Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography A, 1185 (2008) 49–58
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Simple method for the prediction of the separation of racemates with high-performance liquid chromatography on Whelk-O1 chiral stationary phase

Alberto Del Rio *, Johann Gasteiger

Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany

Received 29 August 2007; received in revised form 13 December 2007; accepted 7 January 2008
Available online 19 January 2008

Abstract

A simple method for the prediction of whether or not a racemate can be separated on a Whelk-O1 chiral stationary phase has been developed. In this approach, molecules are represented by counting the number of atom types of the neighbors spheres of the chiral center. A decision tree is then used to decide based on a few of these atom count descriptors whether a given racemate can be separated. High values of correct prediction were obtained, namely with more than 94% for training sets and of about 90% for cross-validation results. The same rate of correct prediction was also obtained on an external data set. The descriptors can be rapidly and easily retrieved by just counting the atom types around the chiral center by inspecting the chemical diagram of the molecule. Furthermore, the decision tree model can be applied through the use of a small set of rules that eventually predicts whether or not a racemate is separated. Due to its computational simplicity, the procedure is of interest for experimentalists that need to make rapid assessment of the separation without having to program or input complex formulas.

Keywords: Chemoinformatic predictions; Topological descriptors; Whelk-O1; Chiral HPLC; Chiral separations; Separation factor; Chiral recognition; ChirBase; Decision trees

1. Introduction

In the last few decades chemoinformatics, molecular modeling and quantum chemistry techniques have been successfully used to address chirality related problems [1,2]. These calculations constitute the basis of reliable and interesting results in the field of chiral recognition [3,4]. While the success of molecular modeling and first principle calculations can be fully acknowledged only when fairly small molecular structures are involved, chemoinformatic procedures have gained great importance due to the availability of an increasing number of experimental data [5,6].

These issues are of great importance in the context of the increased interest and availability of chiral chromatography techniques. Chiral HPLC has gained a great deal of valuable experience not only for the analytical and preparative separations of enantiomers but also for the investigation of chiral recognition processes. Indeed, chiral HPLC is today the most widely used technology to separate racemates as shown by the thousands of articles published each year on the subject and collected in the ChirBase database [5,7–9].

With molecular modeling calculations in which host–guest interactions are directly taken into account it is possible to thoroughly tackle problems related to the separation of racemates such as the prediction of the enantioselectivities and the assignment of the absolute configuration as well as the elucidation of the mechanism of enantioselective recognition [3,10–16]. Unfortunately, these computational techniques can be applied only in a few cases because often the stationary phases have a complex chemical structure. This makes the calculation expensive in terms of computation time and in some cases not reliable.

To circumvent these problems it is possible to perform calculations in which only the ligand structures are considered. In this sense, the main efforts of the scientists in the past years went into building relationships between some molecular descriptors of

* Corresponding author. Present address: Molecular Modelling & Drug Design Lab, Dipartimento di Scienze Farmaceutiche, Università di Modena e Reggio Emilia, Via Campi 183, 41100 Modena, Italy. Tel.: +39 059 2055122; fax: +39 059 2055131.
E-mail address: alberto.delrio@unimore.it (A. Del Rio).

0021-9673/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2008.01.034

the compounds to be separated and the experimental data available. Different models suitable to predict the experimental data are already present on the literature [17–26]. Other approaches based on probability rule [27] and factorial design [28] were also successfully developed. Most of these interesting studies calculate some molecular descriptors and, with the aid of more or less complex modeling techniques, build models that can be applied to the prediction of the results with new compounds. The most critical interest is in establishing if it is possible to achieve a separation of given racemates with given experimental conditions by modeling, for instance, the HPLC separation factor. The separation factor (α) is defined as the ratio k_2/k_1, where k_2 and k_1 are the retention factors of the second and first eluted enantiomers, respectively [29].

As previously shown by Del Rio et al. the separation of racemates can be related to achiral molecular descriptors [3,17,30]. This is true because the power of separation is a result of the differential binding of the two enantiomers with the chiral stationary phase (CSP) that relates to the constitution of the racemates and not to the intrinsic chirality of the two enantiomers.

In this paper we describe a very fast and intuitive tool that can be easily applied by experimentalists to determine whether a given racemate can be separated on a Whelk-O1 chiral stationary phase. This CSP, conceived by Pirkle et al. in the 1990s [31–33], has been studied with several experimental techniques [34–38] as well as with X-rays [39] and NMR [40,41] to elucidate the mechanism of enantioselective recognition.

Despite these mechanisms are better understood for Whelk-O1 CSP in respect to all other CSPs available, few is known on how to predict if a separation will take place or not for a given analyte and no previous example of very simple computational model has been yet conceived.

2. Experimental

2.1. Molecular descriptors

Several experimental and theoretical studies have already shown that the ligands must have minimum structural requirements for the interaction points with the stationary phase in order to achieve a chiral separation [1,10]. Chemoinformatic studies were carried out to try to keep track of all these interactions and predict the experimental enantiosectivities [17,24–26]. However, in most cases the resulting models, while insightful and robust, are of limited practical interest for the experimenter that need an easy and immediate tool to establish whether a compound is separable or not. Moreover, many existent ad-hoc descriptors are not always easy to be interpreted and have limited applicability when used with diverse or many data sets.

In order to keep track of the minimum requirements to achieve a chiral separation in an extremely simple way we investigated new achiral descriptors based on counting atoms around the stereogenic center. The main idea behind these new descriptors is to calculate atomic indices at different bond distance from the stereogenic center.

Fig. 1 shows this idea with the example molecule 1.

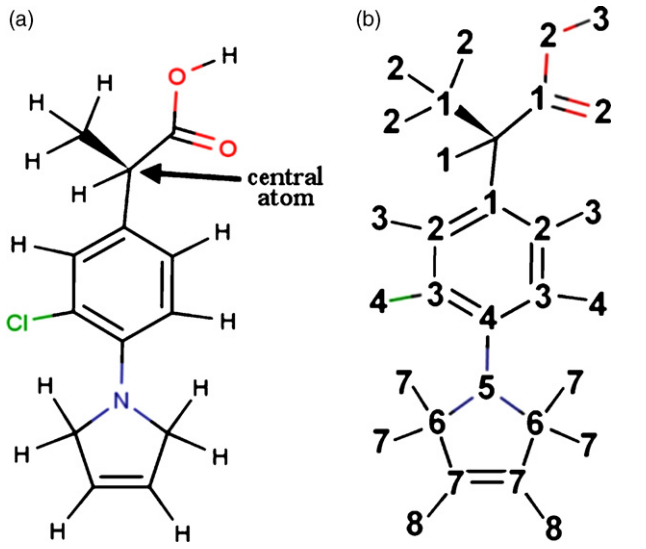

Fig. 1. Example molecule (1) with the central atom shown (a) and the bond distance considered (b).

At first, the stereogenic center is identified (Fig. 1a). At present, the method deals only with molecules having central chirality with one stereocenter.

Starting from the chiral center, the neighboring atoms are labeled following the respective bond distances (Fig. 1b). That is, atoms directly bonded to the chiral center have bond distance 1, those connected to the directly bonded to the chiral center have distance 2, and so forth.

In order to give a faithful account of the topology of all the molecules of the data set we considered atoms up to 10-bond distances away from the chiral center. This bond distance is necessary to unambiguously differentiate each molecule and obtain different sets of descriptors for molecules that are different. Thus, in the example of Fig. 1 all the atoms of the molecule 1 are considered since the farthest atoms around the chiral center have a bond distance of 8.

In a data mining study it has been shown that each CSP needs certain points of preference on the ligand to achieve a chiral separation [30]. For instance, a CSP like Whelk-O1 requires the ligand to have one acceptor and one donor of a hydrogen bond as well as aromatic and lipophilic sites.

In order to reproduce the effects of these interaction sites at atomic level we considered different atom types that were used in combination with the bond distances of Fig. 1b. These atom types are summarized in Table 1.

Table 1
Atom types considered

Description of the atom type	Label used
Total atom count	ac
Hydrogen atom count	H
Carbon atom count	C
Nitrogen atom count	N
Oxygen atom count	O
Sulphur atom count	S
Halogens atom count	Hal