Chinese Journal of Chemical Engineering 26 (2018) 1570–1580

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Article

On the database-based strategy of candidate extractant generation for de-phenol process in coking wastewater treatment☆

Ran Xu [1,2,3], Yuehong Zhao [1,2,3,*], Qingzhen Han [1,2], Xinyu Liu [1,2], Hongbin Cao [1], Hao Wen [1,2]

[1] Division of Environmental Engineering and Technology, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China
[2] State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 9 November 2017
Received in revised form 15 December 2017
Accepted 31 January 2018
Available online 9 February 2018

Keywords:
Existing commercial compounds database
De-phenol extractant
Database-based generation strategy
Environmental risk

ABSTRACT

A database-based strategy of candidate generation was proposed for molecular design of new de-phenol extractants following the idea of finding new applications of existing commercial compounds. The strategy has the advantage that the environmental, safety and health risks of candidate compounds are known and controllable. In this work, the Existing Commercial Compounds (ECC) database and special combined search strategy were developed as the base for the proposed CAMD method following such idea, and molecules for phenol extraction used in coking wastewater treatment were selected from the ECC database. The candidate solvents cover the following categories: ketones, esters, ethers, alcohols, anhydrides and benzene compounds, which are consistent with the de-phenol extractants commonly used in the industry or experiment. The compounds with higher partition coefficient and selectivity than widely used methyl isobutyl ketone (MIBK) are mainly ketones. 26 obtained molecules show higher partition coefficient and selectivity than MIBK, which are suggested to be further investigated by experiment. Furthermore, analysis of these potential molecules may present the effective functional groups as the initial group set to generate new molecular structures of de-phenol extractants. The results show that the proposed method enables us to efficiently generate chemicals with benefits of less time, less economical cost, and known environmental impact as well.

1. Introduction

Phenols are a series of organic pollutants with strong biological toxicity and exist in the wastewater from petroleum, petrochemical, coal conversion, phenol-producing, and phenolic resin industries [1]. Extraction is used as an effective method of gathering and recovering phenols and other organic pollutants in wastewater, where benzene, octanol, isopropoxypropane, methyl isobutyl ketone (MIBK), pentyl acetate, ethyl ethanoate, isopropyl ethanoate, *etc.* are commonly used extractants [2]. The efficiency improvement of the extraction with lower loss of extractants is, therefore, focused on the design and screening of new extractants.

Computer-aided molecular design (CAMD) [3] is becoming the most potential methodology for design and selection of chemicals in the last two decades, besides the traditional experimental synthesis and screening. CAMD is a method to determine the structures of feasible chemical compounds that meet the target properties. It combines property prediction methods with computer-assisted search in the design of various chemical products. Great progresses on the property prediction models and methodology of CAMD, generate-and-test [4,5] and mathematical programming approach [6–9], were made with applications for chemical development in a variety of fields, such as extractants [10], absorbents [11], crystallization [12] and reaction solvent selection [13,8], refrigerants [14], catalysts [15] and polymers [16].

CAMD generates a list of chemical candidates with reasonable accuracy within a moderate time scale. However, new candidates may have unknown environmental, safety and health risks, such as toxicity, body contact, emergency risk, and potential health effects. Assessment of these unknown risks of a chemical before its production and application will take much time and efforts. In Europe, the cost of registering chemicals to comply with REACH could exceed 2.1 billion Euros, based on about 30000 substances [17]. In China, the reporting period of the new chemical substances will take 8 to 56 months [18].

In addition, the choice of functional groups for candidate structure generation is highly dependent upon the experience of researchers [4], which brings uncertainties in the type and total number of initial functional groups, and the complexity of structure generation. The

☆ Supported by the National Natural Science Foundation of China (Grant No. 2156112001) and National Basic Science Data Sharing Service Project (DKA2017-12-02-05).
* Corresponding author at: Division of Environmental Engineering and Technology, Institute of Process Engineering, Chinese Academy of Sciences, No.1, 2nd North Street, ZhongGuanCun, Beijing 100190, China.
E-mail address: yhzhao@ipe.ac.cn (Y. Zhao).

https://doi.org/10.1016/j.cjche.2018.01.014
1004-9541/

preselection of functional groups proposed by Song [19] for extractant design is, based on the interactions between functional groups and solvent selectivity, helpful to limit the type of initial functional groups.

In this work, a strategy of candidate generation is attempted in molecular design of new de-phenol extractants, by selecting potential molecules from the existing commercial compounds database. The advantage of this strategy is that the candidate molecules selected from the existing commercial compounds database have full information on the environmental, safety and health risk. In other words, the strategy may present the possibility of finding new applications of existing commercial compounds. Furthermore, analysis of these potential molecules may present the effective functional groups as the initial group set to generate new molecular structures of de-phenol extractants.

2. Candidate Extractant Generation from the Existing Commercial Compounds Database

2.1. Database-based generation strategy

In order to avoid unknown risk of new candidates generated from CAMD, we proposed a database-based generation strategy of candidate extractants. The strategy is based on the existing commercial compounds to design/select new de-phenol extractants. The existing commercial compounds have the advantage that the environmental, safety and health risks of candidate compounds are known and controllable, such as toxicity, body contact, emergency risk, and potential health effects. The environmental, safety and health impact of all these compounds can be obtained from many open databases *via* CAS number, such as the Material Safety Data Sheet (MSDS) [20].

By collecting the molecules in chemical inventories, the Existing Commercial Compounds (ECC) database was built. The proposed ECC database was detailed along with its structural information, property estimation methods, and estimation procedure *via* the group contribution methods and group match toolkit. By setting multiple property criteria for extraction, a list of potential molecules with high de-phenol performance may be generated. Therefore, the strategy also has the possibility of finding new applications of existing commercial compounds.

The promising molecules will be ranked and compared in terms of the extraction performance. Due to the rich known data of storage, transportation, personnel protection, accident emergency, waste disposal, *etc.*, candidate molecules generated from existing commercial compounds database can be used directly for experimental verification and further industrial production, which will save much economic and time cost. By analyzing the candidate molecular structures, the effective functional groups were obtained, which are useful as initial groups to generate new molecular structures of extractants.

The removal of phenol from coking wastewater by liquid–liquid extraction is the key step for such waste water treatment. In order to find potential molecules for phenol extraction process, a framework following database-based generation strategy was proposed as shown in Fig. 1. It can be carried out by 3 steps: (1) Development of the existing commercial compounds database; (2) generation of candidate de-phenol extractants; and (3) analysis on the generated candidate extractants. The details are given in the following parts.

2.2. Development of the existing commercial compounds database

In order to obtain the compounds with known environmental, safety and health risk, the ECC database was developed. The data sources of commercial compounds are from the inventory of existing chemical substances, involving the European Inventory of Existing Commercial Chemical Substances (EINECS), Toxic Substances Control Act Chemical Substance Inventory (TSCA Inventory), and Inventory of Existing Chemical Substances Produced or Imported in China (IECSC) [21]. The three inventories are the main inventories covering most of the chemicals commonly used in regions and countries.

However, the inventory of existing chemical substances only contains CAS number and chemical substance names, and cannot meet the requirement of extractant selection. In this paper, we used CAS number as input to compound registration system to extract the corresponding structure information (Mol files [22]), which is then inserted into the database for further property estimation for other data needed. The overview of the chemical inventories is listed in Table 1. Finally, a total of 34195 compounds were obtained. The distribution of chemical compounds in three inventories is shown in Fig. 2.

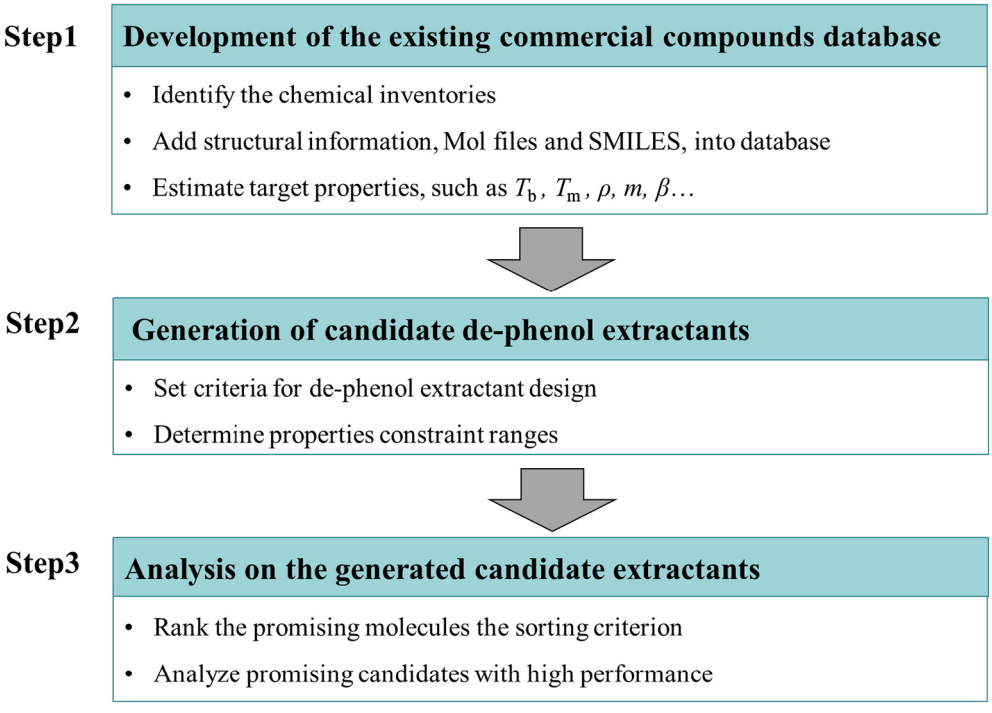

Fig. 1. Framework for candidate extractant generation.