Journal of Chromatography A, 1507 (2017) 53–62

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Use of dual-filtering to create training sets leading to improved accuracy in quantitative structure-retention relationships modelling for hydrophilic interaction liquid chromatographic systems

Maryam Taraji [a], Paul R. Haddad [a,*], Ruth I.J. Amos [a], Mohammad Talebi [a], Roman Szucs [b], John W. Dolan [c], Christopher A. Pohl [d]

[a] Australian Centre for Research on Separation Science (ACROSS), School of Physical Sciences-Chemistry, University of Tasmania, Private Bag 75, Hobart 7001, Australia
[b] Pfizer Global Research and Development, CT13 9NJ, Sandwich, UK
[c] LC Resources Inc., 1795 NW Wallace Rd., McMinnville, OR 97128, USA
[d] Thermo Fisher Scientific, Sunnyvale, CA, USA

ARTICLE INFO

Article history:
Received 28 March 2017
Received in revised form 17 May 2017
Accepted 18 May 2017
Available online 23 May 2017

Keywords:
QSRR
Prediction accuracy
Dual-filtering
Retention prediction
Similarity searching
HILIC mechanism

ABSTRACT

The development of quantitative structure retention relationships (QSRR) having sufficient accuracy to support high performance liquid chromatography (HPLC) method development is still a major issue. To tackle this challenge, this study presents a novel QSRR methodology to select a training set of compounds for QSRR modelling (i.e. to filter the database to identify the most appropriate compounds for the training set). This selection is based on a dual filtering strategy which combines Tanimoto similarity (TS) searching as the primary filter and retention time (t_R) similarity clustering as the secondary filter, using a database of pharmaceutical compound retention times collected over a wide range of hydrophilic interaction liquid chromatography (HILIC) systems. To employ t_R similarity filtering, correlation to a molecular descriptor is used as a measure of retention time. For the retention time of a compound to be modelled a relationship between experimental chromatographic data and various molecular descriptors is calculated using a genetic algorithm-partial least squares (GA-PLS) regression. The proposed dual-filtering-based QSRR model significantly improves the retention time predictability compared to the diverse, global, and TS-based QSRR models, with an average root mean square error in prediction (RMSEP) of 11.01% over five different HILIC stationary phases. The average CPU time for implementing the proposed approach is less than 10 min, which makes it quite favorable for rapid method development in HILIC. In addition, interpretation of the molecular descriptors selected by this novel approach provided some insight into the HILIC mechanism.

Crown Copyright © 2017 Published by Elsevier B.V. All rights reserved.

1. Introduction

The motivation behind the use of quality by design (QbD) principles in conjunction with HPLC method development has been the desire to develop more robust and reliable analytical methods with minimal time and resource effort [1–5]. To this end, theoretical approaches have been employed for method development to propose different retention models that relate retention time/factor of an analyte to properties of the stationary phase, the eluent, and the analyte itself [6–8]. An important computational approach in predicting retention times in HPLC method development is the quantitative structure-retention relationships (QSRR) method, which correlates the retention time of an analyte to its chemical structure [9].

Although QSRR modelling methods have been used for more than 40 years, the retention time prediction accuracy of such models is not often sufficient to make them favorable for HPLC method development purposes [9]. As a consequence, QSRR methodology is still an active research area. A key component that can potentially lead to improvement in QSRR modelling accuracy is the incorporation of appropriate molecular descriptors into the models [5]. Numerous different kinds of molecular descriptors have been reported in QSRR studies, e.g. physicochemical properties [10–12], solvation descriptors [13,14], quantum-chemical indices [15], 2D autocorrelation indices [16], GEometry, Topology, Atom-Weights AssemblY (GETAWAY) descriptors [17] and gonane topological

* Corresponding author.
E-mail address: Paul.Haddad@utas.edu.au (P.R. Haddad).

http://dx.doi.org/10.1016/j.chroma.2017.05.044
0021-9673/Crown Copyright © 2017 Published by Elsevier B.V. All rights reserved.

weighted fingerprints [18]. In addition, several feature selection approaches to capture the most informative molecular descriptors with the goal of producing more predictive models have been reported [6,19,20]. Comparison studies demonstrate that the well-known genetic algorithm (GA) method performs better than other possible feature selection mechanisms [20,21].

Another strategy to enhance the level of confidence in QSRR methodology is the use of the concept of molecular similarity in compound-classification *prior to* QSRR modelling. The essence of the similarity concept is that filtering the database to identify those compounds which are most structurally similar to the target analyte, is also likely to identify those compounds which exhibit similar chromatographic properties to the target analyte leading to more accurate modelling of retention time [22]. Classification according to similarity has proven to be a powerful tool in quantitative structure-property (activity) relationship (QSPR/QSAR) analysis enabling biomarker discovery, mechanistic studies, drug development, and technological evaluations in medicinal and pharmaceutical industries [23–28]. However, the use of the molecular similarity concept for QSRR modelling was seldom reported before Wang et al. [29] presented a compound classification method based on logD profile similarity, resulting in enhanced elution order prediction in acidic and basic chromatographic conditions. Muteki et al. [5] have also assessed the reliability of QSRR prediction and found that QSRR methodology based on compound classification significantly improved retention time prediction in comparison with the models derived from the whole dataset.

Previous work from our group [30] has demonstrated that application of the federation of local models strategy, which involves scanning a database to find those molecules that are most structurally similar to the target analyte and constructing a local model for each target compound based on its top ranked similar molecules, may help to improve the prediction accuracy of QSRR models. This compound-classification-based QSRR strategy successfully utilised Tanimoto [31] cluster analysis to predict retention times of studied test probes in a HILIC database using an amide column [32]. However the further application of the proposed modelling approach in datasets collected from other HILIC stationary phases or other target analytes showed varying degrees of success, possibly due to the complex retention mechanisms at play in the HILIC mode [33].

A comparison of global modelling (using the whole dataset for model production), modelling based on Tanimoto similarity (TS) clustering, and modelling based on a newly proposed retention time (t_R) similarity clustering method applied to a HILIC dataset, has shown that while Tanimoto clustering shows an improvement in error compared to the global model, retention time clustering is by far the most successful method [34]. However, retention time clustering is unable to be applied to a real-life situation because the retention time of the compound under investigation is not known, and so far, no method has been found to successfully utilise retention time clustering as part of a dual-filtering approach.

In this study a novel dual-filtering-based QSRR modelling strategy has been applied successfully to a range of HILIC systems. The proposed dual filtering approach involves selecting structurally similar training neighbours to a target molecule according to calculated Tanimoto pairwise values, followed by further filtering according to t_R similarity found by utilising the correlation of molecular descriptors to retention time. The application of the proposed dual-filtering-based QSRR modelling approach is illustrated by the prediction of retention time for various analytes on HILIC stationary phases, utilising a GA coupled with PLS for variable selection. By using our dual filtering approach, reliable and accurate GA-PLS models have been established over a wide range of HILIC datasets. The performance of the GA-PLS models derived from both diverse and global datasets, and TS-based QSRR models, is also compared to dual-filtering-based QSRR models. Finally, in order to obtain some insight into the HILIC mechanism, the selected molecular descriptors in the dual filtering process have been investigated.

2. Experimental section

2.1. Sample preparation

Analytical grade adrenaline, noradrenaline, isoproterenol, salbutamol, dopamine, tyramine, synephrine, 3-methoxytyramine, norfenefrine, normetanephrine, *N*-methylephedrine, octopamine, salicylic acid, 4-hydroxybenzoic acid, 2,3-dihydroxybenzoic acid, 2,4-dihydroxybenzoic acid, 3,5-dihydroxybenzoic acid, benzoic acid, 3-amino-4-hydroxybenzoic acid, 3-aminobenzoic acid, vanillic acid, syringic acid, 2-methoxybenzoic acid, p-toluic acid, 3-hydroxybenzoic acid, 2,5-dihydroxybenzoic acid, 3,4-dihydroxybenzoic acid, 4-aminobenzoic acid, 4-aminosalicylic acid, 2′-deoxyadenosine, 2′-deoxycytidine, 2′-deoxyguanosine, adenosine, cytidine, guanosine, inosine, thymidine, uridine, 2′-deoxyuridine, acyclovir, guanine, xanthine, caffeine, theophylline, theobromine, hypoxanthine, 1,3-dimethyluric acid, 5-sulfosalicylic acid, mandelic acid, nicotinic acid, p-toluenesulfonic acid, 4-hydroxybenzenesulfonic acid, tropic acid, 4-aminophenylacetic acid, tryptophan, 2-phenylethylamine, phenylalanine, benzyltrimethylammonium (BTMA) chloride, phenyltrimethylammonium (PTMA) chloride, labetalol, nadolol, propranolol, adenine, uracil, thymine, cytosine, pindolol, alprenolol, satolol, atenolol, 4-nitrophenyl-β-D-glycopyranoside, uric acid, vidarabine and tyrosine were purchased from Sigma–Aldrich (St. Louis, MO, USA) and the standards of fenotrole, ritudrine, metaproterenol, isoxuprine, terbutaline, phenylephrine, methoxamine, 5-methylsalicylic acid, 2′,3′-dideoxyadenosine, 3′-deoxyguanosine, 5-methyluridine, 3′-deoxythymidine, 2′-deoxyinosine, pentoxyphylline, diphylline, 7-hydroxyethyl-theophylline, 1-methyluric acid, 1-methyl-guanine, 9-methyl-guanine, 3,7-dimethyluric acid, 7-methyl-xanthine, 1,7-dimethyluric acid, proxyphylline and 1,3,7-trimethyl uric acid were purchased from Santa Cruz Biotechnology Inc. (CA, USA). Acetonitrile and methanol of HPLC grade were supplied by VWR International (Melbourne, VIC, Australia) and Sigma–Aldrich (St. Louis, MO, USA), respectively. Formic acid (FA) and ammonium formate (NH_4FA), both of analytical grade, were obtained from Sigma–Aldrich (St. Louis, MO, USA). 18.2 MΩ Milli-Q water produced using a Millipore Gradient water purification (Millipore, Bedford, MA, USA) system, was used to prepare mobile phase and sample solutions.

2.2. Standard solutions

Standard stock solutions (1000 μg mL^{-1}) of each analyte were obtained by dissolving an appropriate amount in the appropriate solvent. For β-adrenergic agonists and β-blockers methanol was used, with the exception of adrenaline and 3-methoxytyramine, which were prepared in acidified methanol (0.5% 1 M formic acid in methanol); for benzoic acids and nucleosides the standard solutions were prepared in acetonitrile-water (90:10) solution. 0.01N NaOH solution was used to dissolve uric acids and standard solutions of xanthines standard solutions were prepared in water, with the exception of 1-methyl-guanine and guanine which were prepared in aqueous formic acid (1% v/v). The aqueous solutions of guanine, xanthine, 7-methyl-xanthine, theobromine, hypoxanthine, 1-methyluric acid, uric acid and vidarabine were centrifuged and the supernatant used as the stock solution. The standard solution of the rest of the compounds was obtained in acetonitrile-

Download English Version:

https://daneshyari.com/en/article/5135147

Download Persian Version:

https://daneshyari.com/article/5135147

Daneshyari.com