QSAR modelling for predicting adsorption of neutral, cationic, and anionic pharmaceuticals and other neutral compounds to microalgae *Chlorella vulgaris* in aquatic environment

Chul-Woong Cho [1], Yufeng Zhao [1], Yeoung-Sang Yun[*]

School of Chemical Engineering, Chonbuk National University, 567 Baekje-dearo, Deokjin-gu, Jeonju, 54896, Chonbuk, South Korea

ARTICLE INFO

ABSTRACT

Environmental fate or transport of pharmaceutical waste depends on the adsorptive interactions of pharmaceuticals with various environmental phases e.g. soil, sediment, microalgae, and bacteria etc. Therefore, it is important to understand these adsorptive interactions. As part of the study, we studied the adsorptive interaction of 30 chemicals with microalgae, i.e. *Chlorella vulgaris*, because it is ubiquitous and its surface area occupies a high proportion in aquatic environments. For this study, isotherms between *C. vulgaris* and 30 micropollutants in neutral and ionic forms (i.e. 15 cations, 5 anions, and 10 neutrals) were experimentally measured, and their adsorptive affinities were then theoretically predicted based on the concept of the linear free energy relationship. For modeling, the dataset was divided into a training set and a test set, where the training set was used for model development and the test set was performed for model validation. This process was repeated ten times. Finally, we suggested one model which has high predictability in R^2 of 0.96 and standard error (SE) of 0.17 log unit for the training set, R^2 of 0.818 and SE = 0.217 log unit for the test set, and R^2 of 0.926 and SE of 0.169 log unit for the total dataset. Moreover, it was found that dispersive force, H-bond basicity, molecular volume, and electrostatic interaction of anion significantly contribute to the model developed based on the entire dataset. Here, dispersive and hydrophobic interactions (proportional to the magnitude of molecular size) are main attractive forces, while the rest cases are repulsive. In addition, it was found that the adsorption property of the surface of *C. vulgaris* differs from those of Gram negative bacteria *Escherichia coli* and dissolved organic matters in an aquatic environment.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Pharmaceutical waste is an increasing environmental issue because generally pharmaceuticals are hardly treatable in wastewater treatment plant (Clara et al., 2005; Tijani et al., 2013; Verlicchi et al., 2012; Zhang et al., 2008) and they are released to the environments via several routes. The released chemicals can play a role as toxicants to microorganisms, and can become a threat to human and animals. In fact, many studies have reported that some pharmaceuticals have toxicity (Vasquez et al., 2014; Bruce et al., 2010; Khan and Nicell, 2015; Leung et al., 2013) and endocrine-disrupting effects (Catanese et al., 2015; Kiyama and Wada-

Kiyama, 2015; Petrie et al., 2015). Therefore, for safe treatment and protection against risk, it is essential to understand the environmental properties of pharmaceutical waste. However, since new pharmaceuticals have been continuously introduced, achieving the environmental properties of pharmaceuticals is a considerable challenge for environmental scientists and engineers because experimental determination is time and material consuming and labor intensive. Therefore, the development of a solid and robust prediction model is essential.

Among the environmental properties, the adsorption behaviors of chemicals in environmental media are important to understand because the adsorption properties affect their environmental properties (Girardi et al., 2011). In fact, if such micropollutants are released into nature, they can be adsorbed on the surface of environmental media such as bacteria, soil, and microalgae etc., and depending on the adsorption of the environmental phases, the environmental properties of the chemicals can be determined. For

example, if micropollutants are adsorbed in soil and/or sediment, their transport might change depending on the movement of soil and sediment. Moreover, the adsorption of chemicals on microalgae leads to chemical concentration and the chemicals are then supplied to higher organisms e.g. zooplankton and fish, thus possibly increasing the bioconcentration factor.

Among the interactions between chemicals and environmental media, adsorption on microalgae is important because microalgae are ubiquitous in surface and sea water and have a high proportional surface area of 0.98–2.2 m^2/mg of dried cells (Khoshmanesh et al., 1997), which is two or three times larger than that of soil organic matter (i.e. 0.56–0.80 m^2/mg; Chiou et al., 1990). Moreover, microalgae can be also used to remove emerging contaminants in wastewater treatment (e.g. Matamoros et al., 2016; Escapa et al., 2017; Villar-Navarro et al., 2018). However, the adsorption property has not be sufficiently investigated.

Several groups have characterized the adsorptive interactions between environmental phases and micropollutants, as summarized in Table 1. For the study, they were experimentally measured the adsorptive interaction with several environmental phases (e.g. dissolved organic matters, soils, and soil-like sorbents), based on two different methods: batch experiment and chromatography analysis. The previous studies also presented several prediction models based on the linear free energy relationship (LFER) model. Pool and Pool (1999) and Nguyen et al. (2005) studied the adsorption of chemicals onto soil using a chromatographic system and data collection, respectively. Similarly, Endo and Schmidt (2006), Endo et al. (2008, 2009), and Bronner and Goss (2011) examined the sorption properties of the soil-like materials i.e. coal tar and peat. Niederer et al. (2006, 2007) and Neale et al. (2012) estimated partition behaviours of chemicals onto humic or fulvic acids. Recently, our group presented sorption properties between ionic and neutral pharmaceuticals and Gram negative bacteria *Escherichia coli* (Cho et al., 2017). However, in most of the previous works, only neutral chemicals were studied, which indicates that ionic compounds cannot be applied to their prediction models. Actually, this constitutes a significantly large application gap because around 70% of chemicals are ionized in the natural pH range (Bryan and Huggett, 2012).

In this study, we therefore present the adsorptive interactions of neutral and ionic micropollutants on microalgae and develop an accurate prediction model. For this study, we selected the chemicals that currently are considered environmentally problematic i.e. pharmaceuticals and environmental hormones. Firstly, the adsorption affinities between them were experimentally measured

by isotherm experiments, and the measured values were then compared with previous presented models for predicting octanol-water partitioning coefficients, and adsorption onto dissolved organic matters and *Escherichia coli*. Moreover, we developed a new LFER model. Subsequently, the adsorption mechanisms between pharmaceuticals and microalgae were explained based on the developed model and experimental results.

2. Materials and methods

2.1. Sorbent preparation

The microalgae *Chlorella vulgaris* used were cultivated in 250 mL Erlenmeyer flasks including slightly modified OHM medium (Fabragas et al., 2000) comprising of (in g/L) KNO_3 0.41, Na_2HPO_4 0.03, $MgSO_4 \cdot 7H_2O$ 0.246, $CaCl_2 \cdot 2H_2O$ 0.11, (in mg/L) Fe(III)cit-rate·H_2O 2.62, $CoCl_2 \cdot 6H_2O$ 0.011, $CuSO_4 \cdot 5H_2O$ 0.012, Cr_2O_3 0.075, $MnCl_2 \cdot 4H_2O$ 0.98, $Na_2MoO_4 \cdot 2H_2O$ 0.12, SeO_2 0.005 and (in μg/L) biotin 25, thiamine 17.5 and B12 15. For light source, cool-white fluorescent lights (approximately 50 $\mu mol\,m^{-2}\,s^{-1}$) were used in a shacking incubator at $25 \pm 1\,°C$. After cultivation, the cells were washed with distilled (DI) water more than five times, and then were freeze-dried for the following adsorption tests, because the freeze-drying method can be used to preserve cell constituents without rupturing the cell wall (Becker, 1994).

2.2. Chemicals

Pharmaceuticals, environmental hormones, and other chemicals were purchased from Sigma-Aldrich (Seoul, Korea). The selected chemicals have structural diversities as shown by their LFER descriptors: E: 0.73 to 2.55; S: 0.87 to 7.7; A: 0 to 4.33; B: 0 to 4.38; V: 0.89 to 3.81; J^+: 0 to 4.16; J^-: −0.188 to 2.99. Their stock solution was prepared with distilled water and kept in a refrigerator (4 °C) to avoid light decomposition. Acetonitrile and phosphate buffer were purchased from Honeywell Burdick & Jackson (MI, USA) and Sigma Aldrich LTD (Seoul, Korea), respectively.

2.3. Isotherm experiment, calculation of adsorption affinity, and kinetic experiment

While the isotherm experiments for each of the target compounds were performed at a low level of concentration, the concentration was dependent on water solubility and the linearity of the isotherm. The isotherm experiments were performed within

Table 1
LFER modeling studies on adsorption of chemicals on typical environmental phases.

	Types of chemicals	Species of chemicals	Environmental phase (media)
Nguyen et al. (2005)	Unhalogenated monoaromatics Halogenated hydrocarbons PAHs, Polar Chemicals	Neutral	Soil/sediment organic matter (SOM)
Endo et al. (2008)	Polar and non-polar compounds	Neutral	Coal tar
Endo and Schmidt (2006)	Organic compounds	Neutral	Coal tar
Endo et al. (2009)	Polar and non-polar compounds	Neutral	Peat (Pakokee peat), lignite
Niederer et al. (2007)	Polar and non-polar compounds	Neutral	10 hydrate humic or fulvic acids
Bronner and Goss (2011)	Pesticides, pharmaceuticals, hormones	Neutral	Peat (Pakokee peat)
Pool and Pool (1999)	Polar compounds	Neutral	Wet-soil
Niederer et al. (2006)	Polar and non-polar compounds	Neutral	Leonardite humic acid
Neale et al. (2012)	Polar and non-polar compounds	Neutral	Dissolved organic carbon (Suwannee river fulvic acid, Aldrich humic acid)
Cho et al. (2017)	Pharmaceuticals, hormone, other organic compounds	Neutral and ions	Gram negative bacteria *Escherichia coli*

the concentration of 2.5–100 μM. First, 0.01 g of biomass was brought into contact with 30 mL solutions with different concentrations of compounds in 50 mL conical tube (Corning, USA). The biomass solution was then agitated in a shaking incubator at 120 rpm for 24 h at $25 \pm 1 °C$. In control experiment without microalgae, it was observed that there is no hydrolysis or adsorption effect of the tested chemicals. During the incubation, the pH was carefully adjusted at 6.5 ± 0.2 by using 0.01, 0.1, and 1 M HCl and NaOH solution, which were used to determine the actual natural pH value in an environment. At pH 6.5, most pharmaceuticals were totally ionized as anionic or cationic forms, since the pKa values of cationic pharmaceuticals are higher than 8.7 and those of anionic pharmaceuticals are lower than 4.5 (see Table 2). The equilibrium concentration was calculated using the following mass balance (Pehlivan et al., 2012) below.

$$q_e = \frac{(C_i - C_e)V_{sol}}{M}$$

where q_e is equilibrium concentration (μmol/g), and C_i and C_e means initial and final concentrations (μmol/L), respectively. V_{sol} is the working volume (L), and M is the mass of adsorbent (g).

Here, the adsorption affinity was estimated by linear fitting because the isotherm data points were nearly linear at trace level of concentration (Hörsing et al., 2011; Zeng et al., 2009). Except for the verapamil, all isotherm results were analyzed using more than 3 data points. The experiments were duplicated.

In order to determine when adsorption equilibrium was reached, kinetic experiments of three compounds, i.e. diphenhydramine, amitriptyline, and clomipramine, were performed. As shown in Fig. S1, when 0.033 g of C. vulgaris were brought into contact with 50 μM of the compounds, the sorption equilibrium states were attained within 1 h. The detailed explanation is given in supporting information.

2.4. Analytical condition

For quantitative analysis of the chemicals, a high performance liquid chromatography (HPLC) with dual pump, UV detector, and temperature controllable column incubator was applied. For that, 1 mL of sample was taken from the batch and filtered using 0.2 μm of syringe filter based on cellulose acetate (Advantec®, Japan). The column in the stationary phase was C18 (150 mm×4.6 mm) (YMC Co. LTD, Japan). The injected volume of the sample was 10 μl. Most chemicals were detected at 211 nm using an ultraviolet (UV) detector, except for tetraphenylborate which was detected at 265 nm. The mobile phase condition was dependent on the chemical species, but most of the chemicals i.e. amitriptyline, chlorpromazine, clomipramine, desipramine, diphenhydramine, propafenone, verapamil, thioridazine, caffeine, imipramine, promethazine, doxepin, nortriptyline, quinidine, carbamazepine, 2-phenylethanol, acetophenone, benzyl alcohol, bisphenol A, diethyl phthalate, ibuprofen, diclofenac, 1-naphthol, hydrocortisone, nitrobenzene, naproxen, 3,4,-dichlorobenzoate, and flurbiprofen were detected at isocratic condition with 0.4 mL/min of acetonitrile and 0.4 mL/min of buffer solution with 30 mM phosphate buffer (NaH₂PO₄) solution. The remaining cases were as follows: a mixture of 0.3 mL/min acetonitrile (ACN) and 0.5 mL/min buffer for propranolol; and a mixture of 1.0 mL/min and 0.5 mL/min for tetraphenylborate.

2.5. Model development

For the development of the prediction model, linear free energy relationship (LFER), as one of classical quantitative-structure relationship concepts, can be used because it has well-defined descriptors and can support highly accurate predictability (Abraham

Table 2
LFER descriptors of the tested chemicals and their measured and predicted log K_d values on C. vulgaris. The pKa values of chemicals were obtained from Pubchem (http://www.chemspider.com/).

	pKa	K_d	Log K_d	R^2	log P by Eq. (2)	log K_d by Eq. (12)	log K_d by Eq. (5)	E	S	A	B	V	J⁺	J⁻	Ref.
Amitriptyline+	8.8	4.73	0.67	0.98	−1.08	0.74	0.75	1.77	3.72	1.63	0.00	2.42	2.49	0	Abraham and Austin (2012)
Chlorpramazine+	9.2	12.44	1.09	0.97	−0.83	0.97	0.98	2.01	3.65	1.84	0.00	2.43	2.49	0	(Abraham and Austin (2012)
Clomipramine+	9.2	9.70	0.99	0.94	−0.36	0.83	0.83	1.64	3.78	1.71	0.00	2.55	2.37	0	Abraham and Austin (2012)
Desipramine+	10.0	4.73	0.67	0.96	1.51	0.62	0.59	1.47	3.62	2.79	0.00	2.28	1.45	0	Abraham and Austin (2012)
Diphenhydramine+	8.7	0.49	−0.31	0.93	−1.71	−0.06	−0.08	1.16	3.94	1.46	0.00	2.21	2.24	0	Abraham and Austin (2012)
Procaine+	8.8	0.47	−0.33	0.85	−2.07	−0.25	−0.25	2.06	6.57	4.09	0.00	2.78	2.36	0	Abraham et al. (2013)
Propranolol+	9.5	1.04	0.02	0.98	−2.50	−0.02	−0.03	1.69	4.31	2.07	0.00	2.17	2.43	0	Abraham and Austin (2012)
Verapamil+	8.7	1.50	0.18	0.99	0.33	0.25	0.27	1.66	7.70	4.33	0.00	3.81	2.40	0	Abraham and Austin (2012)
Thioridazine+	9.4	38.92	1.59	0.82	1.05	1.55	1.58	2.55	4.52	3.33	0.00	2.92	2.30	0	Abraham and Austin (2012)
Imipramine+	9.2	2.73	0.44	0.96	1.22	0.44	0.41	1.00	3.66	2.02	0.00	2.42	1.61	0	Abraham and Austin (2012)
Promethazine+	8.8	3.59	0.55	0.98	−2.16	0.58	0.59	1.90	3.77	1.51	0.00	2.30	2.70	0	Abraham and Austin (2012)
Doxepin+	9.8	1.84	0.26	0.99	−1.62	0.44	0.44	1.80	4.07	1.96	0.00	2.34	2.45	0	a
Nortriptyline+	9.7	3.59	0.55	0.99	−0.74	0.55	0.54	1.75	3.79	2.16	0.00	2.28	2.18	0	a
Quinidine+	8.6	0.61	−0.22	0.90	−7.44	−0.29	−0.26	2.32	5.81	1.23	0.00	2.57	4.16	0	a
Propafenone+	9.3	2.11	0.32	0.97	−1.19	0.11	0.11	1.68	5.69	3.08	0.00	2.85	2.38	0	Abraham and Austin (2012)
Caffeine	−	0.11	−0.96	0.94	−0.06	−0.91	−0.96	1.50	1.63	0.00	1.29	1.36	0	0	Abraham (1993)
Carbamazepine	−	1.47	0.17	0.89	2.26	0.32	0.31	2.15	1.90	0.50	1.15	1.82	0	0	Abraham et al. (2013)
2-phenylethanol	−	0.22	−0.66	0.87	1.42	−0.45	−0.53	0.81	0.91	0.30	0.64	1.06	0	0	Abraham (1993)
Acetophenone	−	0.48	−0.32	0.95	1.58	−0.30	−0.37	0.81	1.06	0.00	0.49	1.01	0	0	Abraham (1993)
Benzyl alcohol	−	0.25	−0.60	0.94	1.20	−0.50	−0.58	0.80	0.87	0.33	0.56	0.92	0	0	Abraham (1993)
Bisphenol A	−	1.92	0.28	0.99	3.22	0.40	0.38	1.61	1.56	0.99	0.94	1.86	0	0	Abraham (1993)
Diethyl phthalate	−	0.50	−0.30	0.99	2.57	−0.20	−0.24	0.73	1.40	0.00	0.86	1.71	0	0	Abraham (1993)
1-Naphthol	−	4.61	0.66	0.99	2.92	0.62	0.57	1.52	1.05	0.61	0.37	1.14	0	0	Abraham (1993)
Hydrocortisone	−	0.49	−0.31	0.95	1.68	−0.54	−0.52	2.03	3.49	0.71	1.90	2.80	0	0	Abraham (1993)
Nitrobenzene	−	1.46	0.17	0.99	1.82	−0.09	−0.16	0.85	1.14	0.00	0.27	0.89	0	0	Abraham (1993)
Ph₄B-	−	3.53	0.55	0.81	4.16	0.71	0.73	1.95	2.72	0.18	1.15	2.70	0	−0.19	Abraham and Acree (2010c)
3,4-Dichlorobenzoate -	3.6	0.11	−0.97	0.97	−1.26	−0.77	−0.68	1.10	3.04	0.00	2.49	1.16	0	2.11	Abraham and Acree (2010b)
Ibuprofen-	4.3	0.11	−0.96	0.99	−1.62	−1.26	−1.14	0.88	3.50	0.08	3.31	1.76	0	2.42	Abraham and Acree (2010c)
Naproxen-	4.4	0.09	−1.04	0.99	−2.11	−1.03	−0.88	1.66	5.07	0.02	3.11	1.76	0	2.43	Abraham and Acree (2010c)
Diclofenac-	4.4	0.58	−0.24	0.94	−1.59	−0.64	−0.45	1.96	5.31	0.03	3.35	2.00	0	2.62	Abraham and Acree (2010c)

[a] LFER descriptors of the chemicals were calculated using equations for protonated compounds by Abraham and Acree Jr. (2010).bib_Abraham_and_Acree_2010a

and Acree, 2016). Moreover, the developed model can help us interpret the chemical meanings of sorbent-sorbate interactions on a molecular basis. The LFER model (Abraham and Acree, 2016) is shown below:

$$\text{Log } K_d = eE + sS + aA + bB + vV + j^+J^+ + j^-J^- + c \qquad (1)$$

where the dependent variable log K_d is the adsorptive interaction, i.e. adsorption affinity between sorbents and the surface of *C. vulgaris*. The capital letters refer to the LFER descriptors of the solute, as follows: E - excess molar refraction [cm^3 mol^{-1}/10]; S - dipolarity/polarisability [dimensionless]; A and B - hydrogen bonding acidity [dimensionless] and hydrogen bonding basicity [dimensionless]; V - McGowan volume [cm^3 mol^{-1}/100]; and J^+ and J^- - ionic interactions of cations and anions. Small letters ($e, s, a, b, v, j^-,$ and j^+) are system-dependent parameters. The system parameters can be determined by multiple linear regression analysis. This model was successfully applied for characterizing some adsorbents i.e. ion-exchange resins (Jadbabaei and Zhang, 2014; Zhang et al., 2014), which can be used for the removal of anionic and cationic micropollutants.

2.6. Statistical analysis

To determine the system parameters of Eq. (1) and their statistical variables, multiple linear regressions were performed using SPSS (12.0K Windows) software. For the fitting of Figures and data calculation, Sigma-plot (version 10.0) was used. In order to check the prediction acceptability, the 'Xternal ValidationPlus Tool'(Roy et al., 2016), developed by Prof. Kunal Roy and his colleagues was applied, the estimation of which involves checking the mean absolute error (MAE) based criteria of 95% of the dataset after removing 5% high residuals. The criteria of prediction (Roy et al., 2016) can be considered as follows:

   i) Good level: MAE ≤0.1 × training set range, and MAE +3 × η (standard deviation of the absolute error values of test set) ≤ 0.2 × training set range
   ii) Moderate level: it was considered when the predictions are not under either of the two conditions for bad prediction.
   iii) Bad level: MAE >0.15 × training set range, or MAE +3 × η > 0.25 × training set range.

MAE and η calculation can be performed using the following equations.

$$\text{MAE (mean absolute error)} = \frac{1}{n} \times \sum \left[Y_{obs} - Y_{pred} \right]$$

$$\eta \text{ (standard deviation)} = \pm \sqrt{\frac{\sum \left(\boldsymbol{Y_{obs}} - \overline{\boldsymbol{Y}}_{\boldsymbol{pred}} \right)^2}{\boldsymbol{n}}}$$

where Y_{obs} and Y_{pred} represent the measured and predicted log K_d values, respectively. The method for the MAE estimation is well explained on the homepage of Prof. Roy (http://dtclab.webs.com/software-tools and http://teqip.jdvu.ac.in/QSAR_Tools/) (Roy et al., 2016).

2.7. Previous LFER models for predicting environmental properties

In order to find some correlative environmental properties, the experimentally measured log K_d values were correlated with log P (octanol-water partitioning coefficient) (Abraham and Acree, 2010c), log K_{oc} (organic matter-water partition coefficient)

(Nguyen et al., 2005), and log K_d (Escherichia coli-water partitioning coefficient) (Cho et al., 2017). The environmental properties were calculated based on previous models i.e. Eq. (2)–(4).

$$\text{Log P [dimensionless]} = 0.088 + 0.562 \text{ E} - 1.054 \text{ S} + 0.034 \text{ A} - 3.460 \text{ B} + 3.814 \text{ V} - 3.023 \text{ J}^+ + 2.580 \text{ J}^- \qquad (2)$$

$$\text{Log } K_{oc} = 0.14 + 1.1 \text{ E} - 0.72 \text{ S} + 0.15 \text{ A} - 1.98 \text{ B} + 2.28 \text{ V} \qquad (3)$$

$$\text{Log } K_d \text{ [L/g]} = -0.89 + 0.46 \text{ E} - 0.38 \text{ S} + 0.12 \text{ A} - 1.61 \text{ B} + 0.79 \text{ V} + 0.063 \text{ J}^+ \qquad (4)$$

3. Results and discussion

3.1. Isotherm results

The adsorption between sorbent and sorbate has two main properties: adsorption affinity and adsorption capacity. Adsorption affinity literally refers to the adsorption affinity between a chemical and an adsorbent at low concentration, while adsorption capacity refers to the adsorption ability of adsorbents for a certain sorbate. Since the concentration of micropollutants present in aquatic environments is generally at a trace level, only the adsorption affinity between *C. vulgaris* and the chemicals was estimated at low concentration range (i.e. 2.5–100 μM). Moreover, an experimental pH value was fixed at 6.5, which is used to reflect the neutral pH value. Since we used HCl and NaOH solutions for adjusting the pH value, we needed to check the salt-effect, whereby we selected three compounds including a neutral compound (caffeine), an anion (diclofenac), and a cation (desipramine) and performed one-point check depending on the different NaCl concentrations from 0 to 60 μM. The result showed that the three compounds have similar uptake values, as shown in Fig. S2. Here, the isotherm value was estimated by calculating the slope value based on a linear fitting between the adsorption uptake (μmol/g) and the equilibrium concentration (μmol/L). Some example fittings are shown in Fig. 1. The estimated adsorption affinity values in the log unit ranged from −1.04 to 1.59. In general, it was shown that among the chemical species tested, the cationic chemicals have high adsorption affinity values, while anionic chemicals have low adsorption affinity values. Especially, anions with a small molecular structure
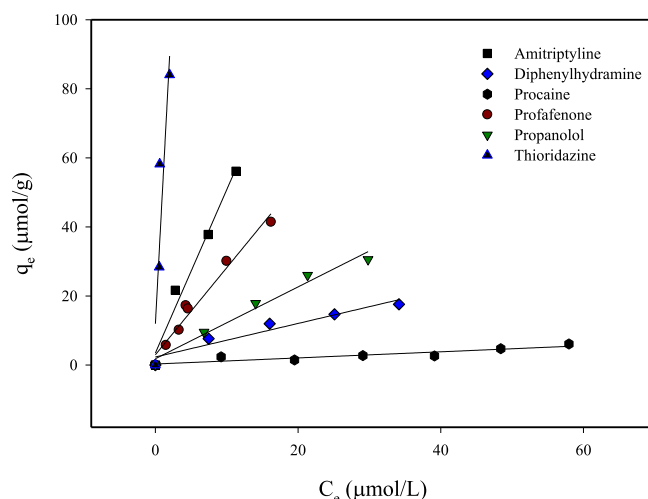

Fig. 1. Linear fittings between adsorption uptake (q_e, μmol/g) and equilibrium concentration (C_e, μmol/L) for calculating the adsorption affinity of chemicals.

showed negligible adsorption affinity on *C. vulgaris*, while large molecular structures with negative charge were slightly adsorbed onto the surface of *C. vulgaris*, although their adsorption was at a low level among the estimated values, except for tetraphenylborate in the anionic form. In cationic compounds, thioridazine has the highest adsorption affinity value in the log unit, which is 1.59.

3.2. Correlation between adsorption affinity and physicochemical properties of chemicals

To investigate the similarity of previous prediction models with the adsorption affinities of micropollutants on the surface of *C. vulgaris* (log K_d), we compared the measured log K_d values with the octanol-water partitioning coefficient (log P), the adsorption affinity of chemicals on bacteria, and dissolved organic matters.

First, we correlated between the log K_d and log P values since it is known that log P has a high correlation with the environmental properties of chemicals. Here, for comparison, we used calculated log P values, based on Eq. (2), because the log P values of most of the ionic compounds have not yet been measured. The comparison between log K_d and log P (Fig. 2) shows that generally the log P values have a poor relationship with the log K_d values of the entire dataset, whose R^2 value was 0.046. However, when correlating log P values with only the log K_d of neutral compounds, a remarkable correlation was shown at R^2 of 0.726. This indicates that, while the hydrophobicity of neutral compounds strongly contributes to the correlation, in the case of ions, their hydrophobicity may not be the main factor for explaining the adsorption property because ions are hydrophilic substances due to their charges.

Next, the log K_d values were compared with partitioning behaviors (log K_{oc}), which were determined based on the partitioning coefficient of chemicals between water and organic matters in sediments and soil (Nguyen et al., 2005). In comparison, since the model was developed based on the use of neutral compounds, the ionic terms of cations and anions were not considered and the combination of only five LFER descriptors (E, S, A, B, and V) were used. The correlation between the log K_{oc} calculated by Eq. (3) and the measured log K_d of chemicals, given in Table 2, shows a remarkable relationship, whose R^2 is 0.59 [see fitting (1) in Fig. 3]. However, this relationship may not be due to some relationship, but to the difference among the values in anionic, neutral, and cationic compounds. Meanwhile, when individually considering cationic compounds or neutral compounds, they have remarkable
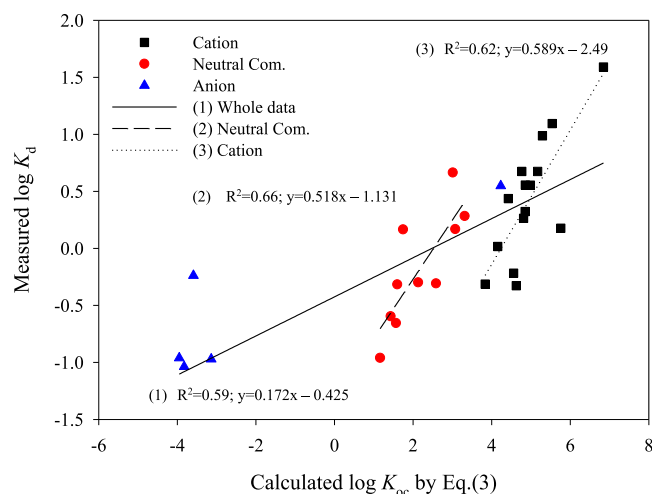

Fig. 3. Correlations between measured log K_d values and predicted log K_{oc} values by Eq. (3).

correlations within R^2 of 0.62 and 0.66, respectively, and their equations are $y = 0.589x-2.49$ and $y = 0.518x-1.131$, respectively [see fitting (2 and 3) in Fig. 3]. It is shown that the fittings have similar slope values (0.589 vs. 0.518) yet different y axis values. This might be due to the missing cationic term (J^+) in Eq. (3). In addition, anions were not discussed because the number of datasets was insufficient.

Next, the LFER model previously developed by our group (Cho et al., 2017) (i.e. Eq. (4)) for predicting the adsorptive interaction between *E. coli* and chemicals was applied to correlate the log K_d values onto *C. vulgaris*. The previous model was developed based on only cationic and neutral compounds because the surface of *E.coli* has no adsorptive interaction with anions. In correlations between the measured log K_d on *C. vulgaris* and the calculated log K_d by Eq. (4), it was observed that both properties have a relationship at R^2 of 0.513. The main reason for the correlation may be the difference of log K_d values according to the chemical species i.e. anionic, neutral, and cationic compounds. Meanwhile, in correlations of cations or neutral compounds with the calculated log K_d values (*E. coli*), both species have several remarkable relationships at R^2 of 0.84 and 0.51, respectively (Fig. 4). In addition, the anion species shows a poor
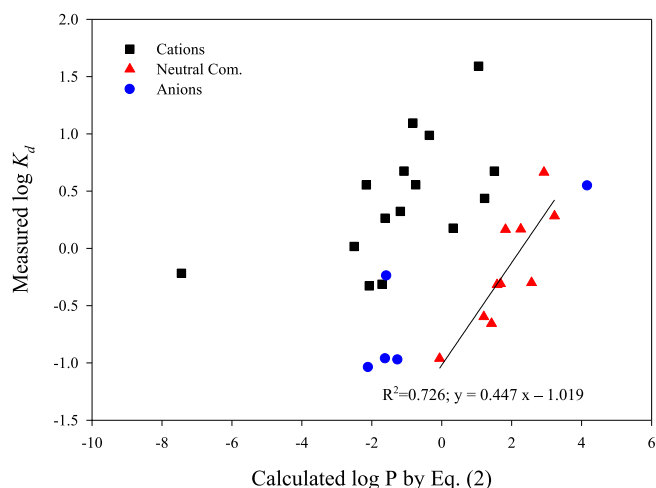

Fig. 2. Correlation between measured adsorption affinity (log K_d) and calculated log P values [by Eq. (2)] of neutral, anionic, and cationic chemicals.
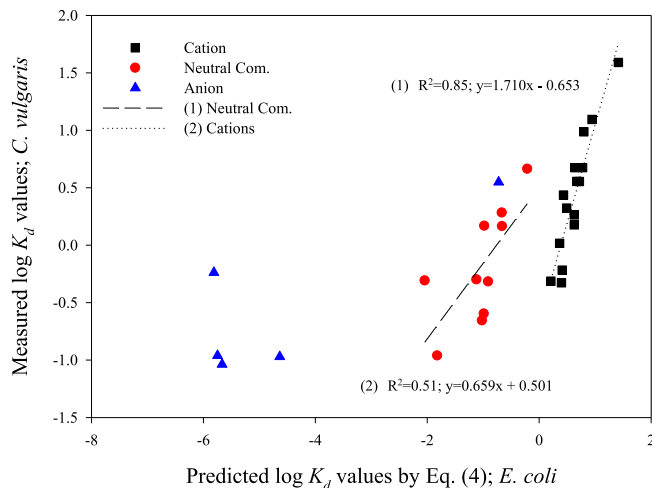

Fig. 4. Comparison between measured log K_d on *C. vulgaris* values and predicted log K_d values on *E. coli* by Eq. (4).

correlation.

3.3. Development of a new LFER model for predicting log K_d

As explained above, the previous prediction models show several good agreements for cations and/or neutral compounds. However, overall, the prediction models cannot predict the log K_d values on *C. vulgaris*. Thus, to achieve better predictability, a new prediction model should be developed. Therefore, we attempted to directly correlate the log K_d values of the targeted micropollutants on *C. vulgaris* with their corresponding LFER descriptors. For the modelling, the experimentally obtained dataset was randomly divided into a training set and a test set, the numbers of which are 20 and 10, respectively. Firstly, the LFER descriptors of chemicals in the training set and their measured log K_d values were used for model development. To determine the system parameters (e, s, a, b, v, j^+, and j^-), multiple linear regression was performed, and the test set and the total data set were then used to validate the predictability of the developed model. This prediction study was repeated ten times. The estimated system parameters and the statistical results of the models are given in Table 3.

As shown in Table 3, the obtained results depended on the selected data set. As seen in Table 3, all models have high predictabilities within R^2 of 0.936–0.970 for each training set. Among these models, model 7 has the highest correlation for its training set, which was estimated based on the R^2 value. The R^2 value and standard error (SE) are 0.970 and 0.150 log unit, respectively. The predictabilities for the corresponding test set and total data seems also acceptable, which are R^2 of 0.847 and SE of 0.194 log unit, and R^2 of 0.946 and SE of 0.156 log unit, respectively. Especially, the validation result for total data was the best among the ten models.

For further verification of the acceptability of the model, we employed the MAE-based validation method by Roy et al. (2016) because the correlation based estimation can be misleading. The criteria of the MAE-based estimation are explained in the Materials and Methods section. The validation can be simply performed by attaching the measured and calculated values to an excel sheet and adding the training set response mean and range. For a detailed explanation, see Roy et al., (2016). Surprisingly, the validation results shown in Table 4 indicated that model 7 was not selected as the best prediction model. The program estimated it as moderate. Because the data range of the selected training set for model 7 is narrow, i.e. 2.628 (maximum value − minimum value), the MAE+3*standard deviation value (i.e. 0.584) increases to more than 0.2* of the training set range (i.e. 0.525) and MAE value (i.e. 0.151). Therefore, the predictability level is "Moderate". The statistical results according to the MAE-based criteria of the 10 models are given in Table 4. Meanwhile, the validation software suggested seven models classified as "Good", which are models 2–5 and

Table 4
Statistical results from MAE-based validation studies for the ten models in Table 3.

	Training set (100% data)		Test set (95% data)			MAE-based result
	Mean	Range	MAE	SD(η)	MAE + 3 × η	
Model 1	0.065	2.066	0.172	0.042	0.554	Bad
Model 2	0.051	2.066	0.120	0.0325	0.412	Good
Model 3	0.100	2.628	0.117	0.0330	0.413	Good
Model 4	0.185	2.628	0.111	0.037	0.444	Good
Model 5	0.059	2.628	0.374	0.0301	0.4083	Good
Model 6	0.172	2.628	0.179	0.0456	0.589	Moderate
Model 7	0.121	2.628	0.151	0.0481	0.584	Moderate
Model 8	0.124	2.628	0.132	0.0293	0.396	Good
Model 9	0.056	2.628	0.152	0.0358	0.495	Good
Model 10	0.080	2.628	0.129	0.0404	0.492	Good

8–10, which also have high predictabilities for the training set, test set, and total data set. Here, among the selected models, model 8 has the highest predictability (R^2 and SE are 0.941 and 0.163 log unit, respectively).

To understand the adsorptive mechanism between the neutral and ionic micropollutants and *C. vulgaris*, we re-estimated the system parameters and statistical results using the total dataset without dividing into the training set and test set. This reflected a greater range of chemical structures of the model. By performing multiple linear regression, the results could then be obtained as follows:

$$\text{Log } K_d \text{ (} C. vulgaris\text{)} = -1.149 + 0.876 \text{ E} - 0.514 \text{ S} - 0.150 \text{ A} - 1.764 \text{ B} + 1.465 \text{ V} - 0.419 \text{ J}^+ + 1.786 \text{ J}^- \tag{5}$$

$$R^2 = 0.950, \text{ SE} = 0.170 \text{ log unit}, R^2_{adj} = 0.934, N = 30, F = 59.2$$

where R^2 was 0.950 and the standard error was 0.170 log units. Here, except H-bonding donor term, all of the LFER descriptors used for modeling have lower probability (*p*) values than 0.05, indicating the used descriptors are within 95% of a confidential level. The *p* value of term A was 0.06, which is almost within the acceptable range. Moreover, the adjusted R^2 value (R^2_{adj}) i.e. 0.934 has a similar R^2 value, proving that the combination of the selected descriptors is reasonable. The fitting between the predicted and measured log K_d values is given in Fig. 5.

Based on Eq. (5), the adsorption phenomena may be explained. First, system parameters for excess molar refraction, molecular volume, and anionic interaction have a positive sign, while those of dipolarity/polarizability, hydrogen bonding acidity and basicity, and cationic interaction terms are negative signs. This indicates that with an increase in the magnitude of the former three terms i.e. E, V, J^-, the adsorptive affinity will increase while the latter's will have

Table 3
Estimated system parameters and their statistical results from the LFER model for predicting adsorptive interaction between *C. vulgaris* and chemicals.

| Eq.		System parameters								Training set			Test set			Total set		
		c	e	s	a	b	v	j+	j-	R^2	SE	N	R^2	SE	N	R^2	SE	N
(5)	Model 1	−1.447	0.981	−0.530	−0.208	−1.860	1.664	−0.500	1.886	0.963	0.151	20	0.900	0.252	10	0.934	0.172	30
(6)	Model 2	−0.966	0.704	−0.521	−0.123	−1.600	1.428	−0.342	1.587	0.941	0.187	20	0.925	0.226	10	0.934	0.172	30
(7)	Model 3	−1.099	1.014	−0.420	−0.193	−1.861	1.313	−0.524	1.700	0.946	0.185	20	0.922	0.229	10	0.935	0.171	30
(8)	Model 4	−1.175	0.998	−0.423	−0.263	−2.018	1.481	−0.578	1.849	0.936	0.182	20	0.939	0.205	10	0.939	0.166	30
(9)	Model 5	−1.499	0.900	−0.578	−0.154	−1.705	1.673	−0.391	1.810	0.969	0.169	20	0.837	0.186	10	0.936	0.170	30
(10)	Model 6	−1.112	0.919	−0.449	−0.267	−1.969	1.511	−0.530	1.768	0.965	0.141	20	0.893	0.264	10	0.932	0.176	30
(11)	Model 7	−0.942	0.809	−0.537	−0.176	−2.022	1.513	−0.427	2.087	0.970	0.150	20	0.847	0.194	10	0.946	0.156	30
(12)	Model 8	−1.013	0.845	−0.515	−0.135	−1.750	1.419	−0.417	1.706	0.953	0.176	20	0.921	0.210	10	0.941	0.163	30
(13)	Model 9	−1.323	0.861	−0.513	−0.016	−1.502	1.387	−0.372	1.581	0.967	0.168	20	0.779	0.254	10	0.929	0.179	30
(14)	Model 10	−1.185	0.801	−0.564	−0.052	−1.643	1.501	−0.395	1.790	0.959	0.166	20	0.906	0.230	10	0.939	0.166	30
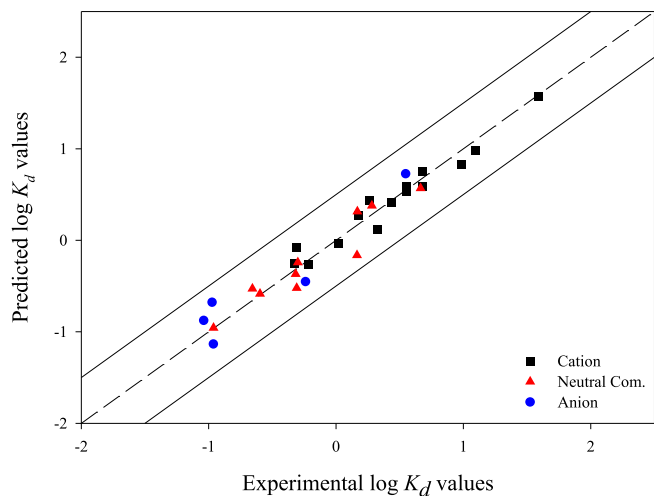
Fig. 5. Correlation between measured and predicted [by Eq. (15)] log K_d values of chemical on *C. vulgaris*.

an opposite trend. However, the opposite trend was shown when correlating between the single J$^-$ with log K_d values; it has a negative sign for its system parameter. The difference may be because the J$^-$ term acts as complementary role to the H-bonding basicity term since both terms are internally correlated at R^2 of 0.885. Therefore, it should be understood that the anionic term has a negative correlation with the measured log K_d values. Among the seven terms, H-bond basicity (B) is the highest correlated descriptor, the correlation factor of which was 0.388. Next, adding the induced dipole term (E) to B helps to enhance the prediction of the log K_d value to R^2 of 0.544. The addition of dipolarity/polarizability (S) to the combination of B and E slightly increases the predictability with R^2 of 0.581. Next, the further addition of ionic interaction terms for the anion and molecular volume (V) increases its prediction accuracy to R^2 of 0.893. By addressing the J$^+$ term, the accuracy of the prediction model increased, the R^2 of which was 0.941. Finally, whole descriptors with an inclusion of H-bonding acidity could achieve the highest prediction model with R^2 of 0.950.

4. Conclusion

In this study, we showed that neutral and ionic micropollutants can have an adsorptive interaction with the surface of microalgae *C. vulgaris* and developed its prediction model based on the linear free energy relationship. To investigate the adsorptive interaction, adsorption affinities between 30 micropollutants including several pharmaceuticals in neutral or ionic forms and microalgae *Chlorella vulgaris* were experimentally determined, their values were correlated with several different properties i.e. octanol-water partitioning coefficient (log P) and adsorption properties in water-organic matters and water-bacteria. In comparison, it was clearly observed that the overall log K_d on *C. vulgaris* could not be well explained by previous prediction models. Therefore, we developed new LFER models using training sets and test sets, and was validated with the MAE based criteria. Among these models, model 8 in Table 3 was suggested as superior, which has R^2 = 0.953 and SE = 0.176 log unit for training set. It could predict its corresponding test set in R^2 of 0.921 and SE of 0.210. Moreover, when applying model 8 for predicting the total dataset (i.e. 30 chemicals), its predictability (R^2 = 0.941 and SE = 0.163 log unit) is very similar to the result (R^2 = 0.950 and SE = 0.170 log unit) developed using the entire dataset of Eq. (5). Based on the developed model, the adsorptive mechanisms were interpreted such that there were

various important molecular interaction potentials contributing to the prediction model, including dispersive force, dipolarity/polarizability, H-bond basicity, molecular volume, and electrostatic interaction of anion. Here, dispersive force and molecular volume play the role of attractive forces while H-bonding basicity, dipolarity/polarizability, and H-bonding forces have an opposite role. Although the prediction model showed that the anionic term acts as an attractive force, it actually has a repulsive interaction because J$^-$ term acted as a complementary parameter to the H-bonding basicity term in the determination of the system parameters of Eq. (5). Therefore, it was auto-scaled multiple linear regression.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.watres.2018.12.033.

References

Abraham, M.H., 1993. HYDROGEN-BONDING. 31. Construction of a scale of solute effective or summation hydrogen-bond basicity. J. Phys. Org. Chem. 6 (12), 660–684.

Abraham, M.H., Acree, W.E., 2010a. Equations for the transfer of neutral molecules and ionic species from water to organic phases. J. Org. Chem. 75 (4), 1006–1015.

Abraham, M.H., Acree, W.E., 2010b. Solute descriptors for phenoxide anions and their use to establish correlations of rates of reaction of anions with iodo-methane. J. Org. Chem. 75 (9), 3021–3026.

Abraham, M.H., Acree, W.E., 2010c. The transfer of neutral molecules, ions and ionic species from water to wet octanol. Phys. Chem. Chem. Phys. 12 (40), 13182–13188.

Abraham, M.H., Acree, W.E., 2016. Descriptors for ions and ion-pairs for use in linear free energy relationships. J. Chromatogr. A 1430, 2–14.

Abraham, M.H., Austin, R.P., 2012. The effect of ionized species on microsomal binding. Eur. J. Med. Chem. 47, 202–205.

Abraham, M.H., Acree, W.E., Fahr, A., Liu, X.L., 2013. Analysis of immobilized artificial membrane retention factors for both neutral and ionic species. J. Chromatogr. A 1298, 44–49.

Becker, E.W., 1994. Microalgae: Biotechnology and Microbiology. Cambridge University Press, UK. Cambridge.

Bronner, G., Goss, K.U., 2011. Predicting sorption of pesticides and other multi-functional organic chemicals to soil organic carbon. Environ. Sci. Technol. 45 (4), 1313–1319.

Bruce, G.M., Pleus, R.C., Snyder, S.A., 2010. Toxicological relevance of pharmaceuticals in drinking water. Environ. Sci. Technol. 44 (14), 5619–5626.

Bryan, W.B., Huggett, D.B., 2012. Human Pharmaceuticals in the Environment: Current and Future Perspectives. Springer, New York, NY, USA.

Catanese, M.C., Suvorov, A., Vandenberg, L.N., 2015. Beyond a means of exposure: a new view of the mother in toxicology research. Toxicol. Res. 4 (3), 592–612.

Chiou, C.T., Lee, J.F., Boyd, S.A., 1990. The surface-area of soil organic-matter. Environ. Sci. Technol. 24 (8), 1164–1166.

Cho, C.W., Park, J.S., Zhao, Y.F., Yun, Y.S., 2017. Quantitative analysis of adsorptive interactions of ionic and neutral pharmaceuticals and other chemicals with the surface of *Escherichia coli* cells in aquatic environment. Environ. Pollut. 227, 8–14.

Clara, M., Kreuzinger, N., Strenn, B., Gans, O., Kroiss, H., 2005. The solids retention time - a suitable design parameter to evaluate the capacity of wastewater treatment plants to remove micropollutants. Water Res. 39 (1), 97–106.

Endo, S., Schmidt, T.C., 2006. Prediction of partitioning between complex organic mixtures and water: application of polyparameter linear free energy relationships. Environ. Sci. Technol. 40 (2), 536–545.

Endo, S., Xu, W.J., Goss, K.U., Schmidt, T.C., 2008. Evaluating coal tar-water partitioning coefficient estimation methods and solute-solvent molecular interactions in tar phase. Chemosphere 73 (4), 532–538.

Endo, S., Grathwohl, P., Haderlein, S.B., Schmidt, T.C., 2009. LFERs for soil organic carbon-water distribution coefficients (K-OC) at environmentally relevant sorbate concentrations. Environ. Sci. Technol. 43 (9), 3094–3100.

Escapa, C., Coimbra, R.N., Nuevo, C., Vega, S., Paniagua, S., Carcia, A.I., Calvo, L.F.,

Otero, M., 2017. Valorization of microalgae biomass by its use for the removal of paracetamol from contaminated water. Water 9, 312–320.

Fabragas, J., Dominguez, A., Regueiro, M., Maseda, A., Otero, A., 2000. Optimization of culture medium for the continuous cultivation of the *Haematococcus pluvialis*. Appl. Microbiol. Biotechnol. 53, 530–535.

Girardi, C., Greve, J., Lamshoft, M., Fetzer, I., Miltner, A., Schaffer, A., Kastner, M., 2011. Biodegradation of ciprofloxacin in water and soil and its effects on the microbial communities. J. Hazard. Mater. 198, 22–30.

Hörsing, M., Ledin, A., Grabic, R., Fick, J., Tysklind, M., Jansen, J.l.C., Andersen, H.R., 2011. Determination of sorption of seventy-five pharmaceuticals in sewage sludge. Water Res. 45 (15), 4470–4482.

Jadbabaei, N., Zhang, H.C., 2014. Sorption mechanism and predictive models for removal of cationic organic contaminants by cation exchange resins. Environ. Sci. Technol. 48 (24), 14572–14581.

Khan, U., Nicell, J., 2015. Human health relevance of pharmaceutically active compounds in drinking water. AAPS J. 17 (3), 558–585.

Khoshmanesh, A., Lawson, F., Prince, I.G., 1997. Cell surface area as a major parameter in the uptake of cadmium by unicellular green microalgae. Chem. Eng. J. 65 (1), 13–19.

Kiyama, R., Wada-Kiyama, Y., 2015. Estrogenic endocrine disruptors: molecular mechanisms of action. Environ. Int. 83, 11–40.

Leung, H.W., Jin, L., Wei, S., Tsui, M.M.P., Zhou, B.S., Jiao, L.P., Cheung, P.C., Chun, Y.K., Murphy, M.B., Lam, P.K.S., 2013. Pharmaceuticals in tap water: human health risk assessment and proposed monitoring framework in China. Environ. Health Perspect. 121 (7), 839–846.

Matamoros, V., Uggetti, E., Garcia, J., Bayona, J.M., 2016. Assessment of the mechanisms involved in the removal of emerging contaminants by microalgae from wastewater: a laboratory scale study. J. Hazard. Mater. 301, 197–205.

Neale, P.A., Escher, B.I., Goss, K.U., Endo, S., 2012. Evaluating dissolved organic carbon-water partitioning using polyparameter linear free energy relationships: implications for the fate of disinfection by-products. Water Res. 46 (11), 3637–3645.

Nguyen, T.H., Goss, K.U., Ball, W.P., 2005. Polyparameter linear free energy relationships for estimating the equilibrium partition of organic compounds between water and the natural organic matter in soils and sediments. Environ. Sci. Technol. 39 (4), 913–924.

Niederer, C., Goss, K.U., Schwarzenbach, R.P., 2006. Sorption equilibrium of a wide spectrum of organic vapors in Leonardite humic acid: modeling of experimental data. Environ. Sci. Technol. 40 (17), 5374–5379.

Niederer, C., Schwarzenbach, R.P., Goss, K.U., 2007. Elucidating differences in the sorption properties of 10 humic and fulvic acids for polar and nonpolar organic chemicals. Environ. Sci. Technol. 41 (19), 6711–6717.

Pehlivan, E., Altun, T., Parlayici, S., 2012. Modified barley straw as a potential biosorbent for removal of copper ions from aqueous solution. Food Chem. 135 (4), 2229–2234.

Petrie, B., Barden, R., Kasprzyk-Hordern, B., 2015. A review on emerging contaminants in wastewaters and the environment: current knowledge, understudied areas and recommendations for future monitoring. Water Res. 72, 3–27.

Poole, S.K., Poole, C.F., 1999. Chromatographic models for the sorption of neutral organic compounds by soil from water and air. J. Chromatogr. A 845 (1–2), 381–400.

Roy, K., Das, R.N., Ambure, P., Aher, R.B., 2016. Be aware of error measures. Further studies on validation of predictive QSAR models. Chemometr. Intell. Lab. Syst. 152, 18–33.

Tijani, J.O., Fatoba, O.O., Petrik, L.F., 2013. A Review of pharmaceuticals and endocrine-disrupting compounds: sources, effects, removal, and detections. Water, Air, Soil Pollut. 224 (11).

Vasquez, M.I., Lambrianides, A., Schneider, M., Kummerer, K., Fatta-Kassinos, D., 2014. Environmental side effects of pharmaceutical cocktails: what we know and what we should know. J. Hazard. Mater. 279, 169–189.

Verlicchi, P., Al Aukidy, M., Zambello, E., 2012. Occurrence of pharmaceutical compounds in urban wastewater: removal, mass load and environmental risk after a secondary treatment-A review. Sci. Total Environ. 429, 123–155.

Villar-Nararro, E., Baena-Nogueras, R.M., Paniw, M., Perales, J.A., Lara-Martin, P.A., 2018. Removal of pharmaceuticals in urban wastewater: high rate algae pond (HRAP) based technologies as an alternative to activated sludge based processes. Water Res. 139, 19–29.

Zeng, Q., Li, Y., Gu, G., 2009. Effect of temperature on the sorption of 17α-ethinylestradiol to aerobic and anaerobic sludges. In: International Conference on Energy and Environment Technology, pp. 294–297.

Zhang, Y.J., Geissen, S.U., Gal, C., 2008. Carbamazepine and diclofenac: removal in wastewater treatment plants and occurrence in water bodies. Chemosphere 73 (8), 1151–1161.

Zhang, H.C., Shields, A.J., Jadbabaei, N., Nelson, M., Pan, B.J., Suri, R.P.S., 2014. Understanding and modeling removal of anionic organic contaminants (AOCs) by anion exchange resins. Environ. Sci. Technol. 48 (13), 7494–7502.