Journal of Chromatography A, 1467 (2016) 206–213

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Toward structure-based predictive tools for the selection of chiral stationary phases for the chromatographic separation of enantiomers

Robert Sheridan[a,*], Wes Schafer[b], Patrick Piras[c,*], Kerstin Zawatzky[b], Edward C. Sherer[a], Christian Roussel[c], Christopher J. Welch[b,*]

[a] Department of Structural Chemistry, Merck Research Laboratories, Rahway, NJ, USA
[b] Department of Process Research & Development, Merck Research Laboratories, Rahway, NJ, USA
[c] Aix Marseille Université, CNRS, ISM2 UMR 7313, 13397 Marseille, France

ARTICLE INFO

Article history:
Received 18 February 2016
Received in revised form 17 May 2016
Accepted 20 May 2016
Available online 20 May 2016

Keywords:
QSAR
Chiral stationary phases
Chiral chromatography
Chromatographic method development
screening
Databases

ABSTRACT

ChirBase, a database for the chromatographic separation of enantiomers containing more than 200,000 records compiled from the literature, was used to develop quantitative structure activity models for the prediction of which chiral stationary phase will work for the separation of a given molecule. Constructuion of QSAR models for the enantioseparation of nineteen chiral stationary phases was attempted using only analyte structural information, leading to the producton of self-consistent models in four cases. These models were tested by predicting which in-house racemic compounds would and would not be resolved on the different columns. Some degree of success was observed, but the sparseness of data within ChirBase, which contains enantioseparations for only a subset of molecules on a subset of columns under a variety of conditions may limit the creation of effective models. Augmented data sets gleaned from automated chromatographic method development systems deployed in academic and industrial research laboratories or the use of models that take other factors such as solvent composition, temperature, etc. into account could potentially be useful for the development of more robust models.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The chromatographic separation of enantiomers has become a valuable and widespread tool since the commercialization of the first chiral stationary phase (CSP) in 1981. [1] Predicting whether or not a particular column will resolve the enantiomers of a given molecule has always been something of a hit or miss affair, although a number of chiral recognition models lend some predictability to these efforts, oftentimes associating different functional group classes or particular molecular features with a likelihood of enantioseparation on a given column. For example, α-amino acids and some α-hydroxy acids are often well resolved on Davankov ligand exchange CSPs [2] in the presence of Cu^{2+}, racemates bearing a primary amine near the stereocenter are often resolved on various crown ether CSPs [3] and racemates containing electron deficient π-systems are often well resolved by CSPs containing electron rich π-systems – and *vice-versa* [1].

Despite these useful rules of thumb, method development for chromatographic enantioseparation has usually proceeded by trial and error. When only one CSP was available in the marketplace, empirical testing was a fairly easy proposition, but as the number of commercial CSPs grew, the task of method development became increasingly difficult, sometimes requiring a week of work or longer. In response, two major innovations occurred as the number of available CSPs began to proliferate: 1) a commercial database (ChirBase) containing the abstracted data for published chromatographic separations was developed to aid in method development [4] and automated column screening using column selection valving was developed [5,6].

Beginning in the late 1980s, Christian Roussel and co-workers began to develop ChirBase, a database of chiral chromatographic separations. The database, which now comprises more than 200,000 entries, is widely used as a tool for analytical and preparative chromatographic method development. Structure searches help researchers to identify chromatographic methods to apply to known (or structurally similar) compounds that appear in the database. In contrast to bibliographic databases that contain only literature citations, ChirBase contains enantioseparation data and conditions. The aim of the database is to provide a unique compilation and organization of the data that is critical for per-

* Corresponding authors.
E-mail address: christopher_welch@merck.com (C.J. Welch).

http://dx.doi.org/10.1016/j.chroma.2016.05.066
0021-9673/© 2016 Elsevier B.V. All rights reserved.

forming and repeating chromatographic enantioseparations. The data included in ChirBase are extracted from the printed and online scientific literature. Most of these data are obtained from regular peer-reviewed journals which constitute the primary and secondary literature. Tertiary sources and the 'grey' literature (books, technical reports, conference proceedings, dissertations and chiral column manufacturer brochures and websites) are also included.

Each ChirBase record contains different categories of information related to sample, CSP, bibliographic reference, experimental conditions, retention and separation data. More specifically each entry is characterized by the following fields:

- Description of the data source (authors, journal, year, volume, pages).
- Description of the sample (chemical structure, IUPAC name, molecular weight, molecular formula, type of chirality and/or number of stereogenic centers).
- Column characteristics (chiral stationary phase name and chemical structure, commercial name, dimensions, particle size).
- Description of conditions (mobile phase, chemical names and structures of solvents and additives, flow-rate, gradient information, temperature, injection amount, scale of the separation, pH).
- Chromatographic data (retention times, retention factors, enantioselectivity, resolution and order of elution of enantiomers).

It should be noted, however, that ChirBase can record only what has been reported in the literature, and not all information is available for each experiment.

The Chirbase database is unique, with no similar database application or solution storing such a huge amount of data: 200,000 separations over 90,000 unique compounds, 1800 unique CSPs. Another advantage of ChirBase is its ability to export compound structures and associated chiral HPLC data. Chemoinformatic methods can thus be directly applied to the compound file being exported. Historically preferred strategies used in these laboratories include data mining [7], virtual libraries [8], regression analysis [9] or enantiophore studies [10].

Automated CSP screening to facilitate chromatographic method development for the separation of enantiomers involves sequential investigation of a set of top-performing CSPs with standard gradients to identify suitable methods for chromatographic enantioseparation. Generally, rote execution of a script is performed, without any consideration of analyte structure. The approach is generally successful and has been widely implemented in laboratories worldwide. However, experimental cycle time is somewhat long, typically requiring overnight analysis, although recent instrument and CSP developments have reduced this to a few hours [11]. In a push to further reduce the time to result, various parallel screening approaches have been developed. [12–16] It has often been noted that the addition of artificial intelligence and feedback algorithms for instrument control to conventional screening setups could afford a more rapid cycle time, [17] with significant time savings being possible if individual chromatograms could be terminated as soon as the enantiomers of interest have eluted or if the entire screen could be halted as soon as a result matching target specifications (enantioselectivity, resolution per unit time, or *etc.*) is obtained. In addition, the long understood fact that different CSPs perform well for different compound families or structural types suggests that a flexible script specifying CSPs and conditions that are based on structural features of the analyte may prove advantageous. In other words, with proper software, considerations of analyte structure may afford a speed advantage in identifying a suitable CSP.

Structure-based categorization of analytes by general characteristics (acids, bases, neutrals, *etc.*) or functional group class (amines, carboxylic acids, amino acids, amides, *etc.*) has long been considered in selecting CSPs for enantioseparation, especially prior to advent of automated CSP screening [18]. This approach has also enjoyed renewed interest [19]. The *de-novo* prediction of enantioseparation *via* docking analysis or other approaches has been investigated for decades, [20–23] although calculation times are long relative to experimental screening, and predictive capability is generally modest, which may stem from the difficulty of accurately predicting the small energy differences for formation of the enantioselective adsorbates that underlie chromatographic enantioseparation (successful separations show a typical energy difference between diastereomeric adsorbates of only 0.05–0.25 kcal/mole, which is generally within the error ranges of docking-based computational approaches).

Quantitative structure activity relationship (QSAR) models potentially allow the prediction of enantioseparation performance based on the presence or absence of certain structural features and properties, without a requirement to understand the exact mechanism of enantioseparation. Once a QSAR model is developed, estimations for new structures can be carried out very rapidly, enabling the near real time assessment that is required for next generation artificial intelligence approaches for CSP selection and method development. Descriptor importances (or weights) from the model allow interpretation of molecular parameters that influence the predicted experimental endpoint. QSAR models must be calibrated against a large amount of data. In this study the potential for using reported data in ChirBase to develop QSAR models of chromatographic enantioselectivity on different CSPs that would allow an estimate of chromatographic separability for an unknown compound based solely on molecular structure is investigated.

2. Experimental

2.1. Data set selection

All methods were applied to molecule data sets exported from ChirBase. As the data from ChirBase are collected from many different sources, it was imperative as a first step to ensure data selection such that the entries are homogeneous enough to be processed in a QSAR analysis.

Data used in this study have been selected by the following criteria:

- A clearly defined structure is assigned to the molecule to be separated.
- Only a single stereocenter is present in the molecule. Compounds having two or more stereocenters are excluded to avoid deviant effects due to the possible presence of other stereoisomers. In addition, molecules with special chirality (planar or axial chirality) were excluded based on previous studies showing these compounds sometimes exhibit specific binding to CSPs and therefore lead to incorrect inferences [24].
- The enantioselectivity (α) value for the separation was reported, or can be estimated using retention times and existing analytical conditions. Discussion of the derivation of α is below.
- The emphasis was on isocratic conditions (100% of the entries) with commercially available CSPs (80% of the entries) although custom-modified commercial CSPs were included as well.

From this selection, a data set of 134,000 entries was prepared. The size of the sample is large enough torepresents a substantial proportion of the published data about chiral HPLC performance for different compounds.

Download English Version:

https://daneshyari.com/en/article/5135623

Download Persian Version:

https://daneshyari.com/article/5135623

Daneshyari.com