Table S6: compilation of QSPRs for reaction with O3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Endpoint** | **Equation and/or model specifications** | **R2** | **Specifications of training/test set** | **References** |
| log(kO3) Second-order rate constants | log(kO3) = 6.863 - 5.818log AMW + 4.711nArOH, MLR | 0.681 | 22/33 pharmaceutically active compounds, hormones, EDCs, pesticides, flame retardants, etc. | Jin et al. (2012) |
| log(kO3) Second-order rate constants | log(kO3) = 6.658 - 7.574log(AMW) + 2.390nAB + 4.010nArOH, PLS | 0.728 | 22/33 pharmaceutically active compounds, hormones, EDCs, pesticides, flame retardants, etc. | Jin et al. (2012) |
| log(kO3) Second-order rate constants | log(kO3) = 2.906 + 1.033t2 - 0.878t3, PCA and stepwise MLR | 0.455 | 22/33 pharmaceutically active compounds, hormones, EDCs, pesticides, flame retardants, etc. | Jin et al. (2012) |
| log(kO3) Second-order rate constants | log(kO3)(<2) = 2.327 - 2.876log(AMW) and log(kO3(>-2) = 7.747 - 4.171log(AMW) + 2.382nArOH, stepwise MLR, classification and PLR | 0.964 | 22/33 pharmaceutically active compounds, hormones, EDCs, pesticides, flame retardants, etc. | Jin et al. (2012) |
| log(kO3) Second-order rate constants | log(kO3)(<2) = -0.534 - 0.097t2 - 0.31t3 and log(kO3)(>-2) = 4.612 + 0.486t2 - 1.158t3, PCR, classification and PLR | 0.929 | 22/33 pharmaceutically active compounds, hormones, EDCs, pesticides, flame retardants, etc. | Jin et al. (2012) |
| ln(kO3) Second-order rate constants | ln(kO3) = 2.940 - 47.079f(0)x, Several molecular descriptors calculated, multiple linear regression | 0.586 | 28/5 various compounds among dyes, fertilizers, pharmaceutical and refinery wastewater pH = 4 | Zhu et al. (2015) |
| ln(kO3) Second-order rate constants | ln(kO3) = 1.965 - 48.711f(0)x - 3.273q(C)min, Several molecular descriptors calculated, multiple linear regression | 0.682 | 28/5 various compounds among dyes, fertilizers, pharmaceutical and refinery wastewater pH = 4 | Zhu et al. (2015) |
| ln(kO3) Second-order rate constants | ln(kO3) = 1.978 - 95.484f(0)x - 3.350q(C)min + 38.221f(+)x, Several molecular descriptors calculated, multiple linear regression | 0.763 | 28/5 various compounds among dyes, fertilizers, pharmaceutical and refinery wastewater pH = 4 | Zhu et al. (2015) |
| ln(kO3) Second-order rate constants | ln(kO3) = 2.489 + 21.117f(0)n - 126.519f(0)x - 3.852q(C)min + 48.501f(+)x, Several molecular descriptors calculated, multiple linear regression | 0.802 | 28/5 various compounds among dyes, fertilizers, pharmaceutical and refinery wastewater pH = 4 | Zhu et al. (2015) |
| ln(kO3) Second-order rate constants | ln(kO3) = 43.765 + 0.195(DBE) - 0.28(WPSA) - 0.855(IP), Several molecular descriptors calculated, multiple linear regression | 0.832 | 55/28 pharmaceuticals, personal care products and organic solvents | Sudhakaran and Amy (2013), Mamy (2014) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = 2.29 × (EHOMO) + 24.15, LR, and QC HF/6-31G;  log(kO3) = 2.69 × (EHOMO) + 27.37, LR and QC HF/6-311++G\*\*;  log(kO3) = 2.40 × (EHOMO) + 18.88, LR and QC B3LYP/6-31G;  log(kO3) = 2.98 × (EHOMO) + 23.08, LR and QC B3LYP/6-311++G\*\* | 0.94; 0.94; 0.95; 0.94 | 35 phenols | Lee et al. (2015) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = 3.18 × (EHOMO) + 32.43, LR and QC HF/6-31G;  log(kO3) = 3.41 × (EHOMO) + 34.60, LR and QC HF/6-311++G\*\*;  log(kO3) = 3.28 × (EHOMO−n) + 24.84, LR and QC B3LYP/6-31G;  log(kO3) = 3.48 × (EHOMO−n) + 27.19, LR and QC B3LYP/6-311++G\*\* | 0.85; 0.86; 0.82;  0.84 | 16 anilines | Lee et al. (2015) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = 3.40 × (EHOMO) + 31.77, LR and QC HF/6-31G;  log(kO3) = 4.17 × (EHOMO) + 38.22, LR and QC HF/6-311++G\*\*;  log(kO3) = 3.80 × (EHOMO) + 25.78, LR and QC B3LYP/6-31G;  log(kO3) = 4.65 × (EHOMO) + 31.81, LR and QC B3LYP/6-311++G\*\* | 0.87;  0.9;  0.8;  0.8 | 17 mono- and dialkoxybenzenes | Lee et al. (2015) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = 2.77 × (EHOMO) + 30.09, LR and QC HF/6-31G;  log(kO3) = 2.70 × (EHOMO) + 29.34, LR and QC HF/6-311++G\*\*;  log(kO3) = 2.34 × (EHOMO) + 20.10, LR and QC B3LYP/6-31G;  log(kO3) = 2.17 × (EHOMO) + 19.56, LR and QC B3LYP/6-311++G\*\* | 0.95;  0.98;  0.997;  0.99 | 4 trimethoxybenzenes | Lee et al. (2015) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = 2.12 × (EHOMO) + 20.12, LR and QC HF/6-31G;  log(kO3) = 2.45 × (EHOMO) + 23.13, LR and QC HF/6-311++G\*\*;  log(kO3) = 2.20 × (EHOMO−n) + 15.58, LR and QC B3LYP/6-31G;  log(kO3) = 2.29 × (EHOMO−n) + 16.64, LR and QC B3LYP/6-311++G\*\* | 0.82;  0.81;  0.74;  0.65 | 40 benzene derivatives | Lee et al. (2015) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = 1.13 × (ENBO,C−C(π)) + 16.57, LR and QC HF/6-31G;  log(kO3) = 1.32 × (ENBO,C−C(π)) + 18.54, LR and QC HF/6-311++G\*\*;  log(kO3) = 1.37 × (ENBO,C−C(π)) + 15.14, LR and QC B3LYP/6-31G;  log(kO3) = 1.67 × (ENBO,C−C(π)) + 17.74, LR and QC B3LYP/6-311++G\*\* | 0.84;  0.82;  0.85;  0.84 | 45 olefins | Lee et al. (2015) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = 0.77 × (ENBO,LP‑N) + 14.26, LR and QC HF/6-31G;  log(kO3) = 0.83 × (ENBO,LP‑N) + 15.80, LR and QC HF/6-311++G\*\*;  log(kO3) = 0.65 × (ENBO,LP‑N) + 10.02, LR and QC B3LYP/6-31G;  log(kO3) = 0.85 × (ENBO,LP‑N) + 12.35, L B3LYP/6-311++G\*\*; | 0.76;  0.83;  0.67;  0.81 | 59 aliphatic amines | Lee et al. (2015) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 10.8 + 0.92EHOMO, semi-empirical QC method and LR | 0.84 | 24 several pesticides | Hu et al. (2000), Mamy (2014) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 0.79 EHOMO + 9.2, semi-empirical QC method and LR | 0.998 | 4 heterocyclic N-pesticides | Hu et al. (2000), Mamy (2014) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 0.37 EHOMO + 5.9, semi-empirical QC method and LR | 0.91 | 8 organonitrogen pesticides | Hu et al. (2000), Mamy (2014) |
| log(kO3) Second-order rate constants (kO3) | log k(O3)= 0.97 EHOMO + 11.3, semi-empirical QC method and LR | 0.92 | 4 phenolic pesticides | Hu et al. (2000), Mamy (2014) |
| log(kO3) Second-order rate constants (kO3) | log k(O3)= -0.116 EHOMO + 0.351 ELUMO + 1.510, semi-empirical QC method and LR | 0.97 | 8 phenoxyalkylacitic pesticides | Hu et al. (2000), Mamy (2014) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 3.53 (+-0.25) - 3.24 (+-0.69) SIGMA(s+o,m,p), linear regression | 0.81 | 24 non-dissociated phenols | Lee and von Gunten (2012) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 8.80 (+-0.16) - 2.27 (+-0.30) SIGMA(s+o,m,p), linear regression | 0.96 | 13 Dissociated phenols | Lee and von Gunten (2012) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 7.15 (+-0.25) - 1.54 (+-0.42) SIGMA(s -o,m,p), linear regression | 0.85 | 14 Anilines | Lee and von Gunten (2012) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = -0.04 (+-0.38) - 3.35 (+-0.26) SIGMA(s+p), linear regression | 0.93 | 50 Benzene derivatives | Lee and von Gunten (2012) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 6.18 (+-0.13) - 0.49 (+-0.03) SIGMA(s\*), linear regression | 0.86 | 48 Olefins | Lee and von Gunten (2012) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 6.13 (+-0.21) - 1.00 (+-0.12) SIGMA(s\*), linear regression | 0.86 | 54 Amines and amine derivatives | Lee and von Gunten (2012) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 3.19 (+-0.57) + 0.76 (+-0.11) logkaromatic (ClO2), linear regression | 0.94 | 16 Aromatics | Lee and von Gunten (2012) |
| log(kO3) Second-order rate constants (kO3) | log k(O3) = 4.50 (+-0.38) + 0.96 (+-0.11) logkaromatic (HOCl), linear regression | 0.94 | 24 Aromatics | Lee and von Gunten (2012) |
| log(kO3) reaction rate constants | log k(O3) = 3.760 + 0.017CMA;  log k(O3) = 4.325 - 0.018CMA - 2.623qH+;  log k(O3) = 4.656 + 0.015CMA - 1.684ELUMO - 3.057qH+;  log k(O3) = 4.881 + 0.015CMA - 2.970ELUMO - 3.397qH+ - 0.0260mu, several descriptors and regression, (PLS) regression | 0.627;  0.768;  0.791;  0.8 | 39 aromatic compounds | Jiang et al. (2010) |
| log(kO3) Second-order rate constants (kO3) | log(kO3) = a – 8.0s | - | 9 (CH3)n-C6H5-n-OH compounds | Gurol and Nekoulnalnl 1983, Jin et al |
| log(kO3) Second-order rate constants | log(kO3) = a – 3.1s+ | - | 7 Xn-C6H6-n compounds | Hoigné and Bader 1983a, Jin et al |
| log(kO3) Second-order rate constants | log(kO3) = a – 2.81s | 0.988 | 3 Xn- C6H5-n-C3H7ON2 compounds | Benitez et al., 2007, Jin et al |
| log(kO3) Second-order rate constants | log(kO3) = 8.9 – 2.4SIGMA(s+), linear regression | 0.96 | 13 Xn-C6H5-n-OH compounds (anionic species), triclosan and other substituted phenols | Suarez et al., 2007, Jin et al |
| log(kO3) Second-order rate constants | log(kO3) = 3.4 – 3.4SIGMA(s+), linear regression | 0.94 | 7 Xn-C6H5-n-OH compounds (neutral species), triclosan and other substituted phenols | Suarez et al., 2007, Jin et al |
| KappO3 apparent second-order rate constant at a given pH | kappO3 = k1a + k2(1 - a) a=1/(1+(10^-pKa/10^-pH)) | - | Triclosan (1) | Suarez et al. (2007) |
| log(kO3) Second-order rate constants | log k(O3) = 9.551 + 10.248 q-, MLR | 0.535824 | 20/6 substituted phenols | Liu et al. (2010) |
| log(kO3) Second-order rate constants | log k(O3) = 8.034 + 7.700 q- + 0.007 α, MLR | 0.638401 | 20/6 substituted phenols | Liu et al. (2010) |
| log(kO3) Second-order rate constants | log k(O3) = 9.751 + 9.154 q- + 0.015 α - 0.005 S00, MLR | 0.7569 | 20/6 substituted phenols | Liu et al. (2010) |
| log(kO3) Second-order rate constants | log k(O3) = 8.769 + 5.017 q- + 0.0171 α - 0.007 S - 3.581 ELUMO, MLR | 0.826281 | 20/6 substituted phenols | Liu et al. (2010) |
| log(kO3) Second-order rate constants | log k(O3) = 9.124 + 6.195 q- + 0.016 α - 0.007 S• - 3.090 ELUMO, MLR | 0.842724 | 20/6 substituted phenols | Liu et al. (2010) |
| ln(kO3) Second-order rate constants | ln k(O3) = 0.899 - 44.713f(+)n, Several molecular descriptors calculated, multiple linear regression | 0.562 | 27/5 various compounds among dyes, fertilizers, pharmaceutical and refinery wastewater pH = 7 ± 0.2 | Zhu et al. (2014) |
| ln(kO3) Second-order rate constants | ln k(O3) = 0.050-46.231f(+)n + 2.632q(C)x, Several molecular descriptors calculated, multiple linear regression | 0.649 | 27/5 various compounds among dyes, fertilizers, pharmaceutical and refinery wastewater pH = 7 ± 0.2 | Zhu et al. (2014) |
| ln(kO3) Second-order rate constants | ln k(O3) = 2.452–46.811f(+)n + 3.517q(C)x + 11.734EHOMO, Several molecular descriptors calculated, multiple linear regression | 0.723 | 27/5 various compounds among dyes, fertilizers, pharmaceutical and refinery wastewater pH = 7 ± 0.2 | Zhu et al. (2014) |
| log(KappO3) Apparent rate constant | log kHOCl = 0.99 (± 0.08) log kO3 – 4.47 (± 0.63), linear regression | 0.964 | 24 aromatic compounds (mostly phenols) | Deborde and von Gunten (2008) |
| KappO3 apparent second-order rate constant at a given pH | kappO3 = k1a + k2(1 - a) a=1/(1+(10^-pKa/10^-pH)), linear regression | - | Metoprolol, Atenolol, Acetbutolol | Brenner et al. (2009) |
| % Removal efficiency | Removal efficiency(%) = 72.46(+-11.8) + 18.55(+-2.78)EHOMO-1 + 12.42(+-1.48)ELUMO+1 - 46.37(+-5.58)Hardness - 22.04(+-2.88)AEI, Computed molecular descriptors were combined with physical properties, multiple linear regression algorithm | 0.994 | seven estrogens (17a-estradiol, 17b-estradiol, 17a-dihydroequilin, 17a-ethinyl estradiol, estriol, estrone and equilin) | Rokhina et al. (2012) Criticized by Sabljic (2013) |
| % Removal efficiency | Removal efficiency(%) = 83.95(+-1.28) + 0.32(+-0.03)ELUMO+1 + 0.00265(+-0.0027)MW - 0.79407(+-0.093)EHOMO + 1.59(+-0.099)ELUMO, Computed molecular descriptors were combined with physical properties, multiple linear regression algorithm | 0.997 | five progestins (levonorgestrel, gestodene, trimegestrone, medrogestone and progesterone) | Rokhina et al. (2012) Criticized by Sabljic (2013) |
| % removal by ozonation | % Removal=67.3+0.0506 (Pi-surface area)+5.2 (#metabolites)+4.34 (#rtvFG)−0.114 (WPSA), Monte Carlo (MC) statistical mechanics simulations to generate 3D molecular descriptors and multiple linear regression analysis | 0.858 | 62 range of compounds including APIs, herbicides and pesticides | Lei and Snyder (2007), Sudhakaran et al. (2013) |
| % percentage removed by chlorination (HOCl/OCl) | % chlorine removal = 106.8 + 0.791%(ozone removal) + 7.89(#rtvFG) + 4.80(QPlog Pow) + 0.175(FISA) - 15.0(IP), several 3D molecular descriptors and physicochemical properties for the development of multiple linear regression analysis | 0.705 | 73 Endocrine-disrupting compounds (EDCs) and pharmaceuticals and personal care products (PPCPs) | Lei and Snyder (2007) |
| ln(removal%AOP) Removal efficiency by O3/H2O2 | ln rml(AOP) = 9.77–0.63 (ELUMO–EHOMO) - 0.194 (EA) + 0.02 (#ring atoms), Several molecular descriptors calculated, multiple linear regression | 0.902 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in CRW (Colorado River) | Sudhakaran et al. (2012) |
| ln(removal%O3) Removal efficiency by O3 | ln rml(O3) = 12.45–0.95 (ELUMO–EHOMO) - 0.32 (MON), Several molecular descriptors calculated, multiple linear regression | 0.866 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in CRW (Colorado River) | Sudhakaran et al. (2012) |
| removal%AOP Removal efficiency by O3/H2O2 | %rml (AOP) = 318.23–26.52 (ELUMO–EHOMO) - 6.11 (EA) + 0.41 (#ring atoms), Several molecular descriptors calculated, multiple linear regression | 0.922 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in ORW (Ohio River) | Sudhakaran et al. (2012) |
| removal% (O3) Removal efficiency by O3 | %rml (O3) = 310.31–25.91 (ELUMO–EHOMO)–8.64 (EA) - 2.11 (#X) + 0.66 (#ring atoms), Several molecular descriptors calculated, multiple linear regression | 0.915 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in ORW (Ohio River) | Sudhakaran et al. (2012) |
| ln(removal%AOP) Removal efficiency by O3/H2O2 | ln rml(AOP) = 15.17–1.33 (ELUMO–EHOMO) - 0.56 (EA) + 0.06 (#in56), Several molecular descriptors calculated, multiple linear regression | 0.862 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in PRW (Passaic River) | Sudhakaran et al. (2012) |
| ln(removal%O3) Removal efficiency by O3 | ln rml(O3) = 18.15–1.63 (ELUMO–EHOMO) - 0.33 (MON), Several molecular descriptors calculated, multiple linear regression | 0.887 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in PRW (Passaic River) | Sudhakaran et al. (2012) |
| removal%AOP Removal efficiency by O3/H2O2 | %rml (AOP) = 286.67–22.41 (ELUMO–EHOMO) - 52.68 (O/C), Several molecular descriptors calculated, multiple linear regression | 0.854 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in SRW (water matrices and synthetic water prepared from Suwannee River) | Sudhakaran et al. (2012) |
| removal% (O3) Removal efficiency by O3 | %rml (O3) = 345.54–29.18 (ELUMO–EHOMO) - 5.59 (EA), Several molecular descriptors calculated, multiple linear regression | 0.862 | 36?/22 pharmaceuticals, personal care products and organic solvents Measurements performed in SRW (water matrices and synthetic water prepared from Suwannee River) | Sudhakaran et al. (2012) |
| removal%AOP Removal efficiency by O3/H2O2 | ANN | 0.8039 | 73/19/8 (%) pharmaceuticals, personal care products and organic solvents Measurements performed using various river water matrices | Sudhakaran et al. (2012) |
| removal% (O3) Removal efficiency by O3 | ANN | 0.7833 | 68/19/13.2 (%) pharmaceuticals, personal care products and organic solvents Measurements performed using various river water matrices | Sudhakaran et al. (2012) |
| % elimination | MLR | 0.836 | 100 structurally diverse pharmaceutical ingredients | Lee (2014) |
| % elimination | MLR | 0.866 | 100 structurally diverse pharmaceutical ingredients | Lee (2014) |