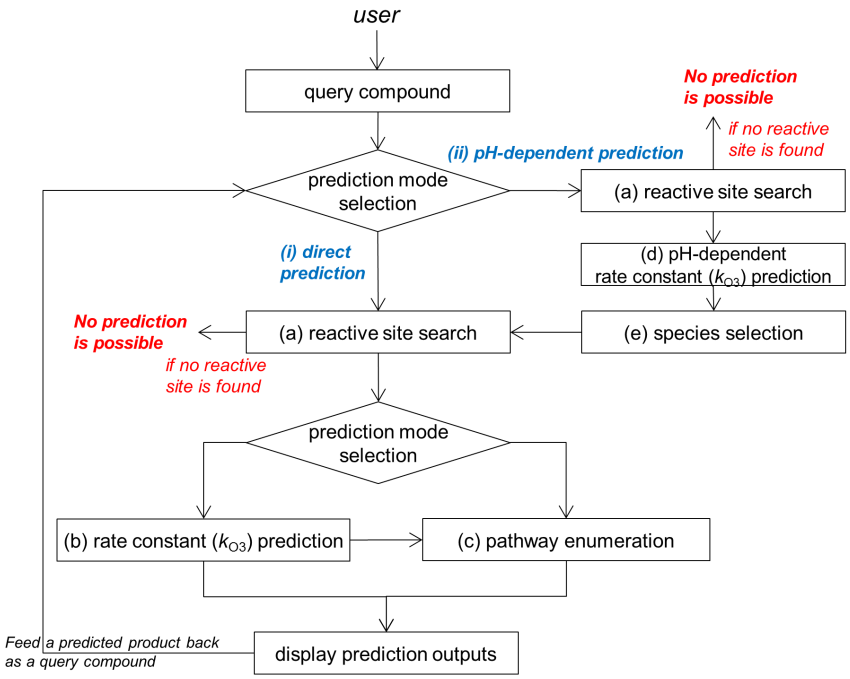


Fig. 1 Workflow of the developed prediction platform to predict kinetics and transformation products for the reactions of ozone with organic compounds.

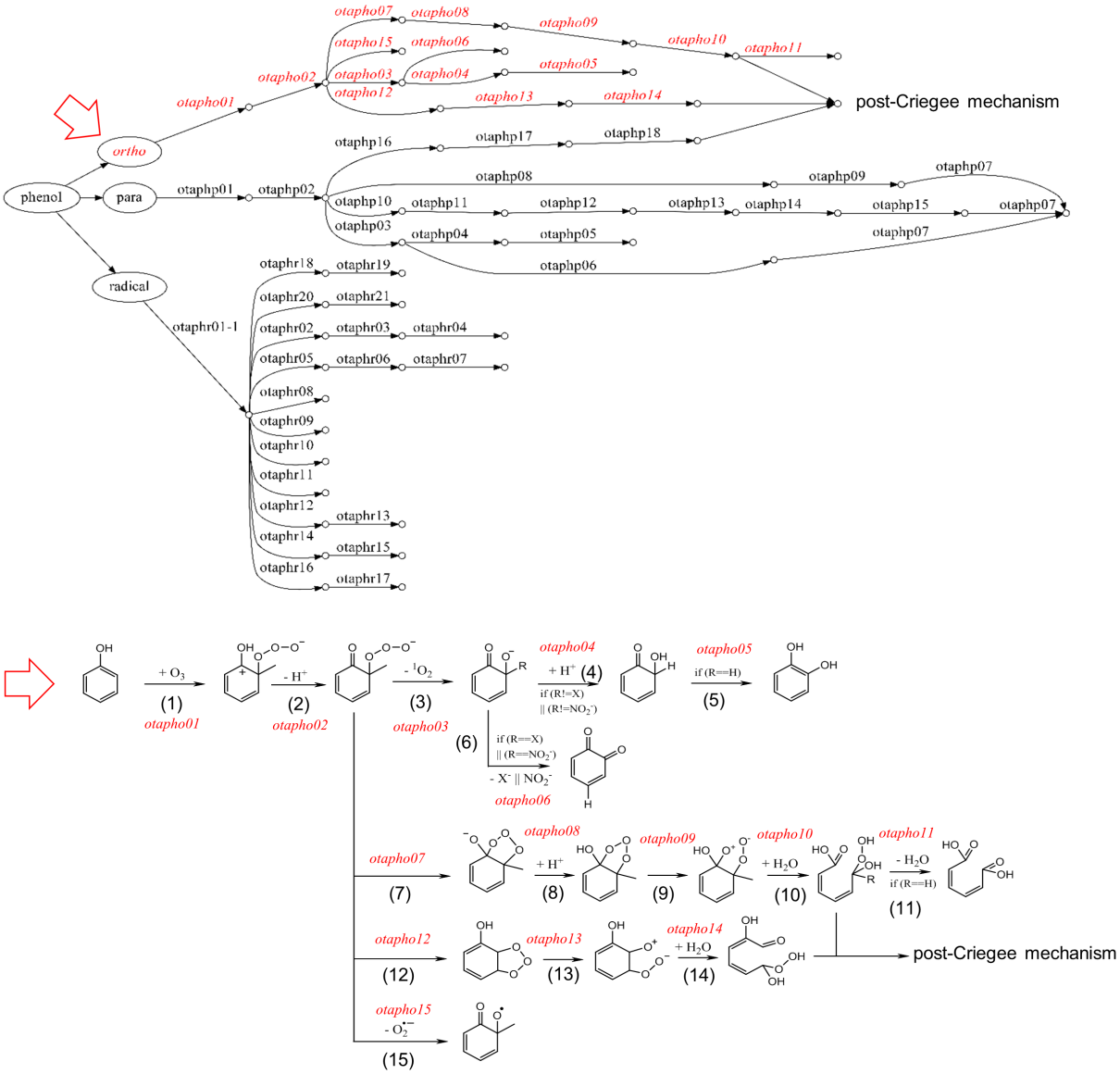


Fig. 2 Reaction pathway enumeration (top) for a phenolic moiety and the reaction pathways (bottom from Scheme S1†) defined for the ortho position of phenol reacting with ozone. The corresponding pathways for "para" and "radical" are given in Schemes S2† and S3†. Logical operators such as '==' (is equal 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. The pathways for the post-Criegee mechanism, which deals with various ensuing reactions such as hydrolysis and fragmentations, are presented in Scheme S11†. Note that further pathway predictions for muconic-type products is to be implemented by feeding such products back as an input structure.

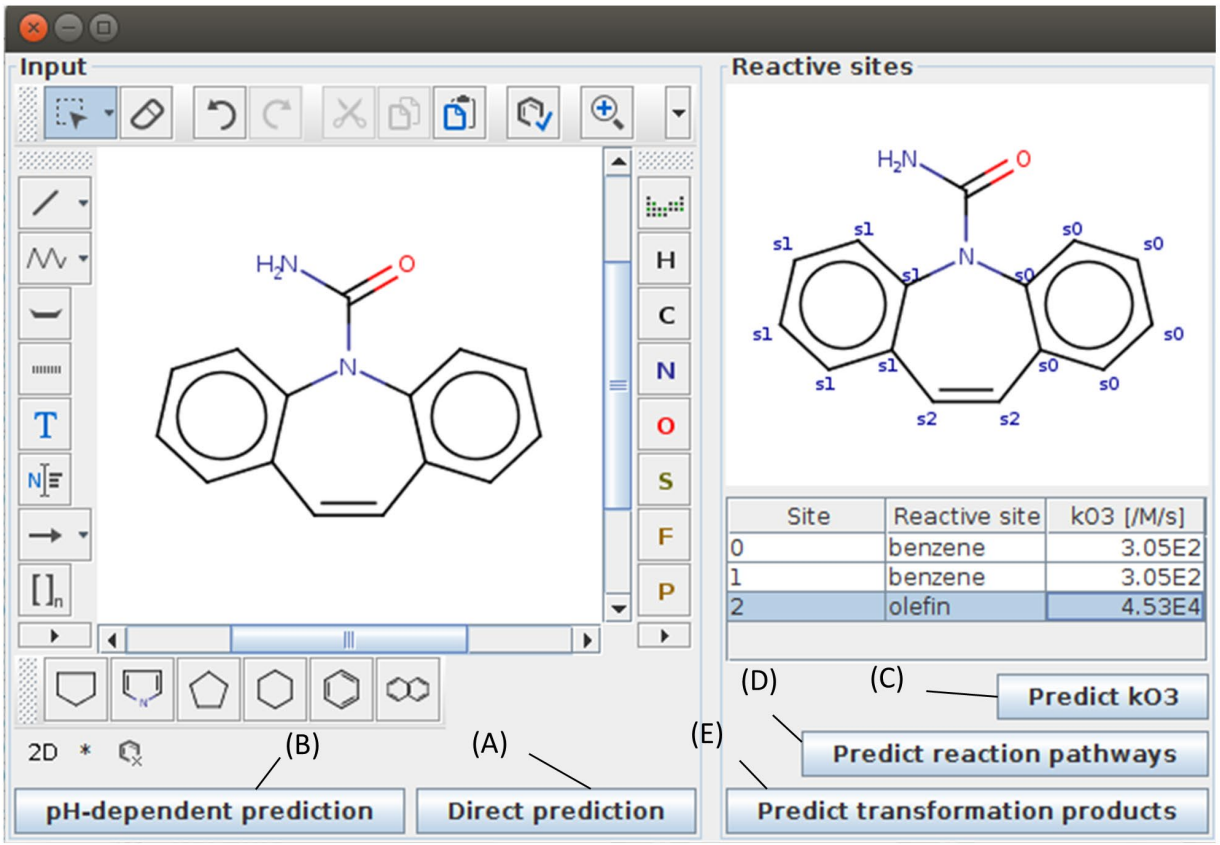


Fig. 3. The main window for the developed kinetics and pathway prediction system. In the left panel of the main window, the user can choose between (A) direct prediction and (B) pH-dependent prediction. By selecting A, the right panel of the main window is activated and further predictions, (C), (D), or (E), can be conducted.

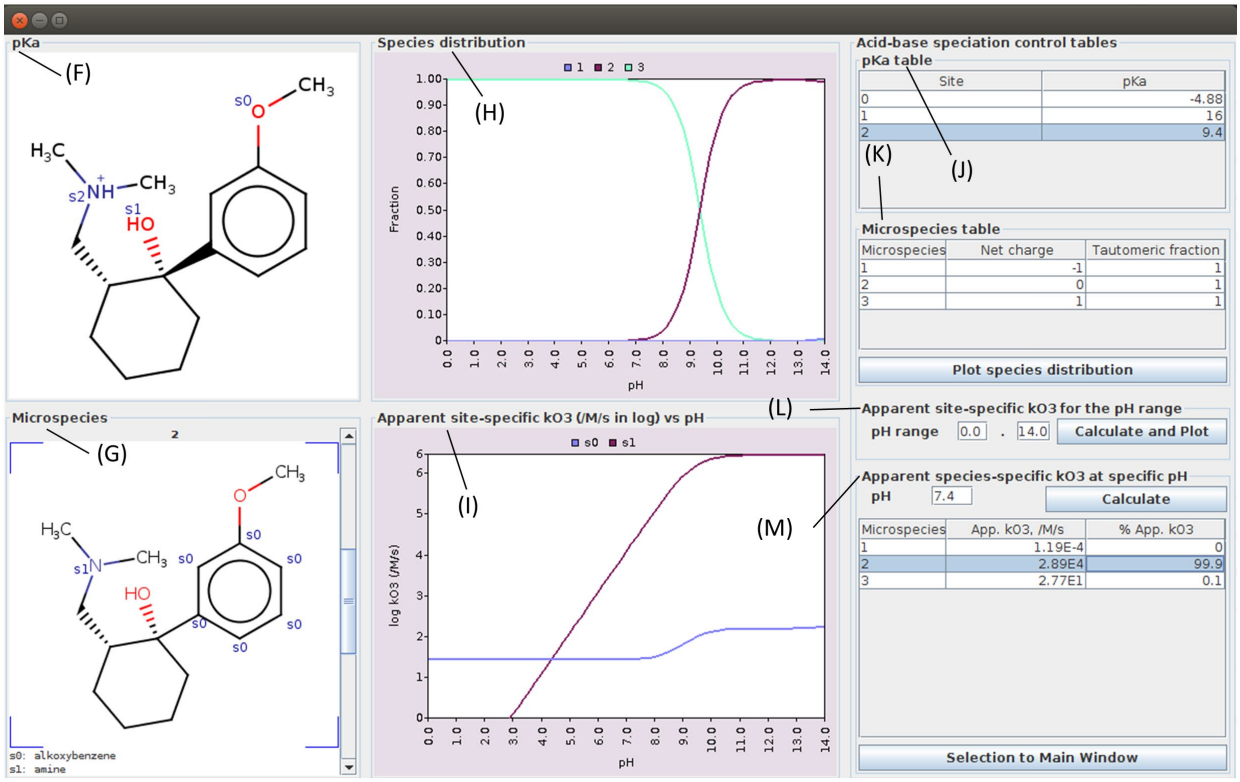
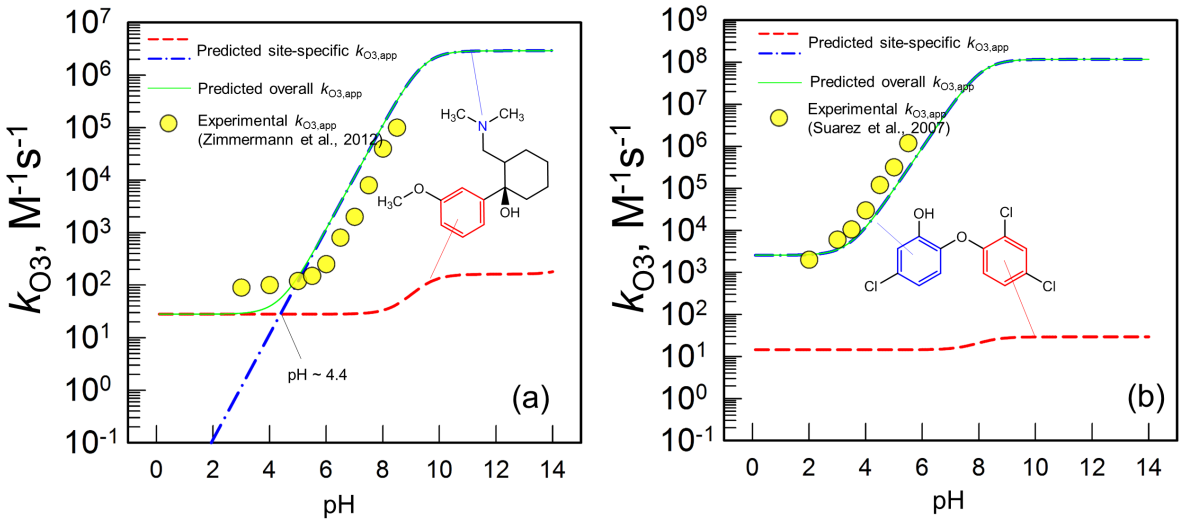


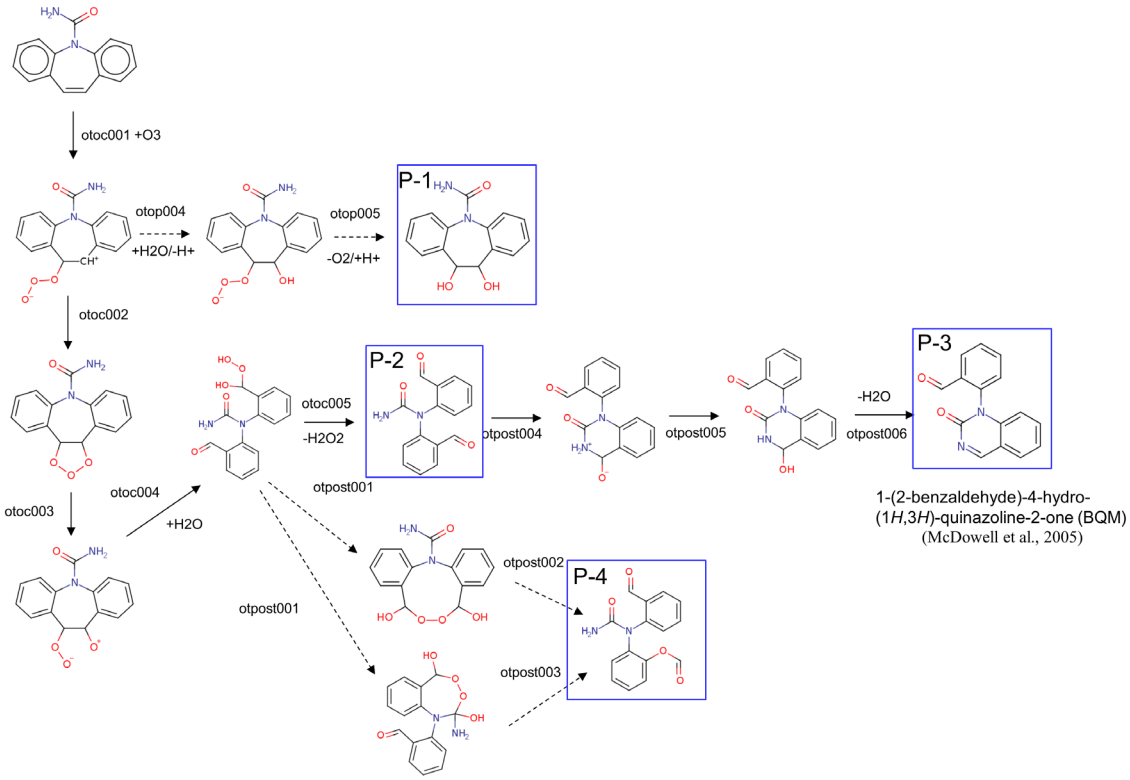
Fig. 4. Window for the pH-dependent prediction of the kinetics of tramadol oxidation by ozone. Panel (F) shows the identified ionizable groups of the *query compound* for which pK_a-values are given in the Table (J). Panel (G) shows individual species of the *query compound* and the identified reactive sites. Panel (H) is a species distribution plot created by choosing 'Plot species distribution' in the control Table (K). The species distribution is drawn based on the presented pK_a (J) and tautomeric fraction(*f*) (K). The pH-dependent site-specific k_{O3} prediction can be implemented for a user-defined pH-range (L), and is presented in panel (I). In Table (M), apparent species-specific k_{O3} are predicted at a specific user-defined pH and the user can export (by 'Selection to Main Window') the species of interest to the main window in Fig. 3 for further predictions. While the pK_a-value of -4.9 for the methoxy group was used as predicted by Marvin (Chemaxon. Ltd), the predicted pK_a-values of 9.2 and 13.8 for the amino and the alcohol group, respectively, were replaced by 9.4[55] (experimental) and 16 (typical pK_a range for an alcohol group), respectively.



417

418 Fig. 5. Comparison between predicted and experimental k_{O3} as a function of pH for (a) tramadol and (b) triclosan.
419 The experimental data for tramadol and triclosan are from Zimmermann et al., 2012[32] and Suarez et al., 2007,[56]
420 respectively.

421



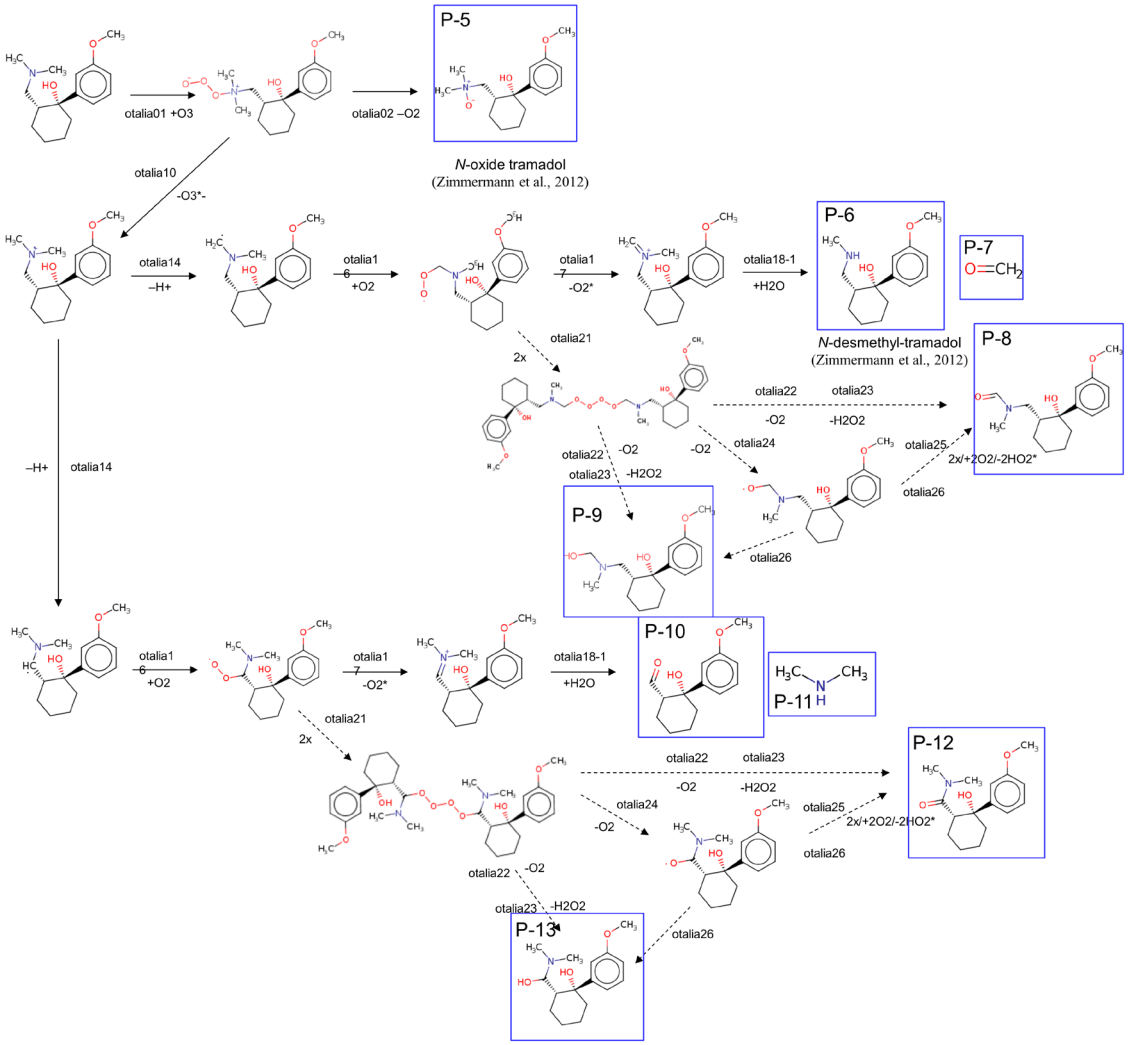
422

423 Fig. 6. Predicted pathways for the reactions of carbamazepine with ozone. The dashed arrows indicate that the
424 pathways are less likely. Predicted potential final products are presented in blue boxes. Product (P-3) was re-
425 portedas the major transformation product in experiments.[8]

426

427

428



429

430 Fig. 7. Predicted pathways for the reactions of tramadol with ozone. The dashed arrows indicate that the pathways
431 are less likely. Predicted potential final transformation products are presented in blue boxes. The products P-5 and
432 P-6 were reported to be the major experimental transformation products.

433

434

References

1 M. Lee, S. G. Zimmermann-Steffens, J. S. Arey, K. Fenner and U. von Gunten, *Environ. Sci. Technol.*, 2015, **49**, 9925–9935.

2 C. von Sonntag and U. von Gunten, *Chemistry of ozone in water and wastewater treatment: From basic principles to applications*, IWA publishing, 2012.

3 Y. Lee, L. Kovalova, C. S. McArdell and U. von Gunten, *Water Res.*, 2014, **64**, 134–148.

4 Y. Lee, D. Gerrity, M. Lee, A. E. Bogeat, E. Salhi, S. Gamage, R. A. Trenholm, E. C. Wert, S. A. Snyder and U. von Gunten, *Environ. Sci. Technol.*, 2013, **47**, 5872–81.

5 M. M. Huber, A. Göbel, A. Joss, N. Hermann, D. Löffler, C. S. McArdell, A. Ried, H. Siegrist, T. A. Ternes and U. von Gunten, *Environ. Sci. Technol.*, 2005, **39**, 4290–4299.

6 J. Hollender, S. G. Zimmermann, S. Koepke, M. Krauss, C. S. McArdell, C. Ort, H. Singer, U. von Gunten and H. Siegrist, *Environ. Sci. Technol.*, 2009, **43**, 7862–9.

7 M. M. Sein, M. Zedda, J. Tuerk, T. C. Schmidt, A. Golloch and C. von Sonntag, *Environ. Sci. Technol.*, 2008, **42**, 6656–6662.

8 D. C. McDowell, M. M. Huber, M. Wagner, U. von Gunten and T. A. Ternes, *Environ. Sci. Technol.*, 2005, **39**, 8014–8022.

9 D. B. Mawhinney, B. J. Vanderford and S. A. Snyder, *Environ. Sci. Technol.*, 2012, **46**, 7102–11.

10 M. M. Huber, T. A. Ternes and U. von Gunten, *Environ. Sci. Technol.*, 2004, **38**, 5177–5186.

11 M. C. Dodd, D. Rentsch, H. P. Singer, H.-P. E. Kohler and U. von Gunten, *Environ. Sci. Technol.*, 2010, **44**, 5940–5948.

12 H. Mestankova, K. Schirmer, B. I. Escher, U. von Gunten and S. Canonica, *Environ. Pollut.*, 2012, **161**, 30–35.

13 H. Mestankova, B. Escher, K. Schirmer, U. von Gunten and S. Canonica, *Aquat. Toxicol.*, 2011, **101**, 466–473.

14 Y. Lee, B. I. Escher and U. von Gunten, *Environ. Sci. Technol.*, 2008, **42**, 6333–6339.

15 H. Mestankova, A. M. Parker, N. Bramaz, S. Canonica, K. Schirmer, U. von Gunten and K. G. Linden, *Water Res.*, 2016, **93**, 110–120.

16 J. Reungoat, B. I. I. Escher, M. Macova, F. X. X. Argaud, W. Gernjak and J. Keller, *Water Res.*, 2012, **46**, 863–72.

17 J. Margot, C. Kienle, A. Magnet, M. Weil, L. Rossi, L. F. de Alencastro, C. Abegglen, D. Thonney, N. Chèvre, M. Schärer and D. A. Barry, *Sci. Total Environ.*, 2013, **461**–**462**, 480–498.

18 B. I. Escher, N. Bramaz and C. Ort, *J. Environ. Monit.*, 2009, **11**, 1836.

19 N. Cao, M. Yang, Y. Zhang, J. Hu, M. Ike, J. Hirotsuji, H. Matsui, D. Inoue and K. Sei, *Sci. Total Environ.*, 2009, **407**, 1588–1597.

20 Y. Schindler Wildhaber, H. Mestankova, M. Schärer, K. Schirmer, E. Salhi and U. von Gunten, *Water Res.*, 2015, **75**, 324–335.

21 D. Stalter, A. Magdeburg, M. Weil, T. Knacker and J. Oehlmann, *Water Res.*, 2010, **44**, 439–448.

22 D. Stalter, A. Magdeburg and J. Oehlmann, *Water Res.*, 2010, **44**, 2610–2620.

23 A. Magdeburg, D. Stalter and J. Oehlmann, *Chemosphere*, 2012, **88**, 1008–1014.

24 U. von Gunten, *Water Res.*, 2003, **37**, 1443–67.

25 Y. Lee and U. von Gunten, *Environ. Sci. Water Res. Technol.*, 2016, **2**, 421–442.

26 Y. Lee and U. von Gunten, *Water Res.*, 2012, **46**, 6177–6195.

27 E. Barron, M. Deborde, S. Rabouan, P. Mazellier and B. Legube, *Water Res.*, 2006, **40**, 2181–9.

28 E. Mvula and C. von Sonntag, *Org. Biomol. Chem.*, 2003, **1**, 1749–1756.

29 M. K. Ramseier and U. von Gunten, *Ozone Sci. Eng.*, 2009, **31**, 201–215.

30 M. Deborde, S. Rabouan, P. Mazellier, J.-P. Duguet and B. Legube, *Water Res.*, 2008, **42**, 4299–4308.

31 E. Mvula, S. Naumov and C. von Sonntag, *Environ. Sci. Technol.*, 2009, **43**, 6275–6282.

32 S. G. Zimmermann, A. Schmukat, M. Schulz, J. Benner, U. von Gunten and T. A. Ternes, *Environ. Sci. Technol.*, 2012, **46**, 876–84.

33 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.

34 J. Benner and T. A. Ternes, *Environ. Sci. Technol.*, 2009, **43**, 5086–5093.

35 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.
36 *MarvinBeans/JChem suite (16.2.29.0)*, Chemaxon(http://www.chemaxon.com), 2016.
37 D. Weininger, *J. Chem. Inf. Model.*, 1988, **28**, 31–36.
38 F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 73–78.
39 E. D. Glendening, C. R. Landis and F. Weinhold, *J. Comput. Chem.*, 2013, **34**, 1429–1437.
40 G. Schaftenaar and J. H. Noordik, *J. Comput. Aided. Mol. Des.*, 2000, **14**, 123–134.
41 J. James, C. A.; Weininger, D.; Delany, *Daylight Theory Manual SMARTS Theory*, http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html, 2000.
42 E. L. Schymanski, H. P. Singer, P. Longrée, M. Loos, M. Ruff, M. A. Stravs, C. Ripollés Vidal and J. Hollender, *Environ. Sci. Technol.*, 2014, **48**, 1811–1818.
43 J. S. Binkley, J. A. Pople and W. J. Hehre, *J. Am. Chem. Soc.*, 1980, **102**, 939–947.
44 M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro and W. J. Hehre, *J. Am. Chem. Soc.*, 1982, **104**, 2797–2803.
45 K. D. Dobbs and W. J. Hehre, *J. Comput. Chem.*, 1986, **7**, 359–378.
46 K. D. Dobbs and W. J. Hehre, *J. Comput. Chem.*, 1987, **8**, 861–879.
47 K. D. Dobbs and W. J. Hehre, *J. Comput. Chem.*, 1987, **8**, 880–893.
48 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.
49 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 553–586.
50 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 616–641.
51 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 520–552.
52 T. A. Halgren and R. B. Nachbar, *J. Comput. Chem.*, 1996, **17**, 587–615.
53 M. M. M. Huber, S. Canonica, G.-Y. Park and U. von Gunten, *Environ. Sci. Technol.*, 2003, **37**, 1016–1024.
54 G. Schaftenaar and J. H. Noordik, *J. Comput. Aided. Mol. Des.*, 2000, **14**, 123–134.
55 M. Pospíšilová, M. Polášek and V. Jokl, *J. Pharm. Biomed. Anal.*, 1998, **18**, 777–783.
56 S. Suarez, M. Dodd, F. Omil and U. von Gunten, *Water Res.*, 2007, **41**, 2481–2490.