Journal of Chromatography A, 1216 (2009) 2540–2547

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Theoretical characterization of gas–liquid chromatographic stationary phases with quantum chemical descriptors

Eufrozina A. Hoffmann[a], Zoltan A. Fekete[b], Robert Rajkó[c], István Pálinkó[d], Tamás Körtvélyesi[a,*]

[a] Department of Physical Chemistry, University of Szeged, Rerrich B. tér 1., H-6720 Szeged, Hungary
[b] HPC Group, University of Szeged, Szikra u. 2., H-6725 Szeged, Hungary
[c] Department of Mechanical and Process Engineering, University of Szeged, P.O. Box 433, H-6701 Szeged, Hungary
[d] Department of Organic Chemistry, University of Szeged, Dóm tér 1., H-6720 Szeged, Hungary

ARTICLE INFO

Article history:
Received 24 November 2008
Received in revised form 7 January 2009
Accepted 12 January 2009
Available online 15 January 2009

Keywords:
QSPR
Stationary phase
McReynolds constants
Quantum chemical
Semiempirical
Chemometrics

ABSTRACT

Quantitative structure–property relationship (QSPR) solvent model has been developed for the McReynolds constants (prototypical solutes) on 36 gas–liquid chromatographic stationary phases. PM6 semiempirical quantum chemical calculations combined with conductor-like screening model (COSMO) has been utilized. From 276 descriptors considered, forward stepwise variable selection, followed by best subset selection, yielded linear regression models containing six purely quantum chemical and two hybrid, topologically based descriptors. Internal (leave-one-out and bootstrap) as well as external validation methods confirmed the predictive power of these structure-driven models across all 10 McReynolds constants, with 40 Kováts-index units overall root-mean-square prediction error estimate.

1. Introduction

Chemometrics methods have long been utilized in chromatography, and their application is getting ever more widespread [1–3]. Various tools have been employed, such as quantitative structure–property relationships (QSPR) [4–22], artificial neural network (ANN) [23,24], cluster analysis [25], in order to address the two major aspects of handling the large body of chromatographic data. One is to predict chromatographic behavior of compounds and the other to organize and rationalize measurements. Even today, when most stationary phases routinely used have long been commercially established, novel stationary phase materials are still being developed. For example, recent advances in chiral GC [26–29] make inroads to enantiomeric analysis. Ever more complicated analytes, such as peptides, DNA or peptide nucleic acids (PNAs) [29,30] and their derivatives, have become amenable to GC.

Until recently the most widely used phenomenological analysis of data on gas-chromatographic stationary phases [31,32] was based on McReynolds constants [33], which relate retention indices of a standard set of prototypical molecules (see Table 1). In this system 10 solutes had been selected (with the original intent of broadly representing various types of interactions: dispersion, dipole and induction interactions, donor–acceptor forces and hydrogen-bonding). The numerical value of each McReynolds constants is taken from the difference of the Kováts retention indices of the corresponding prototypical compound on a given phase I_{p}^{ξ} and that on squalane I_{sq}^{ξ} (the standard apolar reference), according to the general expression Eq. (1).

$$\xi \equiv I_{\mathrm{p}}^{\xi} - I_{\mathrm{sq}}^{\xi} \quad (\xi = x, y, z, u, s, h, i, k, l, m) \tag{1}$$

The utility of the McReynolds system was further extended recently by Zenkevich and Makarov [34], who showed how stationary phases consisting of binary mixtures can be characterized by predicting their McReynolds constants.

The McReynolds system has been criticized from a theoretical point of view [32,35–40], and even a stern call for its abandonment was issued [38]. If considered a solvation model, then four major objections can be raised against it (following the critique of Poole et al. [31,36–39,41,42]):

1. Some of the prototypical solutes are too weakly retained on some stationary phases.
2. The contribution of interfacial adsorption is ignored.
3. The retention index differences are composite terms; their magnitude depends largely on the retention of the *n*-alkanes (Kováts-index standards).

---

[*] Corresponding author. Tel.: +36 62 544 612; fax: +36 62 544 652.
E-mail address: kortve@chem.u-szeged.hu (T. Körtvélyesi).

0021-9673/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2009.01.026

Table 1
McReynolds test compounds.

	Compound	Designation
1	Benzene	*x*
2	*n*-Butanol	*y*
3	2-Pentanone	*z*
4	1-Nitropropane	*u*
5	Pyridine	*s*
6	2-Methyl-2-pentanol	*h*
7	1-Iodobutane	*i*
8	2-Octyne	*k*
9	1,4-Dioxane	*l*
10	*cis*-Hydrindane	*m*

4. Each prototypical solute expresses several intermolecular interactions simultaneously, rather than being dominated by a single one. Therefore the McReynolds constants cannot be related to clearly separated fundamental properties.

Despite these deficiencies the conceptual simplicity had given McReynolds constants enduring popularity among many GC practitioners. There exists a large body of gas–liquid partitioning data available in this standard form, since practically any of the literally thousands of GLC stationary phases ever used have its McReynolds constants measured (or at least the first five ones typically published in commercial compilations, see *e.g.* the current Supelco catalog [43]).

Contemporary theoretical characterization of GC stationary phases has settled on employing a so-called solvation parameters method; this is a linear free energy relationship [44,45] originally developed by Abraham and co-workers [35,45–48]. This system of five fitted solvent-specific constants, each with a well-defined physical meaning, is applicable to many partitioning processes besides GC. It also offers a straightforward approach to the calculation of phase selectivity (system constants). Being based on absolute rather than relative retention data, the method embeds the full thermodynamics of solvation – free of ambiguities introduced by arbitrarily chosen reference data, such as those of *n*-alkanes retention involved in the Kováts-indices. There have been molecular based models to predict solute Abraham descriptors, see *e.g.* Ref. [45] for a review. The chemical information content of the five experimental Abraham parameters has been shown to have a large overlap with a set of five purely computational descriptors (COSMOments) from Klamt's COSMO-RS method [45]. In the past decade other developments occurred in the area of solvation parameters theory, as well; these include the new model by Laffort et al. [49–51], and a retention regression system updated by Rohrschneider [32,40], who have also analyzed his method against the solvation parameters model.

From the modeling point of view a dataset of McReynolds constants has the advantage of containing directly determined experimental values, with well-established estimates of error (about 20 index units for polar phases). For this reason, in the current investigation of quantum chemical descriptors our first pick for dependent variables is the 10 McReynolds constants. It is a purely pragmatist choice: this collection is treated merely as a source of retention data. In the future we intend to carry out similar QSPR studies on the fitted data from the more advanced solvation parameter models, as well. Other related further work [52] considers absolute retention data directly; this can be done either by taking retention volumes themselves, or via back-calculating them after separate modeling of *n*-alkane retention behavior *vs.* carbon number (the "***b*** constant", *i.e.* the slope of the calibration line used in determining the Kováts-indices [33], is published for many of the stationary phases).

Most of the chemometrics analyses published so far on retention data considered properties of the solute molecules. In a previous paper by some of us (R. Rajko and T. Kortvelyesi) [6], quantum chemical descriptors for the prototypical solutes were utilized for correlating their McReynolds constants. Conversely, in this contribution properties of stationary phases are analyzed in relation to structure-driven, quantum chemically based molecular descriptors for solvent molecules. To the best of our knowledge, this is the first time that this type of investigation has been published on McReynolds data. A comprehensive search was carried out to find good stationary phase models for all 10 McReynolds constants as dependent variables, utilizing a large selection of prospective independent variables derived from quantum chemically determined and characterized structures. Successful selection of such models is a step toward developing predictive computational GC ("*chromatography in silico*"), similarly to the impressive molecular-based results in HPLC by Hanai et al. [7,53].

2. Computation methodology

2.1. Retention data

Published McReynolds data for 37 molecular (non-polymeric and non-ionic) GLC stationary phases were collated from the literature [33,54–57]. An effort was made to have as wide data coverage as possible, particularly concerning phases of either the highest or lowest polarity: all available data were considered which included the full set of all 10 McReynolds constants for a stationary phase. Table 2 contains the names, codes and McReynolds constants of these stationary phases.

In the following analysis the average of the first five McReynolds constants, designated *MR5a*, is included as a separate dependent variable besides the McReynolds constants themselves. It was found a useful working variable for model selection. We note in passing that either this average, or the corresponding sum, is called McReynolds polarity in the literature, even though actual column polarity is now recognized as a multi-dimensional feature [58].

2.2. Descriptor calculation

All descriptors considered in this study were generated from our quantum chemical calculations. Starting geometries of stationary phase molecules were generated in the molecular graphics program (ACD/ChemSketch) [59], and pre-optimized with molecular mechanics (the force-field based [60,61] three-dimensional optimizer of ChemSketch). Where available, experimental or precalculated structures available via the PubChem [62] database were used as well. Full geometry optimization was carried out for isolated (gas-phase) molecules by the PM6 semiempirical quantum chemical method [63], as implemented in MOPAC2007 [64]. The convergence criterion was typically a gradient norm of 0.1 kcal/mol.

From the converged MOPAC output the following quantities were extracted: coordinates and Mulliken charges of each atoms; distances and bond orders of each atom pairs; dipole moment; isotropic average polarizability; energies of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO); largest positive charge on any H atom, and largest negative charge on any atom.

Furthermore, our custom processing scripts (available from <http://www.staff.u-szeged.hu/~fekete/descript/>) took the MOPAC-generated connectivities, bond orders and lengths, as well as charges to generate various weightings to graph walk-count based descriptors. These are based on the idea of Rucker and Rucker [65,66], who advocated using the counts of walks as graph theoretical (two-dimensional, 2D) descriptors suitable for QSPR variables. Here we introduce an extension of this concept, by including weights to the edges of the molecular graphs, derived