•

Desalination and Water Treatment

www.deswater.com

1944-3994/1944-3986 © 2010 Desalination Publications. All rights reserved doi: 10.5004/dwt.2010.987

A QSAR (quantitative structure-activity relationship) approach for modelling and prediction of rejection of emerging contaminants by NF membranes

V. Yangali-Quintanilla^{a,b*}, A. Sadmani^{a,b}, M. Kennedy^{a,b}, G. Amy^{a,b}

^aUNESCO-IHE, Institute for Water Education, Westvest 7, 2611AX Delft, The Netherlands Tel. +31 15 215 1745; Fax +31 15 215 2921; email: v.yangaliquintanilla@unesco-ihe.org ^bDelft University of Technology, Stevinweg 1, 2628CN Delft, The Netherlands email: v.a.yangaliquintanilla@tudelft.nl

Received 8 May 2009; accepted 29 October 2009

ABSTRACT

Principal component analysis (PCA) and multivariate regressions were used to find a quantitative structure-activity relationship (QSAR) model equation that combines interactions between membrane characteristics and solute properties for predicting rejection. An internal experimental database that accounts rejections of contaminants by two nanofiltration membranes (NF-90, NF-200) was used to develop the QSAR model equation. Membrane characteristics related to hydrophobicity (contact angle), salt rejection, and surface charge (zeta potential [ZP]); compound properties describing hydrophobicity (log K_{ow}, log D), polarity (dipole moment), and size (molar volume, molecular length, molecular depth, equivalent width, molecular weight); and operating conditions (flux, pressure, cross flow velocity) were identified and evaluated as candidate variables for rejection prediction. Subsequently, using the QSAR model, rejection predictions were made for an external experimental database. Measured rejections were compared against predicted rejections to determine the best model; an acceptable R^2 (0.93) correlation coefficient was found for the best model with a standard deviation of error of 7% for predicted rejections. In conclusion, a general QSAR model equation was able to model and predict rejections of emerging contaminants during nanofiltration.

1. Introduction

Emerging contaminants in water, also referred as micropollutants, are a group of organic compounds that are present in water environments. Emerging organic contaminants derive from pharmaceuticals, endocrine disruptor compounds, personal care products and solvents and products used in manufacture industries. The presence of emerging contaminants in water environments has been reported by various studies [1–3]. However, the concern of their presence

*Corresponding author

grows when emerging contaminants are also detected in supplies of drinking water [4]. Actual concentrations of micropollutants in drinking water are not a threat for human adults but certainly they may affect the development of human embryos [5]. Aschengrau et al. [6] demonstrated that perchloroethene contamination of public drinking water in Massachusetts influenced the occurrence of congenital anomalies among children whose mothers were exposed around the time of conception. Nanofiltration (NF) and reverse osmosis (RO) are proven technologies that are able to remove most of the emerging contaminants but their efficiency is highly dependent on variables related to membrane

January

Presented at the conference on Desalination for the Environment: Clean Water and Energy, 17–20 May 2009, Baden-Baden, *Germany. Organized by the European Desalination Society.*

Table 1 List of emerging contaminants with physicochemical properties

Name	Molecular weight (g/mol)	Acid pKa 20° Ca	log K _{ow} d	Log D ^a (pH 7)	Dipole moment (debye) ^b	Molar volumec (cm ³ /mol)	Molec. length (nm) ^c	Molec. width (nm) ^c	Molec. depth (nm) ^c	Equiv. width (nm) ^c	Classification ^e
Acetaminophen	151	10.2	0.46	0.23	4.55	120.90	1.14	0.68	0.42	0.53	HL-neutral
Phenacetine	179	N/A	1.58	1.68	4.05	163.00	1.35	0.69	0.42	0.54	HL-neutral
Caffeine	194	N/A	-0.07	-0.45	3.71	133.30	0.98	0.87	0.56	0.70	HL-neutral
Metronidazole	171	N/A	-0.02	-0.27	6.30	117.80	0.93	0.90	0.48	0.66	HL-neutral
Phenazone	188	N/A	0.38	0.54	4.44	162.70	1.17	0.78	0.56	0.66	HL-neutral
Sulfamethoxazole	253	5.7	0.89	-0.45	7.34	173.10	1.33	0.71	0.58	0.64	HL-ionic
Naproxen	230	4.3	3.18	0.34	2.55	192.20	1.37	0.78	0.75	0.76	HP-ionic
Ibuprofen	206	4.3	3.97	0.77	4.95	200.30	1.39	0.73	0.55	0.64	HP-ionic
Carbamazepine	236	N/A	2.45	2.58	3.66	186.50	1.20	0.92	0.58	0.73	HP-neutral
Atrazine	216	N/A	2.61	2.52	3.43	160.07	1.26	1.00	0.55	0.74	HP-neutral
17 β-estradiol	272	10.3	4.01	3.94	1.56	232.60	1.39	0.85	0.65	0.74	HP-neutral
Estrone	270	10.3	3.13	3.46	3.45	232.10	1.39	0.85	0.67	0.76	HP-neutral
Nonylphenol	220	10.3	5.71	5.88	1.02	236.20	1.79	0.75	0.59	0.66	HP-neutral
Bisphenol A	228	10.3	3.32	3.86	2.13	199.50	1.25	0.83	0.75	0.79	HP-neutral

^a ADME/Tox Web Software.

^b Chem3D Ultra 7.0.

^c Molecular modeling Pro.

^d Experimental database: SRC PhysProp Database.

^e HL = Hydrophilic, HP = Hydrophobic, log $K_{ow} \ge 2$ were hydrophobic; log $K_{ow} < 2$ were hydrophilic; ionic if ionic speciation occurred at pH 7, otherwise neutral.

characteristics, physicochemical compound properties, water chemistry and operating conditions [7-9]. The substantial number of variables makes the quantification of removals of emerging contaminants by NF/ RO membranes a difficult task. In that sense, this publication uses a quantitative structure-activity relationship (QSAR) approach to facilitate modelling of removals of emerging contaminants for an internal data set and subsequently predict removals for external data, only NF membranes are considered in this approach. Sawyer et al. [10] defines QSAR as a method that relates an activity of a set of compounds quantitatively to chemicals descriptors (structure or property) of those compounds. The principle of QSAR is to relate an activity to a function of structure and conditions of a process as described by hydrophobicity, steric properties and other attributes. Previous studies in drug discovery, medicinal chemistry and RO membranes have applied QSAR for the development of models to find relationships between membranes and organic compounds [11–13].

2. Experimental and methodology

Emerging organic contaminants were selected on the basis of their occurrence in water environments impacted by wastewater treatment plant effluents and

their physicochemical properties arranged in an adequate range for classification. Two thin film aromatic polyamide composite NF membranes were selected for this study (NF-200 and NF-90, Dow-Filmtec). The list of emerging contaminants with their respective physicochemical properties is presented in Table 1. The pharmaceutical compounds (caffeine, sulfamethoxazole (SFM), acetaminophen, phenacetin, phenazone, carbamazepine, naproxen, ibuprofen, metronidazole) and endocrine disrupting compounds (17β-estradiol, estrone (E1), bisphenol A, nonylphenol, atrazine) were purchased from Sigma-Aldrich (Schnelldorf, Germany). Potassium chloride, sodium hydroxide, hydrochloric acid and magnesium sulphate anhydrous were purchased from J.T. Baker (Deventer, Netherlands). 17β-Estradiol, E1, bisphenol A and nonylphenol were analysed by gas chromatography/mass spectrometry GC/MS after solid phase extraction (SPE) and silylation. Acetaminophen, phenacetine, phenazone, carbamazepine, naproxen, ibuprofen were analysed by liquid chromatography LC/MS-MS after SPE; and caffeine, metronidazole and SFM were also analysed by LC/MS-MS after SPE; however, the SPE material, separation column and gradient differed from the rest of pharmaceuticals. Concentrations of atrazine were determined using microplate enzyme-linked immunoabsorbent assay (ELISA) kits (Abraxis LLC,



Fig. 1. Scheme of experimental setup.

Warminster, PA). The detection limit was 10 ng/L per compound (except for atrazine, 0.04 μ g/L). The uncertainty of estimates was of $\pm 15\%$ (also for atrazine) as reported by TZW (Karlsruhe, Germany), the laboratory that performed the analyses. More details of analysis protocols were previously reported [14–16].

The experimental setup consisted of two filtration SEPA CF II cells and cell holders in parallel, two hydraulic pumps, a 60L stainless steel tank, a positive displacement pump, a frequency driver, a chiller/heater, control needle valves, pressure gauges, flow meters, a proportional pressure relief valve, stainless steel tubings, a digital balance and, a computer for flow rate data acquisition. A scheme of the experimental setup is presented in Fig. 1. The experiments were conducted in a recycle mode in which permeate and concentrate were recirculated into the feed tank for the first 72 hours (a pre-equilibration period); then, permeate was collected within the next 24 h. The feed solution of all the experiments contained a cocktail of 14 compounds (concentration ranging from 6.5 to 65 μ g/L). The operating pressures were in the range of 276 to 482 kPa and, the fluxes between 4.32 and 30.22 L/m^2 h. All the experiments were carried out at a controlled temperature of 20 °C, pH 7 and ionic strength of 10 mM as KCl.

The pH of the solutions was measured using a calibrated Metrohm 691 pH-meter (Metrohm AG, Herisau, Switzerland); the electrical conductivity and temperature were measured with a WTW Cond 330i (WTW GmbH, Weilheim, Germany) portable conductivity meter. Membranes were characterized to determine magnesium sulphate salt rejection (SR) at standard conditions specified by manufacturers (2,000 mg/L, 25 °C, pH 8, 1034 kPa and recovery of 15%). To determine the hydrophobicity of membranes, contact angles of clean and fouled membrane surfaces were measured with CAM200 optical contact angle meter (KSV Instruments, Finland) at Delft University of Technology; to measure contact angle, the sessile drop method was used. Surface charge, in terms of zeta potential (ZP), of clean and fouled membranes was quantified using ELS-8000 ZP analyzer (Otzuka Electronics, Japan). The ZP analyses were performed at pH 7 and ionic strength of 10 mM KCl.

3. Results and discussion

To develop the QSAR model, 106 rejection cases (defined as internal experimental data set) of emerging contaminants by NF-90 and NF-200 considering 21 variables were analyzed with principal component analysis (PCA) and multivariate regressions. The variables considered as compound descriptors were molecular weight (MW), solubility, log K_{ow}, log D, dipole moment, molar volume, molecular length, molecular width, molecular depth and equivalent width; variables describing membrane characteristics were molecular weight cut-off (MWCO), pure water permeability (PWP), SR, charge of the membrane as ZP, and hydrophobicity as contact angle (CA); variables describing operating conditions were operating

rubic L	Ta	ble	eΖ	2
---------	----	-----	----	---

Data	rango	of mom	rano ch	aractoristics	operating	conditions	and ro	inctions
Data	range c	и шени	name ci	laracteristics,	operating	contantions	anu re	jections

Variable	Unit	Min. value	Max. value	
Molecular weight cut-off (MWCO)	Da	200	300	
Pure water permeability (PWP)	L/m ² /day/kPa	0.86	2.23	
Salt rejection (SR) ^a	_	0.96	0.98	
Zeta potential (ZP)	mV	-48.04	-10.78	
Contact angle (CA)	0	39.30	58.00	
Pressure (P)	kPa	275.79	482.63	
Cross flow velocity (v)	cm/s	0.50	4.46	
Back diffusion mass transfer coefficient (k)	cm/s	2.70E-04	5.99E-04	
Flux (J)	$L/m^2 h$	4.32	30.22	
Hydrodynamic ratio (Jo/k)	_	1	2	
Recovery	%	3	8	
Rejection	%	17.70	99.00	

^a 2,000 mg/L MgSO₄, 25 °C, recovery 15%, pressure 1034 kPa, pH 8.

pressure (*P*), cross flow velocity (*v*), back diffusion mass transfer coefficient (*k*), flux (*J*), ratio of pure water permeation flux J_0 and back diffusion mass transfer coefficient (J_0/k) and recovery. The range of values for membrane characteristics, operating conditions and rejections is presented in Table 2.

Three components were extracted (Table 3) after PCA with SPSS Statistics 16.0. It means that there are 3 components that define our initial database of 21 variables with 11 variables describing three relations namely membrane/operating-conditions (comp. 1), hydrophobicity (comp. 2) and size (comp. 3). The final 3 components accounted for 89.3% explanation of total variance as can be seen in Table 4. The size of the component loadings (correlation coefficients between the variables and the components they represent) higher than 0.33 represent each component in Table 3.

Table 3 Rotated component loadings

	Component				
	1	2	3		
I	.979				
PWP	.967				
SR	.949				
ZP	936				
k	.936				
v	.880				
Length		.951			
log K _{ow}		.930			
$\log D$.867			
eqwidth			.972		
depth			.910		

Variables with large loadings indicate that they are representative of the component, while small loadings (less than 0.33, in blank) suggest that they are not. The reason for using the ± 0.33 criterion is that if the value is squared, the squared value represents the amount of the variable's total variance accounted for by the component. Therefore, a component loading of 0.33 denotes that approximately 10% of the variable's total variance is accounted for by the component [17].

After PCA, multiple linear regression (MLR) and multiple non-linear regression (NLR) were used to develop the QSAR model. Variables (log K_{ow}, log D, length, depth, eqwidth, PWP, SR, ZP, v, k, J) corresponding to the three components were used to model rejection using MLR and multiple NLR. The best linear equation for the MLR model included variables SR, eqwidth (equivalent width), log D, length and depth; the linear QSAR model (linear equation) presented an R^2 of 0.75. However a multiple NLR equation was able to model rejection with an R^2 of 0.84, the equation for the non-linear QSAR model is

$$rejection = \frac{100}{1 + e^{-\alpha}}$$
(1)

Where α is equal to

 $\alpha = 6.283$ length + 19.377 eqwidth + 108.337 SR - 0.443 log D - 8.112 depth - 119.146

Fig. 2 shows modelling rejection results of the nonlinear QSAR model of Eq. (1) with a confidence interval of 95%, only a few rejection modelling cases (four) for phenacetine (PHN), SFM and E1 were outside of the 95% confidence interval, and all corresponded to the

Comp.	Initial eige	envalues		Rotation sums of squared loadings			
	Total	% Var.	Cum. %	Total	% Var.	Cum. %	
1	5.33	48.47	48.47	5.32	48.39	48.39	
2	3.08	28.04	76.51	2.61	23.72	72.11	
3	1.41	12.82	89.33	1.89	17.22	89.33	

Table 4 Extracted components and total variance explained^a

^a Extraction method: principal component analysis.

NF-200 membrane, the explanation for these cases could be related to the higher MWCO of NF-200 (300 Da) that allowed more/less passage of the compounds or analytical quantification errors (\pm 15%) of concentrations in permeate.

The coefficient α of Eq. (1) that determines its asymptotic behaviour with rejections up to 100% can be mechanistically interpreted. The mechanism of steric hindrance due to size exclusion has been recognized as a main cause of rejection in many studies [18–21]. When α increases by the effect of size, which is explained by the positive coefficients of length and equivalent width, rejection will also increase accordingly. In contrast, α shows a negative coefficient for log D, thus it will decrease rejection, which clearly states that the effect of hydrophobicity lessens rejection due to adsorption and subsequent partitioning mechanisms. It is important to mention that log D is assuming the role of hydrophobicity for neutral and ionic compounds; compounds with high log D values will adsorb to the membrane and partition after saturation. Ionic compounds will present very low or negative log *D* values indicating not adsorption onto the membrane. Hydrophobicity influences rejection after adsorptive interactions with the membrane; this fact has been recognized in some studies [18,22,23]. The role of depth in α will only compensate size exclusion contributions of length and equivalent width in a final α value and subsequent rejection. The equation for α also shows that SR is a parameter incorporating steric/size hindrance and electrostatic repulsion effects related to charge of the membrane and operating conditions. The R^2 (goodness of fit of internal data set) is the most widely used measure of the ability of a QSAR model to reproduce the internal data set, but does not explain its robustness and prediction of an external data set. One technique to evaluate prediction is the leave-oneout cross-validation technique, in which one case at a time is iteratively held-out from the training set and the rest is used for model development and the excluded case is predicted by the developed model [24]. According to Gramatica [24], the predictive power of a model may be estimated by the goodness of prediction parameter Q² leave-one-out (1-PRESS/TSS, where PRESS is the predictive error sum of squares and TSS is the total sum of squares). In general, a $Q^2 > 0.5$ is regarded as good and $Q^2 > 0.9$ as excellent [25]. For the developed QSAR models, the linear model presented a Q^2 leave-one-out of 0.72, and the non-linear model presented a Q² leave-one-out of 0.8. Therefore, after internal cross-validation it was demonstrated that the linear and non-linear QSAR models were suitable to model external rejections. External validation of the linear and non-linear QSAR model was implemented with independent prediction of rejections for an external database of experiments performed with different compounds and by three NF aromatic polyamide membranes. One NF membrane was of the same manufacturer (NF-90, different batch) and the other two NF membranes were of different manufacturers (TS-80, Trisep, and Desal HL, GE Osmonics). Experimental



Fig. 2. Non-linear quantitative structure-activity relationship (QSAR) model of experimental internal database.



◆ Filmtec NF-90 ■ Trisep TS-80 ▲ Desal HL

Fig. 3. Predictions of external database with linear quantitative structure-activity relationship (QSAR) model.

methodologies and details about the external data has been reported by Kim et al. [26], Yangali-Quintanilla et al. [27] and Verliefde et al. [28,29]. The external data set was generated under comparable experimental conditions.

Fig. 3 illustrates results of measured rejections vs. predicted rejections for the linear QSAR model. Fig. 4 illustrates results of measured rejections vs. predicted rejections for the non-linear QSAR model.

In order to determine which of the model was the best model for prediction; an error parameter was determined. The standard deviation of error (STDE), measured as a percentage, provides an unbiased measure of model performance compared to the regression coefficient (R^2) that only measures the regression response between predicted and measured rejections. The R^2 for predictions of the linear QSAR model was 0.88 with an STDE of 9% (Fig. 3). The best model was the non-linear QSAR model, with an R^2 of 0.93 and an STDE of 7% (Fig. 4). It is important to mention that organic compounds with rejections of less than 40% have been identified (labelled) in the figures. Chloroform (CF), perchloroethene (PCE), carbon tetrachloride (CT), 2-methoxyethanol (MET), ethanol (ETH), 2ethoxyethanol (EET), 2-(1H)-Quinoline (QNL), glycerol (GLY) and N-Nitrosodimethylamine (NDMA) were low molecular weight compounds used in external experiments. It can be observed that the model has been able to extrapolate rejections of small compounds not present during the developmental phase of the



◆ Filmtec NF-90 ■ Trisep TS-80 ▲ Desal HL

Fig. 4. Predictions of external database with non linear quantitative structure-activity relationship (QSAR) model.

model. An improved model can be obtained when those compounds are included in the initial data set that defines the model. Although the model has been demonstrated to be valid, the limitation of the model is that it will be valid under comparable experimental conditions and only for aromatic polyamide NF membranes. Nonetheless, its applicability and approach can be of value for the construction of a robust model for pilot and full-scale NF applications.

4. Conclusions

- A QSAR model with physicochemical descriptors, membrane characteristics and operating conditions can be able to model and predict rejection of emerging organic contaminants.
- A non-linear QSAR model showed better performance than a linear QSAR model.
- The accuracy of the models can be further improved with additional experimental data.
- The limitation of the model is that experimental conditions such as pH, pressure, and more importantly the type of membrane used must be comparable.

Acknowledgements

This work was supported by Delft Cluster and EU Techneau Project. We greatly acknowledge the contributions of Tae-Uk Kim and Arne Verliefde for providing data and details of their research. The authors thank Filmtec (Dow Chemical Co.) for donating the membranes, Dr. Jaeweon Cho of GIST (Korea), Dr. Frank Sacher of TZW (Germany) and Dr. Steven Mookhoek of TU Delft (Netherlands) for contributing with analytical results and facilities.

References

- S. Dagnino, B. Picot, A. Escande, P. Balaguer and H. Fenet, Desal. and Wat. Treat., 4 (2009) 93-9.
- [2] D. Dolar, K. Kosutic, D.M. Pavlovic and B. Kunst, Desal. and Wat. Treat., 6 (2009) 197-203.
- [3] S. Zorita, L. Mårtensson and L. Mathiasson, Sci. Total Environ., 8 (2009) 2760-2770.
- [4] M.J. Benotti, R.A. Trenholm, B.J. Vanderford, J.C. Holady, B.D. Stanford and S.A. Snyder, Environ. Sci. Technol., 3 (2008) 597-603.
- [5] F. Pomati, S. Castiglioni, E. Zuccato, R. Fanelli, D. Vigetti, C. Rossetti and D. Calamari, Environ. Sci. Technol., 7 (2006) 2442-2447.
- [6] A. Aschengrau, J. Weinberg, P. Janulewicz, L. Gallagher, M. Winter, V. Vieira, T. Webster and D. Ozonoff, Environ. Health, 8 (2009).
- [7] A.I. Schafer, L.D. Nghiem and T.D. Waite, Environ. Sci. Technol., 1 (2003) 182-188.
- [8] K. Kimura, G. Amy, J.E. Drewes, T. Heberer, T.U. Kim and Y. Watanabe, J. Membr. Sci., 1–2 (2003) 113-121.
- [9] V. Yangali-Quintanilla, A. Sadmani, M. McConville, M. Kennedy and G. Amy, Water Res., 9 (2009) 2349-2362.
- [10] C.N. Sawyer, P.L. McCarty and G.F. Parkin, Chemistry for environmental engineering and science, 5th ed., McGraw-Hill, Boston, 2003.
- [11] S. Ren, A. Das and E.J. Lien, J. Drug Target., 2 (1996) 103-107.

- [12] M. Fujikawa, K. Nakao, R. Shimizu and M. Akamatsu, Bioorg. Med. Chem., 11 (2007) 3756-3767.
- [13] P. Campbell, R. Srinivasan, T. Knoell, D. Phipps, K. Ishida, J. Safarik, T. Cormack and H. Ridgway, Biotechnol. Bioeng., 5 (1999) 527-544.
- [14] F. Sacher, F.T. Lange, H.J. Brauch and I. Blankenhorn, J. Chrom., 1–2 (2001) 199-210.
- [15] F. Sacher, M. Ehmann, S. Gabriel, C. Graf and H.J. Brauch, J. Environ. Monit., 5 (2008) 664-670.
- [16] C. Schlett and B. Pfeifer, Vom Wasser, (1996) 327-333.
- [17] R. Ho, Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS, Chapman & Hall/CRC, Boca Raton, FL, 2006.
- [18] Y. Kiso, Y. Sugiura, T. Kitao and K. Nishimura, J. Membr. Sci., 1– 2 (2001) 1-10.
- [19] B. Van der Bruggen, J. Schaep, D. Wilms and C. Vandecasteele, J. Membr. Sci., 1 (1999) 29-41.
- [20] H. Ozaki and H. Li, Water Res., 1 (2002) 123-130.
- [21] L.D. Nghiem, A.I. Schafer and M. Elimelech, Environ. Sci. Technol., 19 (2005) 7698-7705.
- [22] K. Kimura, S. Toshima, G. Amy and Y. Watanabe, J. Membr. Sci., 1–2 (2004) 71-78.
- [23] L.D. Nghiem, A.I. Schafer and M. Elimelech, Environ. Sci. Technol., 6 (2004) 1888-1896.
- [24] P. Gramatica, QSAR & Comb. Sci., 5 (2007) 694-701.
- [25] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell and P. Gramatica, Environ. Health Perspect., 10 (2003) 1361-1375.
- [26] T.U. Kim, J.E. Drewes, R.S. Summers and G.L. Amy, Water Res., 17 (2007) 3977-3988.
- [27] V. Yangali-Quintanilla, T.-U. Kim, M. Kennedy and G. Amy, Drink. Wat. Eng. and Sci., 1 (2008) 7-15.
 [28] A.R.D. Verliefde, E.R. Cornelissen, S.G.J. Heijman, J.Q.J.C. Ver-
- [28] A.R.D. Verliefde, E.R. Cornelissen, S.G.J. Heijman, J.Q.J.C. Verberk, G.L. Amy, B. Van der Bruggen and J.C. van Dijk, J. Membr. Sci., 1 (2008) 52-66.
- [29] A.R.D. Verliefde, S.G. Heijman, E.R. Cornelissen, G. Amy, B. Van der Bruggen and J.C. van Dijk, Water Res., 15 (2007) 3227-3240.