Journal of Chromatography A, 1398 (2015) 66–72

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Using quantitative structure activity relationship models to predict an appropriate solvent system from a common solvent system family for countercurrent chromatography separation☆

Siân Marsden-Jones[a], Nicola Colclough[b], Ian Garrard[a], Neil Sumner[b], Svetlana Ignatova[a,*]

[a] Advanced Bioprocessing Centre, Institute of Environment, Health and Societies, Brunel University, Uxbridge, Middlesex UB8 3PH, UK
[b] AstraZeneca UK Limited, Alderley Park, Cheshire SK10 4TG, UK

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form 26 March 2015
Accepted 8 April 2015
Available online 17 April 2015

Keywords:
Countercurrent chromatography
Centrifugal partition chromatography
Solvent system selection
Method development
Quantitative structure activity relationships (QSAR)

ABSTRACT

Countercurrent chromatography (CCC) is a form of liquid–liquid chromatography. It works by running one immiscible solvent (mobile phase) over another solvent (stationary phase) being held in a CCC column using centrifugal force. The concentration of compound in each phase is characterised by the partition coefficient (K_d), which is the concentration in the stationary phase divided by the concentration in the mobile phase. When K_d is between approximately 0.2 and 2, it is most likely that optimal separation will be achieved. Having the K_d in this range allows the compound enough time in the column to be separated without resulting in a broad peak and long run time. In this paper we report the development of quantitative structure activity relationship (QSAR) models to predict log K_d. The QSAR models use only the molecule's 2D structure to predict the molecular property log K_d.

1. Introduction

Countercurrent chromatography (CCC) was invented in 1966 by Ito et al. [1]. In CCC, the compounds partition between two immiscible liquids (phases). One phase (stationary) is retained inside the column, which is spun in planetary motion. Whereas the other (mobile phase) is pumped through the column. Separation is achieved as compounds that spend more time in the stationary phase take longer to pass through the column than compounds that spend more time in the mobile phase. CCC has many advantages over traditional liquid–solid chromatography including total recovery of compound; also crude samples containing particulates can be separated and higher loading capacities are tolerated [2,3]. CCC is also reproducible and scalable. A disadvantage of CCC is that the choice of the solvent system is currently based on an analyst's past experience, trial and error or literature analysis. This may mean that systems that would give very well defined chromatography are missed or that large quantities of time and solvents/samples are used to select an appropriate solvent system. Being able to predict the K_d values of target compounds would speed up the method development phase of the CCC process without time consuming, solvent intensive experiments.

There have been previous attempts to computationally predict the partition coefficients of compounds. Hopmann et al. [4] used the software COSMO-RS to predict K_d values using activity coefficients of the upper and lower phases. The conformation of the molecule plays a very important role in the calculation and is computationally expensive to calculate, taking up to 9 h for large molecules.

Another method used the UNIFAC (Universal quasichemical functional-group activity coefficients) model which was developed by Li et al. [5]. This model uses thermodynamics to calculate K_d. A potential disadvantage of this programme is its dependency on group interaction parameters which are often limited.

Ren et al. [6] used NRTL-SAC (non-random two liquid-segment activity coefficient) in combination with UNIFAC and GA (generic algorithm) to predict partition coefficients. This method is not purely computational as some experimental K_d values are needed for the prediction. This is a disadvantage if the compound to be separated is expensive or supply is limited.

The modelling approach that has been investigated in this work is quantitative structure activity relationships (QSARs). QSARs are relationships that are used to predict a molecular property, in this

☆ Presented at the 8th International Conference on Countercurrent Chromatography – CCC 2014, 23–25 July 2014, Uxbridge, United Kingdom.

* Corresponding author. Tel.: +44 1895266911; fax: +44 1895274608.
E-mail address: svetlana.ignatova@brunel.ac.uk (S. Ignatova).

http://dx.doi.org/10.1016/j.chroma.2015.04.020
0021-9673/

case log K_d, from a molecule's structure. The log K_d is predicted instead of the K_d value as this normalises the experimental values. QSARs offer fast computational predictions. They rely on molecular descriptors which can be calculated manually (e.g., the number of hydrogen bond acceptors) or from a number of widely available software packages (e.g. ACD Labs logP/D, Daylight/Biobyte ClogP). As long as a complete set of descriptors is available, QSAR models of the type explored in this work can be run in Microsoft Excel or an equivalent. This type of software is owned by the majority of people so using the model would be convenient. This could allow more automation of CCC, increasing the techniques' appeal especially to industry.

The QSAR models that have been built in this study work with the HEMWat solvent systems. It contains heptane (or hexane), ethyl acetate, methanol and water in varying proportions. It is a very versatile system as changing the proportions of each solvent will change the polarity of the overall system as well as the polarity difference between two phases. This control over the polarity allows the system to be adjusted to optimise the partitioning of many different compounds. In the Brunel University CCC literature database containing 2322 papers, 1121 of these (48%) use HEMWat based solvent systems. The next most commonly used solvent system is based on butanol which was used in 542 papers (23%). This implies that HEMWat is the most commonly used solvent system making it ideal for this proof of concept study [7]. Garrard [8] adopted a numeric labelling scale from 6 to 28 to denote polarity, within which HEMWat 6 was the most polar and HEMWat 28 was the least polar. The HEMWat systems denoted 1–6, contain butanol and not always all four of the other solvents. The QSAR approach can be applied to any solvent system. However, in this work we have chosen to focus on the HEMWat solvent system. Traditionally, QSARs have been developed for much simpler two solvent systems such as octanol/water and cylcohexane/water [9]. Therefore, successfully applying the QSAR methodology to the much more complex HEMWat systems would by analogy show that QSARs are likely to be applicable to all solvent system families. Liquid handling robots are commonly used in industry for log P measurements so could easily be used for fast, accurate partition coefficient measurement. Therefore, measuring K_d values for other solvent system families for QSAR generation would not be too time consuming. This work attempts to develop QSAR models to increase the speed and efficiency of solvent selection in CCC. Through the use of a diverse data set to train each HEMWat QSAR, the aim is that the models will be able to accurately predict log K_d values for a large range of molecules.

2. Experimental

2.1. Materials and chemicals

The solvents used were HPLC grade heptane, ethyl acetate, methanol, acetonitrile and ethanol purchased from Fisher Chemicals (Loughborough, UK). The water used was deionised in house using a Purite Select Fusion purification system (Thame, UK). All compounds were purchased from Sigma Aldrich (Gillingham, UK) (including quality control compound, 2-ethylanthraquinone) with a minimum purity of 95%. Ammonia solution (35%) and TFA (99%) were purchased from Fisher Scientific (Loughborough, UK).

2.2. Apparatus

HPLC analysis was conducted on a HP1100 Agilent system (Stockport, UK) with detection at 254, 260, 275, 295, and 310 nm with a Symmetry C_{18} column (75 mm × 4.6 mm I.D., 3.5 μm), (Waters, USA). An Eppendorf Concentrator 5301 (Hamburg, Germany) was used as a centrifuge at 1400 rpm (240 g) at room temperature. The balance was a Sartorius Mechatronics analytical balance 1601A MP8-1 (Epsom, UK) unit with a range from 0.1 mg to 110 g.

Table 1
Ratios of solvents to make up the selected HEMWat solvent systems (heptane, ethyl acetate, methanol and water).

HEMWat system [7]	Heptane (g)	Ethyl acetate (g)	Methanol (g)	Water (g)
8m	1	9	1	9
14m	3	6	3	6
17m	4	4	4	4
20m	6	3	6	3
22m	6	2	6	2
26m	9	1	9	1

2.3. Preparation of two phase solvent system and determination of log K_d

The predictive ability of the QSAR is dependent on the accuracy of the experimentally determined partition coefficient values used to train the model. Therefore physical factors were controlled to minimise the experimental error. These included temperature and pH which were held constant while the compound was in the two phase system. Once each phase had been sampled, it was diluted in ethanol to remove any matrix effect from the solvent system.

To avoid volume variations when preparing the HEMWat solvent systems due to possible temperature fluctuations six HEMWat solvent systems were made up by mass according to Table 1 [8] using thermostated solvents (at 20 °C for 20 min in a water bath) and the mixtures left overnight to equilibrate at room temperature. Before sampling, the solvent systems were placed in a 20 °C water bath for 20 min. As these solvent systems have been made up by mass as opposed to volume, the final percentage composition is slightly different from the conventional HEMWat systems described by Garrard [8]. Therefore, they have been distinguished by the addition of the letter "m" to the HEMWat numbers.

The six HEMWat systems were chosen as they gave a large polarity range across the whole series. To remove the effect of pH on ionising compounds such molecules were converted to their neutral form, acidic compounds were run in acidified HEMWat (0.1% TFA in water, replacing pure water) and basic compounds were run in basified HEMWat (1% ammonia solution in water, replacing pure water). Compounds with a negative ClogP (octanol/water partition coefficient from Biobyte, Inc., Claremont, CA, USA and Daylight, Laguna Niguel, CA, USA) were dissolved in 1.5 ml of the lower phase of HEMWat until the phase was saturated. Compounds with a positive ClogP were dissolved in the upper phase of the HEMWat system until the phase was saturated. This ensured that the maximum amount of compounds was dissolved in the HEMWat system. The solutions were centrifuged (1400 rpm for 30 s) to remove all particulates from the supernatant. An aliquot of 400 μl of supernatant was mixed and centrifuged with 1400 μl of the alternative phase (1400 rpm for 30 s).

An aliquot of 80 μl of the 1400 μl volume phase and 320 μl of the 400 μl phase were separately diluted using 1 ml of ethanol. To avoid cross contamination, before the lower phase was sampled the remaining upper phase was removed by pipette until no upper phase could be seen on visual inspection. The samples were run on a 10 min gradient method on the HPLC using Symmetry C18 column (4.6 mm × 75 mm, 3.5 μm), at 1 ml/min and 40 °C. Mobile phase consisted of 0.1% aqueous trifluoroacetic acid (solvent A) and acetonitrile (solvent B). The gradient elution programme was as follows: 0–6 min, 10% B; 2–8 min, 80% B.