Chromatography and Mass Spectrometry in Biomedical Sciences, 2

Proceedings of the International Conference on Chromatography and Mass Spectrometry in Biomedical Sciences, Bordighera, Italy, June 20-23, 1982

ANALYTICAL CHEMISTRY SYMPOSIA SERIES — volume 14

chromatography and mass spectrometry in biomedical sciences, 2

Proceedings of the International Conference on Chromatography and Mass Spectrometry in Biomedical Sciences, Bordighera, Italy, June 20–23, 1982

edited by
Alberto Frigerio

Mario Negri Institute for Pharmacological Research, Via Eritrea 62, 20157 Milan, Italy

ELSEVIER SCIENTIFIC PUBLISHING COMPANY
Amsterdam — Oxford — New York *1983*

ELSEVIER SCIENTIFIC PUBLISHING COMPANY
Molenwerf 1
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY INC.
52, Vanderbilt Avenue
New York, NY 10017

Library of Congress Cataloging in Publication Data

International Conference on Chromatography and Mass Spectrometry in Biomedical Sciences (1982 : Bordighera, Italy)
Chromatography and mass spectrometry in biomedical sciences, 2.

(Analytical chemistry symposia series ; v. 14)
Includes index.
1. Chromatographic analysis--Congresses. 2. Mass spectrometry--Congresses. 3. Biological chemistry--Technique--Congresses. I. Frigerio, Alberto. II. Title. III. Series.
QP519.9.C47I55 1982 612'.01585 82-21138
ISBN 0-444-42154-6 (U.S.)

ISBN 0-444-42154-8 (Vol. 14)
ISBN 0-444-41786-9 (Series)

Printed in The Netherlands

PREFACE

The papers published in this volume were presented at the "International Conference on Chromatography and Mass Spectrometry in Biomedical Sciences" held in Bordighera, Italy, in June of 1982.

This Symposium was organized by the Italian Group for Mass Spectrometry in Biochemistry and Medicine in cooperation with the Italian Group for Scientific Research and Study.

The financial support of the National Research Council, C.N.R., Rome, Italy, is gratefully acknowledged.

Milan
December 1982

ALBERTO FRIGERIO
Chairman, Italian Group
Mass Spectrometry in
Biochemistry and Medicine

CONTENTS

A. Frigerio (Editor), *Chromatography and Mass Spectrometry in Biomedical Sciences, 2*

COMBINED GAS CHROMATOGRAPHY - FOURIER TRANSFORM INFRARED SPECTROSCOPY

THOMAS L. ISENHOUR and JOHN C. MARSHALL

Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27514 (U.S.A.)

INTRODUCTION

Gas chromatography (GC) is the most widely used separation method and can be an excellent quantitative tool. However, its utility for qualitative identification is surpassed by mass spectroscopy (MS) and infrared spectroscopy (IR), the two common techniques known to provide the most information about organic molecular structures. Over the last twenty years GC/MS has been highly developed and has probably solved more complex mixture problems than any other analytical method. GC/IR only became feasible with the commercial introduction of medium-priced infrared interferometers (FTIR) in the late 1960's. The development of new semiconductor detectors in the last decade has made GC/IR an important alternative to GC/MS.

This paper describes GC/IR as a new method for the analysis of complex mixtures. A completely automated GC/IR system will be discussed as well as special applications including: 1) on-line library searches based on interferometric data; 2) direct identification of specific compound classes from reconstructed chromatograms; 3) quantitative GC/IR; and 4) the potential of GC/IR coupled directly with MS (GC/IR/MS).

I. A GC/FTIR COMPOUND IDENTIFICATION SYSTEM (1)

GC/IR is a powerful analytical technique that yields much useful information concerning sample composition. Combining the minimal sample amount and sample preparation requirements of gas chromatography (GC) with the great identification power of infrared spectroscopy, GC/IR is capable of rapid physical processing of a sample and of generating large quantities of data concerning the sample. A typical GC/IR run may result in the collection of 1-2 thousand interferograms, potentially requiring days to be reduced to useful information and to be interpreted. Work in this laboratory has been concerned with the development of appropriate computer methods to facilitate the routine analysis of mixtures of complex organic compounds using GC/IR.

Methods have been described for producing sample spectra from a GC/IR experiment using on-the-fly determination of chromatographic peak location (2), and instruments for performing this function are commercially available. With such systems, interferograms from each peak are co-added as they are collected and are stored for later processing. It may, however, be preferable in many cases to store all interferograms collected, singly, on some low cost data storage medium such as magnetic tape or disk, thus preserving a complete record of the experiment for documentary purposes or in order to overcome certain limitations in the peak detection system. Such limitations might arise due to an imprudent selection of a peak threshold or due to the inability of an on-the-fly peak detector to look ahead, and thus might be alleviated by the ability to work with the entire chromatogram. Some recent work has therefore concentrated on the processing of such collections of single scan GC/IR interferograms (3).

The goal of this investigation was to develop an integrated system capable of producing, from such an interferogram set, spectra representing the chromatographic peaks and of performing a search of these spectra against

a library of known spectra. In order to accomplish these goals, the program must be capable of reconstructing a chromatogram from the interferogram set and of identifying those interferograms which represent each peak. The program must then co-add the appropriate interferograms from each peak and transform the results into spectra. Finally, the library must be searched for similar spectra and a listing of the search results produced. All of these operations should be carried out with minimal intervention from the chemist, and yet the chemist should be able to review intermediate results and to exercise control over each step if desired. Also, the program must be compatible with a minicomputer, such as might be found in a normal laboratory environment as part of the spectrometer control system.

Developmental work for this project was performed on a stand alone Data General (Southboro, Massachusetts) NOVA 3/12 minicomputer system with a 10 Megabyte disk pack and an 800 bpi tape transport similar to systems which are used to control current Digilab infrared spectrometers. Several single scan GC/IR interferogram sets were obtained from L.V. Azarraga of the Environmental Protection Agency (EPA), Athens, Georgia, who also supplied a library of 820 vapor phase spectra produced under contract to the EPA by Sadtler Research Laboratories. One of the GC/IR data sets was selected for program testing purposes. This set was produced by the injection into a capillary GC of 0.8 µl of a solution containing 0.5 µg/µl each of bis-2-chloroethyl ether, acetophenone, methyl salicylate, 2,3,5-trimethylphenol, acenaphthene, and 2,4,6-trimethylphenol in chloroform. The solutes were eluted in that order. The column effluent was directed through a 1:3 splitter into a light pipe 30 cm x 2 mm i.d. The light pipe was located within the beam of a Digilab (Cambridge, Massachusetts) FTS-14 infrared spectrometer which recorded the interferograms directly onto magnetic tape. The series of programs developed in the course of this work has evolved into a widely used GC/FTIR analysis package known as the GIFTS system.

II. AN INFRARED SEARCH SYSTEM BASED ON DIRECT COMPARISON OF INTERFEROGRAMS (4)

Current compound identification systems using GC/FTIR data perform numerical comparisons between a series of unknown infrared spectra and a library of reference spectra.

The historical precedent for numerical spectral comparisons lies in the technique of visually comparing spectra in order to assign a structure to an unknown. In a numerical approach, the digital computer serves the purpose of the human eye.

The GC/FTIR experiment yields data in the form of interferograms, however. The Fourier transform must be applied to the interferometric data in order to obtain the desired spectra for numerical comparisons with the library. The computer is used to transform the data to the frequency domain where visual comparisons are possible. A major advantage of computers, however, is that they can perform effective numerical comparisons in cases where visual techniques provide little useful information. A relevant example is the case of the interferogram, which contains the same information that is contained in the spectrum, but in a complex form that defies visual interpretation. A more efficient use of the computer would be to perform numerical comparisons on interferograms, rather than on spectra. By eliminating the Fourier transform, a significant increase can be made in the speed of compound identification.

Interferograms have been used computationally in the reconstruction of gas chromatograms from GC/FTIR data (3). This technique is keyed by two important concepts. The interferogram is treated as a multidimensional vector, thus allowing the application of common vector techniques such as Gram-Schmidt orthogonalization, and the use of metrics such as the dot product and the Euclidean distance. Second, an interferogram segment has

been optimized in order to minimize the number of points in the calculation while maximizing the useful information content of the segment.

These concepts have been adopted in the development of the compound identification system. Each interferogram to be processed is reduced from its original 2048 points to a 100-point segment. This segment is formed by reflecting and co-adding the two 100-point segments starting 60 points on either side of the light burst. The resulting co-added segment is treated as a 100-dimensional vector.

The selection of the interferogram segment was done originally by an empirical study. It is known that due to certain phase relationships that exist in the interferogram, those points located near the light burst reflect sensitivity to broad spectral features, while those remote from the burst show sensitivity to narrow spectral features. Recent work confirms that the selected segment provides the best combination of these features for compound identification purposes. The co-addition technique increases the information content of the segment by a signal-averaging process.

The data from a GC/FTIR experimental run consists of a series of single-scan interferograms. Each interferogram is an infrared scan of the GC effluent at a specific time. By assembling a series of interferograms sampled in this manner, a parallel of the GC detector trace is obtained. A reconstruction of the gas chromatogram allows one to determine which interferograms reflect the presence of the unknown components. It has been found that by co-adding those interferograms which make up each peak in the reconstruction, a set of signal-averaged interferograms results which gives the best representation of the unknown compounds. These interferograms cannot be used directly to identify the unknowns, however, because each interferogram contains both sample and reference information. The reference information must be removed before a set of compound-specific data can be obtained.

In this process, n interferograms are selected from the baseline regions

of the reconstructed chromatogram. These interferograms should contain only reference information. By reducing them to the vector form, a set of n linearly independent reference vectors $\bar{X}_1$, $\bar{X}_2$, ... $\bar{X}_n$ is obtained. These vectors span a Euclidean subspace R^n that serves as a multidimensional representation of the reference information. To define this subspace properly, and as an aid to computational efficiency, a set of n orthonormal basis vectors $\bar{B}_1$, $\bar{B}_2$, ... $\bar{B}_2$, ... $\bar{B}_n$ is formed. A familiar example is the set of three basis vectors for R^3 where $\bar{B}_1 = (1,0,0)$, $\bar{B}_2 = (0,1,0)$, and $\bar{B}_3 = (0,0,1)$. The mathematical details of constructing the vector space that spans the reference information has been reported elsewhere (4).

Several techniques can be employed in the selection of reference interferograms. The simplest method is to form the reference subspace from interferograms taken from the beginning of the experimental run. Normally, 5 to 10 references are taken, forming a (5 to 10)-dimensional reference subspace. Instrumental conditions may vary slightly during a GC/FTIR run, however, and references selected at the beginning of the run may insufficiently characterize the reference information of interferograms collected at the end of the run.

An improvement is found when a separate set of references is taken in the vicinity of each peak to be identified. In cases of overlapped peaks, it may not be possible to select separate references from the region of each peak. In this case, the references should be selected from the region directly preceding the group of overlapped peaks. By taking the references from the vicinity of each peak, an improved characterization of the reference information is obtained.

Using the co-added interferograms of the unknown components, a set of sample vectors is formed. The projection of each sample vector onto the reference subspace defines that component of the sample that corresponds to reference information. The component of each sample vector that is orthogonal to the subspace is a compound-specific vector. This vector is

calculated by orthogonalizing the sample vector to the reference subspace. When the orthogonal vector has been calculated, it is normalized. This corrects for concentration differences that may appear in the form of differing vector magnitudes. Each compound-specific vector should point in a unique direction in 100-dimensional space, and the vectors of related compounds should lie close together in this space.

Each compound-specific vector can be identified by calculating its dot product with each member of a library of known vectors. The dot product of two normalized vectors is equal to the cosine of the angle between them. Thus, the calculation indicates how close together the two vectors lie in 100-dimensional space. If two vectors exactly coincide, their dot product is 1. Two vectors that exactly oppose each other produce a dot product of -1. A useful way to express the dot product is to place it on a scale of 0 to 100% correspondence between the two vectors, where 100% correspondence represents a dot product of 1 and 0% represents -1.

The results in our laboratories of the dot product reconstructions, intralibrary searches, and the formal library searches confirm the applicability of a search system based on interferometric data. It has been demonstrated that the interferograms of chemical compounds can be formed into unique compound-specific vectors. The vectors of related compounds tend to orient themselves in similar directions in 100-dimensional space.

The dot product reconstruction, while not as accurate as a formal library search, does provide a way to detect the presence or absence of specific compounds without first having to determine which interferograms correspond to mixture components. This provides a method for rapidly testing GC/FTIR data in order to determine if a specific feature is present in the mixture. Possible applications of this technique arise in cases where a large number of samples have to be processed quickly, and the user is only interested in certain compounds or compound-types. It may be possible to adapt this technique to operate on-the-fly as data are collected. Current

techniques which operate on-the-fly employ small Fourier transforms to detect absorbances at characteristic frequencies (5). The dot product reconstruction allows the detection of specific compounds without having to perform the transforms.

One of the advantages found in the comparison of infrared spectra is that similar compounds produce similar spectra. The intralibrary searches indicate that the interferograms of related compounds produce similar compound-specific vectors. The library searches indicate that specific mixture components can be correctly identified using interferometric data. This validates the assertion that the computer can make an effective interpretation of the interferogram.

The orthogonalization approach is more efficient than the performance of the Fourier transform. The transform is normally applied to the entire 2048-point interferogram, while the orthogonalization employs only 100 points. In addition, the interferometric approach employs a search based on 100 points, comparing favorably with most spectral searches in terms of the number of points that must be compared. The fact that only 160 points past the light burst are required for the interferogram search technique introduces the possibility of using a short-scan interferogram of less than 2048 points. This would decrease the time required to collect each interferogram, while increasing the amount of data that could be stored.

Combining the GC reconstruction with the interferometric search produces a compound identification system that is entirely removed from the Fourier transform. The reluctance that many users have felt toward converting to the GC reconstruction has been partially due to the fact that Fourier transforms still had to be performed in order to obtain data for the spectral search. The introduction of the interferogram search technique allows the user to analyze GC/FTIR data directly, thereby helping to optimize the allocation of his computing resources.