DATA BASE FOR GC/MS IDENTIFICATION OF TRACES OF ORGANIC IMPURITIES IN ATMOSPHERIC AIR*

Boris V. Ioffe, Igor G. Zenkevich and Andrei S. Bureiko

Organic Chemistry Department, Leningrad University, 199034 Leningrad, U.S.S.R.

A new approach to the GC/MS identification of organic components of the atmosphere is proposed. It is characterized by the use of averaged values of retention indices of different types at different temperatures in combination with several main peaks of mass spectra and allows considerable standard deviations of indices and peak intensities. Hence, it is possible to achieve good interlaboratory reproducibility of results working with simplest computers having a small operative memory. The data base contains averaged eight-peak mass spectra and retention indices on polymethylsiloxane stationary phases for about 600 volatile organic compounds constantly present in atmospheric air or those that should be checked in populated areas.

The main feature of identification of organic microimpurities in atmospheric air is the necessity for the separation of complex mixtures of substances over wide ranges of chromatographic retention parameters including many isomers that cannot be distinguished with the aid of mass spectra. The low reproducibility of signal intensities in mass spectra makes it desirable to use several $(4-6)$ peaks with the highest intensity. This limitation of mass spectrometric information may be compensated for by simultaneous evaluation of retention indices. The optimum solution of this problem is the combination of reduced mass spectra with retention parameters applicable under the conditions of temperature programming. These data banks intended for ecological investigations do not exist, although there are specialized collections of chromatographic and mass spectrometric data for the identification of polymers from the products of their pyrolysis[1], flavor and fragrance substances[2] and oil products[3]. However, all these data bases assume that work should proceed under strictly ordered chromatographic conditions. Thus, for linear temperature programming retention indexes exhibit poor interlaboratory reproducibility, and each laboratory should prepare its own data bases for sufficiently reliable identification (see, e.g. ref.[3]).

The fundamental difference between the data bank proposed here and those used at present is the fact that the retention parameters and (partially) mass spectra contained in it are obtained as a result of averaging and statistical processing of

---

* Paper read at the 46 Congress of Czechoslovak Chemical Society, Pardubice, July 4, 1990.

the literature data and the experimental data of the authors. Moreover, the requirements for the precision of coincidence between indices and signal intensities in mass spectra are reduced to the replacement of intensity values by the sequence of mass numbers. In this case analyses can be interpreted without the necessity for a precise reproduction of chromatographic conditions.

EXPERIMENTAL

The data base formed for the identification of organic components of the atmosphere now contains retention indices and the characteristics of eight main peaks of mass spectra of 583 compounds detected in the open atmosphere[4] or having the values of limiting permissible concentrations in the populated areas of the U.S.S.R. (according to the data for mid-eighties[5]). The data include compounds with molecular mass numbers from 26 to 290. The main mass of them (about 50%) are alkanes, alkenes and aromatic hydrocarbons. The second place (23%) is occupied by oxygen derivatives (alcohols, ketones, aldehydes, carboxylic acids and their esters, anhydrides, lactones and some heterocycles). The fraction of halogen derivatives and nitrogen and sulfur compounds is 16, 6 and 5%, respectively.

Retention indices taken from various literature sources (more than 170 original publications) and measured by us for polydimethylsiloxane stationary phases (SE-30, OV-101, OV-1, SP-2100, SF-96, PMS-100) refer to columns of different types (packed and capillary columns) and temperatures most often differing by 30—50°C. The average standard deviation of the values is ±8 units. The main reasons for this scattering are neglect of the temperature dependence of retention indices and sorption effects which are most important for polycyclic compounds and substances containing active hydrogen atoms. Retention indices are absent for approximately 20% of compounds contained in the bank, so that their identification is based on mass spectra alone.

Mass spectrometric data are represented by the spectra of electron beam under standard conditions (70 eV) and are arranged in the order of decreasing peak intensities. Most spectra presented here (usually eight-peak spectra) are averaged over the data of reference books[6,7] and our measurements. About 50 spectra were obtained by processing of a registry[8]. Peaks with m/z less than 26 were not included regardless of their intensity. When the sets of m/z in different sources exhibited discrepancy, only reproducible main signals were retained (not less than 3—5). The comparison with the spectra of substances to be determined was carried out either from the sequence of mass numbers with allowed deviations of ordinal numbers up to two units or from relative intensities of I, % to within $\pm(1 + 0.3 \cdot I)$.

RESULTS AND DISCUSSION

The checking of the efficiency of this data base consisted in the determination of possible errors of the first kind (i.e., erroneous identification of compound recorded in the data base), errors of the second kind (erroneous identification of compounds not included in the data base) and the unambiguity or the degree of ambiguity of identification. Three variants of identification were considered:

 A) from retention index and the coincidence between the mass numbers of the four main peaks of the mass spectrum with a possible deviation of their ordinal numbers by two units;

B) only from the coincidence between the mass numbers of four main peaks, as in variant A, but without retention indices;

C) from the coincidence between the mass numbers of four peaks and their relative intensities to within $\pm(1 + 0.3\,I)\,\%$.

Kovach indices from independent sources[9,10] were used for checking, and when they were absent, the randomized error to within ± 8 units was introduced into recorded averaged indice values with the aid of a computer. The main peaks were taken from the arbitrarily selected single spectra of the reference book[7]. The results of checking of about 50 compounds for the identification errors of the first kind are given in Table I.

It can be seen from Table I that variant *A* (with the application of retention indices and four main peaks of the mass spectrum) provides the greatest number of unambiguous correct answers. The refusal to use even approximate values of retention indices (to within $= 8$ units) leads to a considerable decrease in the probability of unambiguous individual identification (from 65 to $30-40\%$). Moreover, the fraction of unambiguous answers is much higher for hydrocarbon derivatives then for hydrocarbons themselves. In the latter case the fraction of unambiguous answers attains $84-80\%$ for variants *A* and *C* and 60% for variant *B* which requires the lowest amount of information. The reasons for this great difference between the informativeness of the above variants of identification are the peculiarities of the mass spectra of different classes of organic compounds. It is known that unsaturated hydrocarbons with different positions of the double bond (in our examples they comprised 70% of the total number of hydrocarbons) exhibit very similar mass

TABLE I

Results of GC/MS identification of organic impurities of atmospheric air — a comparison of number of compounds to be checked (N) with number of correct answers (N_{corr}) and with number of errors of the first kind (N_{err})

Compounds	N	N_{corr}						N_{err}		
		A		B		C		A	B	C
		I^a	II^a	I	II	I	II			
Hydrocarbons	27^b	14	11	3	23	5	21	2	1	1
Derivatives	19	16	1	11	8	15	1	2	0	3
Total	46	30	12	14	31	20	22	4	1	4

[a] I read for unambiguous identification, II read for ambiguous identification; [b] including 17 olefin hydrocarbons.

spectra which often cannot be distinguished at all with the aid of the main peaks. The same observation refers to isomeric aromatic hydrocarbons, so that mass spectrometry without additional information is of little value for the individual identification of these compounds. It is not an important defect when it is used in practice in the investigations of polluted environment because for paraffins, naphthenes and olefins there is no normalization of the limiting permissible concentrations of individual hydrocarbons, and only relatively high total permissible concentrations for oil products exist.

One paradoxical feature of the data obtained should be considered. The number of errors of the first kind (in general, it is relatively low) for variants A and C (with a higher amount of the required initial information) is higher than in variant B which requires neither the knowledge of retention indices nor that of relative intensities of the mass spectra being compared. The reasons for this paradoxical relationship are, according to our observations, not the algorithms of comparison themselves but, rather the errors or misprints in the retention indice collections used for checking and very marked scattering of values of peak intensities reported by different authors. In some cases this scattering markedly exceeds the error in I used by us.

The number of errors of the second kind in the attempts to identify the compounds not recorded in the data base according to variant A was only 5·5% (one error in 18 cases). Moreover, the valuation of the intensity of the compared peaks decreases the number of errors to zero. However, without retention indices (when only four peaks of the mass spectrum are used), the number of errors of the second kind increases in variants B and C to 39 and 22%, respectively, so that the evaluation of at least approximate values of retention indices is doubtless preferable.

CONCLUSION

On the whole we consider preliminary results to be satisfactory. The work on the extention of the base of GC/MS data and the development of optimum identification strategy will continue. It is also planned to take into account the official lists of harmful organic impurities that should be checked in the air of other countries.

REFERENCES

1. Tsuchiya Y., Boulanger J., Sumi K.: Chromatographia 14, 154 (1977).
2. Jennings W., Shibamoto T.: Qualitative Analysis of Flavor and Fragrance Volatiles by Glass Capillary Gas Chromatography. Academic Press, New York 1980.
3. Steward E. M., Pitzer E. W.: J. Chromatogr. Sci. 26, 218 (1988).
4. Isidorov V. A.: Organic Chemistry of the Earth's Atmosphere. Springer, Berlin 1990.
5. Bespamyatnov G. P., Krotov Yu. A.: Predelno dopustimye kontsentratsii khimicheskikh veshchestv v okruzhayuschchei srede. Khimiya, Leningrad 1985.

6. *Atlas of Mass Spectral Data* (E. Stenhagen, S. Abrahamsson and F. W. McLaffetty, Eds). Intersc. Publ., New York 1969.

7. *Eight Peak Index of Mass Spectra*. Aldermastone Mass Spectrometry Data Centre, Aldermastone 1970.

8. McLafferty F. W., Stauffer D. B.: *The Wiley/NBS Registry of Mass Spectral Data*. Wiley, New York 1989.

9. Vigdergaus M. S., Semenchenko L. V., Esrez V. A., Bogoslovski Yu. N.: *Kachestvennyi gazokhromatograficheskii analiz*. Nauka, Moskva 1978.

10. Bogoslovski Yu. N., Anvaer B. I., Vigdergaus M. S.: *Khromatograficheskie postoyannye v gazovoi khromatografii*. Izdatelstvo standartov, Moskva 1978.