Short communication

Database of structures and their gas chromatography retention indices, tagged with individual search windows

R. Richmond

Structural and Analytical Chemistry Group, Pre-Clinical Research Department, Sandoz Pharma AG, CH-4002 Basel, Switzerland

Abstract

Experimental calibration methods described earlier allowed the salvage of gas chromatography retention indices from diverse literature sources; these "pseudo-Sadtler" retention indices were paired with searchable structures (carbon connection tables) in a ISIS/Base database called PSEUDO-SADTLER. This database has the data fields: structure, name, synonyms, Chemical Abstracts Service number, retention index, average molecular mass and formula. The retention indices were assigned search windows covering 95% of their expected normal statistical distribution. The database was compacted from a starting point of 3450 entries to 3195 after the elimination of entries that were duplicated or obviously false.

Keywords: Retention indices; Temperature-programmed retention indices

1. Introduction

Many small listings of temperature-programmed gas chromatography (GC) retention indices exist in individual research publications [1,2]. A few larger collections exist e.g., the Sadtler [3], the Pfleger [4] and the International Association of Forensic Toxicologists (TIAFT) compilation [5]. Unfortunately the majority of these indices are of limited inter-laboratory use due to the lack of standardisation. In addition only three collections exist electronically i.e., the program "GC Identifier" from Sadtler in association with Jacques Rijks of Eindhoven University [3], the program "Pro ezGC" from Analytical Innovations, Kettering, OH, USA and the Pfleger spectral library on Hewlett-Packard mass spectrometer work-stations [4]. The first and second have no searchable structures, and no search window tags on their retention indices. The third also has no search windows and the majority of indices were

derived on the now atypical packed column technology rather than on capillary columns.

A large problem lies in the inter-laboratory comparison of retention index values. Capillary column GC can generate retention indices very precisely, assuming standardisation of stationary film polarity, carrier gas flow-rate, stationary phase film thickness and temperature programming rate. However, most published retention index values have not been produced under conditions in which these four parameters were strictly comparable and so their usefulness i.e., in the identification of unknowns, is limited. Initial development work [6,7] to circumvent this problem, involved the successful discovery of "pseudo-Sadtler" GC conditions. The strategy used was the following: one dataset was adopted as a founding core (i.e., the 8°C on OV-1 stationary phase subset of the excellent Sadtler compilation), then experimental conditions were sought after that generated "pseudo-Sadtler" experimental retention index

(I_{exp}) values which approximated the analytically well-defined Sadtler retention index (I_{Sad}) values and so allowed a comparison of unknown structures with those in the Sadtler library [3]. The objective of the presented work was to move away from a simple spread-sheet listing of the retention index data in [6,7] to a more informative structural database format, and use the database's quality control features to eliminate errors from the original published sources.

2. Experimental

2.1. Instrumentation

A Finnigan TSQ-70B quadrupole mass spectrometer was used as a GC detector, running in either 70 eV electron-impact ionisation mode, coupled with a Hewlett-Packard (HP) 5890 series II gas chromatograph. The GC injector was maintained at 250°C, the GC transfer line at 320°C, the source at 150°C and the vacuum manifold at 70°C. The mass range was from m/z 10 to 770 per 0.5 s. The GC program was from 35°C to 320°C at 15°C/min, followed by 10 min isothermally at 320°C. The mass spectrometer electron multiplier was maintained at +1100 V, the conversion dynode at −15 kV and the electron current at 400 μA.

All retention indices were measured with simultaneous injection of n-alkanes standards (C_5 to C_{40}, excluding C_{39}). Average retention indices were determined from six measurements. A HP Ultra-1 capillary column (25 m×0.32 mm I.D. and a film thickness of 0.52 μm) was used for consistency with

the Sadtler compilation [4]. ISIS/Base (Integrated Scientific Information System) was version 2.0 from Molecular Design, San Leandro, CA, USA.

2.2. Retention index calibration

The base/neutrals mixture for Environmental Protection Agency (EPA) method 625 in conjunction with various temperature programming rates (between 15°C and 18°C/min) and various helium carrier gas flow-rates (between 45.0 and 70.0 kPa head pressure) was evaluated to find the best alignment with the 8°C/min on OV-1 subset of the Sadtler compilation i.e., the "pseudo-Sadtler" conditions [6,7]. This base/neutrals mixture (Sigma part number 38,463-1; 41 structures with retention index 952 to 3199 of which 29 are in the Sadtler library), was manually injected splitless, simultaneously with the C_5 to C_{40} n-alkane mixture and the retention indices measured by simple linear interpolation between bracketing n-alkanes.

Table 1 shows the experimental and literature sources for the PSEUDO-SADTLER database. The last row illustrates the number of retained entries with ambiguous *cis–trans* geometry.

Fig. 1 displays the typical shape of the search windows assigned to a set of $I_{corrected}$ values. Specifically, it shows the $I_{Sad-corrected}$ search window's dependency on the I_{Sad} values, that is from the 8°C/min on OV-1 Sadtler dataset.

3. Results and discussion

Two families of retention indices were developed

Table 1
Survey of the PSEUDO-SADTLER database

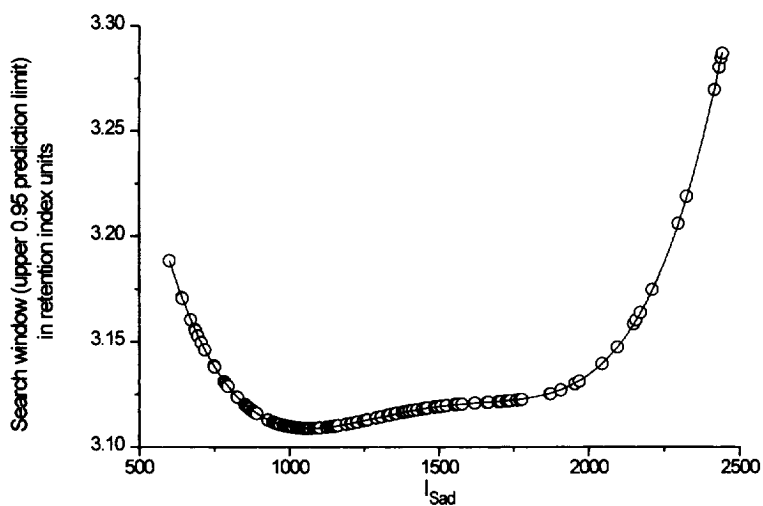|	Sadtler [3]	Jennings and Shibamoto [8]	White et al. [11]	Hayes et al. [10]	Hayes et al. [12]	Richmond [6,7]
Entries' origin	1995	1097	318	115	179	422
Column	Ultra-1	DB-1	Petrocol-DH	OV-101	Ultra-1	Ultra-1
I_{exp}	148	111	97	55	90	422
$I_{corrected}$	1847	986	221	60	89	0
cis/trans ambiguity	20	13	1	0	0	0

Fig. 1. 0.95 prediction limit values versus I_{Sad} values.

prior to the loading of the database. Firstly, the I_{exp} values which were measured under "pseudo-Sadtler" conditions and which are very close to the Sadtler indices (>99.9%), were tagged with ±two times their standard deviation, to be consistent with the 0.95 prediction limits in the corrected entries. Secondly the various "pseudo-Sadtler" $I_{corrected}$ values (e.g., $I_{Sad-corrected}$), derived from correction of literature values were tagged with the 0.95 prediction

limits from the second-order polynomial fits to the respective offset plots [6,7]. This has some similarities to the earlier but cruder approach in [9].

In the final ISIS/Base, various data fields exist, namely: structure, name, synonyms, Chemical Abstracts Service (CAS) number, I_{best}, I_{exp}, average (non-monoisotopic) molecular mass and formula (Fig. 2). Prior to final loading I_{exp} and the various $I_{corrected}$ values were reviewed using a Microsoft
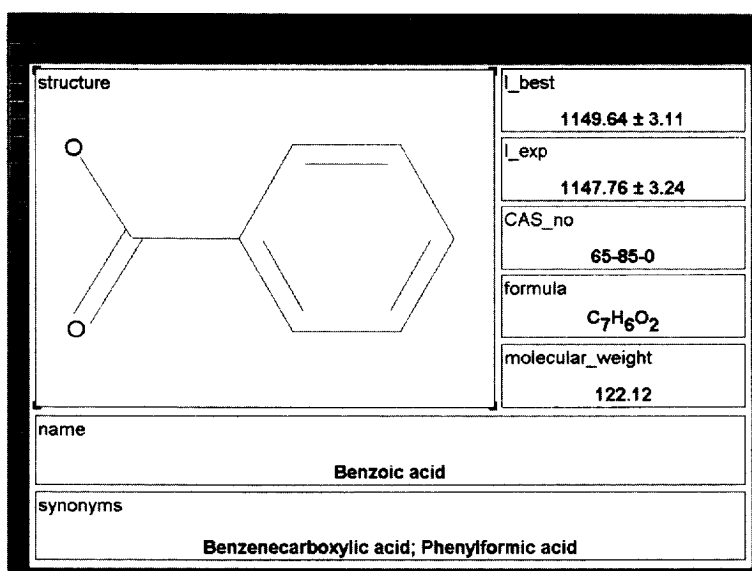

Fig. 2. PSEUDO-SADTLER database display (running under Microsoft Windows 3.11).

Excel table and the one with the narrowest search window was inserted into the I_{best} field. The I_{exp} values were also loaded to show if an I_{best} value originates from an experimental or corrected value. A sorted list of I_{best} values is a useful starting point for comparison with an individual's laboratory values. Interestingly, a significant proportion (29%) of the I_{exp} values reflect polar drug structures, originating from the earlier attempt at incorporation of the Pfleger library [4,6] and this partially compensates for the apolar character of much of the database.

Those entries with I_{exp} or $I_{corrected}$ values outside the range 600.00 to 2960.00 were excluded. The lower limit is due to the fast temperature program used (35 to 320°C at 15°C/min). The higher limit is because temperature programming ends at $I_{exp} = \sim 2960$, with the limiting factor being the temperature stability of the polyimide capillary column support above 320°C.

The I_{exp} search windows (2 times the standard deviation) are generally less than ±3.04 and so are better than the search windows of the various $I_{corrected}$ values. However exceptions exist e.g., benzoic acid with $I_{exp} = 1147.76 \pm 3.24$ compared with an $I_{corrected} = I_{best} = 1149.64 \pm 3.11$ (see Fig. 2) and these above-average search windows widths for I_{exp} values are a quantitative indicator of chromatographic tailing. The use of a statistically derived search window offers more information about peak asymmetry than the asterix notation in the Sadtler collection [3].

One anomaly lies in the I_{exp} values having search windows of ±0.00 (4.5% of the 422 entries). The presented GC–MS experimental system is unable even with an average of six measurements to detect a standard deviation of less than 0.01 for a retention index. This is caused by the electronic limitation on the sampling rate i.e., scanning faster than 800 u per 0.5 s causes mass spectral deterioration. Therefore in the final database all entries that had zero variation for their I_{exp} values were assigned the lowest detectable standard deviation i.e., 0.01.

The ISIS/Base program is superior to a spreadsheet based listing of indices as it prompts when a duplicate structure is uploaded; an entry registration based on alpha-numeric field names e.g., molecular names, allows structure duplication to slip past unnoticed. The calibration techniques in [6,7] also facilitate additional compacting of the database. A

typical entry elimination was: 2-octene in the Jennings and Shibamato database has an $I_{Jenn} = 811$; but has an $I_{White} = 803.37$ (*trans*-2-octene) and an $I_{White} = 810.94$ (*cis*-2-octene). Clearly the *cis* and *trans* isomers are separable and so the ambiguous Jennings and Shimbamoto entry was deleted. However *cis*–*trans* ambiguity survives within PSEUDO-SADTLER (see Table 1) based on the premise that data tagged as ambiguous is better than no data at all for a structure.

This gives a compacted database of 3195 entries of which 422 have I_{best} retention indices originating from experimental (I_{exp}) values [6,7] and 2773 have I_{best} indices originating from corrected literature ($I_{corrected}$) values.

4. Conclusions

The GC calibration techniques exploited to build this PSEUDO-SADTLER database are essentially simple [6,7]. The total of 3195 values is the largest existing collection of "comparable" GC capillary retention indices; each entry is tagged with a statistically derived search window, encompassing 95% of the retention index's expected normal statistical distribution. The justification for PSEUDO-SADTLER is that firstly and most importantly it facilitates the calibrated comparison of GC retention index data derived from diverse sources, secondly the ISIS/Base program allows further quality control in comparison with simpler spread-sheet listings, thirdly it is supplemented with polar drug structures and lastly each entry has a structure that can be searched for and displayed on a computer screen.

References

[1] W.H. Anderson and D.T. Stafford, J. High Resolut. Chromatogr. Chromatogr. Commun., 6 (1983) 247.

[2] B.J. Perrigo, D.J. Ballantyne and H.W. Peel, Canadian Soc. Forensic Sci. J., 17 (1984) 41.

[3] The Sadtler Standard Gas Chromatography Retention Index Library, Sadtler Research Laboratories, Philadelphia, PA, USA, 1984.

[4] K. Pfleger, H.H. Maurer and A. Weber, Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and their Metabolites, 2nd, revised and enlarged edition, VCH, Weinheim, 1992.

[5] International Association of Forensic Toxicologists, Gas Chromatographic Retention Indices of Toxicologically Relevant Substances on Packed or Capillary Columns with Dimethylsilicone Stationary Phase, VCH, Weinheim, 1992.

[6] R. Richmond, J. Chromatogr. A, 724 (1996) 229.

[7] R. Richmond, J. Chromatogr. A, 742 (1996) 131.

[8] W. Jennings and T. Shibamoto, Qualitative Analysis of Flavor and Fragrance Volatiles by Glass Capillary Gas Chromatography, Academic Press, New York, 1980.

[9] A. Stowell and L. Wilson, J. Forensic Sci., 32 (1987) 1214.

[10] P.C. Hayes and E.W. Pitzer, J. Chromatogr., 253 (1982) 179.

[11] C.M. White, J. Hackett, R.R. Anderson, S. Kail and P.S. Spock, J. High Resolut. Chromatogr., 15 (1992) 105.

[12] P.C. Hayes and E.W. Pitzer, J. High Resolut. Chromatogr. Chromatogr. Commun., 8 (1985) 230.