Chemical Engineering Journal 342 (2018) 372–385

Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

Predicting hydrolysis kinetics for multiple types of halogenated disinfection byproducts via QSAR models

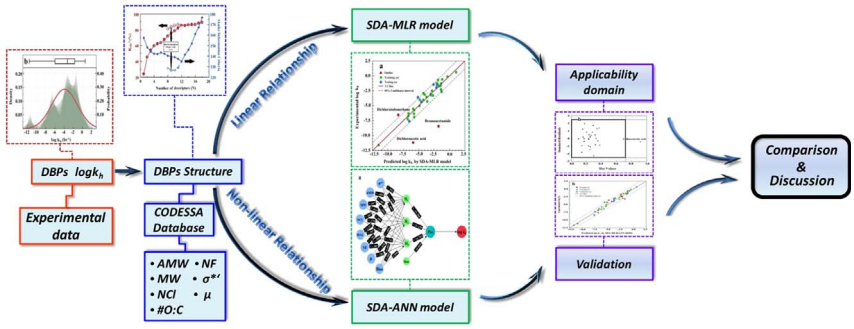
Lei Wang, Baiyang Chen*, Tian Zhang

Shenzhen Key Laboratory of Organic Pollution Prevention and Control, State Key Laboratory of Urban Water Resource and Environment of the Harbin Institute of Technology, Shenzhen 518055, China

HIGHLIGHTS

- Hydrolysis rate constants of 40 DBPs from 7 chemical classes are modeled.
- Unlike before, the QSAR models herein consider multiple types of DBPs together.
- 546 descriptors are evaluated but 7 are ultimately selected via SDA-MLR process.
- Both MLR and ANN models finalized exhibit excellent predictability and fitness.
- ANN outruns MLR mostly, but MLR performs better for three types of DBPs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
DBP
Hydrolysis rate constant
QSAR
MLR
ANN

ABSTRACT

In water and wastewater engineering systems, hydrolysis is one of the most important means in abiotic degradation of disinfection byproducts (DBPs). Enhanced knowledge of hydrolysis of DBPs can help determine the likelihood of DBP occurrence in water and their risks. In order to better understand the roles that functional groups play in the fate and occurrence of DBPs in the environment, this study developed several quantitative structure-activity relationship (QSAR) models to estimate the hydrolysis rate constants for a pool of 40 halogenated, small molecular weight DBPs originating from seven chemical classes. The models are based on the descriptors calculated and filtered by stepwise discriminant analysis with multiple-linear regression (MLR) and artificial neural network (ANN) algorithms, and overall they exhibited better performance than models using conventional descriptors only. The relative importance of selected descriptors are ranked as average mole weight > number of fluorine > polar effect of functional group > mole weight > number of chlorine > dipole moment > the ratio of oxygen to carbon atoms. The squared regression coefficients between predicted and experimental values were 0.939 and 0.976 for the final MLR and ANN models, respectively. While the ANN model demonstrated better performance than the MLR model for all 40 DBPs, the MLR model shows more accurate predictions for haloethanes, haloacetic acids, and haloacetamides. Unlike existing QSAR models, which treat each type of DBP separately, the models developed in this study are unique in considering multiple types of DBPs together; therefore, these models may help researchers to better understand the effects of not only halogens but also functional groups on DBP stability.

* Corresponding author.
E-mail address: chen.baiyang@hit.edu.cn (B. Chen).

https://doi.org/10.1016/j.cej.2018.02.106
Received 4 January 2018; Received in revised form 23 February 2018; Accepted 24 February 2018
Available online 26 February 2018
1385-8947/

Nomenclature

AIC	the Akaike Information Criteria
AMW	average mole weight
ANN	artificial neural network
ARE	absolute relative error
B3LYP	the Becke's three-parameter hybrid with Lee-Yang-Parr's correlation functions
DBP	disinfection by-product
E_{HOMO}	energy of highest occupied molecular orbital
E_{LUMO}	energy of lowest unoccupied molecular orbital
Es	steric effect constant of halogen
Es'	steric effect constant of functional groups
F	Fisher test statistical value
h^*	leverage threshold value
HAA	haloacetic acid
HAcam	haloacetamide
HAL	haloaldehyde
HAN	haloacetonitrile
HE	haloethane
HK	haloketone
HM	halomethane
HNM	halonitromethane
k_h	hydrolysis rate constant
MLR	multiple-linear regression
MW	molecular weight
NBr	number of bromine
NCl	number of chlorine
NF	number of fluorine
NI	number of iodine
OECD	Organization for Economic Cooperation and Development
PRESS	predictive residual sum of squares
$Q_{ext.1}^2$	fitness of the first external validation function
$Q_{ext.2}^2$	fitness of the second external validation function
$Q_{ext.3}^2$	fitness of the third external validation function
Q_{LOO}^2	fitness of leave one out internal validation approach
QSAR	quantitative structure–activity relationship
R^2	squared regression coefficient
$R_{adj.}^2$	adjusted squared regression coefficient
RMSE	root mean square error
$RMSE_{LOO}$	root mean square error of leave one out internal validation approach
$RMSE_{PRESS}$	root mean square error of predictive residual sum of squares
SDA	stepwise discriminant analysis
THM	trihalomethane
σ^*	polar effect constant of halogen
$\sigma^{*\prime}$	polar effect constant of functional group
μ	total molecular dipole moment
#O:C	the ratio of oxygen to carbon atom number

1. Introduction

Since the identification of chloroform as a byproduct of chlorinated water in the 1970s, disinfection by-products (DBPs) have become a hot topic in many research fields [1,2]. Currently, DBPs are often categorized into two groups: traditional, regulated DBPs (e.g., trihalomethanes [THMs] and haloacetic acids [HAAs]) and emerging, non-regulated DBPs (such as iodinated THMs and HAAs, haloketones [HKs], haloaldehydes [HALs], haloacetonitriles [HANs], halonitromethanes [HNMs], and haloacetamides [HAcams]). Based upon a survey in the U.S., iodo-THMs were found to exist at a level of up to 15 ppb at a plant using chloramine as disinfectant; dihalo-HALs were detected at up to 16 ppb at a plant applying chloramine and ozone; and HNMs were observed at up to 3 ppb at a plant employing ozonation followed by chlorination or chloramination [3]. The widespread occurrence of DBPs motivates numerous studies, along with the researches driven by potential toxicities of DBPs. Previous laboratory work has shown that many DBPs in chlorinated water are cytotoxic and genotoxic to Chinese hamster ovary cells [4,5] and some of them are proven to be carcinogenic to animals or possibly carcinogenic to humans [6,7]. Meanwhile, epidemiological studies have demonstrated positive associations between THM concentrations and bladder cancer in men [7,8]. Similarly positive associations between exposure to chlorinated drinking water and risk of disruption of in utero development and adverse birth outcomes have also been reported [6]. Thus, it is important to control the occurrence of DBPs in drinking water and the environment.

Among all DBP transformation processes potentially encountered in surface, ground, and tap waters, hydrolysis is one of the most fundamental mechanisms with numerous implications to the fate and transport of compounds [9–11]. An understanding of hydrolysis of DBPs, especially the rate a toxic chemical transform to a less harmful chemical, is therefore a key to evaluate and regulate DBPs in drinking and waste water. According to the literature, hydrolysis of DBPs has been commonly observed, and the hydrolysis kinetics are sometimes described as pseudo first-order reactions that are likely dependent upon the effects of pH, temperature, initial concentration, and coexisting compounds. Several studies have further investigated the hydrolysis product(s) of DBPs. For example, El Din et al. [12] have identified methanol and halide as the hydrolysis products of trihalomethane under alkaline conditions; Zhang and Minear reported that trihaloacetic acids were hydrolyzed by cleaving the C—C bond to form THMs and halides [13]; and hydrolysis of HALs yielded THMs [14–16]. More recently, Yu and Reckhow [17] demonstrated that the pathways of HAN hydrolysis under neutral or basic conditions mainly resulted in formation of HAcams, which eventually led to HAAs [17,18]. In order to better capture the trends of hydrolysis kinetics, an earlier study summarized several types of DBPs kinetics using quantitative structure-activity relationship (QSAR) models [19]. However, these models are group-specific and only take into account the roles of halogens for each type of DBP, thus lacking the ability to predict the hydrolysis rate constants for other type(s) of DBPs with different functional groups.

In recent years, researchers have become increasingly interested in the uses of QSAR approach in DBP studies to establish relationships between experimental observations and molecular properties [20–24]. The Organization for Economic Cooperation and Development [25] believes that QSAR models have the potential to estimate the environmental and human health risks of various chemicals. However, to date, there are only around fifteen QSAR or linear free-energy models (shown in Table 1) developed specifically for the hydrolysis of DBPs. These models either integrate just one or two quantum-chemical descriptors for one type of DBPs, or involve the uses of geometric descriptors, such as polar ($\boldsymbol{\sigma}^*$) and steric constants (***Es***), to represent the polar and steric effects of DBPs during the QSAR development process (Table 1), but lack consideration of other significant descriptors or other types of structurally similar compounds. Taking HAN as an example [18], the hydrolysis rate constants ($\boldsymbol{k_h}$) of seven HANs were found to increase with $\boldsymbol{\sigma}^*$ and ***Es*** ($R^2 = 0.95$; Table 1), and such trend has been well-modeled in earlier studies authored by Chen [19] and Yu and Reckhow [17]. A similar simulation practice has also been applied to correlate the $\boldsymbol{k_h}$ of trihaloacetic acids with $\boldsymbol{\sigma}^*$ and ***Es*** [13]. However, none of them have accounted for the key functional groups other than halogens, thus these models remain inapplicable by other compounds with different functional groups but identical halogen substitutions. Therefore, it is necessary to develop more advanced models to estimate the hydrolysis kinetics for multiple types of halogenated DBPs simultaneously.