

Computers in Flavor and Fragrance Research

Publication Date: August 22, 1984 | doi: 10.1021/bk-1984-0261.fw001

ACS SYMPOSIUM SERIES 261

Computers in Flavor and Fragrance Research

Craig B. Warren, EDITOR

John P. Walratt, EDITOR

International Flavors and Fragrances

Based on a symposium sponsored by
the Division of Agricultural and Food Chemistry
at the 186th Meeting
of the American Chemical Society,
Washington, D.C.,
August 28–September 2, 1983



American Chemical Society, Washington, D.C. 1984



Library of Congress Cataloging in Publication Data

Computers in flavor and fragrance research.
(ACS symposium series, ISSN 0097-6156; 261)

"Based on a symposium sponsored by the Division of Agricultural and Food Chemistry at the 186th Meeting of the American Chemical Society, Washington, D.C., August 28-September 2, 1983."

Includes bibliographies and indexes.

1. Food—Odor—Analysis—Data processing—Congresses. 2. Flavor—Analysis—Data processing—Congresses.

I. Warren, Craig B., 1939- . II. Walradt, John P., 1942- . III. American Chemical Society. Division of Agricultural and Food Chemistry. IV. American Chemical Society. Meeting (186th: 1983: Washington, D.C.) V. Series.

TP372.5.C66 1984 644'.07 84-14607
ISBN 0-8412-0861-1

Copyright © 1984

American Chemical Society

All Rights Reserved. The appearance of the code at the bottom of the first page of each chapter in this volume indicates the copyright owner's consent that reprographic copies of the chapter may be made for personal or internal use or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per copy fee through the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to copying or transmission by any means—graphic or electronic—for any other purpose, such as for general distribution, for advertising or promotional purposes, for creating a new collective work, for resale, or for information storage and retrieval systems. The copying fee for each chapter is indicated in the code at the bottom of the first page of the chapter.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission, to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

**American Chemical
Society Library
1155 16th St., N.W.
Washington, D.C. 20036**

ACS Symposium Series

M. Joan Comstock, *Series Editor*

Advisory Board

Robert Baker
U.S. Geological Survey

Martin L. Gorbaty
Exxon Research and Engineering Co.

Herbert D. Kaesz
University of California—Los Angeles

Rudolph J. Marcus
Office of Naval Research

Marvin Margoshes
Technicon Instruments Corporation

Donald E. Moreland
USDA, Agricultural Research Service

W. H. Norton
J. T. Baker Chemical Company

Robert Ory
USDA, Southern Regional
Research Center

Geoffrey D. Parfitt
Carnegie-Mellon University

Theodore Provder
Glidden Coatings and Resins

James C. Randall
Phillips Petroleum Company

Charles N. Satterfield
Massachusetts Institute of Technology

Dennis Schuetzle
Ford Motor Company
Research Laboratory

Davis L. Temple, Jr.
Mead Johnson

Charles S. Tuesday
General Motors Research Laboratory

C. Grant Willson
IBM Research Department

FOREWORD

The ACS SYMPOSIUM SERIES was founded in 1974 to provide a medium for publishing symposia quickly in book form. The format of the Series parallels that of the continuing ADVANCES IN CHEMISTRY SERIES except that in order to save time the papers are not typeset but are reproduced as they are submitted by the authors in camera-ready form. Papers are reviewed under the supervision of the Editors with the assistance of the Series Advisory Board and are selected to maintain the integrity of the symposia; however, verbatim reproductions of previously published papers are not accepted. Both reviews and reports of research are acceptable since symposia may embrace both types of presentation.

PREFACE

“WE ARE DROWNING IN INFORMATION but starved for knowledge.” This statement by John Naisbitt in his book *Megatrends* probably best sums up the prevailing theme of the symposium that underlies this book, that is, the use of the computer to convert complex and elusive information into useful knowledge. Our book, *Computers in Flavor and Fragrance Research*, attempts to provide a general view of this impact. The book may be viewed as comprising three sections. The first section, Chapters 1 to 6, deals with the stand-alone use of computers for information handling and number crunching. Chapters 7 through 10 describe the synergistic combination of a dedicated computer and an analytical instrument to create a hybrid analytical technique. Finally, Chapter 11 suggests that we all had better start thinking about the use of robotics as a tool for automating routine laboratory operations.

We chose a general view because the impact of computers on flavor and fragrance research is not limited to a particular area. The advent of the microprocessor has made powerful, inexpensive microcomputers available to the analytical chemist and the sensory scientist alike. These people have connected them to their machines, used them to control robots, and placed them in their sensory evaluation booths. The successful development of inexpensive memory and very fast central processing units, on the other hand, has made very powerful minicomputers available to the computational chemist and the information scientist. These researchers now routinely use the computer to design new functional molecules, design new products, and keep track of huge collections of molecules and associated data.

Although the computer applications presented in this book may appear to be slanted toward the flavor and fragrance field, it should be emphasized that this slant only represents a particular end use of a molecule. The computer applications are valid for all molecules; hence, this book should be of general interest to any researcher whose job it is to develop new molecules or products based on molecules.

CRAIG B. WARREN
JOHN P. WALRADT

International Flavors and Fragrances
1515 Highway 36
Union Beach, New Jersey

March 1984

Application of Microcomputer Technology in Flavor Research

Sensory Evaluation

MARK R. MCLELLAN

Cornell University, Department of Food Science and Technology, New York State Agricultural Experiment Station, Geneva, NY 14456

The advances in microelectronics have brought us the development of powerful microcomputers. These computers can be used in various areas of research including flavor research. A review of the technology is given with specific notes on microcomputer development, design, and languages. A summary of findings from the design and use of the microcomputer as a data collection tool in sensory analysis is reported. Evaluations of this as well as other potential areas of application is included and discussed as related to sensory analysis.

The microcomputer like any other tool needs to be developed and studied for it to perform reliably and consistently. Like few other tools the possible applications of microcomputers has only begun to be discovered. In the field of computers, the pace of development has been phenomenally rapid. It has been roughly akin to going from the Wright Brothers' first flight to the space shuttle in little over ten years. Few other industries have ever matched this relative scale of development. It is certainly reasonable to assume that, with the current advances in large scale and ultra large scale integration and developments within the semiconductor industry in general, computers will eventually affect all aspects of human life.

The definition of a microcomputer

Computers in general have traditionally been distinguished by a number of characteristics. The two foremost characteristics generally cited are, the word size which the computer uses and the speed of operations performed by the computer.

In any computer, the most fundamental unit of information is considered the binary digit, otherwise known as a "bit". In

0097-6156/84/0261-0001\$06.00/0
© 1984 American Chemical Society

script a bit can be represented as either a numerical 1 or numerical 0. However, in the actual circuitry of the computer, this information is represented as the presence of a voltage or the absence of a voltage. Typically, the presence of a voltage is represented at a +5 volt level. The absence of a voltage is generally considered ground potential or 0 volts. The primary unit of information in a computer is a group of bits referred to as a word. An 8-bit word is called a "byte" and this can be considered a universally used unit in the computer industry. In the design of computers, any size of word can be utilized; however, typically, word sizes vary in multiples of 8 bits.

The speed of the computer system is generally cited in terms of the number of machine operations that are capable of occurring within a certain period of time. Many times this is listed as the number of cycles per second (hertz) that the central processor unit can operate at.

Traditionally, a main frame computer was one that operated using 32 or 64 bit words. A minicomputer was one that utilized a 16 bit word and the microcomputer was one that used a 4 or 8 bit word. However, these distinctions in classifications are very quickly disappearing with the further miniaturization of computers, the increasing speed of microprocessors, and specifically the design of more powerful micro and mini computers. Many of the new microcomputers are far more powerful and capable than what just a few years ago was considered a large main frame system.

The birth of the microcomputer can actually be traced back to the development of the transistor. With early electronic devices and computers the system of storing digital information was based on the use of vacuum tubes. These were cumbersome, expensive, and used a tremendous amount of power. They were much faster than relays, however, they were considerably slower than anything produced under today's standards. With the development of the transistor a revolution in the design of computer systems ushered in the time when systems would become smaller and more capable and less costly. The transistor was faster than its predecessor, the vacuum tube, required less power and was much cheaper to develop and produce than the vacuum tube technology.

As development progressed starting in the last 1950's and on into the '70's, a new technology developed which allowed us to build solid state devices containing several components all on the same silicon chip. The term for this technology is integrated electronic circuitry (1). The integrated circuit components essentially became the foundation for the pocket calculator and evolved into what is known today as the microcomputer. This electronic development has rapidly expanded to a point where we now see the use of large scale integration and the development of ultra-large scale integration. This is leading to highly sophisticated equipment available for use by

today's sensory analyst; all one needs to learn is how to communicate appropriately with this new technology.

Communications with a microcomputer

There are a large number of languages that are available for use on microcomputers. A computer "understands" only one language, machine language - binary code. All languages, not binary, must be interpreted or compiled down to machine language, so that computers can understand the instructions. One could use binary to communicate with the computer, however, it would be tremendously burdensome. A more usual means of communicating with the system would be using the more sophisticated or high level languages. BASIC is probably one of the most popular of all computer languages. It can also be considered one of the most versatile. BASIC is an acronym that stands for "Beginners All-purpose Symbolic Instructional Code". There is a version of BASIC available for most every computer system that is marketed today. BASIC does have some drawbacks caused by its inherent lack of structure. It is often said that BASIC programmers have too much freedom to be able to jump around and the result of this type of freedom is the lack of consistency and tremendous complexity in the of development of sophisticated programs. FORTRAN, an acronym which stands for FORMular TRANslation, is a language which can be found on many systems. This was one of the first high level languages to achieve a basic standardization and wide acceptance. It was mainly designed for scientific and mathematical use. In general, the early microcomputers supplied did not have the memory capacity to run FORTRAN programs and therefore had to do with BASIC. With the development of larger memories and ability to run compiled languages, it became possible to run FORTRAN on the smaller systems.

After the widespread use of BASIC and FORTRAN there developed a recognized need to have available structured languages. A structured language is one based on a hierarchy of operations starting with the general and proceeding to the specific. There has been a movement in computer science towards these languages that fully utilize data structures and require users to be very specific in their declaration of all characteristics for a program. Pascal, PL/1, C, and Ada are among the languages of this type and they are quickly replacing older languages for both scientific and business computing. Structured languages, especially Pascal, are becoming major instructional languages in computer science departments. We can certainly expect to see these languages encroach on the domain of FORTRAN, BASIC and COBAL as being the most important language available for use.

The U. S. Department of Defense has decided that Ada shall be the language required on all programming applications dealing with the Army, Navy and Air Force in the near future. Ada was

essentially developed as a derivative of Pascal and it was named after the first computer programmer, Lady Ada Augusta Byron, daughter of the poet Lord Byron. Although we have yet to see any serious marketing effort, of a fully implemented microcomputer based Ada; we should certainly expect to have this language available on most systems in the not too far future.

FORTH can be considered independent of the previously discussed language or what one might consider a Fourth generation of languages. This language is both a language and a computer operating system in one. It is hailed as a giant leap in flexibility and capability (2). FORTH was created by one man, Charles H. More and was first used to control the telescope at Kipp's Peak Observatory. It has since developed a following of programmers and is supported as a good process control language.

Current Applications in Sensory Analysis

Microcomputers as analysis tools

When we speak in terms of analysis tools for sensory evaluation, generally we are talking about the data analysis done using various statistical methodologies. In order to accomplish this type of task one must either turn to a preprogrammed statistical packages or else do specific statistical analysis within ones own programming. There are now available some small statistical packages for microcomputers. However, predominantly these statistical packages have been written by individuals in the field and not by major statistical software vendors. With the development of systems supporting large amounts of memory and sophisticated support software; the development of serious and extensive statistical analysis tools for microcomputers should be fairly straightforward. We are now seeing a very few specific software packages designed for the sensory analysis scientist to help him/her evaluate data from sensory evaluations. An example of this is software being developed by A & N Associates (Lansing, Michigan) for very specific sensory analysis tests including Difference Tests, ANOVA, etc. As we see extended development in this area, we will find software becoming more available and very specialized in the area of support for the sensory analyst.

In this laboratory we are currently in the process of developing and studying an application of rank analysis on microcomputers. This analysis is specifically designed for use in the evaluation of multiple recipes or multiple sample lots for determination of possible rank differences. The resulting output from our software package will consist of an organized listing of judges' rankings, rank totals, rank order of samples based upon rank totals and significant differences where noted. Once fully developed this piece of software would be a tremendous asset to those involved in a comparison of multiple lots and should assist

the researcher in his evaluation of blend-to-blend variations and/or raw product varietal differences. We hope to see further developments in this area of specific software for support of data analysis techniques in sensory analysis.

Microcomputers in data acquisition

In general, we can speak of two specific types of operations using microcomputers in sensory analysis as data acquisition tools. The first would be in a passive mode, that is the system would be used to collect data as entered after the actual evaluation has taken place and the data has been collected on some other form, for example, a paper ballot. In this case the microcomputer is being used as a secondary source of data acquisition. The primary one being the paper ballot. The second mode of operation would be using the microcomputer as the primary and sole data acquisition medium. We would then be relying upon the computer to handle not only the collation and storage of the data but also the interactive means by which the analyst collects data from the panelists.

The use of the microcomputer in a passive mode can be done using any number of techniques. For example, the most common use in this type of secondary data collection phase would be manually punching in the data into the microcomputer. For all intents and purposes this is essentially equivalent to using the system as a data management system or a data analysis system. A second form of this passive data acquisition would be use of the digitizer technology. A digitizer is a system where data is collected using a paper ballot type method and the microcomputer is connected to the digitizer which is then used to reproduce the data from the paper ballot into a digital format as collected by the digitizer. The digitizer itself is essentially an electronic stylus connected to specialized grid network. The grid network acts as a platform where the paper ballot is placed, then the stylus is used to simply touch the paper over each of the premarked points and the grid network underneath the paper identifies where the stylus is touching and what that particular digitized x, y point is. A significant amount of software is required to drive this type of system and to identify the actual data points for what they are.

Another type of passive system has been suggested using automated card reading devices (3). In this case a card is designed to emulate as close as possible the type of sensory evaluation scale required for the particular study. The scale must be denoted in the form of individual marks. Each mark can be read, by the microcomputer using the card reader. The technique is very quick. It can be automated fairly easily and it has been utilized for a number of different types of scales and tests. Although we see it being applied in everything from difference tests to quantitative descriptive analysis, there may

be some biases and problems built in when using this type of marking system for a continuous and semi-continuous scale such as those used for the quantitative descriptive analysis. The scale emulation is done by having a whole series of circles forming the continuous line. The particular circle of interest reflecting the point of interest on the scale must be darkened by a pencil and then that point is read by the card reader. Biases may arise by having discrete points on a line as opposed to a fully continuous scale or line.

The second form of using the microcomputer as a data acquisition tool takes the form of an active or what we call "real-time" data acquisition. By real-time we mean use of the computer for data collection during the actual evaluation by the panelists. The panelists would then interact using various methods with the computer giving results to the computer based on sensory perceptions of the samples.

Thus far there have been two proposed systems for this type of interactive use. The two modes available for this real time data acquisition would be 1) a modified digitizing technique and 2) a direct data entry technique.

In the digitizing system the digitizer would be built into a sensory panel booth. The particular ballot would still be made up in the form of a paper ballot and placed on the electronic grid pallet and then the panelist would evaluate his or her samples and directly mark the paper ballot using the electronic stylus that is located in the booth. Concurrently the computer would monitor the data input from the digitizer and collect the appropriate responses. This application appears to be very successful. There might be considered a number of drawbacks to it in that it requires a fairly heavy investment in hardware, that being the digitizing equipment as well as a sophisticated enough system to control this hardware. It still involves the use of paper ballots, however, in general it is a vast improvement over traditional techniques of data collection in sensory evaluation.

Under the second type of real time data acquisition, we have direct entry methods where essentially you are incorporating an actual terminal into the sensory evaluation booth itself. The electronic terminal would include a data entry keyboard as well as a display screen for receiving results and prompts from the microcomputer (4). This technique has been developed in our laboratories and we feel that it holds tremendous promise in terms of the versatility and the viability for expanding into different areas of sensory evaluation.

Under this type of technique, the system flashed on the display the various questions and scales for use by the panelist to enter his data on. In the prototype that we developed, the software must be programmed for the particular type of test ahead of time. A panelist entered the booth with the samples already

in place and logged onto the computer system. The log on procedure is fairly straightforward where the panelist enters his or her own name and the microcomputer would come back with the response to the person after having checked a small data base for panelist name verification. The microcomputer then randomized the order of the samples that were in front of the panelist. The panelist then responded to the questions for each sample and upon finishing the questions for each of the samples the panelist was told that he was done. The data was then stored in a special file for the purpose of transfer to another computer for data analysis or stored for a later analysis on the same system collecting the data.

The time savings alone based upon this system is enough to warrant further investigation. Based upon a typical quantitative descriptive analysis panel consisting of 18 panels each evaluating 3 samples with 20 questions the time savings over strictly manual methods was approximately 4-5 hours. Compared to digitizing the raw data the time savings was approximately an hour. We feel that this type of direct entry system will offer a tremendous advantage over most other systems in terms of capability, versatility and the development of new applications as well as extensive time savings.

Future Applications

The future applications can generally break down into two specific categories. The first is data analysis by microcomputers. With the increased power of the new systems and extended memory capabilities; we will see the development and application of versions of the traditional large statistical packages available for use on the small systems. With the development of these packages we will have a tremendous capability of doing very extensive data analysis using microcomputer systems. Desk top computers will not be simply be thought of as a personal data base type machine but rather as a major statistical tool that can handle tremendous quantities of data doing highly specific and highly sophisticated data analysis and interpretation.

The second area of progressive development will be in the area of very specialized data collection mediums. We foresee a major extension in the area of direct data input using interactive video terminals built within the sensory evaluation booths. Although this will take a fairly extensive development in the area of software, the gains to be reaped by the development will far exceed the development costs itself. A large area of research has yet to be conducted concerning the application of microcomputers in sensory analysis as a data collection tool. Studies of inherent bias, effects on response freedom etc. will have to be undertaken in order to properly evaluate the computer as a data acquisition tool. Biases based on the way the software

is designed could be very easily incorporated into a sensory project without knowledge by the principle investigator. These type of things should be studied and the people involved in the development of this type of software made aware of it.

We also feel there is room for extensive testing and research in new areas associated with this type of technology. For example, time dependent measurements may be important for various types of sensory evaluation. Previously, to achieve time dependent measurements was a tremendously exhaustive chore and may not have been able to be made using normal taste panel procedures and environments. With this type of application of microcomputers in sensory evaluation we foresee the use of time dependent measurements as a new area of possible development.

Literature Cited

1. Waite, M.; Pardu, M. "Microcomputer Primer"; Howard W. Sams & Co., Inc.; Indianapolis, Indiana, 1977.
2. Dessy, R. E.; Starling, M. K. Am. Lab. 1980, 12(2), 21.
3. Van de Voort, F. R.; Miller, H.; Aston, G. C. Can. Inst. Food Sci. Tech. J. 1981, 14(3), 173.
4. McLellan, M. R. Food Tech. 1983, 37(1), 97.

RECEIVED April 27, 1984

MAECIS

A Computer System for Handling and Analysis of Flavor and Fragrance Molecules

WILLIAM E. BRUGGER

International Flavors and Fragrances, 1515 Highway 36, Union Beach, NJ 07735

MAECIS is a computer program that combines a chemical registry system, a chemical information handling system, and various molecular manipulation capabilities into a single program that is available to the research staff at IFF on an interactive basis. MAECIS provides chemical information retrieval using standard text searching procedures as well as substructure and structure searching. MAECIS also allows users to calculate various molecular properties and a three-dimensional model of any chemical in the system as well as new chemicals yet to be made. MAECIS is designed to be used by the chemist in an interactive manner without the need of a computer expert.

The efficient and proper handling of information are major concerns of all companies doing business in today's world. The ability to retrieve the type of information that allows one to work smarter is a primary driving force behind laboratory and office automation. In companies that deal with chemical processes, information handling systems are more complex due to the need to associate chemical structures with other information. The problem becomes even more complex in the flavors and fragrance industry where it is common to have many chemicals present in a single product as well as one chemical appearing in a number of products. The combination of the above situation with the practice of using trivial names and more than one numbering system for identical products, results in a major chemical information handling problem that only a computer system can solve.

Our solution to this problem was the development of an interactive computer system that can be used by all professionals in the company to locate and manipulate chemical information. This system is called MAECIS, an acronym for Management and Analysis Executive for Chemical Information and Structures. As the name implies, MAECIS is both a chemical information storage system and a chemical structure analysis system.

0097-6156/84/0261-0009\$06.00/0
© 1984 American Chemical Society

MAECIS contains five major sections that are integrated into a single system. At its center is a database management system that handles the storage and updating of data and chemical structures. Associated with this section is an information retrieval section that allows one to search the database for specific information composed of any combination of data and/or substructures. Once information is retrieved, it can be displayed on a computer terminal or sent to a printer for hardcopy output using the display section of MAECIS which includes the ability to display chemical structures in various formats. The final two sections of MAECIS allow for the manipulation of chemical structures and the calculation of various molecular properties that are related to chemical structures.

The majority of MAECIS is written in FORTRAN with a few assembler routines for database management. The program is currently running on a VAX 11/780 computer system and uses Tektronix 4025 graphics terminals as the primary input/output devices. MAECIS is a command oriented program that has a built in prompt mechanism to assist the user with commands. It also has a HELP section that provides the user with detailed online instructions.

DATABASE MANAGEMENT

The database section of MAECIS gives the user access to a large public database or allows him to create a private one. The public database is available for use by any authorized person in a read only mode. The actual maintenance of the data is done by our Technical Information staff who are responsible for the accuracy of the data. To allow a chemist to use the tools available in MAECIS for experimental purposes, a private system is provided. Each chemist can have one or more private files containing "experimental" molecules which can be manipulated in the same manner as those in the public database.

The public database consists of a number of indexed sequential files that contain the data. At the center of the database is the master file which contains an unique International Flavors & Fragrances (IFF) registry number, the Chemical Abstract Service's registry number (if one has been assigned), the molecular weight, various dates associated with discovery and entry of the data, the source of the structure, gas chromatographic retention times, and a series of flags that indicate the availability of various analytical data. Other files in the public database contain chemical names, company code numbers, organoleptic evaluation data, and chemical structure data. The information in these data files are related to the master file by means of the unique registry number that is assigned to each molecule when entered into the system. The public system is setup to allow the entry of a molecule only once. Thus, MAECIS serves as our chemical registry system.

The private database system in MAECIS, on the other hand, is much simpler. It stores only a number and name assigned by the user along with the associated chemical structure. There is no attempt to register the compounds. The owner of a private file has full read and write control over his data.

Along with the public and private databases, there are special public and private files for storing substructures and structure backbones. Backbones are structure elements, e. g., a benzene ring, used to construct more complex chemical structures. Like the public database, the public substructures and backbones can be changed only by authorized people, but the private files are totally controlled by the individual user.

The entry of chemical structures into MAECIS is done using a new version of UDRAW (1) that has been developed for the Tektronix 4025 graphics terminal. The structure is sketched onto the terminal's screen in a two-dimensional graphical representation that all chemist's are familiar with. As the chemical structure is sketched, the computer system is storing its connection table and two-dimensional coordinates. Upon completion of the input, the system checks for any unusual valences or other obvious mistakes in the chemical diagram. If the structure is going into the public database, a search is made of all public structures to make sure that the new structure is unique. If no match is found, the system assigns a new IFF registry number and the user is presented a data form on the CRT screen and asked to enter associated information.

The modification of data already on file is accomplished in a somewhat similar manner. The information in the database is first retrieved through one of the query methods and then altered. In the case of invalid data or an unplausable chemical structure, MAECIS also contains a delete function. MAECIS creates and maintains for the user all of the necessary cross-references between the associated data files.

The chemical connection table(2) in MAECIS stores all the molecular structure information such as atomic valences, atomic symbols, atomic numbers, formal charges, stereo-chemical data, and ring bonding information. Along with the connection table, two-dimensional and three-dimensional coordinates obtained from molecular modeling, are stored for each structure. The storage of three-dimensional coordinates gives the user rapid access to a three-dimensional representation of a public structure without having to go through the molecular modeling procedure. To obtain the three-dimensional coordinates all structures in the public system are automatically modeled in a batch mode and then reviewed for unusual conformations.

INFORMATION RETRIEVAL AND DISPLAY

The information retrieval in MAECIS is accomplished using one of three available commands: SHOW, FIND, or SEARCH. The SHOW command is the simplest one to use and requires only a code number or registry number. It allows the user to retrieve all chemical structures and associated information stored under a particular code number. In most cases this fulfills the user's needs. The FIND command is used for complex searches involving various combinations of multiple data fields. The command also handles substructure searching. Queries such as "find all the structures with a molecular weight between 200 and 250 containing an ester substructure" are handled by the FIND command. Finally, the SEARCH command is used for chemical structure searches. This search takes only seconds and allows the chemist to determine if a particular molecule is already in the database.

The result of any retrieval command in MAECIS is a set of chemical structures which is referred to as the "current set". If a second retrieval is made, the current set is moved into another internal storage area and is referred to as the "prior set". In this way, MAECIS automatically keeps track of at least two retrievals for the user. However, it is often necessary to establish several sets of molecules during a single computer session. In MAECIS, any number of sets can be retained using a SAVE command which allows the user to name and save any set of molecules created by a retrieval command. Any stored set in MAECIS can also be referenced in a subsequent FIND command. Thus, new sets can be generated from old sets through the find command's boolean operations. Sets retained with the SAVE command are in temporary memory. To make them permanent, the STORE and RESTORE commands are available to move sets to and from permanent disk files.

Given a set of molecules, the REVIEW command is used to display the chemical structure and associated information of any member of the set. The review mode allows the user to both page through the current set of molecules, one after another, as well as to jump to a specific entry within the set. The chemical structure currently displayed is referenced by MAECIS as the "current" structure. The user can also designate a second member of the set as the "alternate" structure. This allows for comparison of structures which will be described in the section on chemical structure manipulation.

The REVIEW command draws the chemical structure on one part of the screen, and then writes the associated information on the remainder. To obtain a full screen view of a chemical structure, the DRAW command is used. Options with the DRAW command include atom numbered structures, space-filled diagrams with or without imbedded stick diagrams, stereo pairs, and a special box view where the top, side, and front view of the structure are shown at the same time (see Figure 1). The draw routine can use either

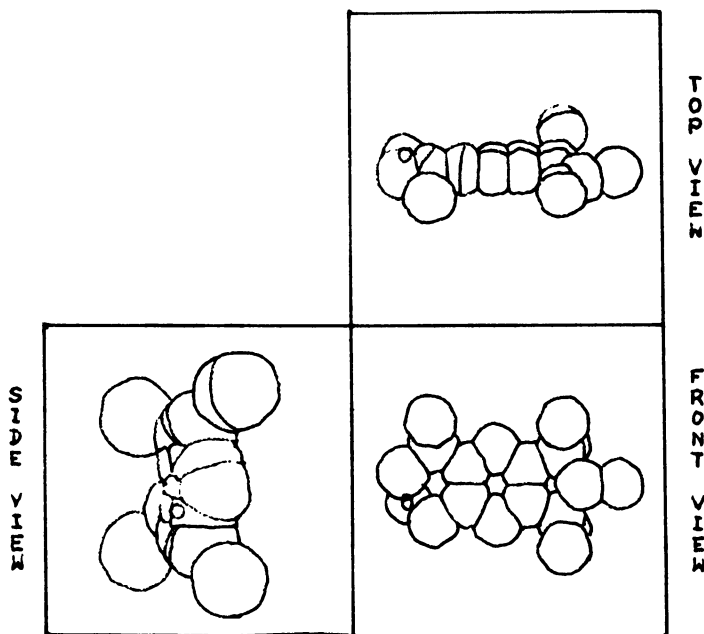


Figure 1. Example of the BOX option of the DRAW command using galaxolide (1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethylcyclopenta g -2-benzopyran) as an example.

two-dimensional display coordinates or the three-dimensional coordinates obtained from molecular modeling. Also, the draw command allows the alternate structure to be drawn with the current structure for a one-to-one comparisons.

MANIPULATION AND CALCULATIONS

Although the two-dimensional representation of a chemical structure is sufficient in chemical information management, the laboratory chemist looking for similarities between a group of chemicals requires additional information obtainable from the three-dimensional representation of the chemical. In MAECIS, all chemicals in public data files are modeled into a low strain conformation (3) and the three-dimensional coordinates are stored with the connection table. When a molecule is retrieved, the user can specify to use the modeled rather than the display coordinates. To assist in the viewing of the three-dimensional structure, the structure can be rotated about any axis or two atoms can be aligned with any screen axis. The DRAW command is then used to view the molecule in its new orientation. In some cases, the chemist may want to move atoms around and then remodel. MAECIS provides these capabilities as well. Hydrogens can be added to the structure or removed. If chiral centers are present in the molecule, the mirror image of the molecule can be obtained with the MIRROR command. In all there are over 10 commands with various options to manipulate three-dimensional chemical structures.

One of the more powerful manipulation commands in MAECIS is COMPARE which allows the current molecule to be superimposed upon the alternate molecule. This is accomplished in the following steps. First, the user specifies the atom pairs in the two structures which are to be overlapped. At least three pairs of atoms must be specified. The program then performs a nonlinear least-squares calculation to minimize the distance between these atom pairs. Finally, the user specifies that the molecules are to be drawn superimposed upon each other. The superimposed molecules can be drawn with any of the options available with the DRAW command. An illustration of this is contained in Figure 2 where two musk odorants are compared. Thus, a chemist can obtain an idea of the structural similarity of any two molecules.

MAECIS also contains a molecular conformation analysis system (4). This system allows the user to generate all possible conformations of the current molecule over a series of single bond rotations. Energy contour maps can be obtained for the various conformations and this allows for the selection of low energy conformations for further manipulation or calculations.

In addition to the manipulation commands, several molecular properties can be calculated for a molecule. The molecular weight is calculated using the exact mass (5) of the most abun-

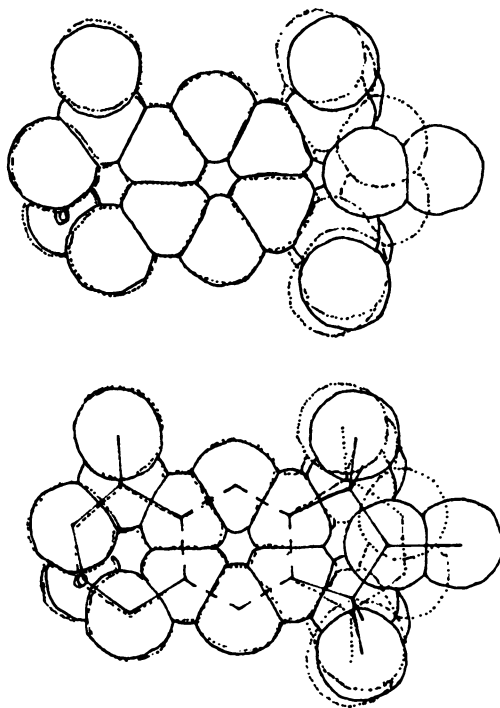


Figure 2. Illustration of the result of a COMPARE operation using galaxolide (1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethylcyclopenta g -2-benzopyran) and Musk-89 (3,4,6,7,8,9-hexahydro-4,6,6,9,9-pentamethyl 1H-naptho 2,3-C pyran) as the example molecules.

dant isotope of each atom in the structure. The molecular volume is calculated using van der Waal radii (6) and either standard or modeled bond distances. Normal, bond, and valence connectivity indices (7) are calculated from the connection table. Also using the connection table, the LogP octanol/water partition coefficients are calculated (8). Finally, the principal moments of the molecule can be calculated using the three-dimensional coordinates in conjunction with the connection table data (9). Since these parameters can be used for doing structure activity analyses, MAECIS provides a method to calculate these parameters for a set of molecules and then to dump them to a computer file for subsequent access by other systems.

SUMMARY

Currently, MAECIS is at the center of our technical information database and contains over 19,000 unique chemical structures. It is used daily by our research staff to locate information on specific products or chemicals. The database structure in MAECIS serves as a bridge between the analytical and organic chemist doing research on specific molecules and those concerned with new products which are usually mixtures of chemicals.

Since its introduction over three years ago, MAECIS has enjoyed wide use for information retrieval. However, the computational parts of MAECIS are still largely unexplored and under utilized. In time this will change as our chemist spend less time looking for information and more time using information.

MAECIS represents the first-generation of chemical informations systems along with many others in various corporations (MMMS at Merck (10); TRIBBLE at duPont (11), MOLY at Rohm and Haas (12), COUSIN at Upjohn (13), and MOLOCH-2 at Searle (14)). Although a large amount of effort has gone into MAECIS, it has more than paid for itself by increasing the productivity of the professional who use it. In the future, new systems will be developed which easily surpass MAECIS in ease of use and satisfaction. However, MAECIS has allowed us to take the first step into computational chemistry and information management.

Literature Cited

1. Brugger, W. E. and Jurs, P. *Anal. Chem.* 1975, 47, 781.
2. Lynch, M. F.; Harrison, J. M.; Town, W. G. "Computer Handling of Chemical Structure Information", Macdonald: London, 1971.
3. Wipke, W. T.; Dyott, T. M.; Verbalis, J. G., Abstract, 161st National Meeting, American Chemical Society, Los Angeles, CA, March, 1971.

4. Hopfinger, J. "Conformational Properties of Macromolecules", Academic Press: New York, 1973.
5. Wapstra, A. H. and Bos, K., Atomic Data and Nuclear Data Tables 1977, 19(3), 177.
6. Bondi, A. J. J. Phys. Chem. 1964, 68, 441.
7. Randic, M. J. Am. Chem. Soc. 1975, 97, 6609.
8. Chou, J. T. and Jurs, P. C. J. Chem. Inf. Comp. Sci. 1979, 19, 172.
9. Stuper, A. J.; Brugger, W. E.; Jurs, P. C. "Computer Assisted Studies of Chemical Structure and Biological Function", Wiley Interscience: New York, 1979.
10. Gund, P.; Andose, J. D.; Rhodes, J. B.; Smith, G. M. Science 1980, 208, 1425.
11. Eaton, D. F. and Pensak, D. A. J. Am. Chem. Soc. 1982, 22, 28.
12. Dyott, T. M.; Stuper, A. J.; Zander, G. S. J. Chem. Inf. Comp. Sci. 1980, 20, 28.
13. Howe, W. J. and Hagadone, T. R. J. Chem. Inf. Comp. Sci. 1982, 22, 28.
14. Hodges, D.; Nordby, D.; Marshall, G., Abstract of the 169th National Meeting of the American Chemical Society, Philadelphia, PA, April, 1975.

RECEIVED May 3, 1984

Computer-Based Molecular Design of Artificial Flavoring Agents

A. J. HOPFINGER and D. E. WALTERS

Searle Research and Development, Department of Medicinal Chemistry, Skokie, IL 60077

Computational chemistry methodology is finding increasing application to the design of new flavoring agents. This chapter surveys several useful techniques: linear free energy relationships, quantitative structure-activity relationships, conformational analysis, electronic structure calculations, and statistical methods. Applications to the study of artificial sweeteners are described.

Implicit to research in the biochemical sciences, of which the food and flavor industries are members, is that biological response, BR, is a function of chemical structure, CS. The goal is to find that CS (compound) which maximizes a desired BR and, simultaneously, minimizes other BRs, especially those directly relating to toxicity. In practice, this involves the synthesis and testing of new chemical entities, and, unfortunately, is governed to a large extent by chance. Application of computer-based molecular modeling is intended to reduce the chance component, and, thereby, to enhance design capabilities.

The design function rests upon recognition and implementation of the relationships shown in Figure 1. In part A we formalize that not only do all BRs map into (depend upon) CS, but so do all physicochemical properties, PPs. Therefore, the CS can, from functional mapping, be considered the common node between the BRs and the PPs. Consequently, a unique function, f_{α} , must exist between each BR and the set of PP. This is defined in α part B of Figure 1.

The key assumption in molecular design is that if a particular BR is of interest, say BR_{λ} , then the corresponding f_{λ} holds for any CS. This is expressed in part C of Figure 1 where CS_m denotes an arbitrary m th chemical structure. The reliability of this assumption increases as the structural homology of the set of CS increases and, also, as the mechanisms of action to realize BR_{λ} over the $[CS_m]$ converge to commonality. This is a formal explanation why quantitative structure activity relationship, QSAR, studies are more successful when applied to explain in vitro activity in a set of congeneric analogs, as opposed to in vivo activity for a set of structurally diverse compounds.

0097-6156/84/0261-0019\$06.00/0

© 1984 American Chemical Society

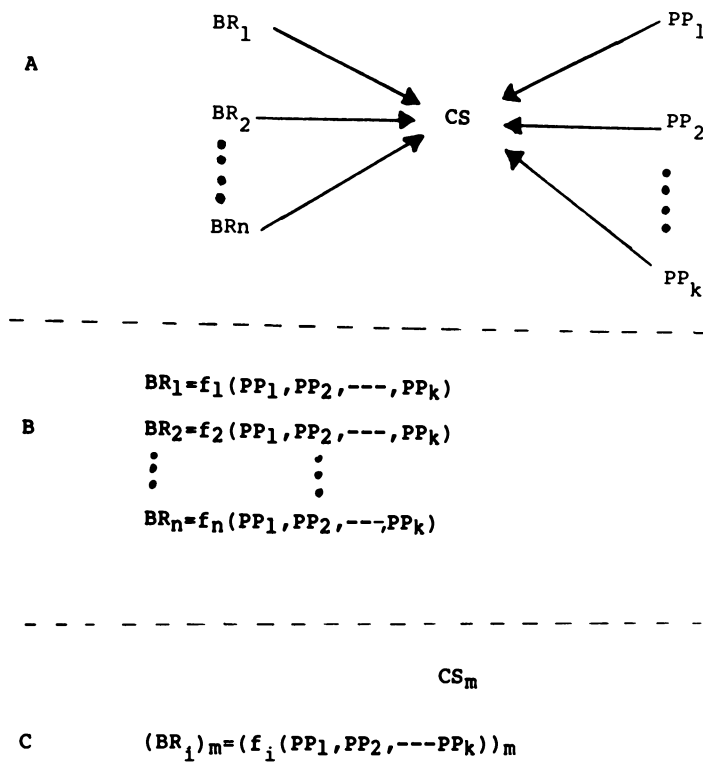


Figure 1. The interrelationships between biological response (BR), chemical structure (CS), and physicochemical properties (PP) that lay the conceptual basis for molecular design.

Methods

1. Additive Property Models

The simplest representation of a molecule, with respect to computing physicochemical properties, is to assume the property to be the sum of the property values of the individual constituent atoms, or groups of atoms. Extensive data bases (1,2) of atomic and group (fragment) property values have been compiled to facilitate implementation of this model. The most notable physicochemical properties employed in QSARs using an additive property model are:

1. log P, the water-octanol partition coefficient, (3)
2. σ , the Hammett constant, (4)
3. MR, molecular refractivity index, (5)
4. $\text{p}K_a$, ionization constant, (6) and
5. E_s , the Taft steric constant. (7)

Log P and MR are considered thermodynamic descriptors, $\text{p}K_a$ a combined thermodynamic and electronic index, and σ an electronic property index. E_s is designed to account for steric effects. Corrections for non-additivity, based upon the chemical bonding topology, have been suggested and used. These include proximity, bond type, ring, and group shape correction features. (8-10)

Molecular connectivity, (11) which is based upon graph theory, (12) is an empirical alternative to an additive model employing physicochemical properties. Descriptors derived from the molecular connection table (chemical bonding topology) employing mathematical functions are considered as potential activity correlates. As such, this approach is completely mathematical and has no obvious physicochemical basis. Its strength is that correlation indices can always be generated (provided one knows how to assign intrinsic relative weights to individual atom types).

2. Hansch Analysis

The most successful, and the most often used, method to construct a QSAR is that of Hansch. (13) This method employs multi-dimensional linear regression analysis to correlate structure to activity in a chemically congeneric set of compounds. The structural features have been traditionally derived from additive property models. However, recent applications of Hansch Analysis recognize any and all molecular descriptors as potential correlates to activity. (14,15) Prominent among this line of thinking is Hansch himself who now freely uses indicator variables as correlates. (16) An indicator variable has a value of 1 if some user-defined property is present in a compound and a value of zero if the property is absent. It is important to point out that Hansch Analysis is based upon a biological action model. By deriving the general QSAR equation associated with the

action model, it is possible to conceptually justify the general usage of any molecular descriptor in a correlation analysis. (17)

3. Conformational Analysis

A drawback to additive property models of molecular structure is that 3-dimensional molecular properties cannot be determined. However, it is clear that conformation, or more generally, molecular shape, can be an important factor in expression of biological activity. Thus, there is a need to be able to determine conformational/shape properties of molecules and to use these properties to develop a QSAR.

There are two principal methods of performing computational conformational analyses: Molecular mechanics(18-19) and quantum mechanics.(20) Molecular mechanics considers a molecule as a set of balls (the atoms), coated with sticking paste, connected by a set of springs (the bonds) according to a prescribed set of valence angles. Quantum mechanics views the molecule as a set of nuclei in space with electrons moving about the nuclei.

The goal of a conformational analysis is to minimize the energy of the molecule as a function of its geometry. The energy minimization can involve different types and numbers of geometric degrees of freedom which include bond lengths, bond angles, and torsional bond rotation angles. An often employed approximation, especially for large molecules, is to hold the valence geometry constant and minimize the energy as a function of only torsional bond rotations.

The unproven, but reasonable, assumption implicit to SAR-directed conformational studies, both experimental and theoretical, is that one of the stable intramolecular conformers is the "active" conformation. A difficulty to applying conformational data in quantitative drug design is selection of conformational features for QSAR development. Moreover, molecular shape properties are preferable features to have available in design studies. Conformation is a component of shape. The properties of the atoms, most notably their "sizes," comprise an additional set of factors needed to specify molecular shape.

Conformational features have been used in some structure-activity studies. Some examples are:

- (1) an interatomic distance within a molecule, (21,22)
- (2) a set of interatomic distances within a molecule, (23,24)
- (3) a set of atomic coordinates within a molecule, (25,26)
- (4) a set of critical intermolecular binding distances. (27)

Molecular shape properties, derived from conformational investigations, have also been used to rationalize commonality and diversity in biological action. These shape properties include:

- (1) molecular volume, (28)
- (2) molecular surface area, (29)

- (3) spatial potential surfaces of a molecule with respect to a test species. (30,31)

A very powerful tool of visualizing three-dimensional molecular properties, including potential surfaces, is computer graphics. (32) Computer graphics is particularly useful in the qualitative comparison of two or more molecules.

A general theory of quantitatively comparing molecular shapes using common overlap steric volume (33-36) and, more recently, descriptors derived from superimposed molecular potential energy fields of pairs of molecules (37) has been derived and tested. This theory allows a "marriage" between Hansch analysis and conformational analysis.

4. Electronic Structure Calculations

Molecular orbital theory provides electronic, as well as conformational, data for inclusion in QSAR development. Electronic properties of a molecule might, in part, control biological activity when chemical reactions are part of the mechanism of biological action. Electronic indices should also be considered in modeling physical interactions in order to provide charge distribution data for estimating binding energies.

Electronic indices often considered in structure-activity studies include;

- a) atomic charge densities,
- b) bond, group, and/or molecular dipole moments,
- c) orbital energy levels, especially the highest occupied molecular orbital, HOMO, lowest unoccupied molecular orbital, LUMO, and their difference, and
- d) orbital wavefunction coefficients.

Several different molecular orbital methods have been used in SAR investigations. These include simple Huckel theory, HT, (38) extended Huckel theory, EHT, (39) CNDO, (40) NDDO, (41) MINDO/3, (42) and PCILO. (43)

5. Statistical Methods

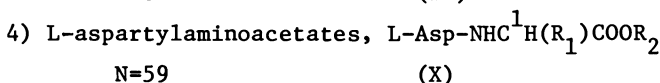
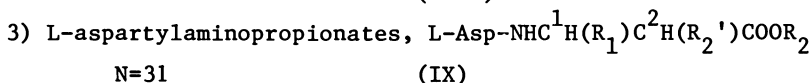
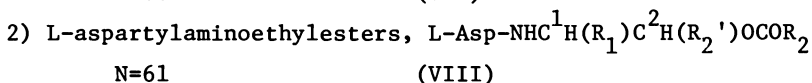
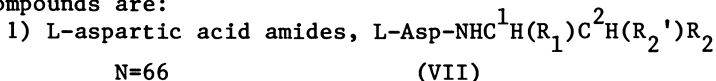
Multidimensional linear regression analysis is the most often employed statistical method for QSARs. This popularity is coupled to the acceptance of the Hansch method for QSAR analyses. The techniques and pitfalls of regression analysis have been well described. (44,45)

Other statistical methods employed in quantitative molecular design investigations include discriminant analysis, cluster analysis, multiple factor analysis, and pattern recognition procedures. (46-48) Pattern recognition may prove particularly useful when the design objective is a complicated profile of biological activities, as opposed to only a maximized potency and minimized toxicity. Gottman (49) and Schiffman (50) have used multidimensional scaling methods to establish interrelationships between psychological sensory profiles and physicochemical properties of flavor agents, especially sweeteners.

Case Study Applications

L-Aspartyl Dipeptide Sweeteners

Iwamura(51) has investigated the structure-sweetness relationship in four classes of L-aspartyl dipeptides using linear free energy descriptors and multi-dimensional regression analysis. In essence, the Hansch methodology was employed. The four classes of compounds are:



N is the number of compounds in each class.

QSARs were generated for each of the four data bases. A representative multi-dimensional linear regression equation is that developed for the L-aspartylaminoethylesters (see Figure 2):

$$\begin{aligned} \log(\text{SP}) = & 0.67\sigma^*(\pm 0.48) + 3.36L_2(\pm 1.03) - 0.29L_2^2(\pm 0.08) \\ & + 4.18W_1(\pm 0.88) - 0.85W_1^2(\pm 0.18) - \\ & 0.53L_1(\pm 0.18) - 11.33 \end{aligned} \quad (1)$$

n=51 r=0.88

In equation (1) the L_i parameter (9) expresses the length of substituent R_i to the rest of the molecule. W_i is the width upward of R_i when one views it from the connecting end along the bond axis defining L_i. The electronic parameter, σ*, was estimated for the structure substituted on the common aspartyl-amino moiety, so that the electronic effect is directed to the peptide bond. Ten compounds of the original 61 do not fit equation (1); hence n=51, not 61.

The analysis, in composite over the four classes of L-aspartyl dipeptides suggests that the electron-withdrawing effect of substituents directed to the peptide bond, and the steric dimensions of the molecules, are important in eliciting the sweet taste. The values of the regression coefficients of the σ* term in the QSAR equations for L-aspartic acid amides, L-aspartylaminoethylesters, and L-aspartylaminopropionates all

are approximately 0.7 suggesting these three classes of dipeptide sweeteners interact in a common manner at the receptor. However, the σ^* regression coefficient for the L-aspartylaminoacetates QSAR is approximately 1.5. This value, along with an examination of the optimum steric parameter values in the QSARs, suggest to Iwamura, that the L-aspartylaminoacetates can better fit to the receptor than the other three classes of compounds.

The QSARs developed by Iwamura provide concise recipes for designing dipeptide sweeteners. However, the uniqueness of an LFE-based QSAR is always an issue. For this case the singularity of the QSARs with respect to the steric parameters L_1 , L_2 , and W_1 is suspect. The values of L_1 and W_1 can be shown to correlate to the lipophilicity of R_1 as measured by log P for 38 compounds with a correlation coefficient of .943. Also L_2 correlates with log P for 40 choices of R_2 with a correlation coefficient of .908. Overall, it is possible to express log (SP) as;

$$\log(\text{SP}) = f(\sigma^*, \log P(R_1), \log P(R_2)) \quad (2)$$

n=36 r=0.87 s=0.24

The practical advantage of equation (1) over equation (2) is that reliable estimates of log P for choices of R_1 and R_2 are available for only 36 of the 51 compounds for which L_1 , W_1 and L_2 can be measured. Nevertheless, care needs always to be taken in establishing the extent of correlation uniqueness among LFE descriptors. This is normally achieved by constructing a correlation matrix among all descriptors considered.

Extension of the Shallenberger AH...B Sweetener Model

Shallenberger and coworkers(52) proposed what is probably the most consistent requirement in sweeteners, namely the A-H...B hydrogen-bond model. An example of this model, as applied by Hopfinger and Jabloner,(53) is shown in Figure 3 for the artificial sweetener P-4000 (4000X sucrose). In this model the A-H group provides a proton donor for a weak intramolecular hydrogen bond to an acceptor B. B can be any of a wide range of electronegative groups. The distance between the proton and B is 2.5-4.0Å, a range of values larger than normally associated with hydrogen bonds (1.7-1.9Å). The problem with the A-H...B model is that many organic compounds, which are not sweet, contain such a grouping. Further, many sweeteners contain multiple A-H and/or B groups which makes it ambiguous in assigning the relevant A-H...B system. For example, the artificial sweetener aspartame contains three possible A-H sites and six possible B atoms, see Figure 4. Thus, the utility of using the A-H...B model in a design mode is limited. A hydrophobic site "located" 5-6Å from the A-H...B system has been suggested(54) as an additional constraining feature necessary for sweeteners. Again a wide range of non-sweet organic compounds meet this additional requirement.

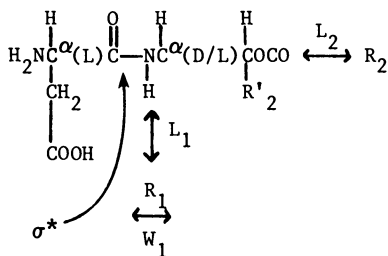


Figure 2. General structure of the L-aspartylaminoethylesters with R_1 , R_2 , L_1 , W_1 , L_2 , and σ^* defined.

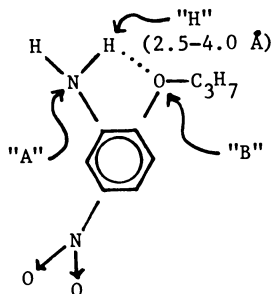


Figure 3. P-4000 with the A-H . . . B system defined.

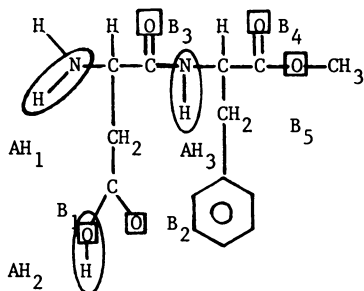


Figure 4. Multiple A-H (ellipses) and B (squares) sites in aspartame.

Therefore, the question arises if there are additional properties related to the A-H...B system that might be characteristic of sweeteners. Hopfinger and Jabloner(53) were struck by the unusual geometry of many A-H...B systems in sweeteners. The geometry in Figure 3 is exemplary. One could envision the formation of bimodal hydrogen bonding involving both H and B with intermolecular, receptor-site acceptors and donors respectively.

These workers proceeded to carry out theoretical conformational analyses of a set of sweeteners, many of which have multiple A-H and/or B sites. The set of compounds analyzed are reported in Table I along with the sweet potency relative to sucrose. A fixed valence geometry molecular mechanics force field was used in the conformational analyses. The conformational search strategy was as follows:

1. Generate all initial conformer states having at least one A-H...B system in which the H...B distance is in the 2.5-5.0Å range.
2. Minimize the energy of each initial conformer as a function of the allowed torsional degrees of freedom.
3. After total energy minimization record/calculate the properties of the A-H...B hydrogen bond including the bond energy.

The most interesting observation of this analysis is that each compound in the data base can exist in at least one stable conformer for which the A-H...B hydrogen bonding energy $E(\text{HB})$ is in the range of -1.5 to -2.5 kcal/mole. Further, most potent sweeteners have an $E(\text{HB})$ near -2.0 kcal/mole. The set of $E(\text{HB})$ are reported as part of Table I, and a histogram plot of sweet potency versus $E(\text{HB})$ for the compounds of Table I is shown in Figure 5. Organic compounds not sweet, but containing an A-H...B system, have a range in $E(\text{HB})$ values of -0.6 to -9.0 kcal/mole. Thus, it is tempting to postulate that an additional necessary, but perhaps not sufficient, condition for sweetness is that the A-H...B system have an $E(\text{HB})$ in the -1.5 to -2.5 kcal/mole range.

One possible interpretation of A-H...B energetics is that $E(\text{HB})$'s weaker than -1.5 to -2.5 kcal/mole may not be able to spatially orient the A-H...B groups properly while $E(\text{HB})$'s stronger than -2.5 kcal/mole may be too great to allow the receptor B...A-H system to compete with the intermolecular behavior.

An obvious questions which arises is what factors are responsible for $E(\text{HB})$? A general A-H...B system is shown in Figure 6. The energy $E(\text{HB})$ is proportional as,

$$E(\text{HB}) \propto \frac{Q_H Q_B}{\epsilon d_{\text{HB}}} f(\theta(\text{AHB})) \quad (3)$$

TABLE I

Relative sweetness, RS, and hydrogen bond energy, E(HB)
of an AH...B system in a set of sweeteners.

Number	Compound	RS	E(HB)kcal/mole
1	P-4000	4000	-1.95
2	APM	180	-1.83
3	phyllodulcin	250	-1.80
4	(D)-tryptophan	35	-1.66
5	(D)-6-Cl tryptophan	1600	-1.85
6	(D)-serine-O-n-C ₃ H ₇	320	-1.72
7	(D)-threonine-O-N-C ₃ H ₇	150	-1.75

$$\begin{array}{c}
 \text{R}_1\text{NHCH}(\text{C}=\text{O})\text{NH}-\text{C}_6\text{H}_4-\text{R}_2 \\
 | \\
 (\text{CH}_2)_n\text{CO}_2\text{H}
 \end{array}$$

	R_1	n	R_2	RS	E(HB)kcal/mole
8	COCF ₃	1	CN	3000	-2.13
9	COCF ₃	1	NO ₂	100	-1.92
10	COCH ₃	1	NO ₂	0	-1.49
11	COCF ₃	2	CN ²	3000	-2.01
12	cyclamate			25	-1.50
13	neohesperidin	DHC		340	-1.79
14	perillartine			225	-1.65
15	acesulfam			130	-1.55
16	saccharin			260	-1.60

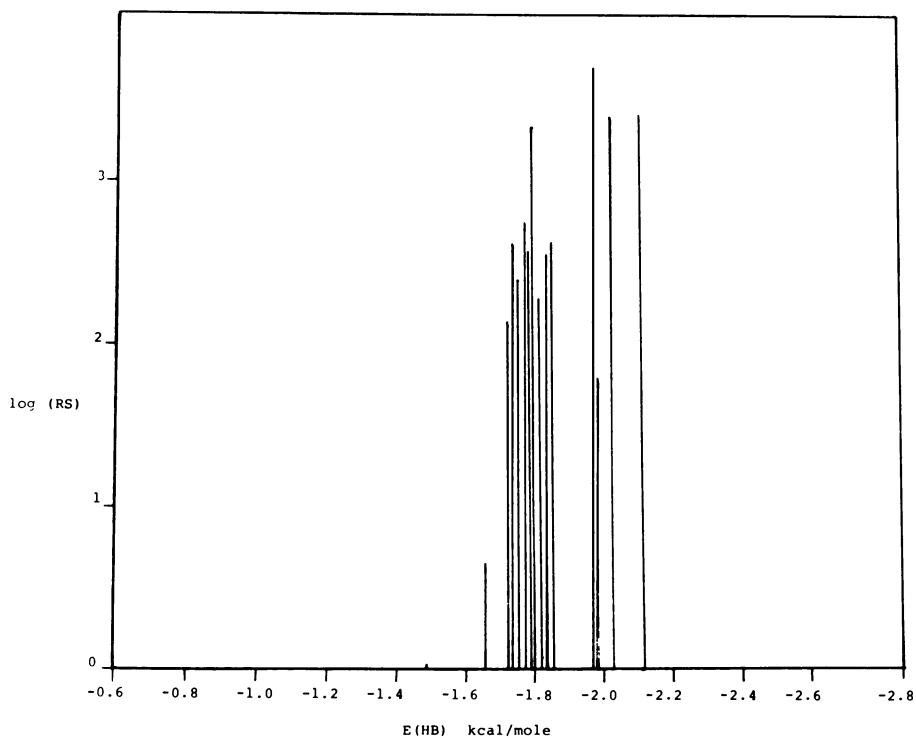


Figure 5. A histogram of RS vs. E(HB) for the set of sweeteners reported in Table I.

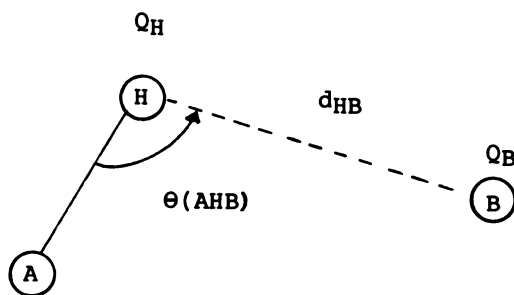


Figure 6. General structure of a hydrogen bond.

Q_H and Q_B are the partial atomic charges on the proton and B atoms, respectively. d_{HB} is the distance between the H and B atoms and $f(\theta(AHB))$ is a scaling function between 0 and 1 depending upon the bond angle $\theta(AHB)$. The fixed molecular dielectric is expressed as ϵ . Equation (3) indicates that four variables control $E(HB)$ and that realization of a small range in values like -1.5 to -2.5 can be due to competitive contributions over the four variables. Nevertheless, it is quite interesting to observe that these four structural variables interplay to elicit a very tight range in $E(HB)$ for the set of sweeteners considered.

Summary and Prospects

The pharmaceutical industry has pioneered in the application of computer-assisted drug design methods in product research. To a significant degree this is a consequence of the direct use of computational chemistry in enhancing the efficiency of the chemical lead optimization process.

For the flavor industry to take significant advantage of quantitative molecular design, more stringent measures of biological responses are necessary. The artificial sweetener area represents an example in flavor chemistry where high quality measures of sweet potency have made it possible for QSAR techniques to be successfully applied. Alternate artificial flavoring areas could be well served by following the sweetener example.

A great amount of methodology in molecular design is now available for artificial flavoring QSAR applications. Much of the methodology is formatted in easy to use software packages. Thus, computer-assisted molecular design can be viewed as a ready to use analytical tool by flavor chemists. The major barrier to using these tools is the lack of familiarity and understanding flavor chemists have with the available software.

Literature Cited

1. C. Hansch and A. Leo, "Substituent Constants for Correlation Analysis in Chemistry and Biology," Wiley-Intersciences, New York, 1979.
2. R. R. Rekker, "The Hydrophobic Fragmental Constant," Pharmacochimistry Library, Vol. 1, Elsevier, New York, 1977.
3. C. Hansch and J. M. Clayton, *J. Pharm. Sci.* 1973, 62, 1.
4. M. Charton, *Chemtech* 1974, 502; 1975, 245.
5. C. Hansch, S. H. Unger, and A. B. Forsythe *J. Med. Chem.* 1973, 16, 1217.
6. P. Seiler, *Eur. J. Med. Chem.* 1974, 9, 663.
7. A. J. Verloop, in "Drug Design," Vol. 3, E. J. Ariens, Ed., Academic Press, New York, 1972, p. 133.
8. A. Leo, P. Y. C. Jow, C. Silipo, and C. Hansch *J. Med. Chem.* 1975, 18, 865.

9. A. Verloop, W. Hoogenstraaton, and J. Tipker, in "Drug Design," Vol. 7, E. J. Ariens, Ed., Academic Press, New York, 1976, p. 165.
10. K. Bowden and K. R. H. Woolridge Biochem. Pharm. 1973, 22, 1015.
11. L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research," Academic Press, New York, 1976.
12. R. Wilson, "Introduction to Graph Theory," Academic Press, New York, 1972.
13. C. Hansch, A. R. Steward, J. Iwasa, and E. W. Deutsch Mol. Pharmacol. 1965, 1, 205.
14. C. Silipo and C. Hansch J. Am. Chem. Soc. 1975, 97, 6849.
15. Y. C. Martin, T. M. Bustard, and K. R. Lynn J. Med. Chem. 1973, 16, 1089.
16. C. Hansch, C. Silipo, and E. E. Steller J. Pharm. Sci. 1975, 64, 1186.
17. A. J. Hopfinger, "Intermolecular Interactions and Biomolecular Organization," Wiley-Interscience, New York, 1977.
18. A. J. Hopfinger, "Conformational Properties of Macromolecules," Academic Press, New York, 1973.
19. D. B. Boyd and K. B. Lipkowitz J. Chem. Ed. 1982, 59, 269.
20. G. A. Segal, Ed. "Methods of Electronic Structure Calculations, Parts A and B," Plenum Press, New York, 1977.
21. L. B. Kier J. Pharmacol. Exp. Ther. 1968, 164, 75.
22. N. S. Ham in "Molecular and Quantum Pharmacology," E. D. Bergmann and B. Pullman, Eds., Reidel-Dordrecht, Holland, 1974, p. 261.
23. D. C. Rohrer, D. S. Fullerton, A. H. L. From, and K. Ahmed in "Computer Assisted Drug Design," E. C. Olson and R. E. Christoffersen, Eds., ACS Monograph Series 112, Washington, D.C., 1979, p. 259.
24. L. B. Kier in "Fundamental Concepts in Drug-Receptor Interactions," J. Danielli, J. Moran, and D. Triggle, Eds., Academic Press, New York, 1970, p. 15.
25. H. J. R. Weintraub and A. J. Hopfinger J. Theor. Biol. 1973, 41, 53.
26. H. Weinstein, B. Z. Apfelder, S. Cohen, S. Maayani, and M. Sokolovsky in "Conformation of Biological Molecules and Polymers," E. D. Bergmann and B. Pullman, Eds., Academic Press, New York, 1973, p. 531.
27. G. M. Crippen J. Med. Chem. 1979, 22, 988.
28. K. Yamamoto J. Biochem. 1974, 76, 385.
29. D. Pensak, Unpublished work, 1978.
30. H. Weinstein Int. J. Quantum Chem. 1975, 2, 59.
31. J. Bartlett and H. Weinstein Chem. Phys. Lett. 1975, 30, 441.
32. N. L. Max, D. Malhotra and A. J. Hopfinger Computers and Chem. 1981, 5, 19.
33. A. J. Hopfinger J. Am. Chem. Soc. 1980, 120, 7196.
34. C. Battershell, D. Malhotra, and A. J. Hopfinger J. Med. Chem. 1981, 24, 812.

35. A. J. Hopfinger J. Med. Chem. 1981, 24, 818.
36. A. J. Hopfinger and R. Potenzzone, Jr. Mol. Pharmacol. 1982, 21, 187.
37. A. J. Hopfinger J. Med. Chem. 1983, 26, 990.
38. A. Streitwieser, "Molecular Orbital Theory for Organic Chemists," Wiley, New York, 1961.
39. R. Hoffman J. Chem. Phys. 1963, 39, 1397.
40. J. A. Pople, D. P. Santry, and G. A. Segal J. Chem. Phys. 1965, 43, 5129.
41. O. Kikuchi Bull. Chem. Soc. Japan 1977, 50, 593.
42. R. C. Bingham, M. J. S. Dewar, and D. H. Lo J. Am. Chem. Soc. 1975, 97, 1302.
43. J. L. Coubeils, Ph. Courriere, and B. Pullman C. R. Acad. Sci., Paris 1971, 272, 1813.
44. P. J. Goodford in "Advances in Pharmacology and Chemotherapeutics," Academic Press, New York, 1973, p. 52.
45. J. G. Topliss and R. J. Costello J. Med. Chem. 1972, 15, 1066.
46. B. R. Kowalski and C. F. Bender J. Am. Chem. Soc. 1973, 95, 586.
47. R. J. Matthews J. Am. Chem. Soc. 1975, 97, 935.
48. C. L. Perrin Science 1973, 183, 551.
49. L. Bottman Psychometrika 1968, 33, 469.
50. S. S. Schiffman Science 1974, 185, 112.
51. H. Iwamura J. Med. Chem. 1981, 24, 572.
52. R. S. Shallenberger and T. C. Acree Nature (London) 1967, 216, 480.
53. A. J. Hopfinger and H. Jabloner in "The Quality of Foods and Beverages," G. Charalambous and G. Inglett, Eds., Academic Press, New York, 1981, p. 83.
54. L. B. Kier J. Pharm. Sci. 1972, 61, 1394.

RECEIVED May 4, 1984

Mathematical Approaches for Quantitative Design of Odorants and Tastants

SUSAN S. SCHIFFMAN

Duke Medical Center, Department of Psychiatry, Durham, NC 27710

Odor and taste quality can be mapped by multidimensional scaling (MDS) techniques. Physicochemical parameters can be related to these maps by a variety of mathematical methods including multiple regression, canonical correlation, and partial least squares. These approaches to studying QSAR (quantitative structure-activity relationships) in the chemical senses, along with procedures developed by the pharmaceutical industry, may ultimately be useful in designing flavor compounds by computer.

It is not yet possible to design a molecule with specific odor (or taste) characteristics because the relations between sensory properties of flavor compounds and their molecular properties are not well understood. As a consequence, the development of compounds with desired flavor qualities has had to rely on relatively tedious synthetic approaches. Recent advances, however, in computer-based methods developed by the pharmaceutical industry to study QSAR (quantitative structure-activity relationships) may ultimately be helpful in the rational design of new flavor-structures with predictable sensory attributes. Results from QSAR studies may also provide insight into the mechanism of the molecule-receptor interaction.

QSAR studies of flavor molecules require two types of data as input:

1. Quantitative measures obtained in psychophysical experiments that describe the sensory properties of the molecules.
2. Physicochemical parameters that provide a description of the molecular properties relevant to the flavor properties.

Understanding the relations between the psychophysical and physicochemical properties can be achieved by a variety of mathematical methods described in this paper.

0097-6156/84/0261-0033\$06.00/0

© 1984 American Chemical Society

Obtaining quantitative sensory data. In the past, the quantification of odor (and taste) quality has relied upon ratings on verbal descriptors. This approach has several drawbacks in that: (a) it is high in experimenter contamination because a priori assumptions must be made about the relevant attributes and (b) there is considerable variability among subjects because of individual differences in interpretation of the meanings and relative importance of the words. The recently developed methodology of multidimensional scaling (MDS) overcomes these drawbacks by using direct measures of similarity as input (1). MDS is simply a mathematical tool that uses experimental judgments of perceived flavor similarity to derive a map that represents the taste or smell qualities of specific molecules as points in a space. Compounds judged similar in taste (or odor) sensation are arranged by MDS procedures close to one another in a resultant spatial map; those dissimilar are located far from one another.

An example from geography illustrates how MDS procedures work. Maps of the United States frequently give tables of intercity distances among major cities. If MDS were applied to all combinations of distances between these cities, the analysis would accurately recover the underlying structure, that is, a map of the cities in proper relationship to one another. The difference between geographical and flavor spaces is simply that we don't know a priori what flavor spaces look like.

There are many ways to obtain distance-like measures among flavor compounds. First, stimuli are generally equated in subjective intensity so that judgments are based on quality rather than intensity. Odorants are diluted in an odorless grade of diethyl phthalate and tastants, in deionized water. Then subjects rate all the $n(n-1)/2$ possible pairs for a set of n stimuli along an undifferentiated 5" line:

exact	most
same	different

Similarity measures can also be obtained using a 10 point scale (e.g., from 1 to 10) or by obtaining confusabilities among triads. Odorants (or tastants) confused most frequently are considered more similar. Missing data designs can be used to reduce the number of actual stimulus presentations.

The similarity judgments are then analyzed by one of a number of specially designed computer programs that have been detailed by Schiffman, Reynolds and Young (1). Some programs, such as INDSICAL, ALSCAL, and MULTISCALE, not only provide multidimensional arrangements of molecules based on their flavor similarity but provide quantitative measures that delineate the individual differences in response for each subject as well. The

maps obtained by MDS can be interpreted by a variety of statistical techniques, including multiple regression, canonical correlation, partial least squares, and a variety of other methods to determine those properties of molecules responsible for their flavor properties; these approaches are described below.

Obtaining relevant physicochemical parameters. The choice of physicochemical parameters to relate to MDS spaces is crucial if the properties found to be mathematically important are indeed appropriate chemical predictors for future design of molecules with desired flavor properties. Unfortunately, we often have no idea what physicochemical properties are indeed important, although many of the parameters described in the examples below as well as those given in Table I (see 2, 3) are probably relevant.

Hydrophobic bonding, characterized by the oil-water partition coefficient, has been implicated in intense sweetness, perception of bitterness, as well as odor intensity. However, it will be shown in Study 2 below that when the concentrations of pyridyl ketones are adjusted to equal subjective intensities, the partition coefficient is not particularly important in discriminating qualitative differences. Short range electrical properties such as dispersion bonding are probably an important source of specificity and may explain why different odor qualities are produced by identical substituents at different positions of a molecule. Dispersion bonding is related to the polarizability of molecules and inversely to the sixth power of their separation. Longer range electrical forces, whether ion-ion, ion-dipole, or dipole-dipole are also likely to be significant. Hydrogen bonding has been implicated in sweet perception. Charge transfer effects such as electron donation by ethers and many aromatic compounds as well as electron acceptance of mercaptans is probably important. Steric repulsion, though not bonding but nevertheless a non-covalent interaction, may also contribute to specificity.

Examples

Four studies are described here that relate physicochemical properties to odor quality as defined by maps derived by multidimensional scaling procedures. The mathematical procedures used to relate the physicochemical properties to the maps are discussed as well.

Study 1: Broad range of odorants. A group of odorants that varied widely in quality and structure were arranged on the basis of odor similarity in a two dimensional space shown in Figure 1 (4). The space derived is two-dimensional with an affectively rather pleasant subset falling to the left and a rather

Table I Relation of parameters to non-covalent bonding (adapted from (2), see (3))

Interaction	Physical properties correlated with strength
<p><u>Hydrophobic bonding</u> (A bond formed because water molecules tend to associate with one another rather than with non-polar molecules. The energy involved in "squeezing out" water is equivalent to a bond.)</p>	<p>Log P (where $P = \frac{[\text{drug}]_{\text{octanol}}}{[\text{drug}]_{\text{water}}}$)</p>
<p><u>Dispersion bonding</u> London forces, induced dipole-induced dipole bond (Although a molecule may not have a permanent dipole, an instantaneous dipole can result from vibration of electrons relative to the nucleus and ultimately induce a dipole in a neighboring molecule. Energy of interaction falls off at $1/R^6$.)</p>	<p>Molar refractivity x ionization potential or molar refractivity alone, Hildebrand's molar attraction constant, or parachor</p>
<p><u>Electrostatic bonding</u> Interactions between charges: ion-ion, ion-dipole, and dipole-dipole (There is more specificity for dipole-dipole interactions than those involving ions.)</p>	<p>Sigma values or charge from molecular orbital calculations</p>

Hydrogen bonding

(As a very small hydrogen atom is the bridge between two electroegative atoms, the interacting molecules can approach sufficiently close to each other to produce an attraction strong enough to be considered a bond, rather than just another dipole-dipole interaction.)

Same as for electrostatic bonding; only for the atom involved in the hydrogen bond

Charge transfer bonding

(When a molecule with a good electron donor comes in contact with a molecule which is a good electron acceptor, the donor may transfer some of its charge to the acceptor.)

Sigma values or E(L_{amo}) and E(homo) from molecular orbital calculations

Steric repulsion

(Intermolecular forces can be repulsive as well as attractive in nature. Two molecules ultimately reach a minimum distance between them as they approach one another, and this distance is the sum of the van der Waals' radii of the interacting groups. Hence, the van der Waals' radius is considered a measure of the effective size of an atom in noncovalent interactions. Van der Waals' radius is correlated with another measure of steric repulsion, E_s.)

van der Waals' radii or Taft E_s values

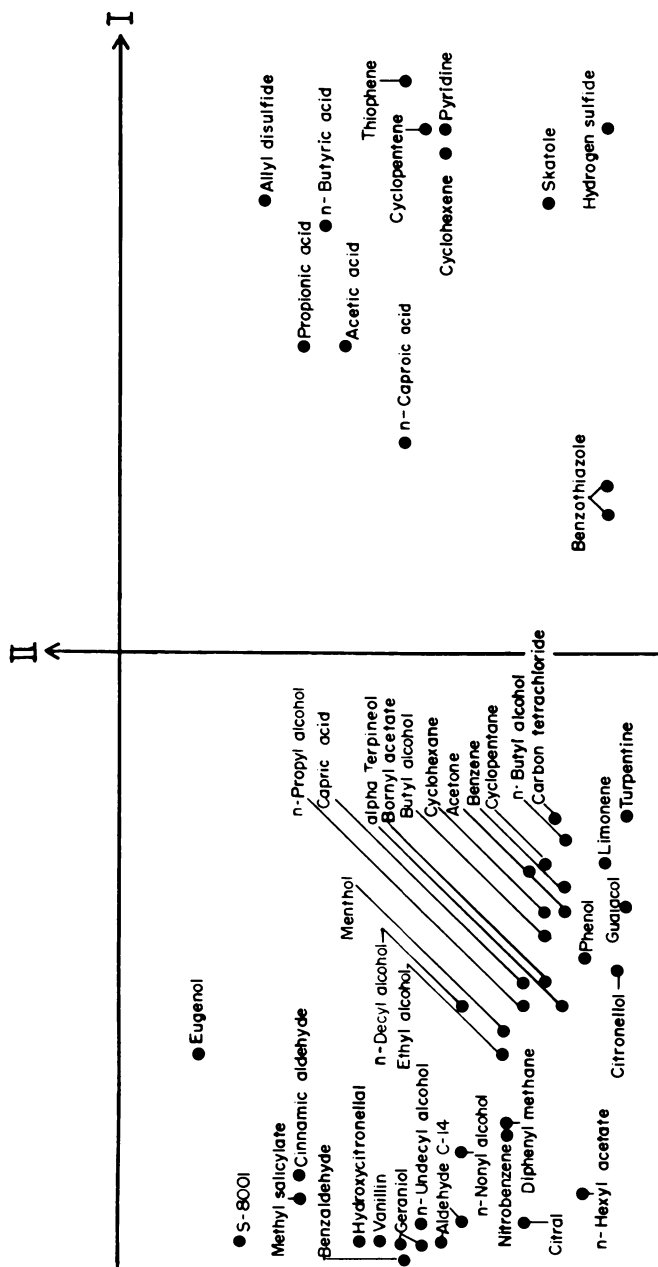


Figure 1. Two-dimensional space for a broad range of odorants. Compounds located close to one another have similar odor quality (from Ref. 4).

unpleasant group to the right. Because no single factor (such as molecular weight, number of double bonds, or dipole moment) could individually account for quality, Schiffman (4) weighted a series of variables shown in Figure 2 and was able to recapture 84% of the interstimulus distances in a two dimensional space.

The mathematical procedure used to maximize the configurational similarity in Figure 1 with a space generated by the weighted physicochemical parameters is based on a least-squares method in which the basic matrix equations are:

$$\begin{aligned} \mathbf{P} &= \hat{\mathbf{P}} + \mathbf{E} \\ \mathbf{P} &= \mathbf{DQ} \\ \mathbf{P} &= \mathbf{DQ} + \mathbf{E} \end{aligned}$$

where: n is the total number of stimuli, and \mathbf{P} is an $n(n-1)/2$ column vector whose elements p_{ij} represent all the interstimulus distances between stimulus i and stimulus j . The proximity measures based on weighted physicochemical parameters are given by $\hat{\mathbf{P}}$, an $n(n-1)/2$ column vector. \mathbf{D} is an $[n(n-1)/2]$ by k scalar distance matrix whose elements $d_{(ij)k}^2$ are the squared differences between stimulus i and j for each k physicochemical parameter. The weights for k physicochemical parameters are represented by \mathbf{Q} , a k element column vector of weights. \mathbf{E} is an $n(n-1)/2$ column vector representing the error between the subjective proximities and the proximities based on physicochemical measures.

The error to be minimized is:

$$\frac{\partial \mathbf{E}'\mathbf{E}}{\partial \mathbf{Q}_k} = 0$$

leading to the least squares solution

$$\mathbf{Q} = (\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}'\mathbf{P}$$

This methodology can be helpful in relating strict quantitative measures of olfactory quality with quantitative physicochemical measures. However, the results for this specific study suggest that the range of molecular structures and olfactory quality are too broad to derive practical information for the rational design of molecules with specific flavor qualities.

Study 2: Pyridyl ketones. A narrower range of compounds, a group of pyridyl ketones, was examined by Southwick and Schiffman (5). Two-dimensional cross-sections through the three-dimensional space achieved by the MDS are shown in Figures 3a and 3b. The first dimension separates the three 2-pyridyl ketones that have popcorn-nutty aromas from the six other compounds that are green-vegetable in character.

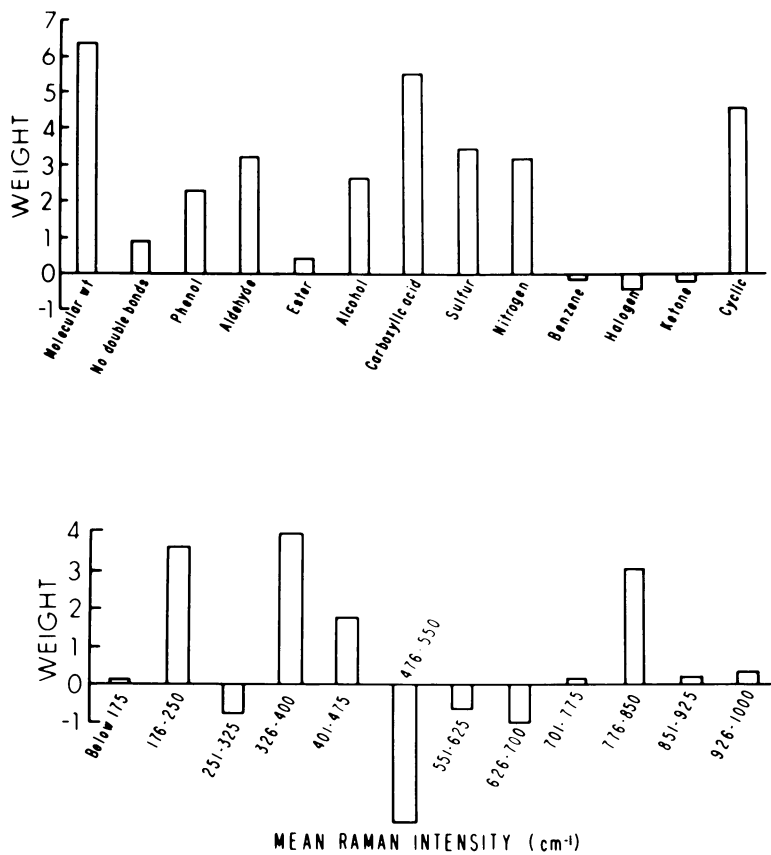


Figure 2. A series of variables and their weights that were used to account for the MDS space in Figure 1.

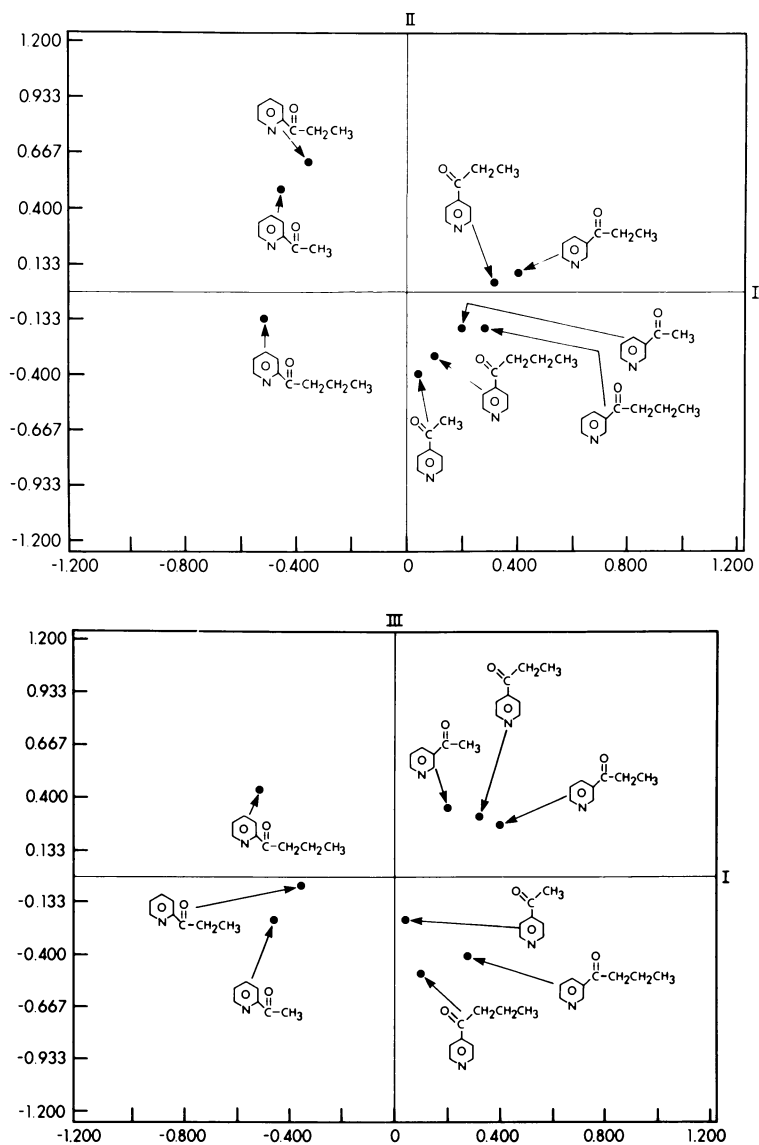


Figure 3. Two-dimensional cross-sections through a three-dimensional space achieved from odor similarity data among pyridyl ketones (from Ref. 5).

Two methods were used to relate a series of physicochemical descriptors to the space in Figure 3, a vector model and an ideal point model (see 1, 6). The vector model assumes that there is a direction through the space that corresponds to increasing amounts of a chemical descriptor. The direction of the vector can be found by linear regression techniques in which a physicochemical variable is regressed over the coordinates of the configuration. Weighted combinations of the configuration coordinates are sought that best explain the variable.

When y_i is the value of the physicochemical parameter for stimulus i , then

$$\hat{y}_i = b_0 + b_1(x_{i1}) + \dots + b_r(x_{ir}) = b_0 + \sum_{a=1}^r b_a (x_{ia})$$

where \hat{y}_i is the value of the variable y_i if the linear relationship were perfect; b_1, b_2, \dots are the regression weights; b_0 is the intercept; x_{ir} is the coordinate for the i th stimulus on the r th dimension. The most common method to solve for the values of b is called least squares regression in which the coefficients are chosen to maximize G where:

$$G = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

The ideal point model is useful when a point in the space can be found that is most like the physicochemical parameter. Thus, the ideal point is the hypothetical stimulus, if it existed, that would contain the maximum amount of the physicochemical attribute. The attribute reaches its maximum at the ideal point and falls off in all directions as the square of the distance from the ideal point. The ideal point is located in an MDS space by a special kind of regression proposed by Carroll (6) that correlates the physicochemical attribute values with the stimulus coordinates and a dummy variable constructed from the sums of squares of the coordinates for each point:

$$\hat{y}_i = b_0 + \sum_{a=1}^r b_a (x_{ia}) + b_{r+1} \left(\sum_{a=1}^r x_{ia}^2 \right)$$

The results of vector and ideal point analyses are given in Tables II and III respectively as analyzed by the program PREFMAP (6, 1). Two kinds of ideal points were found, a positive ideal point for which the relevance of a descriptor decreases with increasing distance of the stimuli from the ideal point and a negative ideal point, for which the relevance increases with increasing distance from the ideal point. The correlations for the vector and ideal point analyses are given in Table IV. It

can be seen that odor quality is highly correlated for both vector and ideal point models with the ratio of the relative intensities of the ions at m/e 106 and 107 obtained in mass spectroscopic fragmentation. Interaction of the ring nitrogen with the substituent in the 2-position is responsible for the fragmentation pathway as well as odor quality. The C_{13} - n.m.r. chemical shifts were also highly correlated with odor in the vector model with the first and second dimensions of the odor quality space. Infrared stretching frequencies, $\log k'$ (a measure of column retention used in high pressure liquid chromatography), and $\log P$ (octanol-water partition coefficient) were only weakly correlated with quality. The high correlations for negative ideal points (unlike positive ideal points) do not provide strong support for a physicochemical parameter.

Table II Direction Cosines of Fitted Physicochemical Vectors

		Dimension		
		1	2	3
$\log P_{\text{oct}}$	1	-0.9966	-0.0101	-0.0815
$\log k'$	2	-0.9294	-0.3662	0.0468
carbonyl shift	3	-0.6541	0.7561	0.0216
carbonyl frequency	4	-0.7974	0.5873	-0.1384
concentration	5	0.6900	-0.7227	0.0399
mass spectrum ratio	6	0.9133	-0.4054	-0.0387

Study 3: Pyrazines. Another method of relating physicochemical parameters to an MDS space, canonical correlation, was employed by Schiffman and Leffingwell (7). In this case pyrazines were arranged in a 3-dimensional space on the basis of odor quality (Figure 4a and 4b). Then a set of descriptors given in Table V were generated by ADAPT, a computer system for automated data analyses by pattern recognition techniques (Stuper and Jurs, 8). The substructures that were used to generate the environment descriptors are given in Figure 5.

Canonical regression (see 1) was used to find linear relationships between the two sets of variables, that is, the stimulus coordinates of the MDS space and the physicochemical attributes. The equations for canonical correlation are:

$$\hat{y}_{ki} = a_{ko} + a_{k1} (y_{i1}) + \dots + a_{kr} (y_{ir})$$

Table III Coordinates of Ideal Points

			Dimension		
			1	2	3
log P _{oct}	1	negative ideal point	-0.11853	0.05568	-0.02633
log k'	2	negative ideal point	-0.11396	0.06849	-0.03036
carbonyl shift	3	negative ideal point	-0.05810	-0.04761	-0.03163
carbonyl frequency	4	positive ideal point	-0.32895	0.18924	-0.06022
concentration	5	positive ideal point	-0.08281	-0.01206	-0.02496
mass spectrum ratio	6	positive ideal point	0.12824	-0.06688	-0.04036

Table IV Correlations

(a)		(b)		(c)
		Ideal Point Model	Vector Model	
log P _{oct}	1	0.9584	0.3933	
log k'	2	0.9489	0.3995	
carbonyl shift	3	0.9725	0.9058	
carbonyl frequency	4	0.4861	0.4730	
concentration	5	0.9499	0.8250	
mass spectrum ratio	6	0.9715	0.9555	

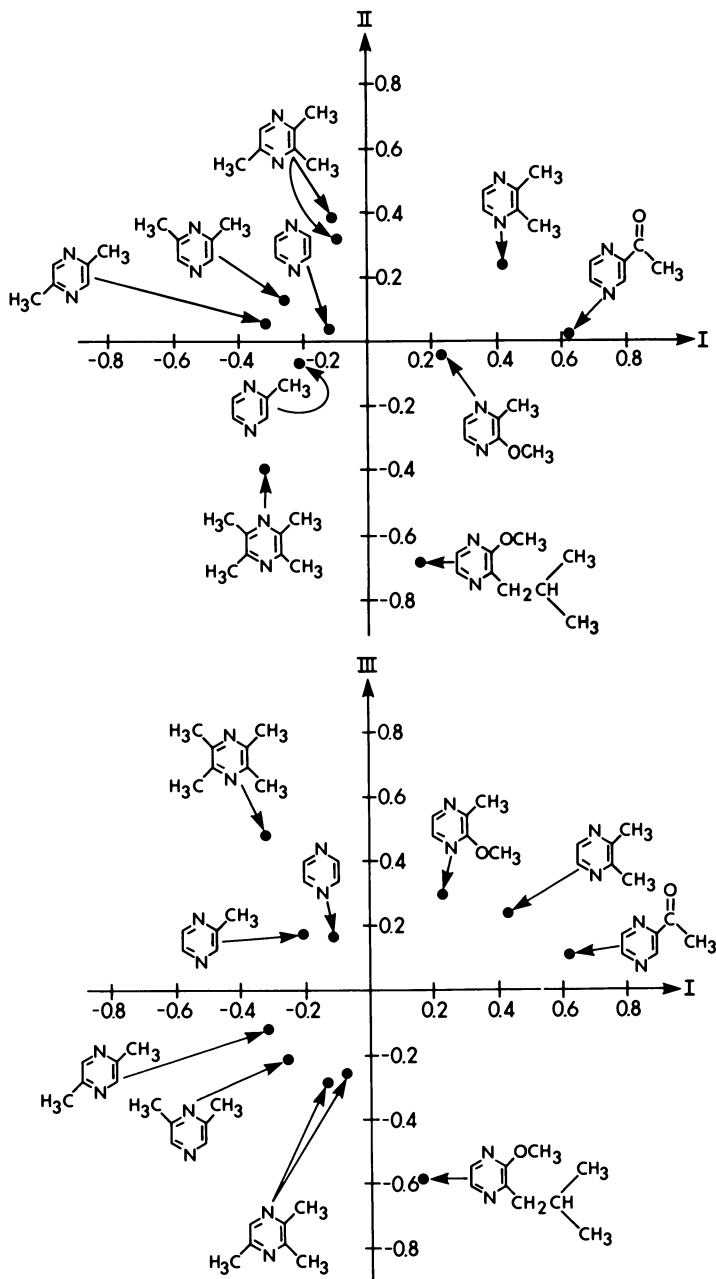


Figure 4. Two-dimensional cross sections through a three-dimensional space for pyrazines (from Ref. 7).

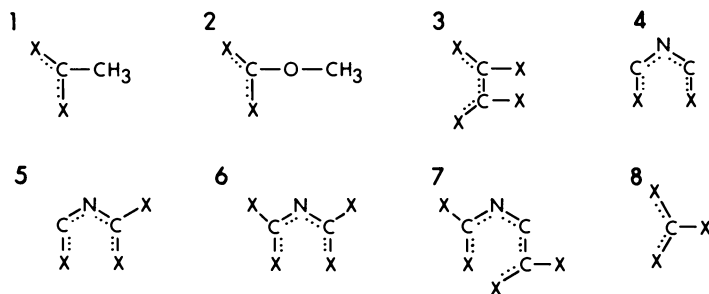


Figure 5. Substructures utilized by ADAPT (8) to generate environment descriptors (from Ref. 7).

Table V Descriptors of the pyrazines generated by ADAPT

1	Number of atoms except hydrogen
2	Number of carbon atoms
3	Number of oxygen atoms
4	Number of bonds
5	Number of single bonds
6	Number of double bonds
7	Molecular weight
8	Path 1 molecular connectivity for all bonds in the structure
9	Path 1 molecular connectivity corrected for rings
10	Path 1 molecular connectivity calculated using the valencies of heteroatoms and corrected for rings
11	Path 2 molecular connectivity
12	Path 3 molecular connectivity
13	Path 4 molecular connectivity
14	Molecular volume
15	Number of substructure 1 (see Figure 5)
16	Environment-substructure 1 (calculates connectivity for substructure 1 and nearest neighbours)
17	Number of substructure 2
18	Environment-substructure 2
19	Number of substructure 3
20	Environment-substructure 3
21	Number of substructure 4
22	Environment-substructure 4
23	Number of substructure 5
24	Environment-substructure 5
25	Number of substructure 6
26	Environment-substructure 6
27	Number of substructure 7
28	Environment-substructure 7
29	Number of substructure 8
30	Environment-substructure 8

and

$$\hat{x}_{ki} = b_{ko} + b_{k1} (x_{i1}) + \dots + b_{kr} (x_{ir})$$

where x_{i1} , x_{i2} etc. are the values of stimulus i on dimensions 1 and 2 of the MDS space just as in multiple regression equations and y_{i1} , y_{i2} are ratings of stimulus i on several physicochemical parameters. The intercepts and weights are solved to maximize the correlation between \hat{y}_{ki} and \hat{x}_{ki} .

Canonical correlation analysis was used to relate small subsets of physicochemical parameters to the MDS space. Small subsets were necessary because in canonical correlation analysis, the number of stimuli should be greater than the number of dimensions and physicochemical parameters combined. The analysis revealed that a linear combination of two ADAPT parameters in Table 5 (number of oxygen atoms and chemical environment of substructure (7)) in addition to a concentration variable accounted for 63% of the arrangement of the pyrazine odor space.

Study 4: A broader range of chemicals fitted by a new technique.

A final means of relating physicochemical parameters to MDS spaces is a new technique called PLS (partial least squares, (8)). In this example, the relation between 19 chemicals and 23 physicochemical parameters was examined (9). PLS, unlike canonical correlation, permits use of more chemical parameters than stimuli. The twenty-three physicochemical variables included molecular weight, functional groups, Raman frequencies and Laffort parameters (see (1)). The Laffort parameters are alpha (an apolar factor proportional to molvolume), rho (a proton receptor factor), epsilon (an electron factor) and pi (a proton donor factor).

PLS is best described in matrix notation where the matrix X represents the calibration matrix (the training set, here physicochemical parameters) and Y represents the test matrix (the validation set, here the coordinates of the odor stimulus space). If there are n stimuli, p physicochemical parameters, and m dimensions of the stimulus space, the equations in Figure 6a apply. The C matrix is an $m \times p$ coefficient matrix to be determined and the residuals not explained by the model are contained in E_p . The X matrix is decomposed as shown in Figure 6b into two small matrices, an $n \times a$ matrix T and an $a \times p$ matrix B where $a \ll n$ and $a \ll p$. F is the error matrix. The computation of T is such that it both models X and correlates with Y and is accomplished with a weight matrix W and a set of latent variables U for Y with a corresponding loading matrix B_y .

The multidimensional space derived by application of the multidimensional scaling to the experimental psychophysical similarity measures is given in Figure 7. The reconstruction of this space based on application of the PLS algorithm is given in Figure 8. The PLS procedure weighted 23 physicochemical parameters resulting in the plot in Figure 8 in which the latent

American Chemical
Society Library
1155 16th St., N.W.
Washington, D.C. 20036

$$Y = XC' + E_p$$

$$\begin{matrix} n \\ \left\{ \begin{matrix} | \\ | \\ | \\ | \end{matrix} \right\} \\ Y \end{matrix} \quad \begin{matrix} p \\ \left[\begin{matrix} | \\ | \\ | \end{matrix} \right] \\ X \end{matrix} \quad \begin{matrix} m \\ \left\{ \begin{matrix} | \\ | \\ | \\ | \end{matrix} \right\} \\ C \end{matrix}$$

$$X = TB + F$$

$$\begin{matrix} n \\ \left[\begin{matrix} | \\ | \\ | \end{matrix} \right] \\ X \end{matrix} = \begin{matrix} n \\ \left\{ \begin{matrix} | \\ | \\ | \end{matrix} \right\} \\ T \end{matrix} \begin{matrix} m \\ \left[\begin{matrix} | \\ | \\ | \end{matrix} \right] \\ B \end{matrix} + \begin{matrix} n \\ \left[\begin{matrix} | \\ | \\ | \end{matrix} \right] \\ F \end{matrix}$$

Figure 6. Left, the physicochemical data for the n stimuli are contained in X . The coordinates in the multidimensional space are in Y . C , the matrix of calibration coefficients, is determined by statistical analysis; right, X is decomposed into T and B plus F , the error matrix.

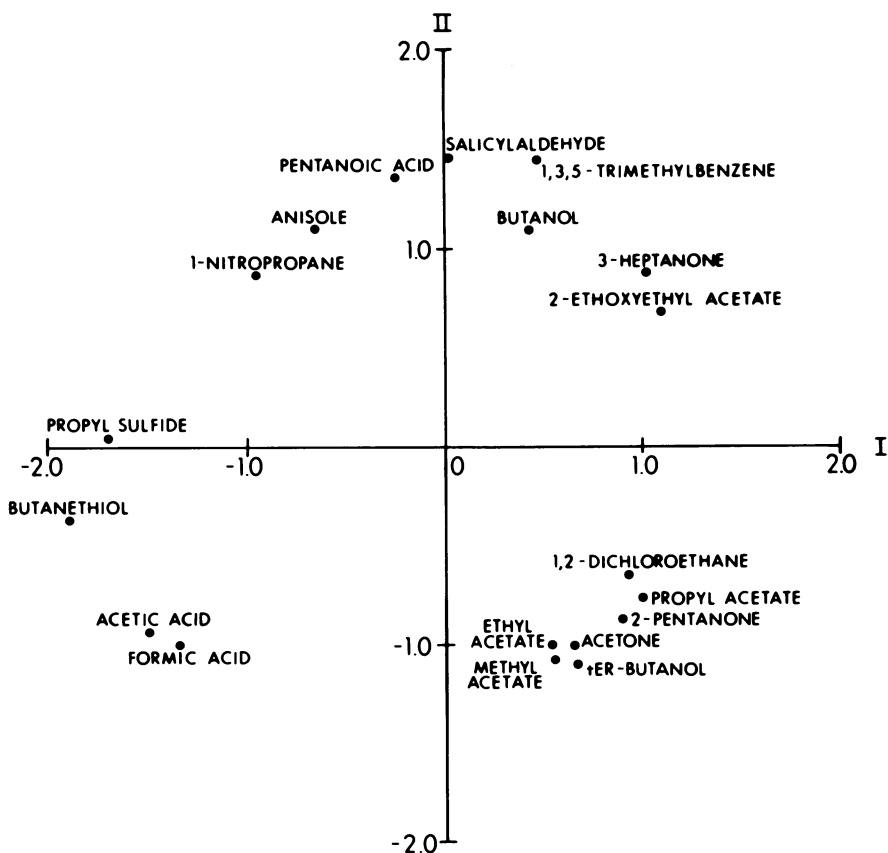


Figure 7. Two-dimensional space representing the similarity among a group of odorants.

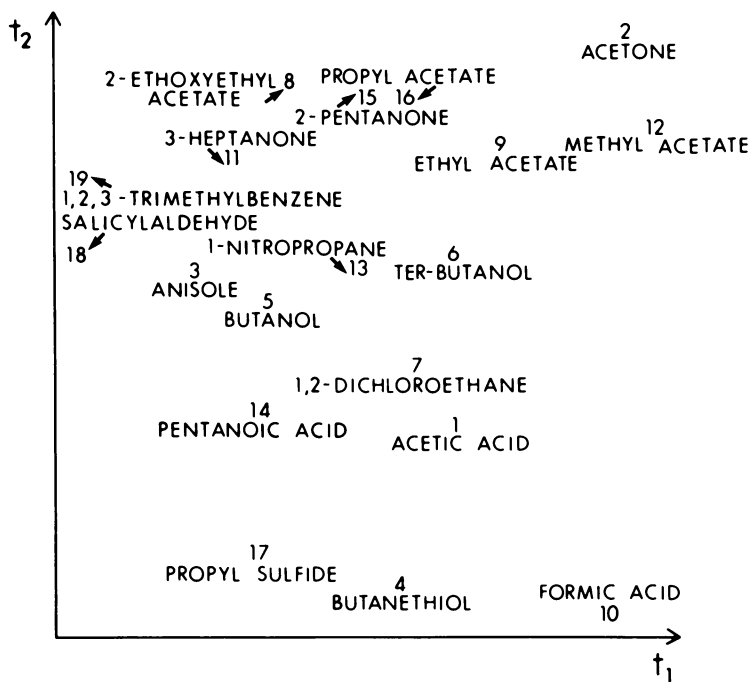


Figure 8. Reconstruction of the space in Figure 7 using PLS (see Refs. 9 and 10).

X variables (t 's) are plotted against one another. It can be seen that the relative arrangements in Figures 7 and 8 are similar (correlation = .70), although the PLS procedure rotated the space.

Summary

These examples are merely modest beginnings at relating geometrical maps that represent similarities in odor quality with physicochemical parameters. It is now necessary to expand these mathematical techniques to more data sets with extensive use of the non-covalent bonding parameters shown in Table 1. Ultimately reliable design of flavor molecules by computer will be feasible as we explore a wider range of molecular types, especially isomers, with the techniques described in this paper.

Literature Cited

1. Schiffman, S. S.; Reynolds, M. L.; Young, F. W. "Introduction to multidimensional scaling"; Academic: New York, 1981.
2. Martin, Y. C. in "Quantitative drug design"; Marcel Dekker: New York, 1978; pp. 33-58.
3. Schiffman, S. S. Chem. Indust. 1983, 3, 39-42.
4. Schiffman, S. S. Science 1974, 185, 112-117.
5. Southwick, E.; Schiffman, S. S. Chemical Senses 1980, 5, 343-357.
6. Carroll, J. D. in "Multidimensional scaling: Theory and applications in the behavioral sciences"; Shepard, R. N.; Romney, A. K.; Nerlove, S. B., Eds.; Seminar Press: New York, 1972.
7. Schiffman, S. S.; Leffingwell, J. C. Pharm. Biochem. & Behavior 1981, 14, 787-798.
8. Stuper, A. J.; Jurs, P. C. J. Chem. Inf. Comp. Sci. 1976, 16, 99-105.
9. Lindberg, W.; Persson, J. A.; Wold, S. Anal. Chem. 1983, 55, 643-648.
10. Schiffman, S. S.; Wold, S.; Sarle, W.; Dunne, W. in "Multidimensional scaling: Theory and applications"; Young, F. W.; Hamer, R., Eds.; in press.

RECEIVED April 27, 1984

Use of Microcomputers for Product Optimization

H. R. MOSKOWITZ and B. JACOBS

Moskowitz/Jacobs, Inc., 1075 Central Park Ave., Scarsdale, NY 10583

Recent developments in microcomputers, sensory analysis and experimental design have made it possible to efficiently evaluate and optimize products. This paper focuses on the conduct and analysis of a study to optimize the flavor constituents of an alcoholic "digestive" liqueur. It illustrates the use of the panel data, and the contributions of the microcomputer as both a tool for gathering data, and as an inexpensive replacement for mainframe computers in statistical computations.

Over the past decade product developers have begun to use methods of experimental design to optimize products. By setting up a grid of products, each systematically varied on a limited set of formula alternatives, the product developer can generate a roadmap, relating formulation/processing inputs, and consumer reactions. The roadmap, embodied as a series of equations relating formulations and ratings, allows the product developer to quickly ascertain the most acceptable product. Furthermore, with the model, or set of equations, it becomes possible to estimate the likely sensory and acceptance profiles that panelists would assign to either the optimum product, or to a product with any pre-specified formulation.

This paper deals with the technique of product modelling, using a case history. The case history illustrates the technique of experimental design, the method of product testing, and the use of the microcomputer for data acquisition, statistical analysis, and modelling/optimization. As such, the paper provides both methodology and substantive data.

0097-6156/84/0261-0051\$06.00/0

© 1984 American Chemical Society

The Product Optimization Problem. Alcoholic liqueurs, specifically digestives, comprise a variety of formula variables which generate mixed sensory signals. On the one hand the components often taste bitter, signalling efficacy. However, the digestive drink must engender a sufficient perception of efficacy (unpleasant taste) with a moderate degree of acceptability. Consumers will not drink a digestive product which tastes "too good," nor will they drink one that tastes "too bad."

In order to optimize the product it becomes important to select that particular set of formula variations which generate the appropriate sensory profile. The technique of experimental design, consumer evaluation, and product modelling play a key role in discovering the proper formulation balance among ingredients to generate the necessary sensory profile.

Stimuli. The stimuli comprised an alcohol base which varied in the level of alcohol. In addition, two other variables systematically changed, Absinthe and Cassia. These two formula variables, in concert with the alcohol base sufficed to generate products having a wide variation in sensory profile, and thus variation in both innate acceptance and communication of "efficacy."

In order to efficiently evaluate the impact of 3 variables, which each can assume many different levels, the researcher can use the technique of experimental design (1). Experimental design provides an efficient, limited array of formula variations which substitute for the many hundreds or thousands of prototypes that the product developer could produce and test. For 3 variables, and recognizing that the variables would interact with each other, a fractional factorial design recommended itself (2). This test design explores pair-wise interactions among formula variables, and permits non-linear response functions. A non-linear response function implies that the relation may follow a parabola, rather than a straight line, with the optimum in the middle of the range, rather than at the upper or lower end.

For a 3 variable case, the product developer created 15 different products, arrayed as shown in Table 1.

Attribute Ratings. Panelists perceive a variety of characteristics in products. When the product developer systematically varies the formulation over a systematically wide range, the panelist may perceive an entire spectrum of different characteristics. One can trace these characteristics back to the specified formulation levels, but only if the panelist has had a chance to evaluate the characteristics by means of scaling. In this study the panelists rated a variety of different attributes after tasting each digestive sample.

TABLE 1EXPERIMENTAL DESIGN - LIQUEUR STUDY

<u>SAMPLE</u>	<u>ALCOHOL LEVEL*</u>	<u>ABSINTHE-TYPE LEVEL*</u>	<u>CASSIA LEVEL*</u>
1	H	H	H
2	H	H	L
3	H	M	M
4	H	L	H
5	H	L	L
6	L	H	M
7	L	M	H
8	L	M	M
9	L	M	L
10	L	L	M
11	M	H	H
12	M	H	L
13	M	M	M
14	M	L	H
15	M	L	L
H	34	10	35
M	29	5	30
L	25	0	20

*Units = Relative Level

The attributes appear in Table 2. Note that they comprise 3 different categories:

Pure sensory attributes, representing degree of perceived intensity.

Liking -- or evaluation of the sample.

Image -- (e.g., appropriateness for an end use).

The panelists used the magnitude estimation scaling procedure (3,4). Magnitude estimation generates ratio scale data, in which a rating of 30 represents twice the perceived intensity as a rating of 15, and half the intensity as a rating of 60. Magnitude estimation has become extremely popular in the scientific community because it allows panelists to use their own frames of reference, without feeling constrained to limit their ratings to a narrow fixed point scale (e.g., 1 - 9).

The panel comprised 26 consumers of digestives familiar with the product category.

Table II. Attributes Evaluated by the Panel - Digestive Liqueur

Sensory Attributes

Overall Flavor Intensity

Aroma Intensity

Perceived Alcohol Taste

Image Attributes

Body

Clean Taste

Unique Product

Efficacious

Acceptance Attributes

Overall Liking

Purchase Interest

Testing Protocol. Owing to the alcoholic nature of the stimulus the testing required 3 days. On each of the days the panelists evaluated 5 samples, over a 2 hour period. They tasted and swallowed the samples in order to obtain a clear perception of the product, from appearance to aroma to taste to consistency in the mouth, and to feeling as they swallowed the product. The extended sessions allowed the panelists to evaluate products in a leisurely manner.

Data Analysis. Part of the benefits of using a micro-computer stem from its ability to act both as a data acquisition device and as a data analysis device. In this study the panelists rated their perceptions by using a pre-coded mark-sense card, shown in Figure 1. The card has 3 positions, allowing the panelists to assign ratings from 0 to 999. (In practice, most respondents never go above 999 in their ratings.) The card also has a position for the "top of scale" rating. The mark-sense card has dedicated positions on the card. By darkening in the appropriate "bubbles" just after the rating, the panelist effectively "keypunches" his/her own data. An inexpensive cardreader (Chatsworth OMR 500) programmed in BASIC and machine language reads the filled out cards, checks that the card contains valid panelist ID, valid product ratings, etc., and then passes the information through the computer, and onto disk storage. In this fashion, the micro-computer (an Apple II Plus) simplifies the data acquisition, making it faster and more economical than the more conventional key punching procedure.

After the machine has read each card, it pre-processes the data by dividing all ratings on the card by the "top of scale" rating. This generates a decimal between 0 and 1. We then post multiply the decimal by 100 to eliminate fractions, and store the data on disk.

Results

Do Panelists Perceive Differences Between Product Prototypes. Prior to building a model of the product via regression equations, the product developer should test the degree to which the panelists perceive differences among the prototypes. If they do not, then the modelling process actually fits "noise" or random differences among the ratings, rather than "signal" or true differences.

Table 3 presents the database for this project, along with the average standard errors for each attribute. Note that the panelists can and do differentiate between the different prototypes, as shown by the differences between the means of two prototypes vs. the standard error of the mean. A quick rule of thumb for differences consists of taking the difference between two samples and dividing that difference by twice the standard error. If the two samples have different standard errors, then divide by the higher standard error of the two.

Building Product Models. The next step in product optimization deals with model building. A model summarizes the relations between formula variables in a succinct, quantitative way. With the use of a computer and packaged software, researchers can easily use the statistical techniques of regression

ID	TOS	CD	PRODUCT	A	B	C	D	E	F	G	H	I	J	K	L	M	N
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01	01
02	02	02	02	02	02	02	02	02	02	02	02	02	02	02	02	02	02
03	03	03	03	03	03	03	03	03	03	03	03	03	03	03	03	03	03
04	04	04	04	04	04	04	04	04	04	04	04	04	04	04	04	04	04
05	05	05	05	05	05	05	05	05	05	05	05	05	05	05	05	05	05
06	06	06	06	06	06	06	06	06	06	06	06	06	06	06	06	06	06
07	07	07	07	07	07	07	07	07	07	07	07	07	07	07	07	07	07
08	08	08	08	08	08	08	08	08	08	08	08	08	08	08	08	08	08
09	09	09	09	09	09	09	09	09	09	09	09	09	09	09	09	09	09

Figure 1. Example of a mark-sense card for use with magnitude estimation scaling. The first field (PAN) corresponds to the panelist's 3 digit ID number. The second field (PROD) corresponds to the 3 digit stimulus number. The third field (TOS) corresponds to the panelist's top of scale. The remaining 3 digit fields allow panelists to use numbers from 0 to 999. To denote negative numbers, the panelists write 0 and darken in the appropriate 1-3 digit number, and darken in the topmost row (xxx). This type of mark-sense card allows panelists to use their own range of numbers (up to an arbitrary top limit of 999). During the data input to the micro-computer, we divide each rating by the panelist's own top of scale (TOS), and post-multiply by 100.

TABLE 3

DATA BASE OF RATINGS**

Product	Alcohol*	Absinthe*	Cassia*	Liking	Purchase	Flavor Intensity	Aroma Intensity	Perceived Alcohol	Body	Clean	Efficacious	Unique
1	34	10	35	71	57	71	62	71	58	83	83	79
2	34	10	20	58	56	58	63	80	51	87	83	70
3	34	5	30	54	69	54	49	75	49	87	83	85
4	34	0	35	47	64	47	47	58	46	77	84	77
5	34	0	20	71	71	71	65	57	43	81	76	72
6	29	10	30	60	64	60	66	54	43	74	79	76
7	29	5	35	52	53	52	64	57	42	76	63	63
8	29	5	30	56	69	56	66	49	45	76	81	73
9	29	5	20	47	62	47	49	56	42	73	68	60
10	29	0	30	52	83	52	66	51	53	81	87	83
11	25	10	35	54	53	54	54	45	44	72	62	57
12	25	10	20	56	80	56	44	41	43	71	70	71
13	25	5	30	40	57	40	45	42	41	68	68	60
14	25	0	35	39	65	39	48	38	41	77	66	66
15	25	0	20	41	59	41	44	39	38	66	56	56
Average				3.8	3.9	2.7	3.4	2.8	3.3	3.9	4.0	3.5
Standard Error												

* = Relative Levels

** = Scale

Liking: -100 = Extreme Disliking; 0 = Neutral; 100 = Extreme Liking
Sensory/Image: 0 = Not At All; 100 = Extreme

analysis to build equations relating variables. The micro-computer has available a wide variety of "canned statistical programs" which the researcher can use for regression analysis. The analysis by regression used the ELF (Econometric Linear Forecasting) package, an integrated set of programs designed for statistical analysis on the Apple II, and written in BASIC. One can purchase software for these analyses quite readily from any of a number of commercial suppliers of micro-computer software.

In order to develop the proper equations relating independent and dependent variables, the product developer and researcher must keep certain known facts in mind:

Generally, formula variables interact with each other. The regression equation used to interrelate formula variables and panel ratings must take these interactions into account. Otherwise, the equations will not validly model the ratings.

Liking often follows a parabolic curve, when the researcher independently varies one formula variable over a wide range (Figure 2). Liking goes up with formula level, peaks and then drops down. The regression equation must contain square terms (quadratic terms), to capture this relation. Just because the researcher allows quadratic terms in the equation does not, however, mean that these terms will appear in the final equation. To do so means that the terms achieve statistical significance. They may or may not. Nonetheless, in the modelling process, the researcher must allow the equation to possess quadratic terms.

Table 4 presents the product models for the different attributes. Note that in the models only the statistically significant terms appear (for a predictor or term in the equation to reach statistical significance). Although one could, in theory, have used a total of 9 predictor variables (3 linear, 3 square, 3 cross terms), most of the product models necessitated fewer terms.

Using the Product Models to Describe the Surface and to Optimize Attributes. The product models represent a short-hand description of the relation between formula variables and consumer attribute ratings. As such, the researcher can use the model instead of the raw data, to estimate the likely attribute rating for any combination of formula inputs. This means that the researcher can systematically explore the different combinations of formula ingredients, and either find the optimum level (e.g., for an attribute such as purchase), or estimate the likely attribute ratings for any formulation.

Optimizing Purchase Interest. In this study interest focused on the specific set of formula variables which would generate the most acceptable product. Acceptability for the case of a "digestive liqueur" primarily consisted of high purchase

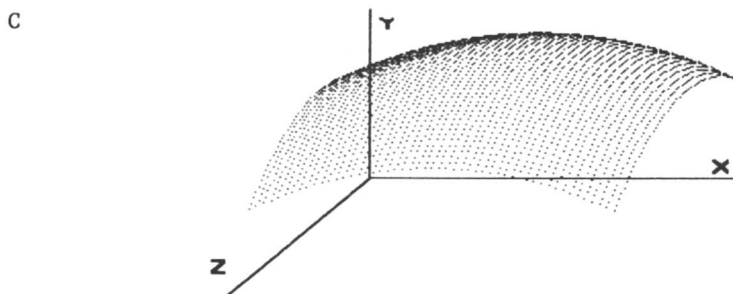
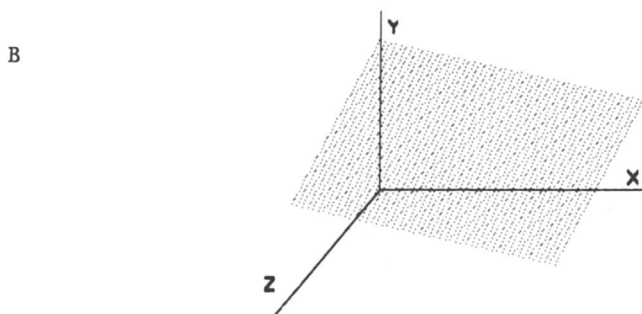
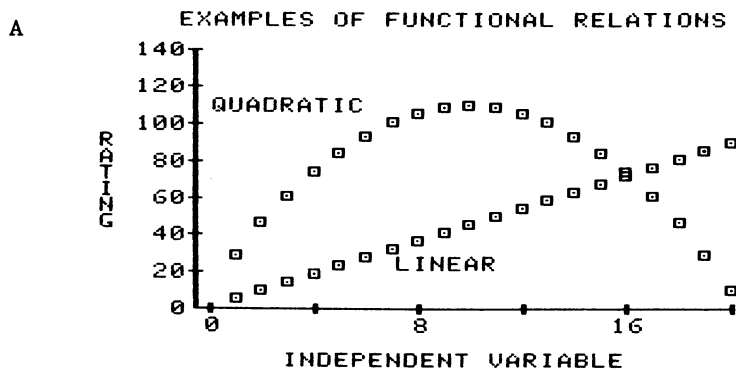


Figure 2. Comparison of linear vs. quadratic (parabolic) function, and planar vs. quadratic surface functions.

- A - Linear vs. quadratic functions
- B - Plane
- C - Quadratic surface

TABLE 4
PRODUCT MODELS INTERRELATING FORMULA VARIABLES AND RATINGS*

Attribute =	Liking	Flavor Intensity	Aroma Intensity	Perceived Alcohol	Body	Clean	Efficacious	Unique
Constant +	293.0	-18.6	-617.2	7.0	65.3	37.9	30.2	115.1
K1 (Alcohol) +	-15.6	1.3	42.4	---	-1.5	---	---	-4.6
K2 (Absinthe) +	-7.8	-6.4	-4.7	---	-4.4	-2.6	-7.2	-6.8
K3 (Cassia) +	---	3.7	3.2	---	---	1.8	3.7	2.3
K4 (Alcohol) ² +	0.2	---	-0.6	0.1	---	0.04	-0.05	---
K5 (Absinthe) ² +	---	0.3	0.3	-0.2	0.1	---	0.3	0.3
K6 (Cassia) ² +	-0.1	-0.1	---	0.03	-0.03	---	-0.1	-0.1
K7 (Alcohol) (Absinthe) +	0.2	---	---	0.1	0.1	0.1	0.1	0.1
K8 (Alcohol) (Cassia) +	0.1	---	-0.1	-0.05	0.1	-0.1	0.1	0.2
K9 (Absinthe) (Cassia)	0.1	0.2	0.1	-0.03	---	---	---	---
Multiple R ²	0.94	0.82	0.83	0.95	0.74	0.82	0.92	0.87
Standard Error/Regression	1.7	5.5	5.3	3.8	3.6	3.3	3.8	4.7

*Rounded to nearest tenth.

NOTE: The models or equations relate formula variables to attributes. Only the statistically significant coefficients appear in the equations. (Viz. absolute value of "T" for the coefficient exceeds 1.) The models show coefficients to the nearest tenth.

interest. In this product category consumers often do not purchase the product which they perceive to have the best taste, because to some extent the product must have an unpleasant bitter taste to signal "efficacy" as a digestive.

One can consider product optimization as a type of "hill climbing." The product models in Table 4 each represent a hill or a surface, with one hill or surface for each attribute. In order to optimize purchase intent, consider the sequence of steps shown in Table 5. Note that the computer attempts to find the highest point on the surface to maximize purchase intent, subject to the constraint that the formula ingredients (alcohol, absinthe, cassia) all must lie within the tested limits, rather than exceeding those limits. The constraints insure that in the study the optimum product represents a value for which there exist actual observations. (It does little or no good to project an optimally acceptable product which has a formulation lying very far away from any existing products. In that case the extrapolation may lose its validity.)

In Table 5 note the different steps or iterations, representing successive attempts to improve the purchase intent score. Each successive attempt represents a new formulation which achieves a higher purchase intent rating than previously. Note the rapid improvements in purchase intent initially, for the first set of iterations. However, as we reach the top of the hill the increments in purchase intent for the product become smaller and smaller. This occurs because at the top of the hill we find less room for improvement.

The results from Table 5 illustrate the most acceptable product, from the point of view of purchase intent. Note that the product models in Table 4 also allow the product developer to estimate the likely sensory attribute profile of the optimum product, which also appears in Table 5. Thus, it becomes possible to rapidly achieve an optimum product, whilst at the same time predict its profile on sensory and image characteristics.

Discussion

The foregoing study illustrates some of the utilitarian benefits of micro-computers for product modelling and product optimization. The computer provides the following specific benefits to the practical analysis of foods (and other consumer products).

1. It allows the product developer and researcher to statistically analyze relatively large amounts of data, rapidly and inexpensively.
2. It permits the development of equations to quantitatively relate formula variables to consumer sensory and acceptance responses.

TABLE 5

EXAMPLE OF HILL CLIMBING OPTIMIZATION TO MAXIMIZE PURCHASE

Best Formulation (Viz Highest Purchase) At Each Iteration

Iteration:	<u>1</u>	<u>6</u>	<u>11</u>	<u>16</u>	<u>21</u>	<u>26</u>
<u>Formulation</u>						
Alcohol	27.8	27.8	28.3	26.7	25.1	25.1
Absinthe	0.45	0.45	0.05	0.00	0.00	0.00
Cassia	25.8	25.8	25.3	26.0	21.3	22.0
<u>Expected Ratings*</u>						
Purchase	87.5	87.5	88.2	90.7	97.7	97.8
1. Liking	43	43	43	46	52	52
2. Sensory	55	55	57	55	54	54
3. Aroma	61	61	63	57	46	47
4. Alcohol Taste	43	43	43	39	35	35
5. Body	46	46	46	46	46	46
6. Clean	75	75	75	74	71	71
7. Efficacious	85	85	86	85	83	83
8. Unique	81	81	82	82	81	82

*Using equation in Table 4 to predict attribute ratings from formula variables.

NOTE:

The search algorithm required 26 iterations or trials to reach the optimum formula. Below the formulation, there appears the profile of expected attribute ratings corresponding to that formulation.

3. It allows the researcher to use those equations in a pro-active manner, to answer "what-if" questions (i.e., what formulation level corresponds to the most acceptable product in the minds of the consumers?).
4. It allows the researcher to place constraints on the model, and optimize the product, subject to constraints. These constraints can represent limitations on the formula inputs, or limitations on the sensory responses, or both.

We should keep in mind that over the past decades researchers have not always had computers available. Yet, for the most part the researchers and product developers have created acceptable foods. One does not need the computer to produce an acceptable product. However, the computer acts as a rapid calculating device. Properly programmed, and using mathematical models which describe the relations among formula variables, the computer can act as an exploratory device. It can search out combinations of ingredients which generate desired responses. We could, as individual researchers and product developers, come up with the same answers. More than likely, however, we would:

Not use experimental design.
Not develop the models.
Not explore the combinations of ingredients in a systematic manner.
Rely on our intuition and concurrent learning to optimize the product.

The computer provides a more efficient tool to accomplish the objectives of product modelling and product optimization. Whereas traditionally, a product developer might have spent 6 months to a year doing the experiments to learn about the response characteristics of a product, the use of experimental design, product modelling and computer evaluation/optimization cuts down that time to weeks or months. This leads to pragmatic benefits, in addition to providing insights about the products.

Literature Cited

1. Box, G.E.P.; Hunter, W.G.; Hunter, J.S.; "Statistics For Experimenters." John Wiley, New York, 1978.
2. Mullen, K.; Ennis, D. *Food Technology*, 1979, **33**, 74.
3. Stevens, S.S. "Psychophysics: An Introduction To Its Perceptual, Neural and Social Prospects." John Wiley, New York, 1975.
4. Moskowitz, H.R. *Journal of Food Quality*, 1977, **1**, 195.

RECEIVED April 23, 1984

Economics of Laboratory Information Management Systems

JOSEPH H. GOLDEN

Laboratory Management Systems, Inc., 301 East 21st Street, New York, NY 10010

Laboratory Information Management Systems (LIMS) have become a widely recognized tool for increasing the productivity and quality of service of the analytical laboratory. Laboratory managers are increasingly being faced with the problem of determining the benefits that LIMS could offer their organization, and of relating those benefits to economic measures which can justify a system's purchase or development. This paper presents an overview of the economics of LIMS. It presents a rationale for identifying sources of economic value to be derived from LIMS, and for estimating their worth. It presents the various factors that contribute to the actual cost of a system and finally, it suggests financial analysis techniques which can be employed to justify system acquisition.

The Laboratory Information Management System (LIMS) has achieved wide recognition as a powerful tool for increasing the productivity and quality of service of the analytical laboratory. Systems have been developed that range from inexpensive microcomputer based systems to half-million dollar or more large, minicomputer based systems. In addition, many firms have already developed or acquired custom systems tailored to their specific needs(1-8).

LIMS Functions

In general, LIMS can perform a basic set of functions which greatly facilitate the operation of analytical laboratories: They can provide for work scheduling, for status checking and sample tracking, for automated entry and processing of analytical test data, for automated report generation, for laboratory data quality assurance, and for data archiving. In addition, they can provide management level reports of work backlog, turnaround time, laboratory productivity, and frequently also provide billing and other administrative information whose compilation would impose a clerical

0097-6156/84/0261-0065\$06.00/0
© 1984 American Chemical Society

cal burden on the laboratory staff. Table I presents examples of the functions that may be encompassed by LIMS. It is unlikely, however, that all of these functions would be incorporated in a single system.

TABLE I. Functions Encompassed by LIMS Technology

Analytical Support Functions

- Instrument interface for data interchange.
- Computational support.
- Report generation.
- Archival and retrieval of data.
- Storage and retrieval of methods and specifications.
- Comparison of measurements to specifications.

Work and Resource Management

- Sample log-in.
- Label generation.
- Assignment and scheduling of work.
- Preparation of work lists.
- Tracking and status of samples.
- Reporting of backlogs.
- Reagent inventory and preparation procedures.

Laboratory Quality Assurance Support

- Generation of audit trails.
- Tracking of blind and round robin samples.
- Automatic checking of tolerances and limits.
- Scheduling and tracking of instrument calibrations.
- Reporting of variances.

Management Support

- Analysis of lab productivity.
- Analysis of turn-around-time.
- Computation of cost-per-analysis.
- Analysis of equipment utilization.

Business Support

- Entry and reporting of labor charges.
 - Charging and billing for customer accounts.
 - Processing and entry of analytical data for finished products.
 - Quality control data for vendor qualifications.
 - Quality control data for incoming raw materials.
 - Data for regulatory agencies (EPA, OSHA, FDA, TOSCA).
-

The technology for implementing LIMS is now well established. The major problem now confronting those responsible for bringing LIMS technology into their laboratories is the need to economically justify the acquisition or development of the desired system. This translates into a need for quantitative measures of the economic impact that LIMS will have on the organization as a whole. This economic impact is defined as the cost and value of laboratory operations and their relationship to corporate revenue. Such a view is required to establish reasonable objectives and budgetary guidelines as well as to estimate the cost-to-value relationship for any specific LIMS proposal.

LIMS in different laboratory environments

Within the flavor and fragrance industry two types of laboratories predominate: the Quality Assurance/Quality Control (QA/QC) laboratory, and the Research and Development (R&D) laboratory.

The QA/QC lab is responsible for the testing of feedstocks, raw materials, process intermediates, and finished goods. In addition, the lab may be responsible for the development of standards for materials, processes, and procedures. Maintenance of the quality (purity, uniformity, concentration, adherence to standards, etc.) of the finished goods, however, is the primary goal. The QA/QC lab is usually characterized by the routine, repetitive nature of its workload. Testing is primarily to specification and, where lot acceptance or rejection is involved, is often on a grade category or pass/fail basis. Data may be archived in order to comply with regulatory directives or for later analyses of trends in material quality or process performance. The QA/QC lab, then, would benefit from LIMS technology which would mechanize the collection and analysis of appropriate data from routine tests.

The R&D lab, on the other hand, is responsible for the development of new products and processes, the improvement of existing ones, and occasionally, the analysis of competing products. Testing is more frequently done by professional, analytical chemists rather than by technicians. The work is often non-routine and method development may be a significant part of the lab's tasks. This results in higher average labor costs than the QA/QC lab. The R&D lab in the flavor and fragrance industry may also have responsibility for sensory testing.

The R&D lab, then, would benefit more from a LIMS system with high flexibility, the ability to structure large empirical data bases, and the ability to support statistical analyses and graphical presentations of the data.

The Common Laboratory Management Problem

Regardless of the laboratory's mission, managers are confronted with a common set of problems: Increases in data volume from increased use of smart instruments and from increased testing and record retention requirements imposed by EPA, FDA, OSHA, and other regulatory agencies; constantly rising operating and material

costs; and ever tightening constraints on staff and material expenditures. These are manifested by increasingly burdensome paperwork, inefficient utilization of resources, and exasperating searches for misplaced samples and data.

Economic Considerations

Since the QA/QC laboratory is concerned with the quality of the firm's products, it can influence product costs and revenues (the cash flow associated with the products). It determines acceptance or rejection of raw materials and feedstocks and assesses their market value for the purchasing department. It frequently initiates the processing of claims against vendors providing raw materials which are below grade specification but are nevertheless used. The lab also may be responsible for process monitoring to determine process parameters which minimize the production of scrap or off-grade product.

The economic impact of LIMS technology in the QA/QC lab is primarily produced by its ability to speed the delivery of dependable information to those responsible for making immediate decisions regarding the purchase, production, and sale of product. Since the QA/QC lab budget may be relatively small compared with product production costs the improvement in laboratory staff productivity which LIMS offers, may be secondary in importance to its more direct impact on product costs and revenues.

The R&D lab has a different function. It contributes to the long term profitability of the firm by developing and perfecting products and processes. While controlling the costs of R&D as a whole is important, the speed at which a specific analytical test can be completed may be more important because it relates to the speed and success with which a project can be completed. This relates to the effectiveness of the lab at its overall mission. The ability of an R&D lab to quickly and successfully develop products and/or processes and, if necessary, to protect them through patent actions, may ultimately impact the firm's market share and its profitability.

Assessing the value of LIMS

Having looked at some general ways in which LIMS can contribute to the corporate well-being, we must next consider how contributions can be put into specific financial terms. In this process the establishment of reasonable expectations for LIMS' economic benefits is to a great extent dependent on understanding the mission of the lab rather than on the technical merits of the system. Furthermore, the assertion that acquiring LIMS technology for a laboratory is economically justifiable must be based on an analysis of how the laboratory contributes to the corporate bottom line. This holds true whether one is trying to determine the scope of a LIMS that is appropriate, or is trying to justify its purchase. In other words, we must consider not only what LIMS will do, but what it is worth. We must examine the specific ways in

which the analytical laboratory's only true product, information, contributes to the corporate bottom line as well as the ways LIMS can reduce the direct cost of laboratory operations.

In the QA/QC lab the primary area to investigate is the time-value of the information it produces. In the case of raw materials analyses, the ability to quickly accept or reject raw material lots may reduce costs associated with having to compensate for inferior materials or of holding materials or having to return them to the vendors. Similarly, the ability to quickly assess the quality and worth of raw materials in a competitive market may insure achievement of maximum value for the firms purchasing dollar. For in-process testing, improving lab performance may contribute to process optimization (saving energy and/or material) or to reducing waste and rework. Here, the speed at which test data can be used to adjust a process can be directly related to the cost of the end product. This may be seen in the reduction of total production time (with accompanying labor) or in the reduction of scrap or off-grade product, or in increasing the effective capacity (salable product per unit time) of the plant itself. Increasing effective plant capacity, for example, might reduce the need for finished goods inventories to accommodate peak demands. This would reduce the cost of inventory space, management, and insurance (it would also result in customers receiving fresher products). Furthermore, in applications where certificates of analysis are required prior to shipping or acceptance, time saved may translate into material holding and/or labor savings (e. g., for truck driver's idle time awaiting authorization to depart).

To this must be added the savings in direct lab productivity to be achieved through elimination of repetitive clerical work and through computerization of inefficient and error prone manual procedures. This productivity improvement usually amounts to at least 10%-20% of total staff costs. The reader can verify this by assessing the percent of staff activity spent manually recording data that could be captured electronically, transcribing and checking data, performing manual computations, searching for test data, looking up test methods and specifications, locating samples, and consolidating data into issuable reports.

In the R&D lab, the economic value of LIMS is more heavily skewed to the productivity area. That is, to cost reduction in the lab itself. Here, the productivity improvement of 10%-20% is of importance because of the larger labor costs of the R&D lab. However, even in the R&D environment, other economic benefits can be discovered although they may be difficult to quantify. The speed at which a product can be perfected and brought to market may be the competitive edge through which the firm secures a lead position in its marketplace. Sometimes this means getting there first; sometimes it means fielding a product in response to a competitor -before the competitor has been able to capture the market.

Thus, the rationale for LIMS technology in the R&D lab is not just an improvement in productivity or in the speed at which particular analyses can be completed, but the ability of the entire R&D

organization to complete projects in less time. When a lab manager can present a system to be acquired or developed as a productivity-improving capital investment that can be related to the firm's market share, then he is speaking in terms that the corporate financial people can understand.

Cost Analysis

Having given an overview of where the benefits of LIMS may originate, let us consider the actual cost of these systems. First, we must look at the total cost, not just the price tag on a purchased system or the development cost of an "in-house" system. This includes the costs of preparing a specification and/or requirements analysis (9), of site preparation, of system purchase or development, of installation, of integration into operations (including training and redundant activities during changeover), of continuing operation and maintenance, and perhaps even of insurance costs.

If the system is developed in-house there is the obvious cost of the development labor, but even if the system is purchased, staff will have to be committed to requirements analysis, liaison with the system vendor, integration of the system into operations, and in-house system support and maintenance. If the system is modest (based on a small minicomputer or super microcomputer), only the acquisition cost may be significant. If the system is a large one (based on a mainframe or super minicomputer) then the cost of site preparation, air conditioning, cable installation and service contracts will have to be considered.

In estimating the real cost of a LIMS, the actual cost may be significantly less than the system's purchase price or apparent development cost due to special tax considerations. First of all, there is an immediate Investment Tax Credit which is currently 10%. This means that a system whose price is, for example, \$100,000.00, would actually cost the firm only \$90,000.00. Then using an accelerated depreciation rate, the system's cost can be written off over 3 or 5 years. If the firm in our example is in a 50% tax bracket, as is typical, the actual savings would be \$10,000 on the ITC plus 50% of \$95,000 (cost less one-half the ITC as per the current tax code) or a total saving of \$57,500. This means that the \$100,000 system would cost only \$42,500 over a three year period. Now suppose that the firm borrows the initial cash outlay to buy the system and is paying a rate of 13% interest. The interest payments are also deductible at the firm's tax rate so the financing cost, offset by its tax deduction, raises the total system cost to roughly \$70,000. The exact figures would depend on the details of the firm's fiscal policies such as whether payments are made monthly, quarterly, etc., whether purchases are financed or bought out of cash reserves, etc.

The point to remember from this analysis is that the actual cash cost of a LIMS system is usually between one-half and three-quarters of its price tag.

Benefit Analysis

After looking at the costs of a system one must next answer the question, "is it worth it?". This is where financial analysis comes into play. Financial analysis is the process through which one decides whether the cost is justified by the anticipated benefits. It is the process of assessing whether a particular system is worth the price.

The benefits discussed earlier fall into two areas: the time-value of information and the improvement in laboratory productivity. Assessing the benefits of the time-value of information requires examination of the total process through which that information will be used. For example, using data such as: the turn-around time, rate at which material is produced, its value, and losses attributable to scrap or rework, one can estimate the savings. Similarly, knowing the average labor, overhead, and General and Administrative (G&A) costs allows one to convert productivity improvements in the lab to dollar values of equivalent labor made free for more productive uses.

Given estimates of both the real cost of bringing LIMS technology into the laboratory, and the dollar value benefits to be accrued from it, there only remains to make a final comparison. Since costs and benefits are not necessarily experienced in the same time frames, the financial analysis methods by which the firm makes capital investment decisions must be used. The LIMS which appears to be desirable from a technical and operational viewpoint must be proven to be a productivity or revenue improving capital investment and not just another overhead expense item.

Application of financial analysis methods

Three methods are used to assess the value of a capital investment. They are cash payback, net present value, and internal rate of return (also known as Discounted Cash Flow-Rate of Return).

Cash Payback

Cash Payback is calculated by a cash flow analysis. The cash flow generated by an investment is the cash value of the benefits it achieves less the cash outlays to pay for the capital investment. Assuming that the system's costs precede its benefits, cash payback is then the time interval until the cumulative benefit equals the cumulative costs. This figure is of major importance for the decision-making process but is not by itself conclusive. A projected fast payback indicates quickly achieved benefits and rapid return of the investment so that it is available to the firm for other uses. A fast payback alone does not insure that the system purchase is sound. That is, no matter how fast it pays for itself, one would desire a return on the investment at least as high as could be attained by some other (usually lower risk) use of that money. The other methods as described below are used to establish this criterion.

Net Present Value

The Net Present Value (NPV) of a capital investment is the equivalent total cash flow generated by all the acquisition's benefits less all the acquisition's costs computed over the life of the system on a year to year basis, adjusted for the value of money as reflected by such factors as finance rates, and projected ("discounted") to the present day. A dollar benefit projected for the system next year would only be worth \$ 0.91 today if that dollar could be earning 10% interest. A net present value of zero means that the acquisition will, over its projected life, just break even and that it is therefore an acceptable purchase. A better than zero NPV would be a high priority purchase since it indicates a real profit.

Internal Rate of Return

The Internal Rate of Return (IRR) is the equivalent interest rate at which the Net Present Value of the acquisition would be zero. Given the projected total cost of the system, and the projected total benefits of the system, both projected back (discounted) to today, it is the interest rate that the investment could sustain and still just break even. Since firms, in general, operate at a point where their incremental cost of money is equal to its incremental earning power, any investment that returns an IRR better than the cost of money is a good investment. Traditionally, the IRR is found by calculating the NPV with different interest factors in a trial and error method until the interest factor is found which drives the NPV to approximately zero.

Table II presents an example of an NPV determination for a hypothetical LIMS, purchased as a package with negligible site preparation and with installation costs included in the purchase. It is to be acquired for a service laboratory primarily supporting R&D activities but with some minimal process monitoring responsibilities. While the result in this example clearly shows a highly worthwhile acquisition, it not unreasonable, based on this author's experience with several different laboratories. The conclusion that a LIMS is a worthwhile acquisition is generally inescapable in all but very small laboratory operations. In fact, the reader is encouraged to try recalculating a NPV using representative figures for his own organization. For further reading on financial management analysis tools, the reader is referred to two works by Weston and Brigham. The first (10) presents theoretical and detailed analytical expositions; the second (11) is a more practical, applications oriented presentation.

TABLE II. Net Present Value - Discounted Cash Flow Analysis

Assumptions: -Cost of LIMS is \$250,000 both depreciated and amortized over 5 years.
 -Cost of money interest factor (I) is 13% and the firm is in a 50% tax bracket.
 -Lab has staff of 60 which, with overhead, costs a total of \$3,600,000 per year.
 -LIMS will produce productivity savings of 5% the first year, 10% the second, and 15% thereafter.
 -LIMS will produce savings from expedited QC testing (reduction of rework and scrap, improved claims processing on out of spec raw materials, and lower inventory maintenance and insurance) of \$25,000 the first year, \$50,000 the second, and \$100,000 thereafter.
 -One senior staff member will be committed to the project full time during the first year. Operations and Maintenance (O&M) will cost \$25,000 annually thereafter.

Formulas: $PVCF(n) = \text{Present Value of Cash Flow for year } n$
 $= CF(n) * PVIF(n)$ where $CF(n)$ is the n-th year's Revenue less Costs

$PVIF(n) = \text{Present Value Interest Factor} = 1/(1 + I)^n$
 where $n = \text{the period (year) number}$

$NPV = \text{Net Present Value} = (PVCF(1)*PVIF(1) + PVCF(2)*PVIF(2) + \dots) - \text{Initial Cost}$

Spreadsheet Computation:

CF Item	Year 1	Year 2	Year 3	Year 4	Year 5
Amortization of principal(-)	50,000	50,000	50,000	50,000	50,000
Interest (-)	32,500	26,000	19,500	13,000	6,500
Project labor(-)	60,000	25,000	25,000	25,000	25,000
Tax Credits (+)	41,250	13,000	9,750	6,500	3,250
Depreciation (+)	47,500	47,500	47,500	47,500	47,500
Productivity (+)	180,000	360,000	540,000	540,000	540,000
Revenue Gain (+)	25,000	50,000	100,000	100,000	100,000
Cash Flow	151,250	369,500	602,750	606,000	609,250
PVIF	0.8850	0.7831	0.6930	0.6133	0.5428
PVCF	133,856	289,355	417,706	371,660	330,701

Payback occurs in the second year.

$NPV = 1,543,278 - 250,000 = \$1,293,278.$

Conclusions

All this discussion may seem a far cry from the problems of running a flavor and fragrance QA/QC or R&D laboratory or for introducing LIMS technology to it. Indeed, if the lab is small, and its needs modest, all that may be needed to justify acquisition of LIMS technology is the knowledge of its value in facilitating laboratory management and increasing its productivity along with realistic estimates of the system cost and payback.

On the other hand, for a large laboratory complex whose problems would necessitate a comprehensive computerization program costing from a quarter to well over three quarters of a million dollars, we are dealing with a more complex decision making process. For large system acquisitions, a thorough analysis of the system's economics will probably be required. In such a case it is best if those responsible for exploring the benefits of LIMS technology, understand the financial decision making process that goes on in their own firm. Not only will this insure a judicious system acquisition, but it will develop, in advance, the specific data that the financial decision makers require and which proves the value of the system not just to the labs, but to corporate profitability as well.

Literature Cited

1. Dessy, R. E., Ed. Analytical Chemistry 1983, 55, 70A.
2. Dessy, R. E., Ed. Analytical Chemistry 1983, 55, 277A.
3. Harder, M. E.; Koski, P. A. American Laboratory 1983, September, 27.
4. Liscouski, J. G. American Laboratory 1983, September, 127.
5. Ouchi, G. I. PC World 1984, 2, 122.
6. Reber, S. American Laboratory 1983, February, 78.
7. Stinson, S. C. Chem. & Engr. News 1983, January 24, 41.
8. Golden, J. H. American Laboratory 1980, November, 111.
9. Robinson, A. L. Science 1983, 220, 180.
10. Weston, J. F.; Brigham, E. F. "Managerial Finance", Seventh Ed.; The Dryden Press; Hinsdale, IL, 1981.
11. Weston, J. F.; Brigham, E. F. "Essentials of Managerial Finance", Third Edition; The Dryden Press; Hinsdale, IL, 1974.

RECEIVED May 11, 1984

Instantaneous Analysis of Fragrances, Flavors, and Other Vapor-Phase Chemicals

Atmospheric Pressure Chemical Ionization Tandem Triple Quadrupole Mass Spectrometry (APCI/MS/MS)

BORI SHUSHAN

SCIEX, 55 Glen Cameron Road, Unit 202, Thornhill, Ontario, Canada L3T 1P2

Atmospheric Pressure Chemical Ionization (APCI) is coupled to Tandem Mass Spectrometry (MS/MS) for the purpose of analyzing vapor phase perfume mixtures. Air-borne fragrances are analyzed directly by APCI/MS/MS without the need for time consuming and potentially adulterating trapping and chromatography steps. Volatile fragrance chemicals have been rapidly identified by this novel technique as they emanate from vials or directly from skin.

Recent analytical methodology in the identification of the chemical components of fragrances and flavors has relied heavily on gas chromatography (GC) and gas chromatography/mass spectrometry (GC/MS). The success of these techniques are certainly well documented(1)and they have provided perfume and flavor chemists with a very good understanding of the chemical composition of complex formulations and mixtures. Despite the inherent sensitivity and wide applicability of conventional GC and GC/MS techniques they will never replace the well trained nose as a means of identifying odiferous components. The obvious advantages that the nose has is its great sensitivity and selectivity to those chemicals which give us smell. Another advantage of the olfactory organ is its ability to sample vapors directly at atmospheric pressure without the handling steps required for GC or GC/MS which may alter chemical composition. A major disadvantage of olfactory analysis, however, is a lack of resolution where all the chemical components reach the "detector" at the same time causing masking of one odor by another. Ideally an analytical technique is required which combines the resolution of GC/MS with the direct, "real-time" analytical capabilities of the nose. This paper will describe a technique for the direct analysis of vapor-phase chemicals which uses an atmospheric pressure chemical ionization (APCI) ion source coupled to a tandem triple quadrupole mass spectrometer. This type of instrumental system

0097-6156/84/0261-0075\$06.00/0
© 1984 American Chemical Society

(the TAGA® 6000 manufactured by SCIEX®) has been shown(2) to be particularly well suited to the detection of organic and inorganic chemicals in ambient air. Vehicle mounted mobile TAGA® 6000 Laboratory Systems and similar fixed-site instruments have demonstrated the ability to detect ppt-levels of a variety of organic compounds from gaseous, liquid and solid matrices using a number of inlets including Direct Air Sample (DAS), Direct Insertion Probe (DIP), Liquid Chromatography (LC) (3), and GC.

In this article we will present data obtained using the DAS inlet, sampling perfume mixtures, emanating from open vials or applied to skin, in order to demonstrate the principles of APCI/MS/MS and its use in research and quality control within the fragrance and flavor industry.

EXPERIMENTAL

The TAGA® 6000 is a quadrupole-based mass spectrometer combining three quadrupole arrays aligned axially (a schematic of the ion optics is shown in Figure 1)(4). The instrument can be configured to accept a variety of inlets and ionization sources however, for the present work the APCI ion source was used. Ambient or purified air can be drawn through this ion source using the DAS inlet at controlled rates of up to 9 l/sec. Trace organics and inorganic vapours are ionized by reagent ions formed within a point-to-plane corona discharge, forming product ions which are pseudo-molecular (indicative of molecular weight). The ions pass from the ion source (at 760 torr) through a small orifice and are focussed into the analyser portion of the mass spectrometer which has a base operating pressure of ca. 10^{-6} Torr. This represents a transfer of ions through a pressure reduction of almost 9 orders of magnitude using a single pumping stage. The vacuum system efficiency is due to a high capacity (60,000 l/sec) cryogenic pumping system consisting of a closed-loop helium-based refrigerator connected to strategically placed cryo-arrays. The pumping system requires only electrical power to operate. The orifice which connects the ion source with the analyser is protected by a stream of ultra-pure nitrogen. This gas curtain keeps ionized molecules, particulate matter and air away from the orifice and high vacuum analyser portion. Consequently the orifice never becomes plugged by particulates and the ion optics are always clean, being exposed only to ultra-pure nitrogen. Ionized molecules, on the other hand, proceed unimpeded through the gas curtain focussed along electrical fields through the orifice and into the analyser.

The positive mode reagent ions consist of protons in various states of hydration $H^+(H_2O)_n$, while the negative mode reagent species are primarily O^- , O_2^- and CO_3^- . In the positive mode proton transfer (1) dominates the CI chemistry i.e.

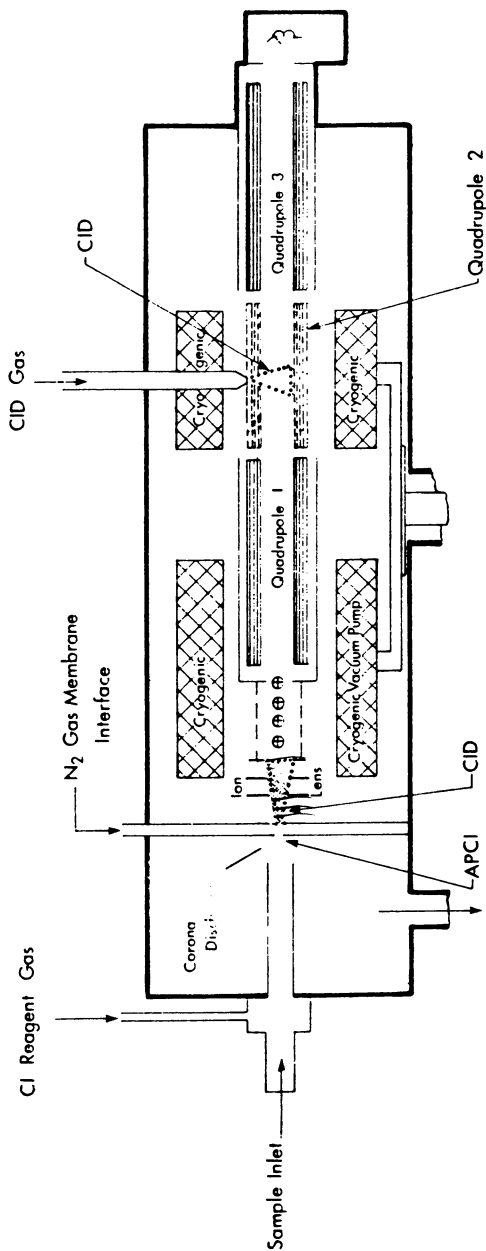
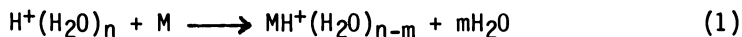


Figure 1. Schematic diagram of the TAGA 6000 APCI/MS/MS instrument.



where M is the trace species to be analyzed. The resulting product ion is further dehydrated by exposure to the ultra-pure N₂ curtain gas and by low energy collisional activation (CID on Figure 1) brought on by the application of slight electrical fields during the free-jet expansion into the analyser portion. The result is an APCI mass spectrum dominated by molecular or pseudo-molecular ions (usually MH⁺) with minimal associative or dissociative product ions. The negative mode reactions are dominated by proton abstraction (2) from Bronsted acids and electron capture (3) by Lewis acids again providing pseudo-molecular and molecular ions respectively.



Because of the chemical complexity of gaseous phase perfume mixtures, it is sometimes necessary to exploit the selectivity of the APCI process. In the positive mode, only those compounds which are more basic than the reagent ion H⁺(H₂O)_n species are ionized. Through the addition of a reagent with higher proton affinities than water the specificity of the ion source is increased. This also applies in the negative mode as well, exploiting relative gas phase acidities and electron affinities of the reagent to the analyte in order to alter ion source specificity. Alternatively, the APCI chemistry can be altered to provide charge transfer as opposed to proton transfer ion reactions. This is accomplished through the addition of a charge transfer reagent (such as C₆H₆) to generate a reagent ion "plasma" which only ionizes those compounds which have an ionization potential (IP) lower than that of the reagent's. Thus, for example, C₆H₆⁺ reagent ions would only ionize compounds with an IP < 9.25eV which includes aromatic hydrocarbons and excludes aliphatics.

Once the ions have been focussed through the orifice and into the analyser portion they are further focussed by electrostatic lenses into the first of three axially aligned quadrupole arrays. The first and third arrays can be operated as mass filters while the central quad (Q2 in Figure 1) functions only as an ion guide, incapable of resolving one mass from another.

Conventional mass spectra are obtained by scanning Q1 while operating Q3 as an ion guide. Three additional scan modes are available to the MS/MS user, they are:

1. Daughter Ion Scan,
2. Parent Ion Scan, and
3. Neutral-Loss Scan

Examples of these scan modes will be described below, however a fundamental description of each will be provided in this section. All three MS/MS scans require Q1 and Q3 resolving and a collision gas (usually Argon) must be present in Q2 at a typical target thickness of ca. 10^{14} cm⁻² or 10^{-4} torr pressure. The nominal ion energy through the Q2 collision cell is ca. 70eV.

1. In order to generate a Daughter Ion Scan, Q1 is set to pass ions of a particular mass-to-charge (m/z) ratio. After these preselected ions fragment in Q2 the fragment- or "daughter"-ions are mass analysed in Q3. The types of fragmentation reactions which occur are indicative of the precursor ion's structure, thus providing a means of identifying the original compound present in the ion source.
2. The Parent Ion Scan is used to generate a spectrum of all those "parent" ions of a preselected daughter ion. Certain classes of compounds fragment to give similar daughter ions; for example, all positively charged molecular ions of aliphatically substituted aromatics fragment to give $C_7H_7^+$ ions at $m/z = 91$. In order to scan for this particular class of compounds using the parent ion scan mode, Q3 is first set to pass ions of $m/z 91^+$. Then Q1 is scanned from $m/z 91^+$ on upward and every time the 91^+ daughter ion is detected a parent ion's mass will be recorded.
3. Just as certain classes of compounds yield characteristic fragment ions during fragmentation, most chemical classes yield typical neutral losses. Such specific neutral losses are detected by scanning both quadrupoles Q1 and Q3 synchronously whereby the mass setting of Q1 is always greater than that of Q3 by the mass of the neutral fragment.

All of the above scan modes are generated under computer control and the data obtained are stored on disk for easy access after acquisition. The data system consists of a DEC PDP-11/23 computer which uses two 10.4 MByte RLO2 hard disk storage devices. The software has been written specifically for MS/MS mixture analysis applications involving high through-put, batch sampling; as well, there exists software packages to handle the other more traditional GC, DIP and LC inlets.

RESULTS AND DISCUSSION

Rose Oil

In order to illustrate the principles of APCI/MS/MS a sample of a Rose Oil was analyzed directly. A vial containing 2 ml of the oil was uncapped and simply placed upright into the glass inlet line of the direct air sampling inlet. Vapours emanating from

the open vial were carried into the ion source on a 5 l.min⁻¹ flow of "zero" air. The resulting spectrum of positive ions is shown in Figure 2 (top). The complexity of the mixture can well be appreciated from this spectrum. By analogy, a gas chromatogram can be equally as complex in terms of the number of discernable peaks, except that a mass spectrum is generated in a matter of a few seconds and GC usually takes appreciably longer. Like the gas chromatogram, each peak in the APCI mass spectrum represents a pseudo-molecular species, which can be instantaneously subjected to MS/MS fragment ion analysis once separated by Q1. In GC/MS, each molecular species is subjected to mass spectral analysis to furnish fragmentation information only after separation from the other components by its retention time. In MS/MS however, each component is instantaneously separated on the basis of its molecular weight and then almost as rapidly undergoes CAD to furnish the representative fragmentation pattern. Thus, the component at m/z 193⁺ of Rose Oil furnishes the MS/MS daughter ion spectrum shown in Figures 2 and 3 (bottom) which was identified as α -ionone by comparison to spectra in our CAD MS/MS library (see Figure 3). Since β -ionone is a natural product of roses⁵ we were able to conclude that the rose oil sample in the open vial was a man-made formulation. This conclusion was based on the MS/MS identification of α -ionone obtained in only a few minutes. Some of the other components similarly identified were diethylphthalate (223⁺), citronellol (157⁺), nerol/geraniol (155⁺), citral (153⁺) and phenyl ethanol (123⁺).

One thing that became apparent in our investigation of this perfume mixture was that the terpenoid family of compounds (terpenes and substituted terpenes) all undergo CAD to form a prominent 81⁺ ion (C₆H₉⁺). A Parent Ion Scan for this ion should furnish a much simplified mass spectrum where only terpenes and their substituted analogues are detected. Figure 4 demonstrates this result where the degree of simplification can be appreciated by comparing the APCI/MS spectrum (top) to the APCI/MS/MS Parent Ion Spectrum for 81⁺ ions shown below it.

Another example of chemical class screening using MS/MS is demonstrated in Figure 5. Again, the simplicity of the APCI/MS/MS neutral loss spectrum of rose oil is evident by comparison to the relative complexity of the APCI/MS spectrum of the same sample shown above it. In this MS/MS neutral loss spectrum we are looking at all those ionic species present which lose neutral water (H₂O; 18 mass units). This would include the oxygenated species present in the sample such as aldehydes and especially alcohols which would tend to lose H₂O most readily i.e.



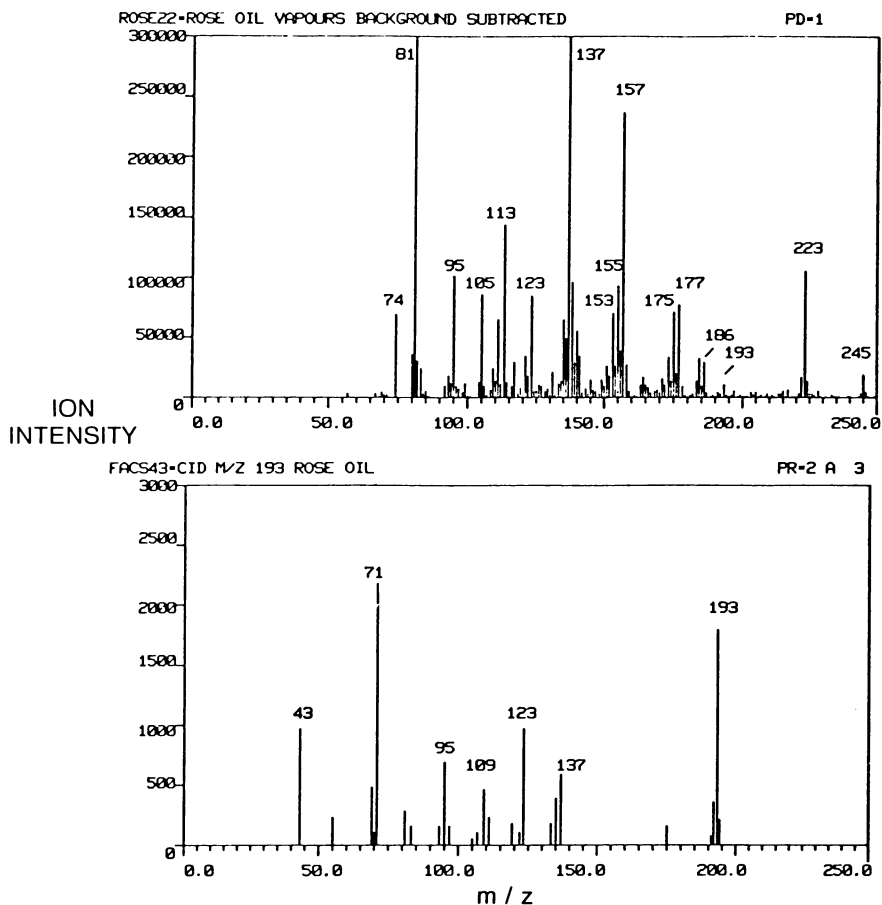


Figure 2. Top, APCI mass spectrum of rose oil vapors emanating from an open vial; bottom, CAD/MS/MS fragment ion spectrum of m/z 193 ions from the rose oil vapors.

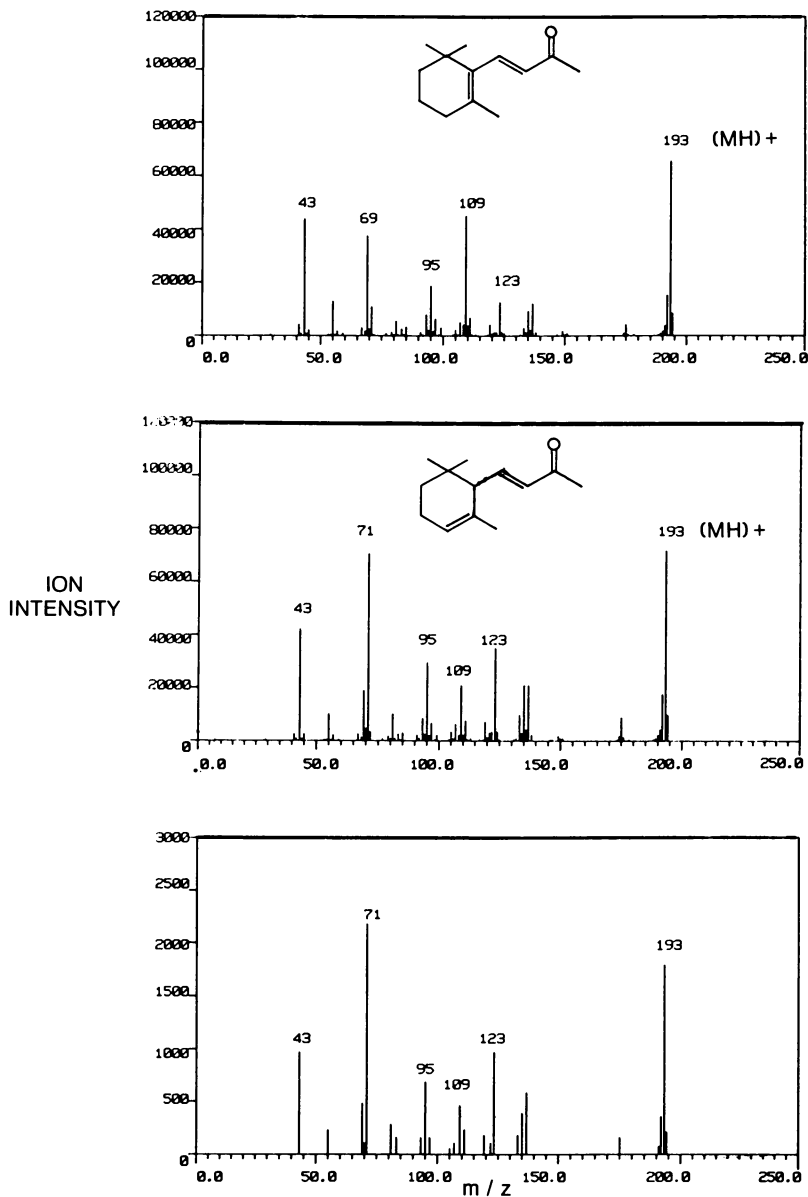


Figure 3. Top and middle, CAD MS/MS fragment ion library spectra of β - and α -ionone, respectively; bottom, CAD MS/MS fragment ion spectrum of 193⁺ ions from the rose oil vapors.

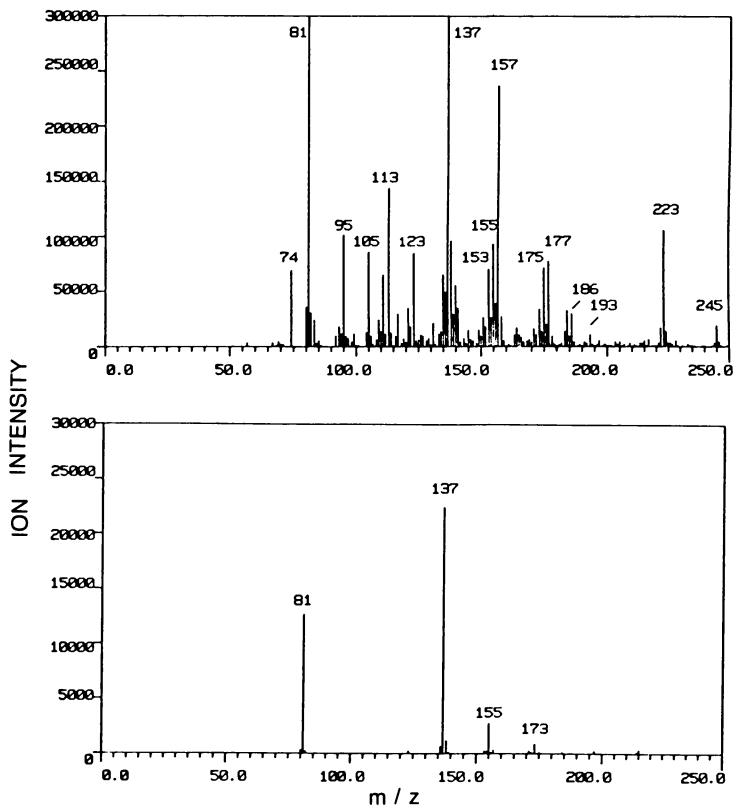


Figure 4. Top, APCI mass spectrum of rose oil vapors emanating from an open vial; and bottom, MS/MS parent ion spectrum for 81 ions ($C_6H_9^+$) accentuating the terpenoid class of compounds in the mixture.

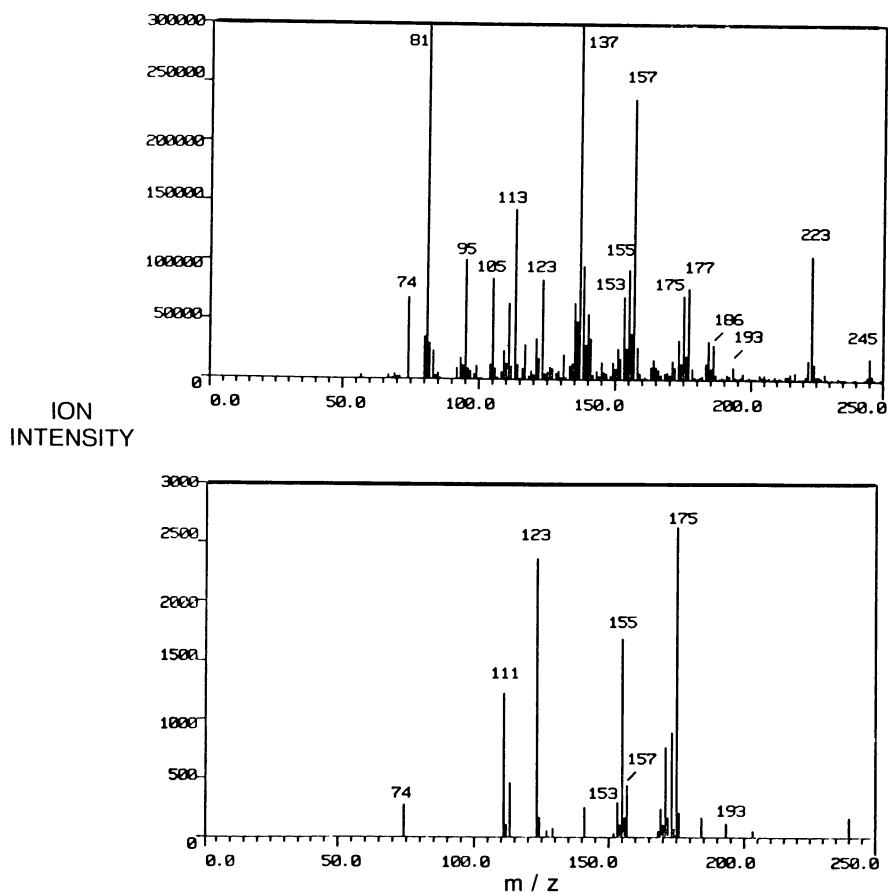


Figure 5. Top, APCI mass spectrum of rose oil vapors; and bottom MS/MS neutral loss scan for ionic species from the spectrum above which lose 18 amu (H_2O) during CAD.

Cutaneous Emissions

One of the most interesting aspects of APCI is its ability to directly sample streams of ambient air into which odor releasing samples can be placed without difficult, time consuming or adulterating pretreatment steps. This capability opens up vistas of analytical research which could never before be easily explored. One such area which shows direct applicability to the perfume industry, was the analysis of chemicals emanating from the skin.

In the first of two examples, negative mode APCI was used to selectively ionize acidic components emanating from the skin. In this short study, seven subjects (3 males and 4 females) placed their arms (after washing with warm water) through a latex rubber air-lock placed over the bottom end of a horizontally mounted glass bell jar. A stream of ultra-pure zero air (5 l. sec^{-1}) swept through the jar carrying volatiles emanating from the hands and arms of the subjects directly into the ion source. Ten to twenty negative mode APCI mass spectra were obtained for each subject. In the negative mode, volatile acids are detected as their pseudo-molecular ions $[\text{M-H}]^-$. Large signals were obtained for the three metabolically related acids; acetic, lactic and pyruvic (59^- , 89^- and 87^- ions respectively). These latter two species have been reported elsewhere⁽⁶⁾ and are metabolites of glucose generated in the Krebs cycle⁽⁷⁾. Figure 6 shows the MS/MS identification of the lactic acid coming off the hand of one of the subjects.

One aspect of this study which interested us was the possibility of differentiating males and females on the basis of their cutaneous emissions. In order to accentuate differences, all spectra from the subjects of one sex were averaged and difference spectra generated between the sexes; these can be observed in Figure 7.

The top spectrum is the result of subtracting the female's from the male's response. In this upper spectrum, one can readily observe a relatively large amount of lactate and pyruvate at 87^- and 89^- . These acids are products of metabolism; males generally have more muscle structure and hence more metabolism than females, therefore the evolution of these compounds in quantities significantly greater from males than from females is not surprising. The other peaks in this top spectrum can be ascribed to reagent ions and some remain unidentified.

The "females minus males" spectrum (Figure 7 bottom) showed a relatively large response to acetic acid as well as an apparently homologous series of acids at higher molecular weight (m/z 115-, 143-, 171- and 199-). This latter group of ions seems to be ascending by 28 mass unit intervals from 87^- (pyruvate) implying a series of compounds differing by multiples of C_2H_4

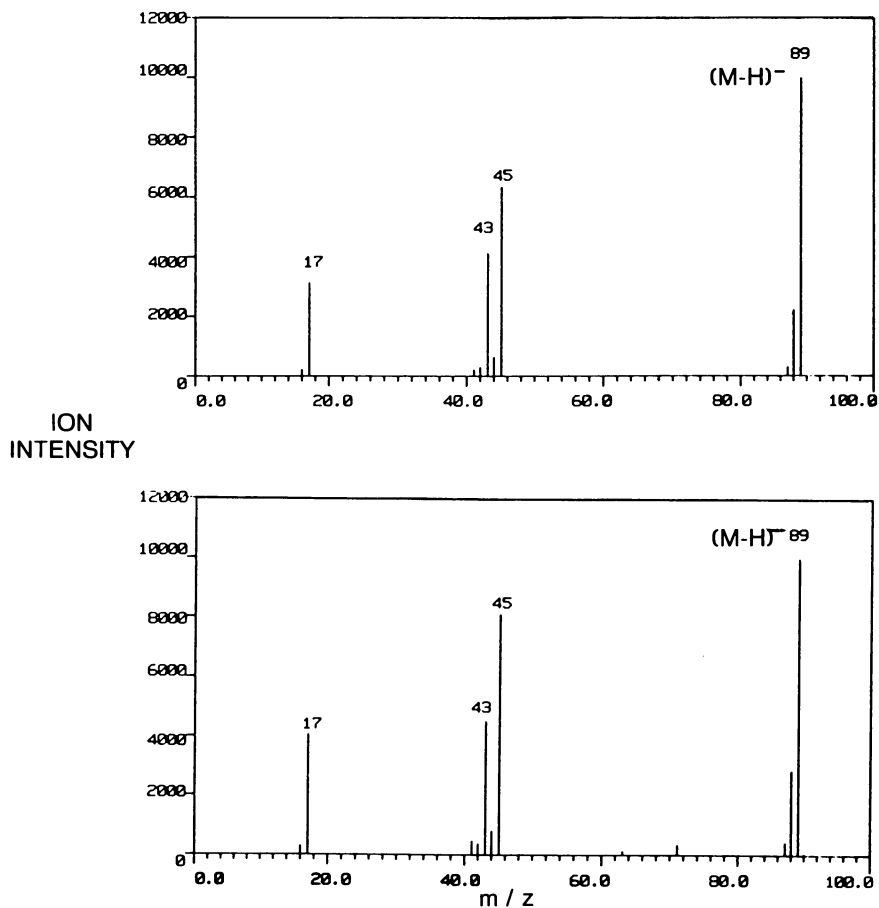


Figure 6. Top, MS/MS fragment ion library spectrum of lactic acid standard; and bottom, MS/MS fragment ion spectrum of lactic acid emanating from the hand.

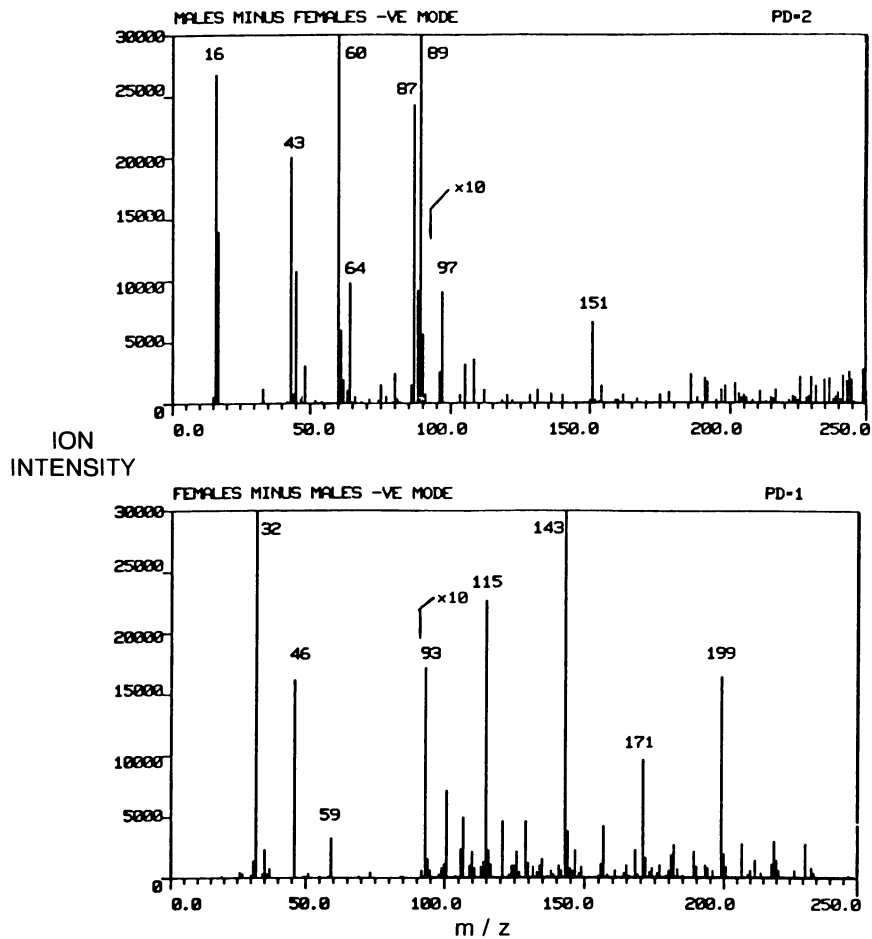


Figure 7. Top, averaged APCI mass spectra of the female subjects cutaneous emissions were subtracted from the average of the male subjects'; and bottom, same as above but in reverse order.

units homologous to pyruvic acid. Thus the signals of 87-, 115-, 143-, 171- and 199-, which are observed in the spectra of all the subjects, have been tentatively identified as pyruvic, levulinic, oxo-heptanoic, oxo-nonanoic and oxo-undecanoic acids respectively.

With the exception of lactate and pyruvate, the differences in levels of these acids emanating from the skin of the subjects did not vary too greatly between our small groups of representatives of the different sexes. The statistical and socio-sexual significance of these results notwithstanding, the above study has demonstrated the chemical complexity of cutaneous emissions even prior to the application of natural or artificially formulated scents and fragrances.

The presence of naturally occurring substances on the skin has recently attracted a great deal of attention(8) especially from perfumers who are even incorporating some naturally occurring steroids suspected of pheromonal qualities into their perfume formulations. The possibility of chemical interaction between the scents applied to skin and those chemicals already present on the skin has lead one perfume manufacturer to claim that their product "smells differently on every woman". In order to test this claim qualitatively, a sample of this perfume product was applied to the washed forearms of 6 female subjects and allowed to evaporate and "react" for two hours. The volatiles emanating from the skin were then tested using the above direct sampling procedure but in the positive mode. The APCI/MS spectra for all of the subjects were added together and averaged and this result was subtracted from each of the subjects in order to accentuate any differences between individuals. A sampling of the results are shown in Figure 8. The large number and intensity of signals in the spectra of subjects 2 and 3 suggests that these two individuals differ substantially from the mean. The mean APCI/MS spectrum (the average of all subject's spectra; not shown) showed many peaks which were prevalent in that of the perfume's shown at the top right. The lower number and intensity of peaks in the spectra from subject 6 tells us that the perfume's components have not undergone the apparent chemical transformation which has occurred to those applied to subjects 2 and 3.

In view of these above cursory observations the topic of scent modification through cutaneous contact warrents a more complete investigation.

Quantitation

Vapor phase components are quantified through the syringe injection of a head space vapours at controlled rates directly into the ion source. Into a syringe barrel are deposited a few drops of

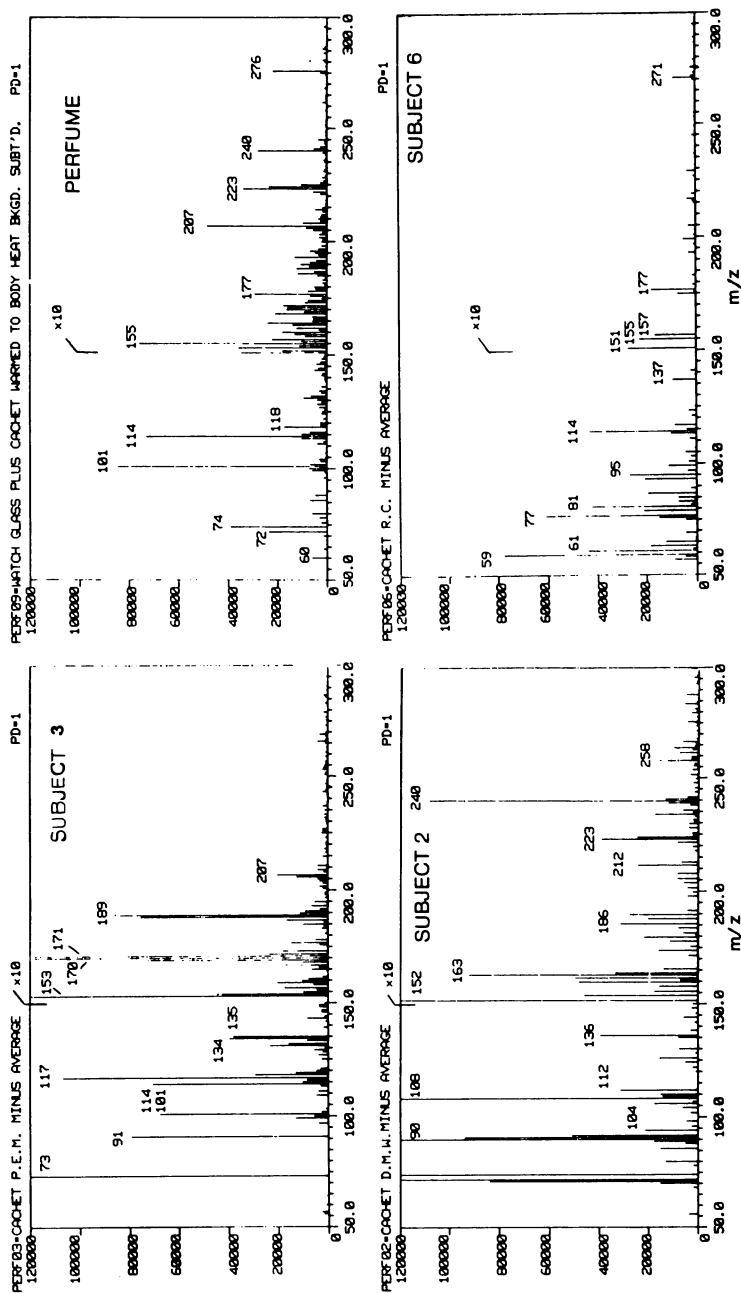


Figure 8. Top right, APCI mass spectrum of perfume vapors on blotting paper two hours after application; and others, APCI mass spectra of various subjects' cutaneous emissions two hours after application of perfume. The average of all subjects' spectra has been subtracted from each to accentuate any differences.

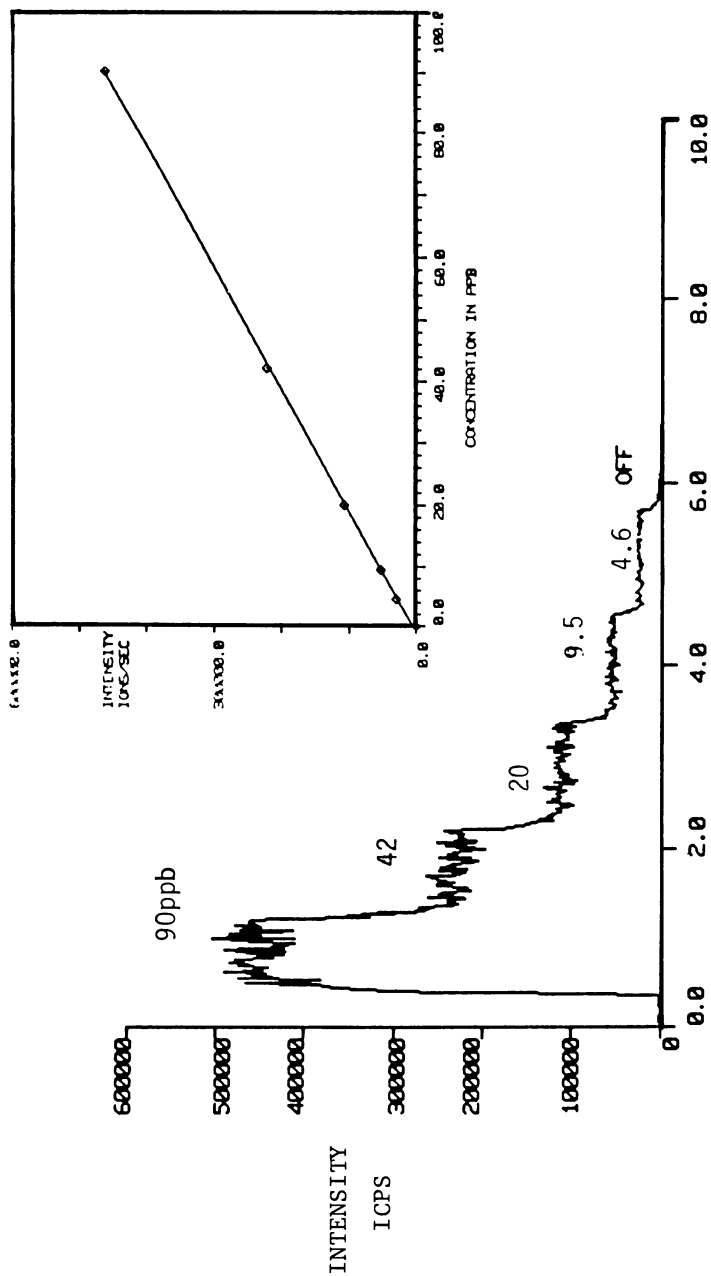


Figure 9. System response to various vapor-phase concentrations of methyl-2-methylbutanoate detected as MH. Inset, calibration curve derived using these data.

the pure chemical, if liquid, or a few crystals, if solid. After equilibration the vapors inside the syringe are injected into a known amount of carrier gas at carefully controlled rates using a syringe pump. The carrier gas plus calibration sample go directly into the ion source. The APCI/MS response to various speeds of injection of methyl-2-methylbutanoate vapours is shown in Figure 9 where the actual amount of this compound (in ppb) is determined from the rate of the syringe pump, the material's vapour pressure and the carrier gas flow. The calibration curve is plotted above the response (insert Figure 9) and is observed to go through the origin with a slope of ca. 5100 ion counts per second/ppb. This response factor can then be applied to peak responses in mass spectra in order to determine the absolute amounts of materials present.

CONCLUSION

This paper has demonstrated that APCI/MS/MS can provide a highly selective and sensitive means of detecting and identifying vapor-phase compounds, even within complex chemical mixtures. The ability to use ambient air as the reagent and carrier gases greatly facilitates the application of such sophisticated computerized mass spectrometry to "real-world" sampling situations such as those encountered in organoleptic research within the fragrance and flavor industries. Combining the specificity of the TAGA® 6000 APCI/MS/MS system with software tailored to the perfume industry (such as multivariate analysis) will provide abilities to objectively assess and discriminate scents which will rival those of that bottom line of organoleptic quality control, the nose.

Literature Cited

1. W.H. McFadden, "Techniques of Combined GC/MS: Applications in Organic Analysis", 1973, John Wiley & Sons.
2. F.W. Karasek, Ind. Res./Devel, December (1978) cover article.
3. J.D. Henion, B.A. Thomson and P.H. Dawson, Anal. Chem., **54** (1982) 451.
4. P.H. Dawson, J.B. French, J.A. Buckley, D.J. Douglas, and D. Simmons, Org. Mass Spectrom., **17**(5) (1982)205.
5. G. Ohloff, Perf. and Flav., **3** (1978) 11.
6. B.A. Thomson, W.R. Davidson and A.M. Lovett, Environ. Health Perspect., **35**(1980)77.
7. A.L. Lehninger, "Biochemistry", 1970, Worth.
8. J.N. Labows, et. al. Steroids, **34**(3)(1974)249.

RECEIVED April 6, 1984

Near IR Reflectance Analysis An Example of Computer-Aided Chemistry

DONALD A. BURNS

Technicon Industrial Systems, 511 Benedict Avenue, Tarrytown, NY 10591

Near Infrared Reflectance Analysis (NIRA) is in use at over 5000 sites for the analysis of multiple constituents in food and other products. The technology is based upon correlation transform spectroscopy, which combines NIR spectrophotometry and computerized analysis of a "learning set" of samples to obtain calibrations without the need for detailed spectroscopic knowledge of factors being analyzed. The computer can obtain spectral characteristics of the analyte (based upon a correlation with data from an accepted reference analysis) without separation of the sample's constituents.

The secret dream of most analytical chemists is to have some sort of magic analyzer into one end of which they can poke their samples and obtain from the other end a final report, in correct concentration units, and without their having to perform such menial tasks as preparing samples or keeping reagent bottles full. Such an instrument may already exist, and this presentation describes one manufacturer's offering (Figure 1) and explains the principles upon which it is based.

Basic Principles

Reference to Figure 2 will simplify what follows. If a light beam of original intensity I_0 passes through a transparent sample, some light will be absorbed and some will emerge at a reduced intensity I . The ratio I/I_0 is defined as the transmittance T , and absorbance Abs is defined as the logarithm of $1/T$ (Figure 2a). If a sample is opaque, it may still absorb light; that which is not absorbed will be reflected at a lower intensity. Again, the ratio I/I_0 may be defined as the reflectance R , and the absorbance can likewise be defined as $\text{Log } 1/R$ (Figure 2b). Indeed, this is

0097-6156/84/0261-0093\$06.00/0

© 1984 American Chemical Society

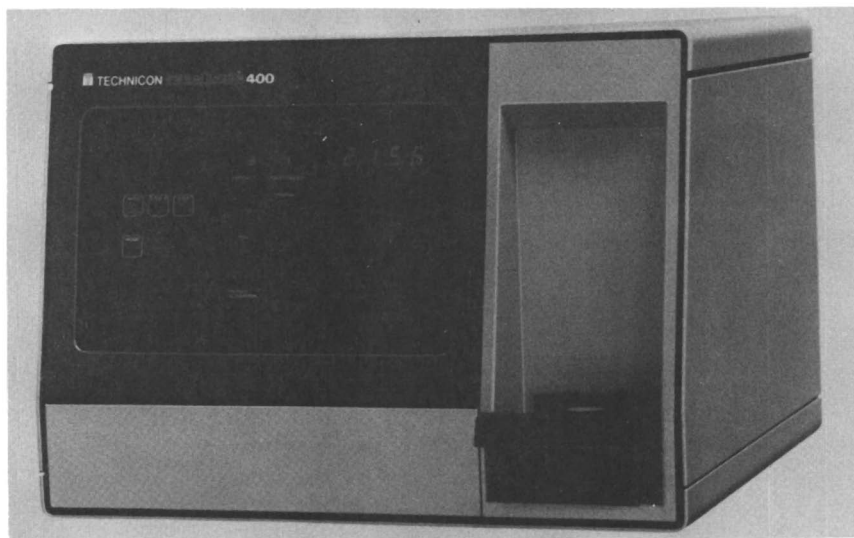


Figure 1. InfraAlyzer 400 system (courtesy of Technicon Industrial Systems).

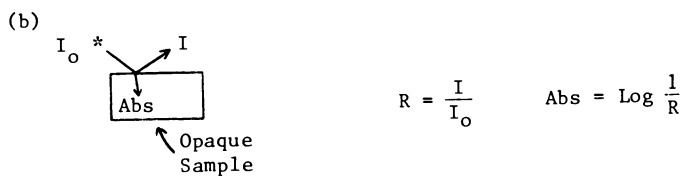
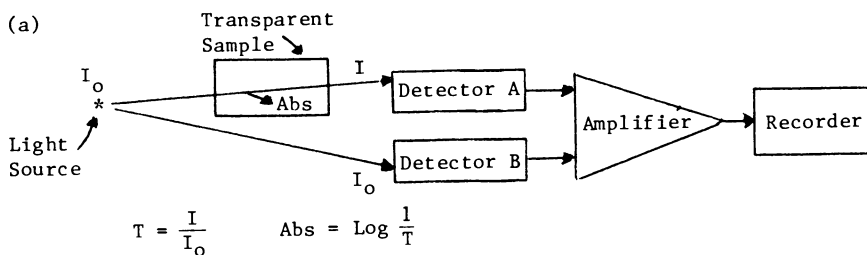


Figure 2. Absorbance: a, via transmittance; and b, via reflectance.

the common designation in the field of Near Infrared Reflectance Analysis (NIRA).

Reflectance

There are two kinds of reflectance: specular and diffuse. The first is best demonstrated by standing in front of a mirror. One sees himself, essentially intact, because the light rays which travel from the subject to the mirror and then to the viewer's eyes have not been changed by the mirror; they contain the same information before and after reflection.

Diffuse reflectance becomes apparent if one stands in front of a wall and hopes to see his own reflection. He cannot, of course, since the light rays traveling from him to the wall will interact with the wall and are subsequently reflected in all directions. In addition, these light rays now carry information about the wall -- its color, its texture, and even its chemical composition. Suitable instrumentation can evaluate the difference between the light before and after diffuse reflectance and provide a route to deducing chemical structure.

One thinks of reflected light as striking a surface and then leaving it, as depicted by R_1 in Figure 3. Actually, the situation is more like R_2 . The light penetrates some distance into the material, this distance being small but not zero. If it is enlarged, one sees some chemistry within the sample -- bonds vibrating according to type of atoms, temperature, etc. This vibration is at a specified frequency, and when light of the same frequency (or an overtone) strikes a bond, there is a potential for energy absorption. As one scans through many frequencies of I_0 he may observe rising and falling levels of I , and such a plot is essentially a spectrum such as will be described later.

To bring this closer to our ordinary experiences, presume that white light is falling upon common lawn grass. What happens? The relative intensity of reflected light is greatest at the color we call green; other colors are either partially or completely absorbed. These colors are more specifically defined in terms of their wavelengths, usually expressed in nanometers (nm); thus, green is 535 nm, and the visible spectrum runs from 400 to about 750 nm (Figure 4). The full decade from 100 to 1000 nm has been re-expressed in powers of ten to show the position of the visible portion of the electromagnetic spectrum. Above the visible region, up to about 10,000 nm, is the classical infrared (IR) region, and that portion nearest to the visible (up to about 2500 nm) is called the Near Infrared (NIR). It is this NIR region which we shall use in the remainder of this publication.

NIR and its Instrumentation

Those analysts who come to NIR from the classical IR region will note that the absorbances are lower by 2-3 orders of magnitude.

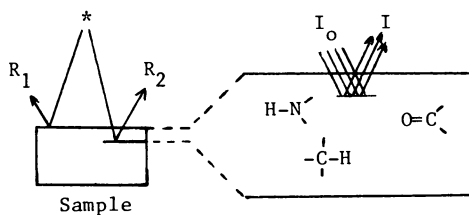


Figure 3. Interaction of light with a sample.

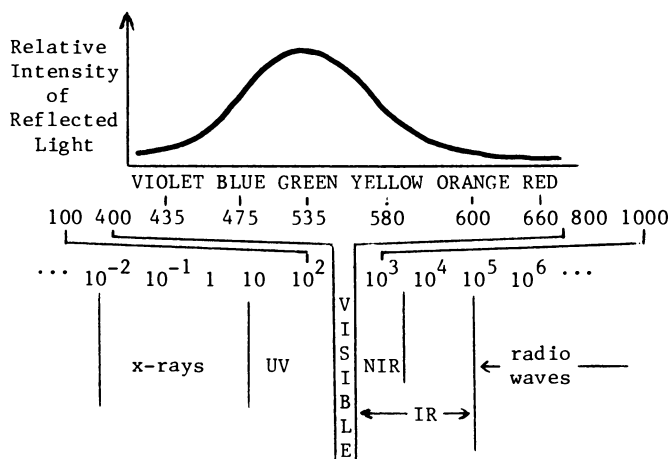


Figure 4. Electromagnetic spectrum, showing position of NIR region.

In addition, they see much broader peaks with considerable overlap and the absence of expected baseline resolution. Clearly, it would require the power of a computer to deconvolve such seemingly useless wiggles. But today's computers cost less, run faster, have more memory, and readily lend themselves to such calculations as are necessary to glean useful information from NIR spectra. Compare the spectrum of water and of a natural product like soybeans (Figure 5): water, with its two fairly well defined peaks, is one of a few exceptions to the above description of NIR spectra, while the soybean spectrum is quite typical of what one sees in the real world.

A middle-of-the-line instrument such as the Technicon InfraAnalyzer 400 system is shown diagrammatically in Figure 6. The light is rendered monochromatic by passage through one of 19 interference filters. The emerging NIR light is directed by a mirror onto the surface of a sample where some light is absorbed, interacts with the sample, and is diffusely reflected in all directions. This light illuminates the interior surface of an integrating sphere and is measured by two detectors. The mirror then tilts to its alternate position, thereby causing the light to miss the sample and fall instead on the wall of the sphere. Thus, the two different mirror positions permit measurements of both I_0 (the reference beam) and I (the sample beam). The computer calculates R and Abs , then proceeds to solve an equation (to be described later) to determine the concentration of various constituents in an unknown sample.

More sophisticated instruments (such as the Technicon InfraAnalyzer 500 system) employ a scanning monochromator so that data can be collected at up to 700 different wavelengths. In most instruments data are collected between 1100 and 2500 nm, although provision is made for observations which begin in the visible region when useful information is there.

Calibration

If only one constituent were present in a sample, and there were no scattering effects, then there would be no possibility of interfering substances and one could likely find a wavelength at which Abs is proportional to concentration $Conc$. Under these conditions (adherence to Beer's Law), a simple equation can be created from a series of known samples by using a standard statistical tool (a least squares best fit) and defining an equation as follows:

$$Conc = F_1 \cdot Abs \lambda_1 + F_0$$

Note the similarity to the equation of a straight line:

$$Y = m \cdot X + b$$

The factor F_1 is the slope, λ_1 is the corresponding wavelength,

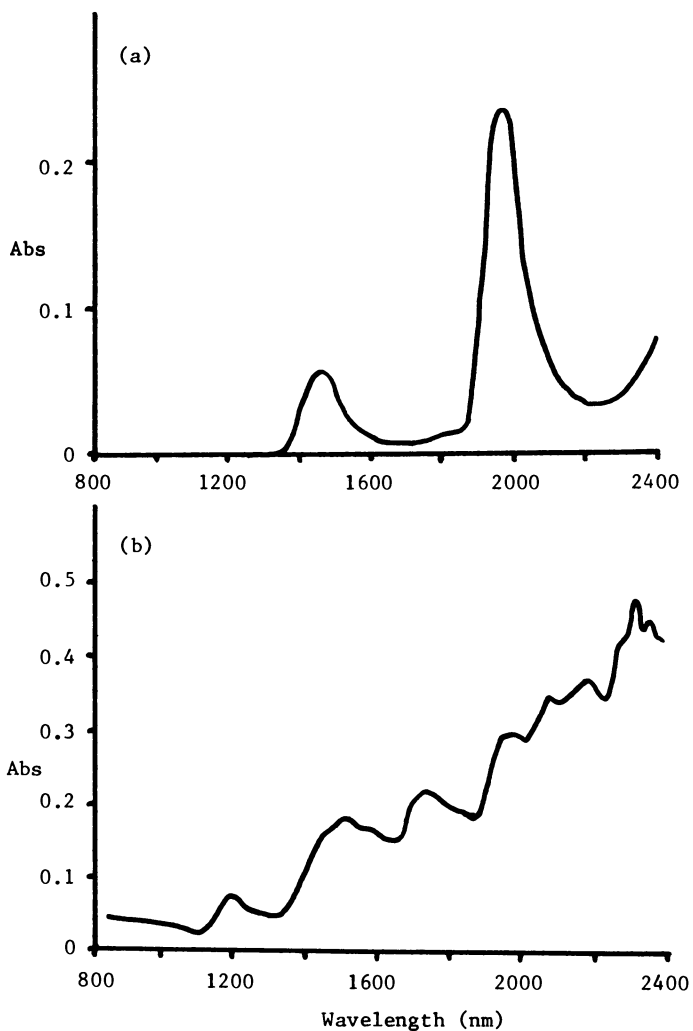


Figure 5. NIR spectra: a, water; and b, soybeans (example of natural product).

and F_0 is the offset (or so-called Y-intercept). Graphically, it looks like Figure 7, where the absorbance \underline{a} of an unknown material is measured, the point \underline{b} on the calibration line directly above \underline{a} is noted, and the concentration is read at \underline{c} .

In the real world samples generally contain more than one constituent, and their spectra usually overlap. If we had a 2-component sample with individual and composite spectra like those in Figure 8, and if we wanted to determine the concentration of J, we would need to make measurements at two wavelengths, j and k. The tail (leading edge) of K overlaps the peak of J, giving rise to a composite absorbance (the solid line) which is higher than J's contribution.

How can this be accomplished? Simply by adding another dimension to the previous illustration. Instead of one Abs line, we define two: Abs λ_1 and Abs λ_2 which define an absorbance plane. On each of these axes we can construct standard curves (essentially two calibration lines) and these will define a calibration plane (see Figure 9).

For any given unknown for which this calibration plane has been defined, one can measure vector \underline{oa} (the absorbance at λ_1 , designated Abs λ_1) and vector \underline{ob} , then the vector sum \underline{od} . Directly above \underline{d} is point \underline{e} on the calibration plane, and the concentration \underline{c} can be read on the Conc axis (at the end of the vector \underline{ec} which is parallel to vector \underline{od}). All this is done by the computer, of course, and the calculations are easily programmed.

But where did the calibration plane come from? By letting the computer use the same statistical tool that it used before: the least squares best fit. Only this time it operates with one more dimension. Initially (in order to calibrate the system) one obtains a so-called "learning set" of standards (samples which have been analyzed by some acceptable reference method). Each sample is then measured at each wavelength and the absorbances are plotted in three dimensional space rather than in the plane of a sheet of graph paper. The residuals (distances from point to plane) are minimized by the least squares algorithm, and the plane which fits best through these points is, by definition, the calibration plane (see Figure 10).

Although relatively easy to picture for one or two wavelengths (i.e. with drawings in two or three dimensions), it is rather difficult for most of us to picture a drawing in four or more dimensions. But the computer has no such difficulty. It can handle the calculations for as many dimensions as are necessary to cope with as many interferences as exist. The resulting hyperplane is then used to determine an unknown constituent by making absorbance measurements at several well-defined wavelengths and solving a simple equation. Such an equation could be

$$\% \text{ Conc} = F_1 \cdot \text{Abs } \lambda_1 + F_2 \cdot \text{Abs } \lambda_2 + \cdots + F_n \cdot \text{Abs } \lambda_n + F_0$$

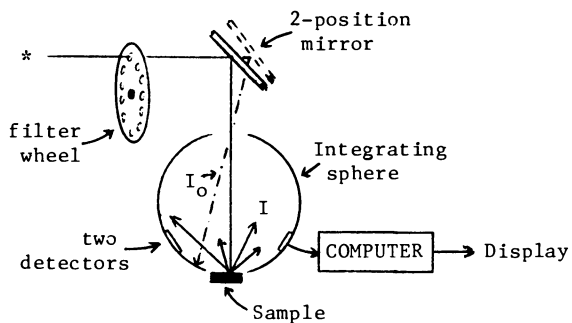


Figure 6. Optical system of InfraAnalyzer 400.

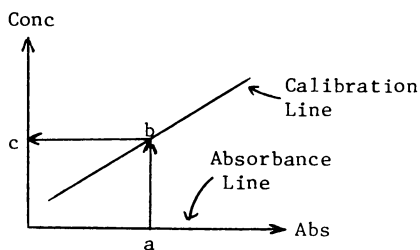


Figure 7. Simplest calibration (single wavelength).

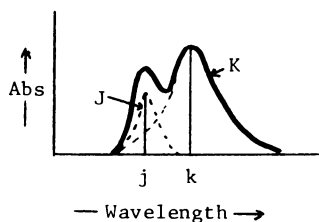


Figure 8. Interference between two spectra.

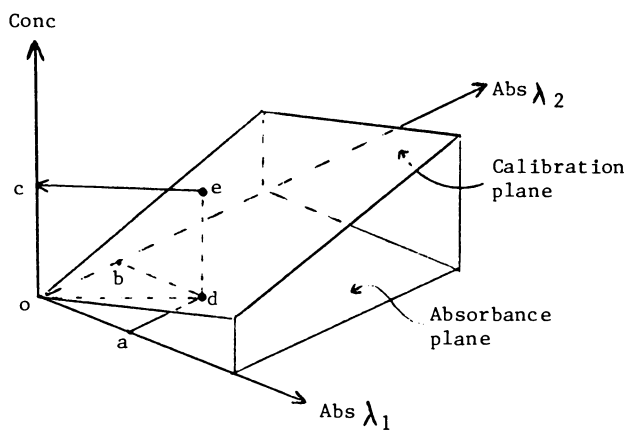


Figure 9. Two-wavelength calibration plane.

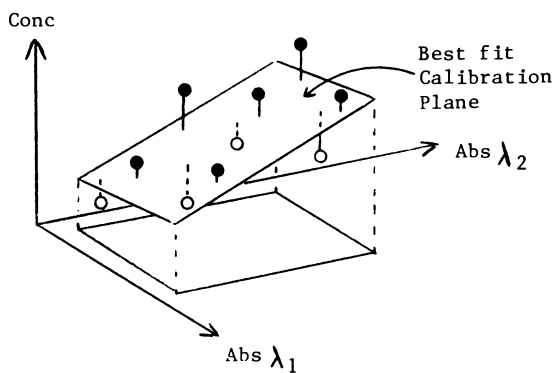


Figure 10. Least squares best fit of plane through points in space.

where n represents the number of wavelengths required to account for all interferences (typically 3-6) and F_0 is a sort of "composite offset" (equivalent to the "b" in the equation $y = mx + b$).

Operation

In operation, samples may be solids or liquids. Most solids are placed in sample cups which have a quartz window on top so the relative position under the integrating sphere is reproducible. Semi-solids are sometimes handled in open cups to simplify cleanup. Liquids are usually injected via syringe into a special cell following passage through a heat exchanger (since temperature plays a bigger role in liquid measurements than it does with solids). A special liquid drawer is interchangeable with the standard solid drawer.

A more sophisticated instrument with a scanning monochromator is frequently used with a learning set to create the equation which will be used for subsequent determinations of constituent concentrations in unknowns. The scanning instrument is quite similar to the filter instrument except that it can make up to 700 measurements on each sample rather than being limited to 19 as on the simpler version with a filter wheel. The computer which accompanies the Technicon InfraAnalyzer 500 system, for example, is a Hewlett-Packard 1000 with a 12" screen, thermal printer, and floppy disc drives. Its program is user-friendly and menu-driven, meaning that most of the time a single keystroke with one finger is sufficient to move logically through the steps required to go from sample scanning to report generation.

Correlation

A set of standards which covers the analytical range of interest will produce a family of spectra in which (it is hoped) the change in absorbance at one or more wavelength will correlate with the changing concentrations of the analyte. In the simplest case, the computer program methodically examines these changes at each wavelength where measurements of absorbance were made (typically 350 data points over the range 1100-2500 nm at 4 nm increments) and identifies that wavelength for which the correlation is highest. Holding this wavelength constant, it then searches for the next best wavelength (i.e., the next highest correlation), then the 3rd best, etc., until no significant improvement is seen. There are other search strategies (such as those involving a test of all combinations of two or more wavelengths), but most are based upon the improvement of the correlation coefficient of repeated regression analyses (an expression with which the reader should be familiar, since it is a most basic calculation in statistics and is fundamental to an understanding of correlation transform spectroscopy).

Calibrations which are generated on the top-of-the-line instrument can be transferred to a less expensive instrument for routine use by less skilled personnel. Computer programs handle all the details, such as a filter transform (to make data collected via monochromator look like it was collected via interference filter), a data reduction (to limit a large data base of as many as 700 wavelengths to only those 19 wavelengths represented by the filters in a smaller instrument), and a calibration (the creation of an equation by examining known standards and maximizing the correlation as described above).

The computer printouts are lengthy and filled with enough statistics to satisfy the most sophisticated user, but for purposes of illustration, we'll limit our example to the highlights of the data from a 3-wavelength search on wheat samples for the constituent protein. It is noteworthy that such ill-defined molecules as proteins can be determined so rapidly, with so little sample preparation (a few seconds of grinding), with such low errors (typically with standard errors of prediction of about 0.2 over the range 9-18 percent protein).

Highlights of Printout

	NUMBER OF WAVELENGTHS		
	1	2	3
STANDARD ERROR	7.11	1.47	0.55
CORRELATION COEFFICIENT	.923	.997	.999
COEFFICIENTS: B(0)	-105	120	90
B(1)	260	1389	775
B(2)		-1391	-962
B(3)			545
WAVELENGTHS	1440	1664	1652
		1824	1820
			1292

As more wavelengths are added, the standard error diminishes and approaches some minimum value, and the correlation coefficient increases and approaches unity. The equation becomes more complicated with the addition of more coefficients so that it may better define the analyte's concentration.

Feasibility Studies with Flavor Constituents

In order to establish the potential of NIRA in any given field, three methodical steps are required: a feasibility study (with low, medium, and high levels of the sought analyte), a full calibration (with 25-30 samples covering the range of analytical interest), and confirmation that the generated equation works in

the real world (by correctly determining a group of about 20 reference samples which were not used in the development of the equation).

The reader must understand the proprietary nature of information obtained from many companies, and the requirement that names and actual constituent levels be arbitrarily scaled to conceal the identities. Accordingly, all values in the following examples have been rescaled to set the lowest levels to ten.

One manufacturer of microencapsulated flavors wished to analyze his product for moisture and oil essence by NIRA and compare results with his reference method. Preliminary results were as shown in Table I and Table II. For the oil essence, the lower relative standard deviation (RSD) for the InfraAnalyzer method indicates a more precise measurement than was obtained with the manual method upon which it was based, although accuracy would depend upon that of the reference method. The slopes are near 1.0, showing good correlation between the two methods. The NIRA method gives slightly higher results with a range (for "A") of 10.8-11.5 compared with 10.0-11.0 for the reference. RSD's are often $1\frac{1}{2}$ times higher than the reference method, so the moisture determinations are equally good. For a feasibility study with such a small data set, these findings are within experimental error and suggest that further work should be done to establish a robust calibration.

In another study, flavor beads were analyzed for carbohydrate and water. In this case, however, the product was made in a variety of colors (white, red, yellow, green, blue) which must be regarded as interferences, since the level of analyte was unrelated to the visible color. It required 25 samples and five wavelengths to account for the irrelevant colors and produce these results:

<u>Level</u>	<u>Range</u>	<u>Correlation Coef</u>
Normal	10-10.8	0.95
Low	10-10.3	0.90

It would normally require more than 25 samples in a learning set to produce a really good equation. Despite this less than optimum number, the search algorithm was able to identify a combination of five wavelengths (from the 19 discrete filters in the InfraAnalyzer 400 system) and generate an equation which could extract carbohydrate and water concentrations from spectra which came from highly colored material. Clearly, the NIRA technique shows promise for this analysis.

Finally, a food processor presented 11 samples for a feasibility study on cheese powder to indicate its usefulness in quantifying caloric content as well as such constituents as moisture, protein, fat, ash, and carbohydrate. The composite absorbance tracings of Figure 11 produced a 3-wavelength equation, the results of which are shown in Table III. The

Table I. Oil Essence

Product	#	Reference Method		InfraAnalyzer Method		Slope
		Level (%)	RSD (%)	Level (%)	RSD (%)	
A	1	10.0	4.2	10.9	10.8	1.04
	2	11.0		11.5	11.3	
	3	10.6		10.9	10.9	
	4	10.9		11.0	11.1	
B	1	11.0	4.8	10.5	10.6	0.97
	2	10.3		9.9	10.0	
	3	10.0		9.6	10.3	
C	1	10.0		9.6	9.6	0.96

Table II. Moisture

Product	#	Reference Method		InfraAnalyzer Method		Slope
		Level (%)	RSD (%)	Level (%)	RSD (%)	
A	1	15.0	16.3	15.0	15.6	1.04
	2	10.0		11.6	10.6	
	3	12.5		12.8	13.1	
	4	12.5		12.5	12.8	
B	1	10.0	17.0	10.8	11.4	1.00
	2	13.3		12.8	13.6	
	3	13.9		13.0	13.0	
C	1	10.0		11.5	11.5	1.15

Table III. Cheese Powder

Constituent	Range	Corr Coef	RSD (%)
Calories	10-10.3	0.81	0.6
Moisture	10-12.9	0.93	3.6
Protein	10-10.3	0.83	0.5
Fat	10-11.3	0.70	3.0
Ash	10-10.5	0.89	4.5
Carbohydrate	10-10.4	0.70	0.9

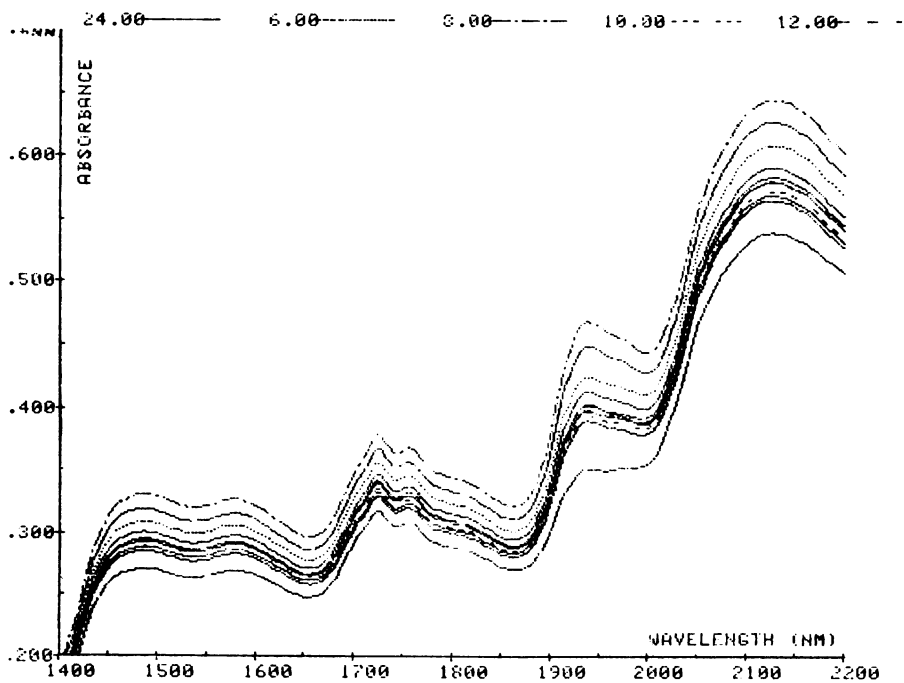


Figure 11. Spectra of cheese powder (composite of 25 samples in learning set).

instrument used here was the InfraAnalyzer 500 system with the scanning monochromator. Except for the carbohydrate analysis, all analyses (even the rather non-descript caloric values) can likely be handled quite well if the standard deviation of the reference method could be improved.

Conclusions

Near Infrared Reflectance Analysis (NIRA) is finding ever increasing use in the food industry as a rapid and non-destructive testing technique. Sample preparation is usually simple, and sometimes isn't required at all. With a good reference method upon which to base the NIRA method, one can expect good correlation, accuracy, and precision. Methods developed on a sophisticated scanning instrument can be used by less skilled personnel on lower cost filter instruments at the most useful sites (e.g., manufacturing plant, QC lab). The calibrations are user-transferable among instruments within the InfraAnalyzer line. There is so much information present in the NIR region that NIRA is probably destined to become a preferred analytical method.

RECEIVED April 6, 1984

Multivariate and Gas Chromatographic Techniques in Flavor Research

M. A. JELTEMA, B. W. GOOD, F. S. HSU, and M. E. PARRISH

Philip Morris U.S.A., Research Center, P.O. Box 26583, Richmond, VA 23261

A substantial effort has been made to develop sound objective measurements for quality evaluation of a consumer product's flavor, aroma and taste. It is apparent that progress in flavor research was dependent on advances in analytical separation techniques as well as the application of computer technology to provide not just data management, but more importantly, direction for further research. The growth and success of this research activity has been tied closely to the development of gas chromatography and more recently to the application of pattern recognition techniques for data analysis and simplification. Input from the analytical, sensory and statistical disciplines have resulted in the development of techniques which are playing an increasingly important role in solving quality assurance and product development problems. The impact of the direction of this research on our work, as well as two specific examples of the use of multivariate analyses techniques for tobacco research will be discussed.

Analytical Data Analysis. The development and commercialization of the gas chromatograph in the mid 1950's had a dramatic effect on flavor research because the technique made it possible to obtain objective measurements of the numerous compounds which made up the flavor of the product under investigation. Data analysis was reasonably simple and straightforward, as the number of resolved peaks was small. However, as chromatographic techniques were refined and high resolution capillary columns and microprocessor controlled GC's were introduced, the use of computers and multivariate analysis techniques have become essential for data analysis and reduction.

0097-6156/84/0261-0109\$06.50/0

© 1984 American Chemical Society

These most recent developments have played an important role in our research investigations on cigarette smoke. As a consumer product, cigarette smoke can be considered a very complex and dynamic flavor system. One facet of our research efforts has been to develop analytical methods for characterizing the cigarette smoke generated from different tobaccos used in the product. For addressing the task of differentiating cigarette types, statistical analyses are necessary for data reduction, pattern extraction and ranking the importance of the gas chromatographic peaks. This approach has been used by researchers in many different areas (1-5). These techniques have been applied successfully in our laboratory to the low boiling volatiles comprising the gas phase (6) and the higher boiling volatiles extracted from a filter pad with Freon-11, followed by trimethylsilylation and automatic sampling onto a fused silica capillary column (7).

Our emphasis continues to be the development of analytical techniques for obtaining reproducible chromatographic capillary profiles which reflect to the greatest extent the possible flavor attributes of cigarette smoke and tobacco, and to become more experienced in the proper use of pattern recognition techniques for data analysis. Recently, however, we also have recognized the need to correlate analytical data to sensory data. Just as pattern recognition has been demonstrated to be valuable in the distinction of samples based on their gas chromatographic profile, pattern recognition also is useful in the analysis of sensory data.

Sensory. Although the basis for multivariate analysis was developed in the early 1900's, its use in sensory analysis is relatively recent. These types of statistics, however, have been valuable in dealing with two fundamental problems which occur in sensory testing. First there are difficulties encountered when one attempts to breakdown complex sensory parameters into single semantic terms which can be rated, and second it is difficult to achieve the goal of every panelist having the same internal understanding of each term. Approaches to minimize these difficulties included: 1) evaluation of semantic terms used by the panel to determine if the variables are unique or can be condensed to a new set of unique variables; 2) evaluation of the panelists use of semantic terms to determine inconsistencies as well as the relative importance of the terms to food quality or discrimination.(8)

While some researchers feel that individual components making up the sensory response need to be characterized (9), others feel that methods dealing with the composite sensation are more appropriate. For this reason methods which do not rely on internal representation of terms have been applied. One such method is multidimensional scaling (MDS), which treats data based on a persons' total perception of the dissimilarity between objects.(10)

One of the major uses of multivariate techniques has been the discrimination of samples based on sensory scores, which also has been found to provide information concerning the relative importance of sensory attributes. Techniques used for sensory discrimination include factor analysis, discriminant analysis, regression analysis, and multidimensional scaling (8, 10-15).

Specific to our research, the multidimensional techniques such as MDS, factor analysis, canonical correlation and regression analysis have been used not only to analyze sensory and analytical data, but also to perform correlation between the two sets of data.

Objective and Sensory Correlations. The wish to correlate physical parameters to sensory parameters is not new. It was first used to correlate subjective and objective measurements of color and texture. It was not until the 1960's that researchers began to correlate flavor with gas chromatographic data, basically due to the lack of good analytical techniques. Since that time, research in the area has accelerated. Although some researchers have been successful in relating the concentration of single components to sensory response (16,17), it has been recognized that for most applications, a multivariate approach is necessary since sensory response is due to interactions of a multitude of components. For instance, in a mixture of 12 compounds there is a possibility of 4095 sources of combinations which could generate a sensory response (15). This problem is compounded by a lack of knowledge of how we respond to sensory mixtures and models to predict such a response. The difficulties of relating objective measurements to sensory scores are compounded further by the fact that the presence of one compound may suppress the perception of another or the sensory response may be due to compounds which are transparent to the analysis, thereby making correlations more difficult to perceive. Also, while analytical response is linear, this is not true of sensory response; therefore linear functions may not always be appropriate (18).

While the problem of relating sensory response to a simple mixture is difficult, this is compounded when efforts are made to relate sensory response to the thousands of components contained in cigarette smoke. As with many food systems, the differences are essentially quantitative rather than qualitative. The use of multivariate techniques are essential since they are designed to deal with all the peaks of a chromatographic profile. Fortunately, many of these components are highly correlated with others, and therefore simpler variables can be extracted through techniques such as factor analysis.

Our efforts to correlate instrumental and sensory data have initially centered around two areas, both of which are somewhat simpler than the prediction of sensory response to cigarette smoke. The first was aimed at developing a model for a quality

assurance (Q.A.) procedure where it is essential to know if a misformulation would fall out of a preset sensory specification. The second dealt with tobacco aroma where we were interested in discovering if differences which could be seen analytically between types of tobacco could be related to differences found by a sensory panel.

Each of these problems is different, therefore the use of different sensory as well as analytical techniques was necessary.

Application to Quality Assurance (Q.A.)

Implementation of a sensory quality control specification can be extremely difficult because production will often exceed panel capabilities. A model which predicts panel test results with even moderate success would be extremely valuable in avoiding the acceptance of poor production lots. For quality control problems, it is often sufficient to establish that a difference has been detected which is great enough to exceed present specifications. It is generally assumed that a relationship (not necessarily linear) exists between percent correct response in difference tests and the actual degree of difference between the stimuli (19). Therefore, in this study replicated triangle tests were used as the sensory input. The samples used for the study included blends of 5 multicomponents where the level of each multicomponent in the blend was individually altered. Modifications ranged from -100% (missing) to +100% (2 times reference level) of the concentration of each multicomponent in the reference mixture.

Sensory. Triangle tests were performed by a panel of ten experienced subjects. Each subject performed ten triangle tests at 60 second intervals so that all subjects evaluated all experimental triangle test comparisons. A triangle test typically consisted of ten randomly placed combinations of AAB and ABB. Panelists were asked to choose the odd sample based on odor. It was found that the initial three replications were part of a learning process by the panelist, therefore the first three test results were not used. As shown in Figure 1, while individual replications (ie. 10 subjects performing a triangle test) varied substantially about the significance level, percent cumulative response showed a consistent mean correct panel response after the first few trials, with little evidence of fatigue.

Analytical. Samples were chromatographed on a Hewlett-Packard 5880A gas chromatograph which was fitted with a 30M fused silica capillary column (DB5) and an automatic sampler. The GC was interfaced to a Hewlett-Packard 3354 Laboratory Automation System (LAS). Raw data was automatically transferred to the LAS where peaks were selected by retention time, integrated and stored in a processed file. Processed data was then transferred

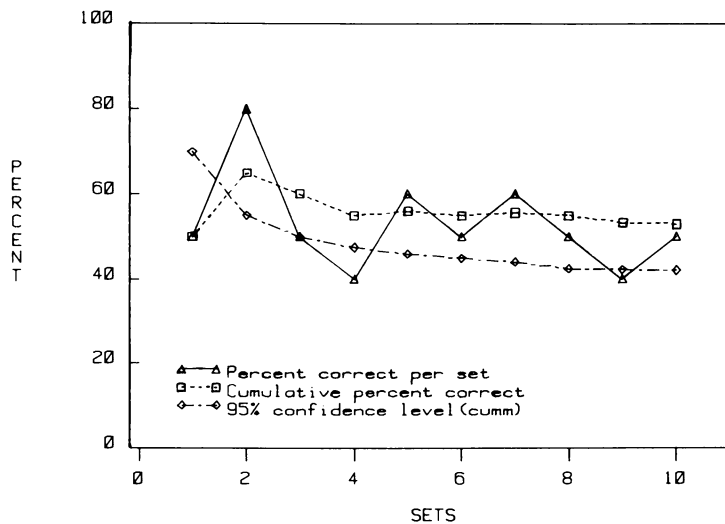


Figure 1. Panel response on triangle tests.

over terminal lines to a DECSYSTEM 2060 for statistical analysis using BMDP programs (20).

Since the sensory data collected involved degree of sample difference from a reference, it was felt that the analytical data should be analyzed in a similar manner. In cases where some peaks making up a multicomponent mixture are known to be specific to that mixture, this is a relatively simple matter. In such cases, the peak areas of the known components can be compared to a reference and average percent difference calculated. However, if it is not possible to pick out peaks that are clearly specific to a single multicomponent mixture, a more sophisticated technique such as factor analysis is required. There are circumstances where all peaks are common to each multicomponent mixture, i.e. qualitatively similar but quantitatively different. Also there are cases where peaks are found only in one of the multicomponent mixtures, but it is not clear to which mixture they belong. In these cases factor analysis is required to extract patterns that are characteristic of the specific multicomponent mixtures. Analytical concentrations of each of the multicomponent mixtures are then calculated as a set of factor scores where each score is directly proportional to the actual concentration of each multicomponent mixture.

Data Reduction. Data reduction was performed in two separate ways: factor analysis and peak area ratios. A typical chromatogram of the component blend used is shown in Figure 2. Factor analysis (20) was performed to reduce and simplify the data. The chromatographic data was thereby reduced from the 31 selected peaks to 5 factors. Each factor was principally comprised of components which originated from one multicomponent mixture but the analysis also indicated which components were found in more than one of the multicomponent mixtures. The factor scores, which were calculated for each sample, gave a composite score for the concentration of each individual multicomponent. Samples were therefore separated by plotting their factor scores (Figure 3). Figure 3 demonstrates the separation that was achieved by plotting samples on factor 1 (x axis) versus factor 2 (y axis). The samples which contained +100% of multicomponent 1 received a high positive score on factor 1 (~3.0), while samples containing -100% of that multicomponent received a high negative score. Likewise, samples which contained +100% or -100% of component 2 received high positive and negative scores on factor 2. Smaller alterations in concentration ($\pm 20\%$) produced lower factor scores (± 0.6). Samples in the center which were not separated from each other included the reference (factor score ~ 0.0) and samples which contained the same concentration of multicomponents 1 and 2 as in the reference. These samples had been altered in the concentration of 3, 4 or 5; however, this did not affect their factor scores for factor 1 and 2.

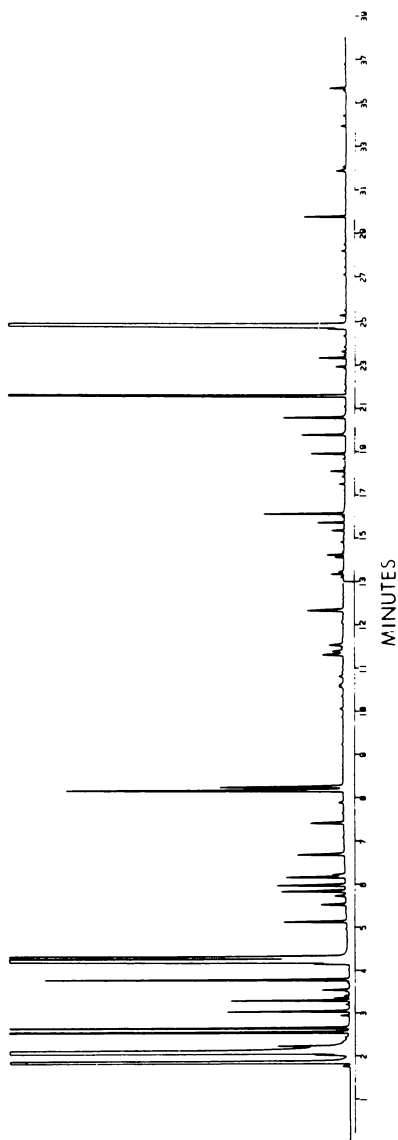


Figure 2. Chromatographic profile of a reference.

Publication Date: August 22, 1984 | doi: 10.1021/bk-1984-0261.ch009

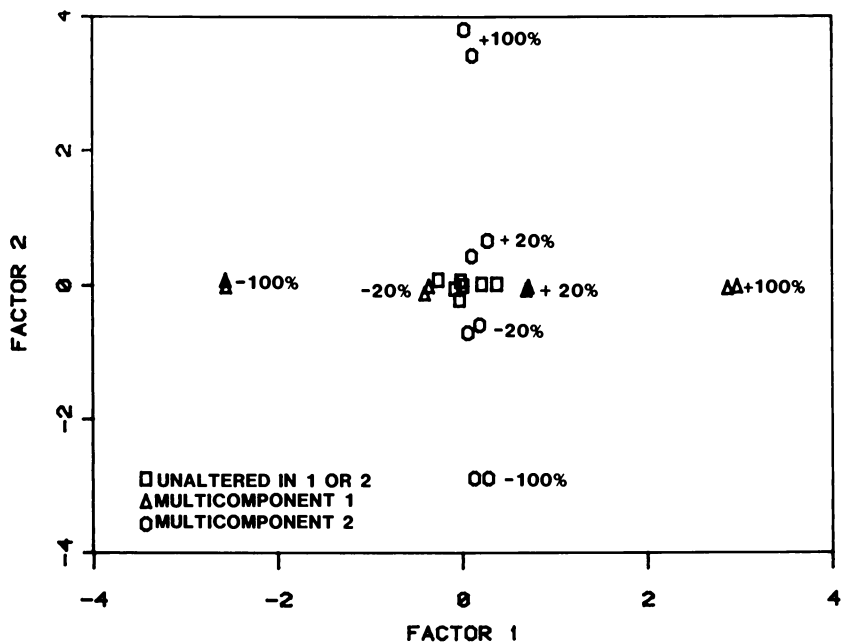


Figure 3. Factor score plots of altered samples.

In the case of this multicomponent mixture, it was known which peaks belonged to which multicomponent. Therefore peak areas of known components could be easily compared to a reference and the average percent difference calculated. This was done for each sample with a simple program. This program was automatically initiated at the time a sample was chromatographed by the HP3354 LAS. An example of the output is shown in Table I. This sample was formulated to contain -100% (or 0%) of a multicomponent C and -50% (or 1/2) the normal concentration of a multicomponent A. From the list of missing peaks and the fraction of individual peaks, this was found to be true. The extra peaks which are listed can be used to pinpoint possible sources of contamination.

Table I. Computer Output from Calculation of Percent Deviation of Each Component Based on a Reference Mixture

** FINGERPRINT SPECIFICATION PROGRAM **
SAMPLE NAME: 50A W/O C

** MISSING PEAKS **

EXPECTED RT	NAME
11.59	C1
13.38	C2

** EXTRA PEAKS **

ACTUAL RT
3.3
15.77

PEAK	TIME	FRACTION	NAME
1	2.55	0.52	A1
2	3.79	0.52	A2
3	11.37	0.52	A3
4	16.19	1.05	B1
5	19.83	1.04	B2
6	20.64	0.21	BC1
7	29.81	0.54	A4

Prediction of Panel Response. It was believed that the best theoretical model to predict sensory response from the analytical data would be a quadratic function (Figure 4), since as the concentration of a multicomponent was either increased or decreased from the reference, more of the panel should be able to differentiate the odd sample from the reference.

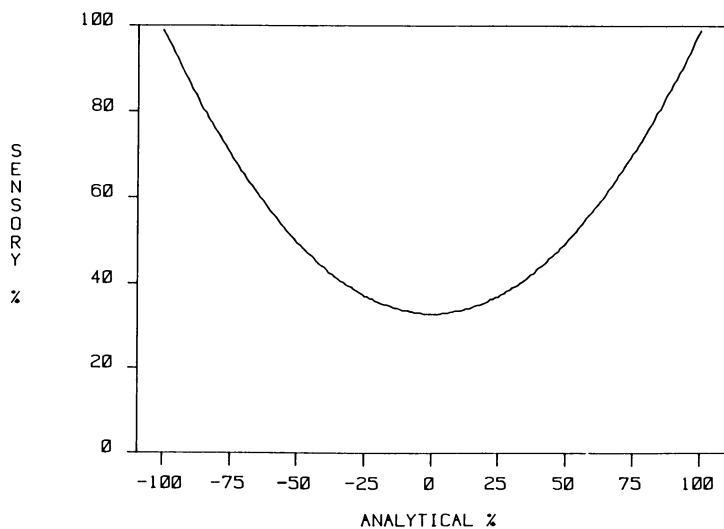


Figure 4. Theoretical panel response to alterations in the concentration of a multicomponent mixture.

The theoretical minimum of this curve is a 33.3% correct response which represents a chance response of selecting the odd sample when a reference from one lot is compared to a reference from another lot.

This prediction using either factor scores or analytical deviation from a reference is shown in Figure 5 for multicomponent 1. The types of curves obtained for the other multicomponents were similar, although the steepness varied according to the sensitivity of the panelists for changes in each multicomponent.

In order to set up an analytical method based on factor scores or analytical deviation in a Q.A. environment, several precautions would need to be followed. For instance, changes in raw materials would need to be monitored and either new reference materials used or factor analyses repeated. Changes in chromatographic conditions would require a system which allowed constant updating to avoid shift in factor scores. While the peak ratio technique is the simpler of the two, factor scores provide a more sophisticated method which can be used where overlap of components does not lend itself to the simpler methods. These two techniques for deriving the concentrations of the multicomponents in a mixture were shown to predict sensory response when one multicomponent was altered. However it also could be used in a multidimensional formula to predict response when more than one multicomponent has been altered.

Application to Product Development

The second study focused on the use of analytical/sensory correlations for a product development application. Since the sensory character of a cigarette is due in part to volatilization products which are distilled off of the tobacco in the rod by the heat of the coal, differences in the odor of tobaccos may correlate to sensory differences upon smoking.

Analytical. Five tobaccos were available in a cased (flavored) and uncased (unflavored) mode: 100% bright, (A) 100% burley, (B) 100% oriental, (C) 33.3% bright/33.3% burley/33.3% oriental and 60% bright/ 30% burley/10% oriental.

A 10 gram sample of conditioned tobacco (23°C, 60% RH, 48 hrs) was placed in a 70°C water-jacketted desorbing chamber. Two liters of helium were passed through the tobacco over a 20 minute period, trapping the volatiles on an air sampling probe containing 500 mg of Tenax 35/60 mesh.

The Tenax containing the tobacco volatiles was placed on a CDS 320 Concentrator equipped with a thermal desorber to accept the 6" x 1/2" probe. A liquid nitrogen cryogenic trap was controlled by the LAS for further focusing. Temperature programming and data acquisitions were performed by the Perkin-Elmer Sigma 1 and the Hewlett-Packard 3354 LAS respectively.

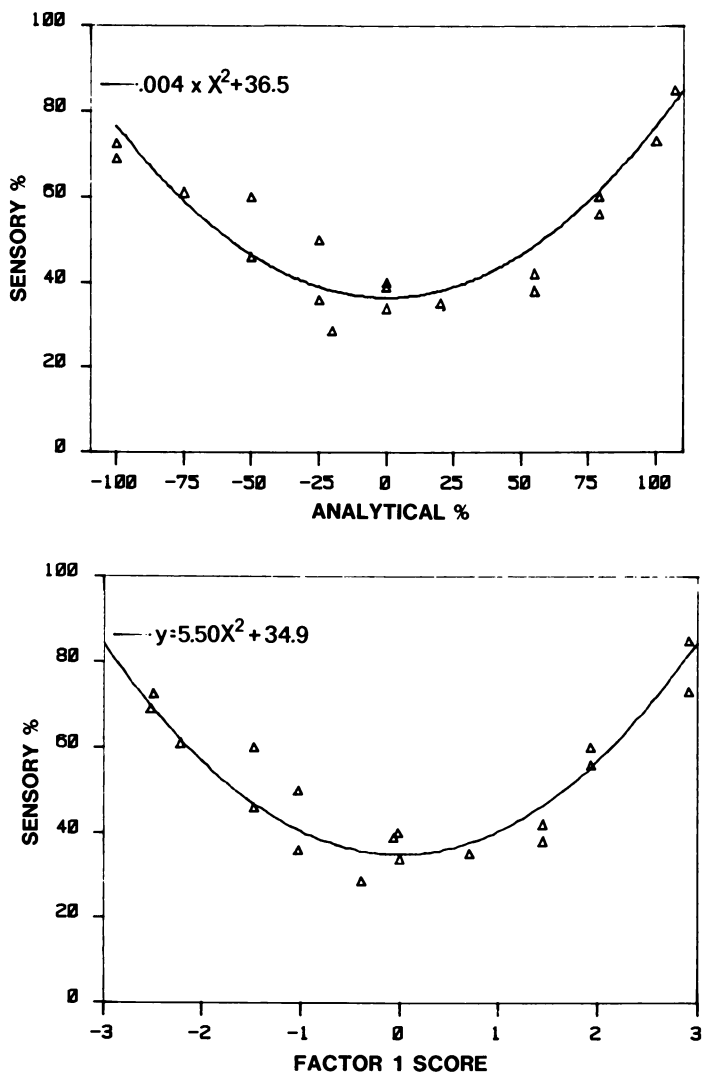


Figure 5. Panel response to alterations in the concentration of multicomponent mixture one. Top, based on analytical deviation from a reference: $R^2 = 0.86$; and bottom, based on factor scores: $R^2 = 0.91$.

Twenty-two peaks, 5 of which were designated as retention reference peaks, were selected arbitrarily without identification based on reproducibility of integration over the entire data set.

Data Reduction. A gas chromatographic profile of the Tenax-trapped volatiles from a 33.3% bright/33.3% burley/33.3% oriental tobacco sample is shown in Figure 6. The 22 selected peaks are designated.

A statistical program, BMDP4M (20) was applied to the 22 peaks from the ten tobacco types and four bright/burley blends individually. The purpose of factor analysis was to find which peaks correlate with each other and what information a peak (or group of peaks) might impart as to tobacco type, blend or casing. In the analysis of the ten tobacco types, five factors were calculated that explained 81% of the variance of the data set. Plots of factor scores in Figure 7 make it easy to see the utility of certain peaks for distinguishing the tobacco types and casings. Factor 1 (derived principally from peaks 11, 13, 15 and 18) explained the greatest variance, but was not useful for classifying the samples in any recognizable order. Factor 2 was clearly indicative of the presence of casing (peaks 6, 7, 8, 10, 19 and 22). Factor 3 (peaks 1, 3, 4, 16 and 20) vs. factor 5 (peaks 2, 17 and 21) offered the greatest classification. Factor 3 appeared to represent the burley character, and factor 5 the bright character (negative axis) and oriental character (positive axis). Factor 4 (not shown-contains peaks 5, 9, 12 and 14) visually offered no separation of the samples, but the factor scores from factors 4 and 5 had a correlation coefficient of 0.74.

Another useful technique is discriminant analysis (BMDP7M). This technique is used to find the function which is able to best separate samples into pre-set groups by maximizing inter-group distances while minimizing intra-group distances. Discrimination of the 10 groups of tobaccos (Figure 8) was achieved with 100% correct classification. Jackknifing, a procedure which uses one sample at a time as a test case and reclassifies using all other cases, showed 87% correct classification. The resulting discriminant plot showed that separation along function one was based primarily on tobacco type. Bright received a high positive score, burley was centered near zero and oriental received a negative score. Function 2 appeared to separate by casing.

Sensory. Multidimensional scaling was used in this study to evaluate the subjective differences between tobacco types. MDS treats data based on a person's total perception of the degree of dissimilarity between objects presented in a pairwise fashion (12). A map is produced which groups the objects by degree of similarity on several dimensions.

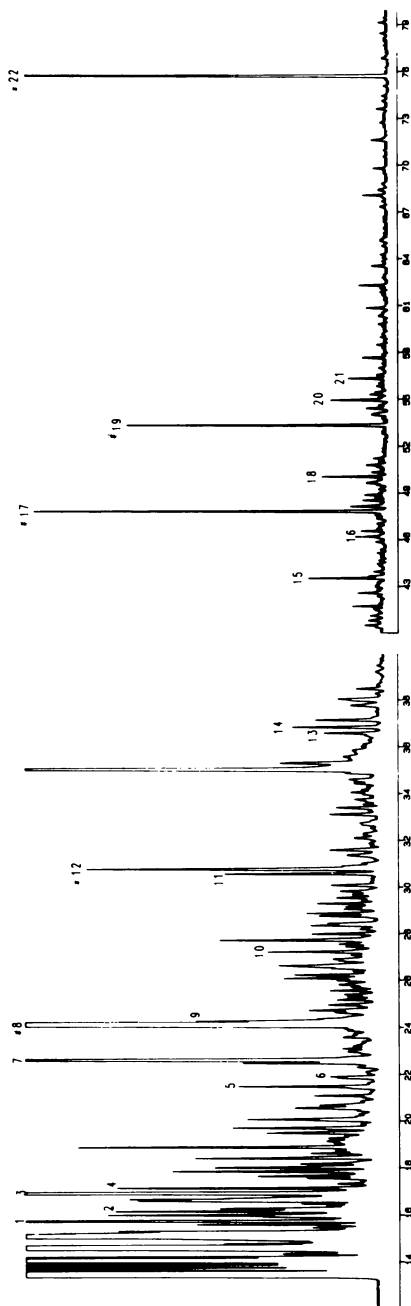


Figure 6. Chromatographic profile of tobacco headspace of the blend: 33.3% bright, 33.3% burley, and 33.3% oriental.

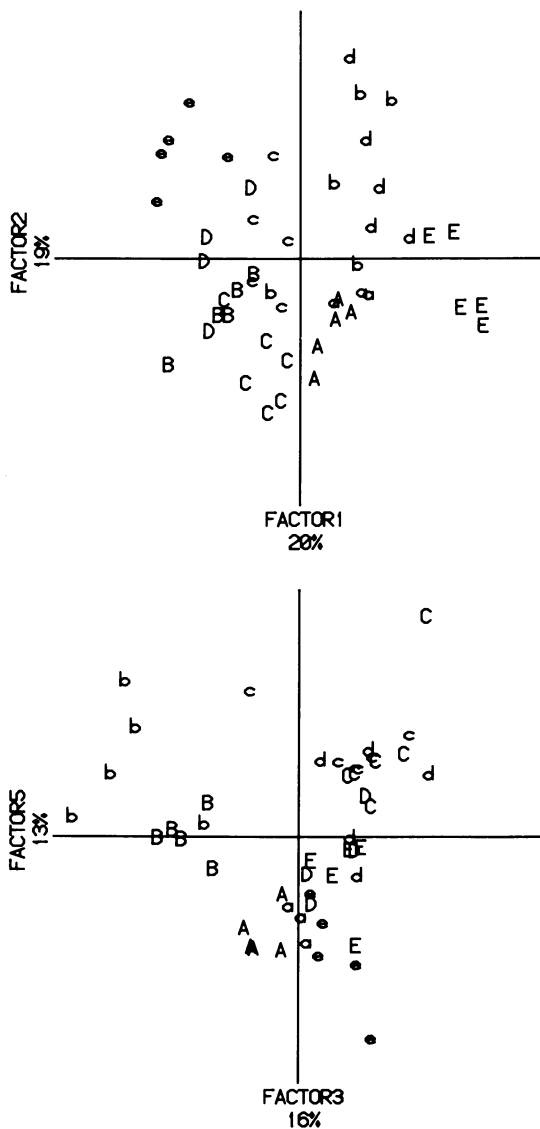


Figure 7. Factor score plots of 100% bright (A,a), 100% burley (B,b), 100% oriental (C,c), 33.3%/33.3%/33.3% (D,d), and 60%/30%/10% (E,e). Uppercase were cased. Lowercase were uncased samples. Top, factor 1 vs. factor 2; and bottom, factor 3 vs. factor 5.

The dimensionality to be used in an MDS solution can be based on several factors, two of which are stress value and ease of interpretation. Guidelines for the number of dimensions chosen is that stimuli should range from 4-6 times the number of dimensions (12). For this study, where 12 types of tobacco were compared, dimensionality should lie between 2-3. Stress is often used by finding an elbow (point at which the curve flattens) in the stress value as it approaches zero. For the 12 types of tobaccos, a solution involving 2 dimensions was chosen, based on both stress values and ease of interpretation.

Twelve panelists (5 females, 7 males) were chosen who performed well on preliminary triangle tests. Panelists were familiarized with samples before testing began by sniffing representative pairs of tobacco samples. Each sample was evaluated against every other sample in a pairwise fashion. Panelists rated samples on degree of difference using a 10 cm line. Two replications were performed. Data from the score cards were entered into the DECSYSTEM 2060 computer via a graphics tablet. Dissimilarity scores were evaluated by MDS using the statistical program ALSCAL 4 (21). Data from each panelist was separately evaluated for stability of response and similarity between panelists.

From this analysis it is clear that dimension 1, which accounted for most of the variation in the data, separated the types of tobacco based on casing (Figure 9). The panelists were more sensitive to presence of casing than to the particular type of tobacco. Uncased burley was found to be very different from cased burley, while cased bright or oriental were similar to their uncased counterpart.

Dimension 2 differentiated samples based on tobacco type. Bright and oriental were found to be most different from each other. Blend 2 (60(A), 30(B), 10(C)) was most similar to bright, while blend 1 (33.3(A), 33.3(B), 33.3(C)) lost bright character and was more similar to burley and oriental. There were more differences in tobacco type between uncased samples than those which had been cased. That is, most of the differences seen between uncased bright and burley were lost when casing was added. Tobacco types from different crop years were very similar indicating more differences due to tobacco type than crop year.

Correlation of Analytical/Sensory Results. Sensory data was correlated with headspace data of tobacco volatiles by factor analysis (BMDP4M) and canonical correlation BMDP6M. Analytical data included factor scores and discriminant analyses scores; sensory data included scores from the two MDS dimensions. Sorted rotated factor loadings of combined sensory/analytical data using factor analysis are shown in Table II. Factor one contained those variables from the analytical and sensory data which related to differences between bright (A), burley (B), and oriental (C) (Figure 10). These included dimension 1 in the

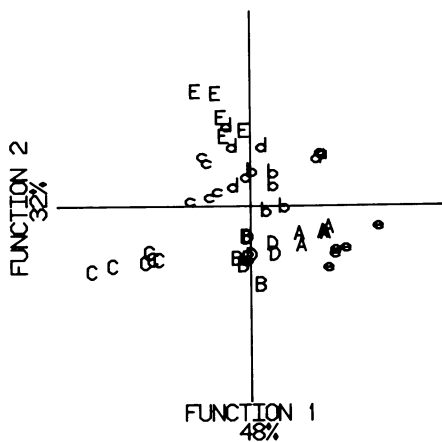


Figure 8. Discriminant function score plots. Sample codes are as in Figure 7.

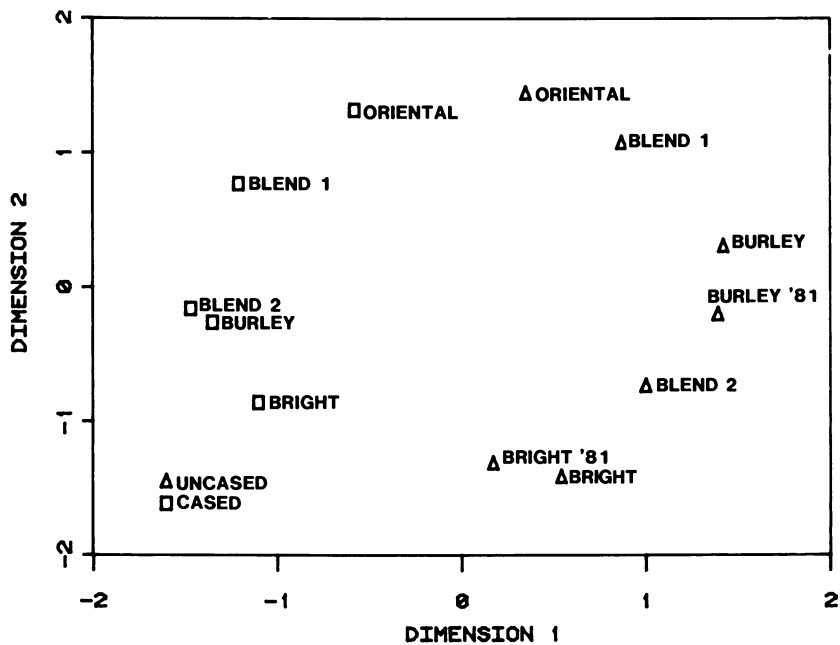


Figure 9. Multidimensional scaling score plots of tobacco smoke. '81 designates 1981 crop year; all others are 1979 crop year.

Table II. Factor Analysis of Analytical/Sensory Correlation of Headspace Data.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ANFAC5	0.960	--	--	--	--
MDSCOM2	0.937	--	--	--	--
ANFAC4	0.789	0.398	--	--	--
DISCRIM1	-0.712	0.527	-0.329	--	--
ANFAC2	--	0.883	--	--	--
MDSCOM1	--	0.836	--	--	--
ANFAC3	--	--	0.845	--	--
DESCFAC4	--	--	--	-0.637	--
DESCFAC5	--	--	0.418	0.628	--
DESCFAC1	--	--	--	0.573	--
DESCFAC3	--	--	--	--	0.913
DESCFAC2	--	-0.456	0.372	--	-0.316
DISCRIM2	--	0.495	0.252	--	--
Variance					
Explained	3.071	2.493	1.368	1.356	1.097

discriminant analysis, analytical factors 4 & 5 and MDS dimension 2. In all of these figures, oriental was most different, while bright and burley were more similar. Blend 1 (33.3/33.3/33.3) was centered close to burley while blend 2 (60/30/10) remained close to bright.

Factor 2 contained those variables which related to casing. These included MDS dimension 1 and analytical factor 2. Plots of factors 1 and 2 (Fig. 10) showed that these analytical/sensory factors contain both tobacco type and casing information.

The second data treatment was based on canonical correlation (BMDP6M) which finds combinations of two sets of data which show high correlation, i.e. what combination of analytical variables (A1-A2) shows a high correlation with another combination of sensory variables (S1-S2) (Table III). Canonical factor 1 (A1 and S1, $R^2=99\%$) contained information relating bright vs burley vs oriental as can be seen by high loadings on those

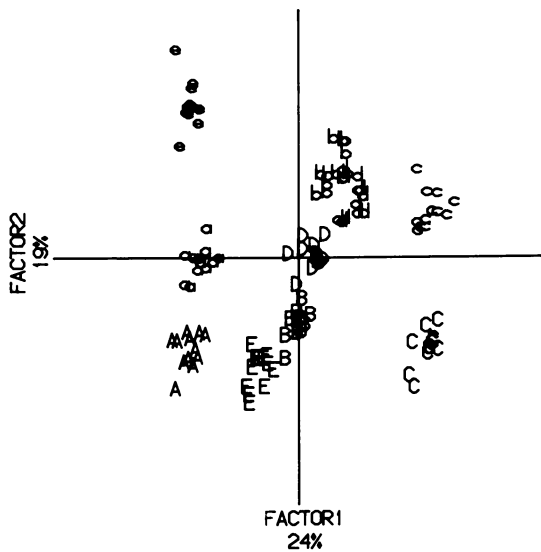


Figure 10. Factor score plots of analytical/sensory data. Labels are as in Figure 7.

analytical and sensory dimensions which discriminated between bright, burley, and oriental (analytical factors 4 & 5, discriminant dimension 1 and MDS dimension 2). Canonical factor 2 (A2

Table III. Canonical Factors from Analytical/Sensory Correlation from Headspace Data.

	Canonical Factors	
	S-1	S-2
MDSCOM1	0.451	0.886
MDSCOM2	0.913	-0.406
	A-1	A-2
ANFAC1	-0.109	0.041
ANFAC2	0.379	0.752
ANFAC3	0.256	-0.300
ANFAC4	0.802	-0.017
ANFAC5	0.845	-0.288
DISCRIM1	-0.467	0.796
DISCRIM2	0.336	0.094

and S2) correlated with an R^2 of 70%. Variables which showed high loadings in these factors correlated to casing information (analytical factor 2, MDS dimension 1).

Both factor analysis and canonical correlation techniques were successful in demonstrating that differences between tobacco type and casing could be detected from both analytical and sensory data, and that those differences found analytically were highly correlated to sensory differences. From this type of data correlation, components can be pinpointed which may be responsible for sensory differences between tobacco types.

Summary

The two research investigations reported here - the sensory quality control specification model and the application of sensory and analytical data for defining differences in tobacco aroma - both demonstrate the usefulness of multivariate analysis techniques for analyzing analytical and sensory data as well as correlating these data. Although these tasks do not compare in complexity to that of the prediction of sensory response to analytical data collected on cigarette smoke, our research to date has revealed no element which indicates that this is an impossible task. In fact, the results of these and similar

investigations (6,7,22) support the proposition that information does exist and is detectable in the analytical data set, which is necessary for the successful correlation and hence prediction of the sensory response to cigarette smoke.

Acknowledgments

The authors wish to thank those people who contributed to the collection of the data used for the analytical/sensory correlations. These include Vivian Willis, Chris Kroustalis, Danny Ennis, Tim Crews, Mary Buckner, Janet Shelton and Duane Watson. Anne Donathan is also acknowledged for her work on the preparation of this manuscript.

Literature Cited

1. Clark, H. A.; Jurs, P. C. *Anal. Chem.* 1975, 47, 374.
2. Hoff, J. T.; Herwig, W. C. *ASBC Journal*, 1975, 34, 1.
3. Saxberg, B. E. H.; Duewer, D. L.; Booker, J. L.; Kowalski, B. R. *Anal. Chim. Acta.* 1978, 103, 201.
4. Blomquist, G.; Johansson, E.; Söderström, B.; Wold, S. *J. Chromatogr.* 1979, 173, 19.
5. McConnell, M. L.; Rhodes, G.; Watson U.; Novotný, M. *J. Chromatogr.* 1979, 162, 495.
6. Parrish, M. E.; Good, B. W.; Jeltema, M. A.; Hsu, F. S. *Anal. Chem. Acta.* 1983, 150, 163.
7. Hsu, F. S.; Good, B. W.; Parrish, M. E.; Clabo, D. A. Jr.; Crews, T. D. *J. High Resoln. Chromatogr. Chromatogr. Commun.* 1982, 5, 658.
8. Ennis, D. M.; Boelens, H.; Haring H.; Bowman, P. *Food Tech.* 1982, 36(11), 83.
9. Powers, J. J. in *Food Flavours, Part A.*; Morton, I.D.; MacLead, A. J., Eds.; Elsevier Scientific Publishing Company: Amsterdam, 1982; pp 121-168.
10. Schiffman, S. S.; Reynolds, M. L.; Young, F. W. *"Introduction to Multidimensional Scaling"*; Academic Press, Inc.: New York, 1981.
11. Williams, A. A. in *"Progress in Flavour Research"*; Land, D. G.; Nursten, H. E., Eds.; Applied Science Publishers LTD: London, 1979; pp 287-305.
12. Pangborn, R. M. in *"Flavour '81"*; Schreier, P., Ed.; Walter de Gruyter & Co.: Berlin, 1981; pp 3-32.
13. Kwan W.; Kowalski, B. R. *J. Food Sci.* 1980, 45, 213.
14. Clapperton, J. F.; in *"Progress in Flavour Research"*; Land, D. G.; Nursten, H. E., Eds.; Applied Science Publishers LTD: London, 1979; pp 1-14.
15. Powers, J. J. *"Sensory Evaluation Methods for the Practicing Food Technologist"*; Johnston, M. R., Ed.; Institute of Food Technologists: Chicago, 1979; p (9) 1-17.
16. Warner, K.; Evans, C. D.; List, G. R.; Dupuy, H. P.; Wads-

- worth, J. I.; Goheen, G. E. J. Am. Oil Chem. Soc. 1978, 55, 252.
17. Schmitt D. J.; Hoff, J. T. J. Food Sci. 1979, 44, 901.
 18. Trant, A. S.; Pangborn, R. M.; Little, A. C. J. Food Sci. 1981, 46, 583.
 19. Amerine, M. A.; Pangborn, R. M.; Roessler, E. B. "Principles of Sensory Evaluation of Food"; Acad. Press N.Y.: 1965; pp 3-32.
 20. "BMDP Statistical Software", W. J. Dixon, Ed., University of California Press, 1981.
 21. Young, F. "ALSCAL 4"; University of North Carolina: Chapel Hill, N.C., 1977.
 22. Jeltama, M. A.; Good, B. W.; Kroustalis, C. S.; Ennis, D. M.; Watson, D. C.; Willis, V. E. Presented at 43rd Annual Meeting of the Institute of Food Technologists, Abstract #134, New Orleans, June 20, 1983.

RECEIVED April 6, 1984

Analytical Flavor Data

Enhancement with Computer Techniques

RICHARD A. DEPALMA

The Procter & Gamble Company, Winton Hill Technical Center, 6071 Center Hill Road, Cincinnati, OH 45224

The joining of flavor analysis by capillary gas chromatography with computerized data handling techniques is a natural combination of information generators with information analyzers. The chromatographic analyses of flavor extracts from food matrices produces hundreds of discrete peaks. A dedicated data system is needed to process this data, and a programmable system is needed to develop software to extract the maximum information from the flavor data. The experience of our laboratory, with the Hewlett-Packard 3357 Laboratory Automation System, will be discussed. The chromatographic/software applications are divided into four categories: standard chromatographic software, display capabilities, file searching, and multivariate analysis.

Computers are pervading all areas of our life and, in this same way, they are pervading all areas of chemical research. In flavor analysis, computers with their information analysis capabilities are a natural ally with separation science which is an information generator. Our laboratory has extensive experience with the Hewlett-Packard 3357 Laboratory Automation System. The specific hardware which we have is given in Table I.

Table I. Hewlett-Packard 3357 Laboratory Automation System

E Series Central Processor
85 Mbyte Fixed Disc
320 Kwords Memory
Magnetic Tape Unit
4 Graphics Terminals
3 Non-Graphics Terminals
Plotter

0097-6156/84/0261-0131\$06.00/0

© 1984 American Chemical Society

The system is actually composed of two sets of software: the RTE-6/VM operating system software which provides all the features which you expect from a real time multi-tasking operating system (high level language programming, disc file access and maintenance, utilities, etc.), and the Laboratory Automation Software which collects data from chromatographic-type instruments, integrates the data, stores the data on the disc, and displays the results in standard report form.

Before going on, the reasons for the acquisition of such a large (expensive) data system needs to be addressed. Experience in our laboratory with chromatographic data has shown that a data system needs the attributes in Table II.

Table II. Data System Requirements

1. Turnkey chromatography capabilities.
 2. Ability to take data from several types of instrumentation (capillary GC, packed column GC, LC, AutoAnalyzer).
 3. Ability to take data from instruments for several vendors.
 4. Ability to take data from 15 instruments and have capacity to grow. (Instruments located hundreds of feet apart.)
 5. On-line storage of chromatographic data, raw data and integrated data.
 6. Ability to program in high level languages to a) display the chromatograms, and b) perform statistical analysis on chromatographic data.
-

At the time of purchase, the HP data system was the only one which met all criteria and our experience with it has not been disappointing. The major advantage of such a system is that each user has the full power of the system available to them at all times, when they are at any terminal, regardless of the number of other users and demands on the system (although the system does slow down when the number of users increases). This availability of computational power is the most significant learning from our experience.

The user software is divided into two sets. The turnkey chromatographic software and in-house software which plots chromatograms and performs statistical analysis. The turnkey software provides data acquisition and integration which leads to a report as shown in Figure 1. The report includes all the standard items of retention time, peak area, identification, etc.

In addition, a BASIC program reformats the report into a form which is directly sent to the submitter (Figure 2). FORTRAN programs can also be used to reformat the data or perform other post-run analysis of the data, such as (a) column resolution checks, and (b) secondary standard analysis.

```

REPORT: 758.21 CHANNEL: 2          FATTY ACID COMPOSITION
SAMPLE: W          INJECTED AT 8:55:42 ON JUL 22, 1983
NORM METHOD: FAC   SEQ: *SEQ02 SUBSQ/SAMP: 1/ 1
SL-WIDTH  MV/MIN  DELAY  MIN-AR  BUNCH
.500      .010    1.00   50      AUTO
SUP-UNK   DVT     ID-LVL  REF-RTW  %RTW  %DIL-F  ISO
NO        0.00   50      .30     5.0    100.00  YES

ACTUAL RUN TIME: 18.858 MINUTES
RUN ABORTED

RT  ITM  FACTOR  AREA  AREA %  NAME
2.20 2.21 .94700  75 BB .075 C14
3.18 3.18# 1.00000  9788 BB 10.281 #C16
4.78 4.79 1.05000  5478 BV 6.042 C18
5.59 5.57 1.04700  34527 UV 37.973 C18-1
6.80 6.71 1.04400  36413 UV 39.933 C18-2
7.30 7.28 1.10000  394 UV .455 C20
8.27 8.22 1.04000  4311 UB 4.710 C18-3
9.37 1.00000  67 BB .070
11.31 11.30 1.15000  381 BB .461 C22

TOTAL AREA = 91434 TOTAL AREA % = 100.000
PROCESSED DATA FILE: *PRC02 RAW DATA FILE: *RAW02
    
```

Figure 1. Standard report from HP3357 system.

*** FATTY ACID COMPOSITION REPORT ***

SAMPLE RUN AT 8:55:42 ON JUL 22, 1983

SUBMITTER:-----
 SAMPLE CODE: W
 METHOD NO. GC-001-0781

LOCATION:-----
 SAMPLE IDENTITY:-----
 INSTRUMENT NO. WH-9217-----

COMPONENT	%	IODINE VALUE/(CALCULATED ON A METHYL ESTER BASIS)
C14	.1	
C16	10.3	
C18	6	
C18-1	38	
C18-2	39.9	
C20	.5	
C18-3	4.7	
C22	.5	
		113.3

ANALYST:-----

Figure 2. Reformatted report.

Graphics

In the area of graphics, the ability to store the entire digitized chromatogram has several advantages. Normally, the chromatogram is displayed on the scale in Figure 3 with most of the peaks off scale. But, since the chromatogram is stored on the disc, it can be called up and re-displayed to the users' own set of scale factors and time window. Figure 3 is a FID trace of fatty acid methyl esters separated on a Silar 10C capillary column which separates the methyl esters according to double bond position and geometry(1).

Figure 4 shows the same chromatogram as Figure 3, with the time window from 15-25 minutes and the scale factor set so the peaks are on scale. From this plot the chromatographic fronting, where the leading edge of the peaks does not rise as rapidly as the trailing edge, is obvious. This is not apparent in Figure 3. This fronting is an indicator of column overload.

Another area of the use of graphics is baseline display for improved quantitation. Every chromatographer has wondered how his data system was drawing the baselines. With this system, the baselines can be visualized even to the point of extreme magnification, as in Figure 5, so the analyst is convinced that the baseline is correct. If it is incorrect, the method can be modified, the data re-integrated, and the chromatogram re-displayed. This interactive process can be repeated in a matter of minutes.

Another display feature is to plot several chromatograms on the same screen. With the scaling and time window features, this is a quick way of comparing chromatograms, as in Figure 6. The chromatograms in Figure 6 represent a fatty acid methyl ester mixture and its isomerized version. The isomerization converts the double bonds to their equilibrium geometrics(2). In this case, the conversion is from predominantly *cis* to predominantly *trans*.

An extension of this is to subtract two chromatograms with the hope that subtle differences would be more apparent in the difference chromatograms as in Figure 7. This difference chromatogram is the result of subtracting the lower chromatogram from the upper chromatogram in Figure 6. Since the lower chromatogram has higher levels of *trans* unsaturation, the negative peaks in the difference chromatogram represent *trans* components and the normal peaks are *cis* components.

The difference chromatogram can then be stored for display or analysis by other programs. Also, baseline rise can be removed in this manner. Other graphic routines are available for drawing calibration curves, as in Figure 8.

Non-graphic routines are also available. These are divided into two categories: aids to chromatography and file searching.

The aids to chromatography include a) resolution calculations on chromatograms of standard mixtures to monitor column performance, b) calculation of Kovats' retention index for help in identifying peaks, and (c) multiple point calibration curves for improved quantitation. The file searching routines access two sets of data. Information (such as molecular formula, molecular weight) is stored on 3100 compounds from the Arctander data(3). This allows a quick computer search through the data which is difficult

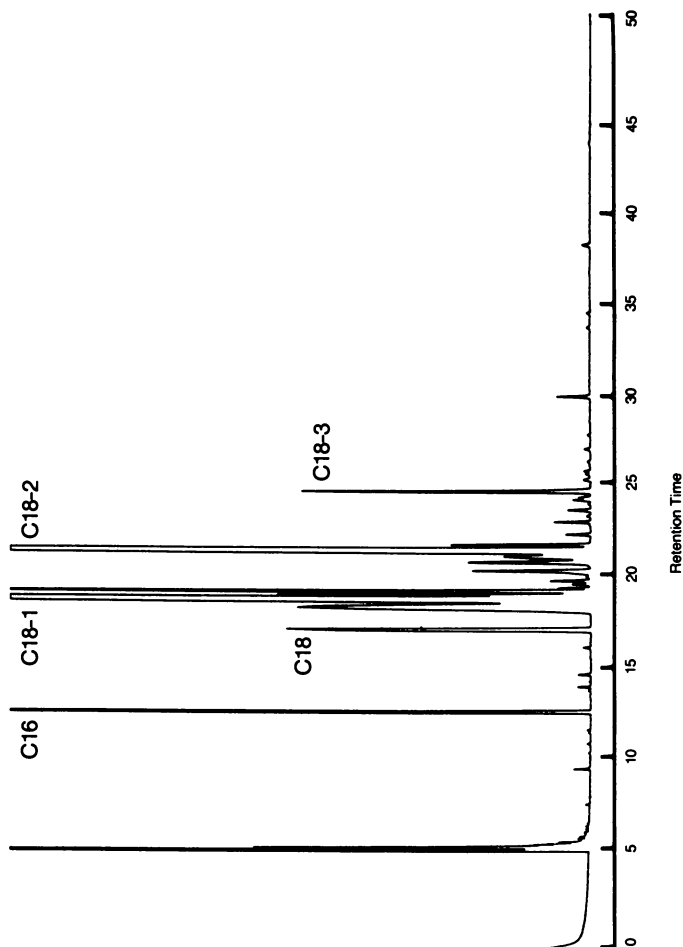


Figure 3. Chromatogram of fatty acid methyl esters separated on a Silar 10C capillary column.

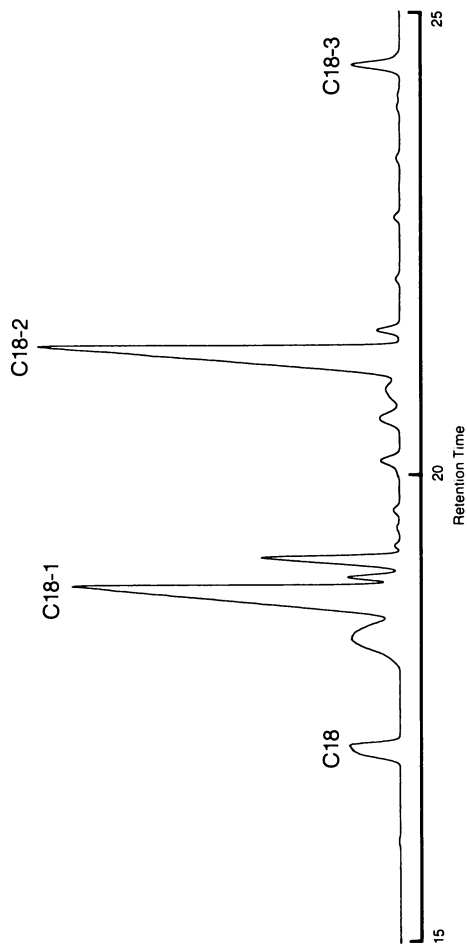


Figure 4. Replot of Figure 3 with different time window and scale factor.

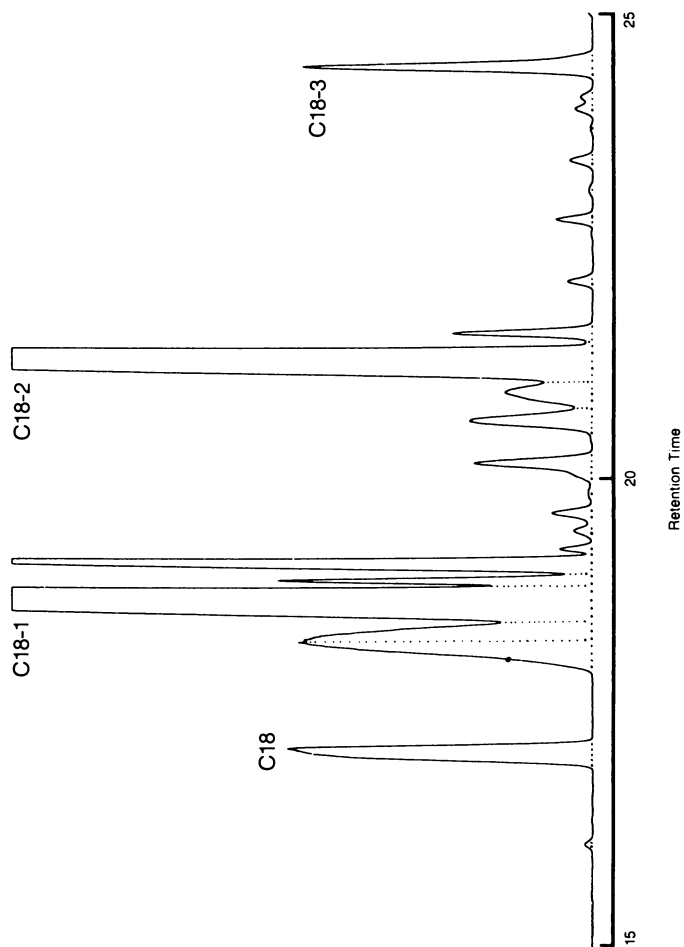


Figure 5. Replot of Figure 3 with baseline (dotted line) shown.

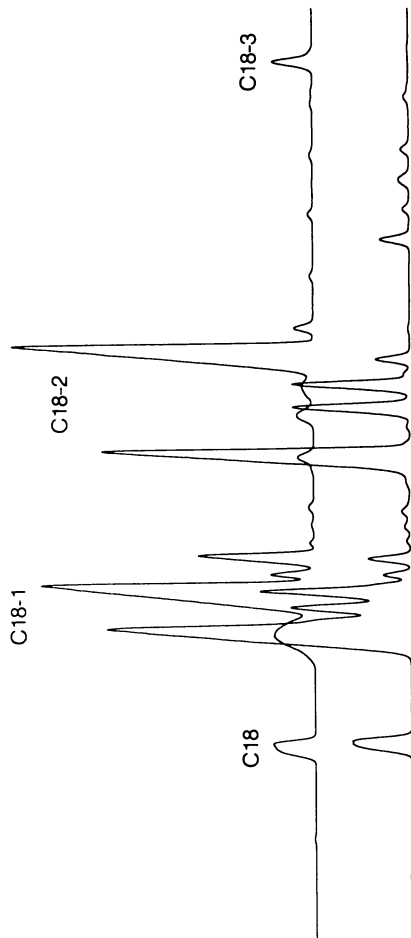


Figure 6. Upper trace, fatty acid methyl esters; and lower trace, isomerized fatty acid methyl esters.

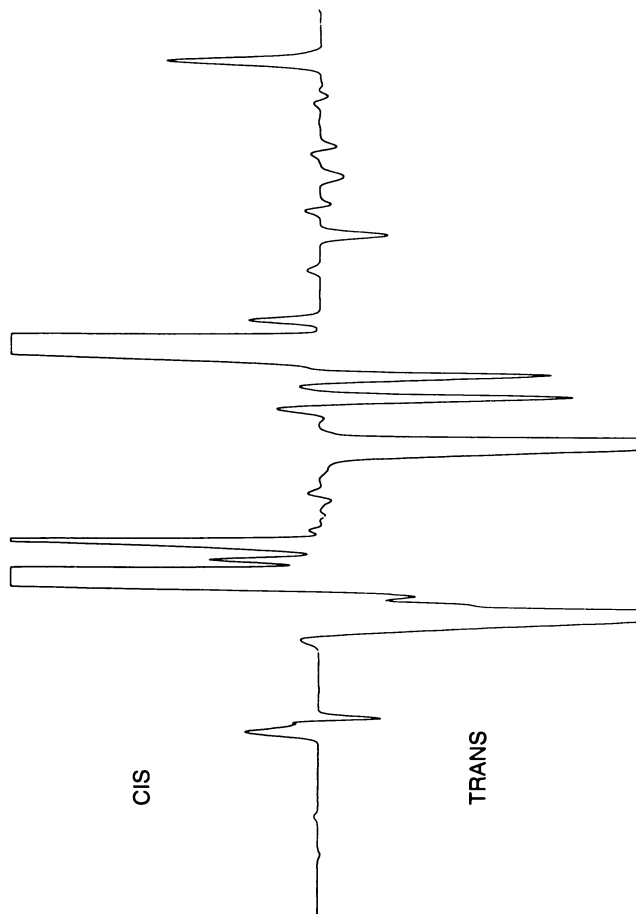


Figure 7. Difference chromatograms from Figure 6 (upper trace minus lower trace).

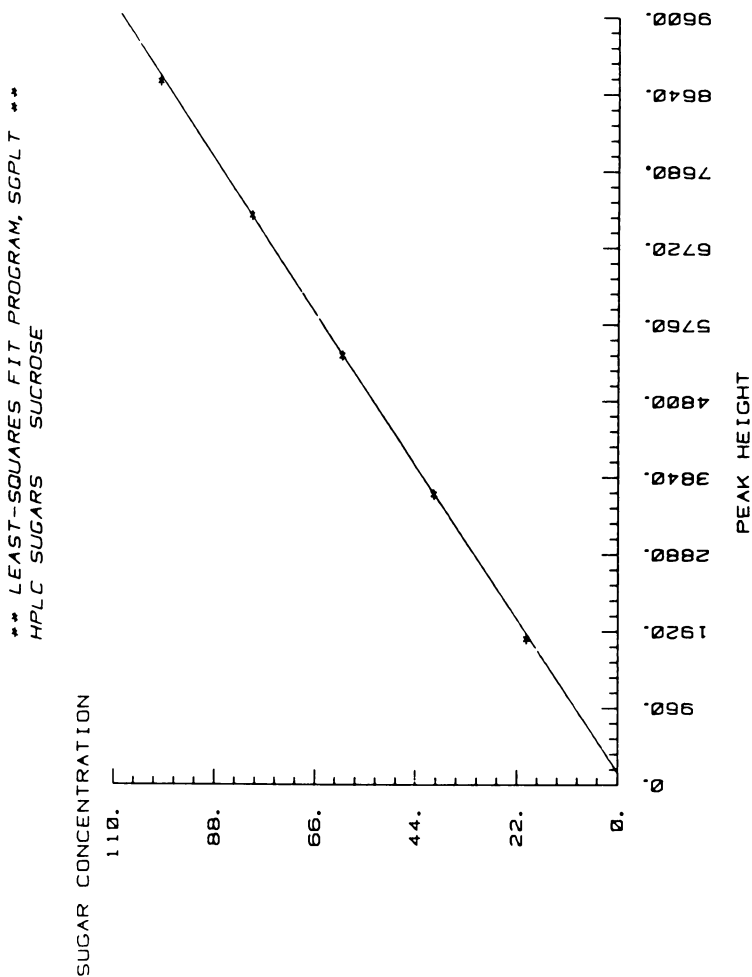


Figure 8. Example calibration curve.

to access by molecular weight since the data is only stored alphabetically in the books. Another set of files is available to users of the system to store literature references by key word. These files are created and maintained by the users, and allow for rapid search by key word for relevant article in a users' filing system.

Multivariate Analysis

Multivariate analysis (MVA) is the collection of statistical techniques which we use to relate product performance (taste panel, processing conditions) data to product composition data (e.g., ppm of extracted volatiles as measured by GC).

As an aid to better understanding MVA, consider this example. To answer the question of which volatiles are generated as a cake is baked, you need an experiment where cakes are baked for various amounts of time, and each cake is extracted and the extract is analyzed by capillary GC. For five bake times, you would have five chromatograms of approximately 200 peaks each, or 1000 peaks total. To find out which peaks are generated by the baking, you could identify every peak by mass spectrometry and then follow their fate in the baking process. But, this would be an extreme burden on the mass spectral support. As an alternative, you could correlate the peak areas with bake time and deduce that those with high correlations are generated by the baking process and then only identify those peaks with high correlations. To perform this correlation, you need to match up the corresponding peaks in each chromatogram, taking into account the slight shifts in retention time. This matched data can then be stored for additional data analysis. Once this match-up is done, the peak areas for each individual component across all chromatograms can be correlated against the product performance data, bake time in this example, to find the high correlators. Also, multiple regression, factor analysis, and cluster analysis can be performed on this data once it is matched up. This match-up procedure is the largest barrier to the routine application of MVA. Manual procedures are so laborious that they preclude the use of MVA as a routine tool in the flavor chemist's lab. The availability of computerized algorithms(5) make this match-up procedure much easier than manual procedures. And, the combination of a computerized algorithm and chromatographic data all on the same computer system, make MVA available for routine application.

The following specific example demonstrates the power of MVA. We were asked to determine which compounds were responsible for several taste attributes of a vegetable. The experiment was designed so nine food samples were analyzed by headspace gas chromatography with the purge and trap technique, and sensory data obtained on the samples. The sensory data consisted of flavor scores of the fruity character of the food. The peak areas were matched up in the nine chromatograms and the correlation performed. The nine peaks with the best individual correlation with the fruity flavor score are shown in Figure 9. This simple correlation analysis has focused further investigation on this small subset of all the peaks in the chromatogram. There is no guarantee that these

Technique: Purge and Trap
(followed by Capillary GC)

<u>Retention Time</u>	<u>Correlation Coefficient</u>
5.83	.88
7.42	.83
8.99	.82
5.34	.76
8.19	.75
13.69	.75
11.33	.75
9.69	.74
5.95	.74

Figure 9. Peaks that correlate with fruity flavor score.

compounds are responsible for the fruity character, but it certainly is a responsible starting point.

A plot of the panel scores versus peak areas for one of the components (Figure 10) shows a negative slope which means that the more of this component which is present, the lower the score. All the peaks which have high correlations have negative slopes which signify that these compounds are off-flavors.

Cluster analysis was also performed on this data. The clustering algorithm groups data according to their similarity as measured by their covariance(4). Figure 11 is the result of the clustering algorithm in dendrogram form when asked to determine the similarity between C1-C3 alcohols, aldehydes and acids based on molecular weight, boiling point and melting point. The X-axis is the degree of similarity. The closer the compounds are joined to the right of the X-axis the more similar they are. In this case, it is easy to see the acids are in a group by themselves, separate from the aldehydes and alcohols.

This same clustering algorithm was applied to the fruity character data with the input information being the peak areas from the nine samples. A portion of the dendrogram is shown in Figure 12. This cluster of similar compounds is the same group of compounds which correlate well with fruity flavors. From this clustering, we believe these compounds are somehow related, either by similar chemistries or similar generation mechanism in the fruity flavor development through natural or processing mechanisms. With this data, these peaks were identified by MS and found to be primarily a group of furans which are oxidation products, and are therefore reasonable off-flavor compounds.

Multivariate analysis is not a panacea for all flavor problems. It is a valuable tool which should be used in conjunction with other sensory and analytical skills to solve flavor problems. The availability of a programmable chromatographic data system makes implementation of MVA straightforward.

Summary

The combination of capillary gas chromatography with a sophisticated data system is a natural combination of information generators with information analyzers. Such a system is needed to enhance flavor research by (a) allowing detailed analysis of acquired chromatographic data through graphics displays, (b) storing important chemical and literature information on-line for rapid retrieval, and (c) performing sophisticated statistical analysis on the data for improved information recovery.

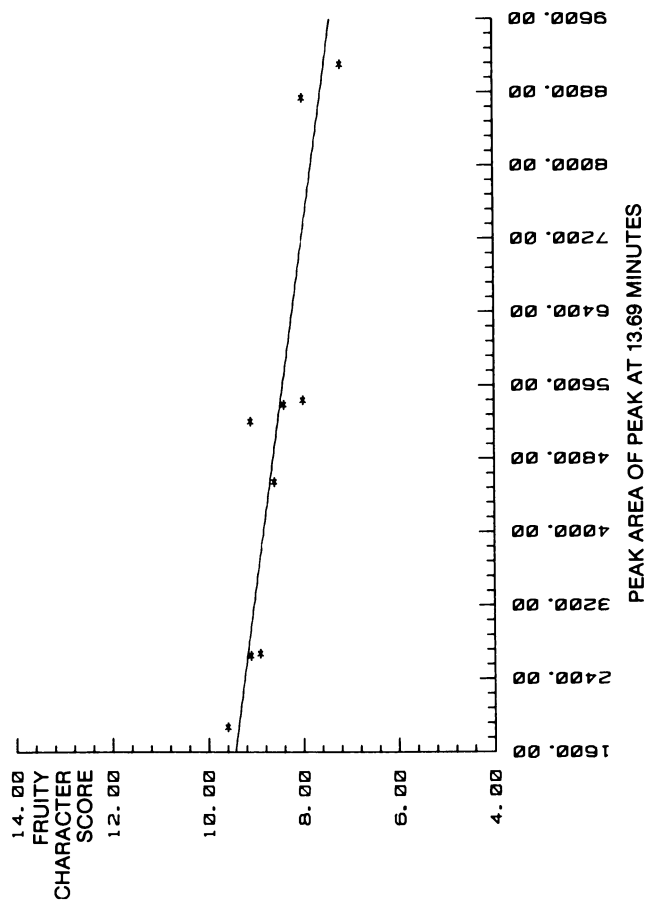


Figure 10. Plot of fruity flavor score vs. peak area.

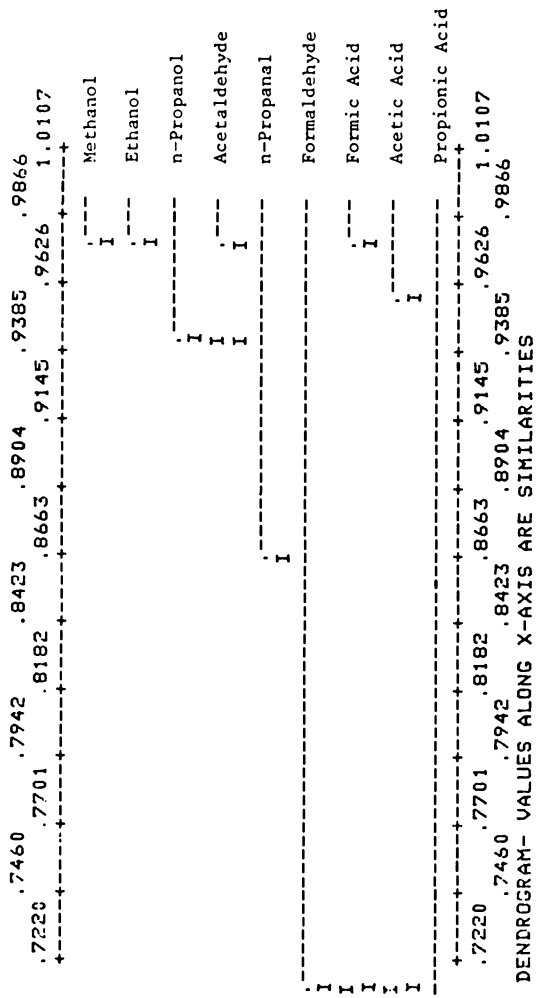


Figure 11. Example dendrogram of the similarity between alcohols, aldehydes, and acids.

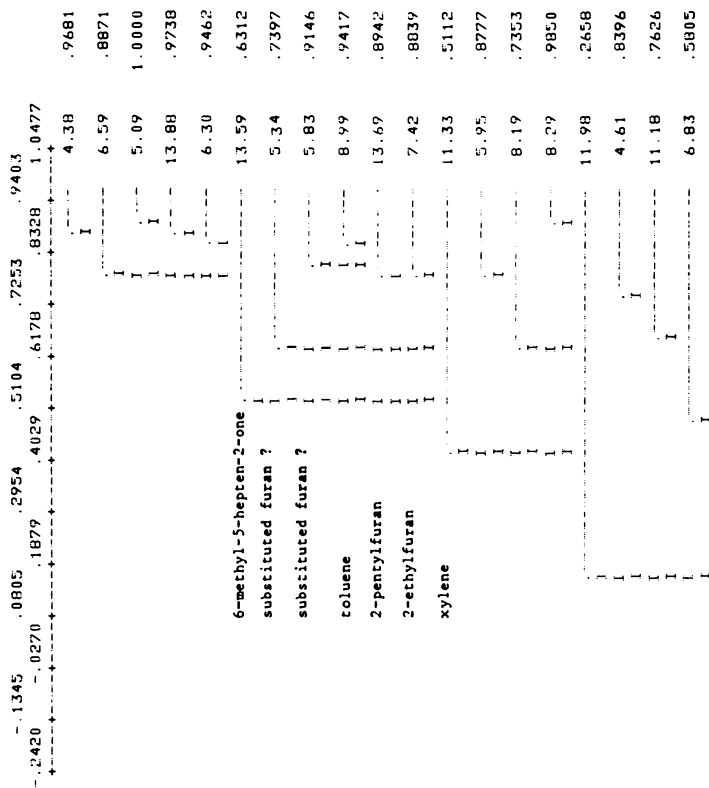


Figure 12. Dendrogram of fruity character data.

American Chemical Society Library
 1155 16th St., N.W.
 Washington, D.C. 20036

Literature Cited

1. Heckers, H.; Dittmar, K; Melcher, F. W.; Kalinowski, H. O. J. Chromatogr., 1977, 135, 93.
2. Gibson, T. W.; Strassburger, P. J. Organic Chem., 1976, 41, 791.
3. Arctander, S. "Perfume and Flavor Chemicals"; Steffen Arctander, Publisher, 1969; Vol. I and II.
4. Davis, J. C. "Statistics and Data Analysis in Geology"; John Wiley & Sons, New York, 1973; p. 467.
5. Mayfield, H. T.; Bertsch W. Computer Appl. Lab., 1983, 1, 130.

RECEIVED May 30, 1984

Experiences in Use of Robotics in an Analytical Research Laboratory

W. J. HURST, K. P. SNYDER, and R. A. MARTIN, JR.

Hershey Foods Corporation Technical Center, 1025 Reese Avenue, P.O. Box 805, Hershey, PA 17033-0805

The use of the Zymate Laboratory Automation System allows the standardization and automation of many routine operations in an analytical chemistry laboratory. It additionally allows for a closing of the analytical automation loop of sample preparation and analysis therefore potentially decreasing the need for personnel with a resultant increase in productivity. These operations include, but are not limited to, weighing, pipetting, diluting, blending, heating, liquid-solid extraction, and filtration. Since our laboratory frequently uses HPLC for the final determination step, those assays were first chosen for automation. Each procedure was subdivided into discrete laboratory unit operations for final inclusion into the Zymate program. Each of these operations was also assigned to a module such as hand, master lab station, or blender. The sequence of operations and modules was then merged to arrive at a final procedure. This final procedure was then "taught" to the robot using a series of user-defined terms which could then be coupled into a program for that sample preparation. Since many of the laboratory operations are the same for many assays, an analyst needs to define only a limited number of terms to be intermixed into a variety of programs. Examples of system layout and method uses will be given. Additionally, data is presented as to the precision available in sample preparation using laboratory robotics.

In the analytical chemistry field, there is increasing emphasis on automation of the laboratory. In many cases implementation of automation is limited to the installation of a laboratory computer which assists in laboratory management, sample flow and

0097-6156/84/0261-0149\$06.00/0
© 1984 American Chemical Society

use of any of a variety of autosamplers for analytical instruments. These autosamplers are tied to the laboratory computer for final data analysis and report generation. While these are steps in automation, there is usually human intervention of some kind involved in sample preparation, therefore leaving a void in the automation loop. The human intervention steps are labor intensive in sample preparation for both service and analytical methods development applications.

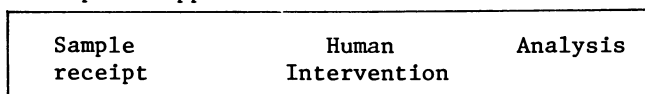


Figure 1. Analytical Sample Flow.

The introduction of laboratory robotics for sample preparation and ultimately methods development serves to close this gap. These laboratory robots bear no resemblance to C3PO and R2D2 of Star Wars fame, but rather they are complex computer controlled units specifically manufactured for use in analytical chemistry and are capable of a large number of tasks. They can be obtained commercially or can be laboratory manufactured (1,2). The initial application in our laboratory was to automate the preparation of samples for a final HPLC determination of sorbate in chocolate syrup.

The basic robot unit used in this study (Zymate, manufactured by Zymark Corporation, Hopkinton, MA) consists of a controller which controls motion through a Cathode Ray Tube and keyboard, and the robot arm which has three dimensional motion. The unit has many auxillary items of equipment including a master lab station for pipetting and diluting, a Vortex unit for sample mixing, a digital balance for automated sample weighing, a heating block and a power/event controller which serves to monitor switch closures and actuate specific equipment. Various holding racks are dispersed in strategic locations around the robot. These racks are for test tubes, pipette tips and HPLC autosampler vials.

The unit also is equipped with an RS-423 interface for data sharing between the robot and the laboratory computer. A printer serves as a means of program documentation and data review. A configuration of the system is shown in Figure 2. The robot unit has a large number of available operations. These include automatic weighing, diluting and other liquid handling, separation (partitioning), mixing, filtration, dispensing, manipulation with hands, data reduction and documentation.

The robot operates by moving to a position in three dimensional space. For one familiar with dealing in the complex methodology, the use of the robot is a humbling experience since all its operations must be described in their most basic movements. For example, if a person is moving a piece of glassware from one point to another and encounters an obstacle such as a shelf or bracket, that person knows to avoid this obstacle. Unless told

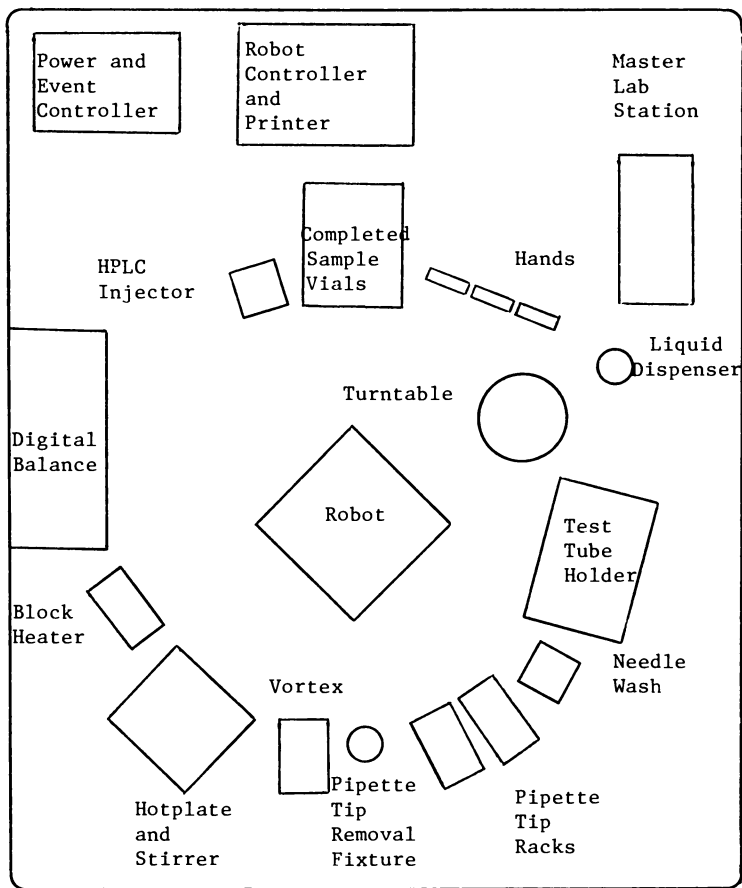


Figure 2. Laboratory Robot Organization

Publication Date: August 22, 1984 | doi: 10.1021/bk-1984-0261.ch011

otherwise, the robot moves from one point to another without avoiding the obstacle which can then result in broken glassware and wounded pride. The robot therefore must be programmed to avoid the obstacle.

The unit is equipped with devices such as remote switches to alleviate problems such as missing the pickup of a test tube or pipette tip. In a sequence of events, the robot would pick up a test tube and move to a location that has a remote switch. It would then press down the switch which would send a response to the controller. The Zymate program and the response from the switch location would indicate whether the robot had picked up the test tube. If the response was positive it would continue with the assay. If negative, it would take other actions that might be programmed into the system. These actions might include another attempt at picking up the test tube or a system shutdown. These actions thus allow safety checks on system operation or status.

Since the robot is computer controlled, all motion must be programmed. In our case the Zymate motion is dictionary driven and operator programmed. It is classified as a "lead-through" type robot in its programming (3). Commands are operator named for inclusion in the system's dictionary. Since the commands are operator named, it is helpful to develop your own applicable language to avoid confusion at a later date. After positions are "taught" and named, they are then archived on a floppy disk for further recall.

Programming is straightforward and uses a combination of the user defined dictionary, math expressions and conditional statements such as IF----THEN or GOTO. It is helpful to develop small programs for common routines such as weighing or diluting operations and then combine these into larger more complex programs. Since many of the laboratory operations are the same for many assays, the analyst needs to define a limited number of terms which can be intermixed into a variety of programs for various sample preparation sequences.

The initial application of the unit was for the automation of a sample preparation method for the determination of sorbate in chocolate syrup. Table I summarizes the data in this study. Figure 3 outlines the sample preparation scheme for this assay.

The unit also has the ability to advance or increment the various test tube and pipet tip holding racks in the programming for automatic sequential operations. Figure 4 gives a typical small program called "Pick up Tube."

The robot, like all new technology, has a steep "effort" curve and requires a substantial initial time investment. It is not a turnkey operation which when plugged is completely operational. In fact, such a concept would defeat the versatility of a laboratory robot. In this study, the time savings are seen as substantial with 10 min./sample for robot preparation versus 20 min./sample for manual preparation. The precision and accu-

- Initialize Robot and Peripherals
- Pick up Sample Container from Holder
- Tare Sample Container
- Remove Sample from Test Tube
- Dispense Sample in Container
- Weigh Sample
- Place Sample and Container Under Dispensing Nozzle
- Dispense Liquid
- Vortex
- Place Sample and Container Back in Hold
- Withdraw Aliquot
- Place in Autosampler Vial for Subsequent Analysis

Figure 3. Outline of Sample Preparation Scheme.

EasyLab Program PICK UP TUBE	
<u>EasyLab Command</u>	<u>Explanation</u>
HAND1POSS	- a "safe" starting position
RACKCLEAR1	- position far enough above the test tube rack in preparation for manipulation and hand descent
HAND150	- open fingers
DOWN13CM	- hand moves down 13 cm, fingers are around the tube
HAND50	- closes fingers to grip tube
RELO	- returns hand to RACKCLEAR1 position (up to 13 cm)
HAND1POSS	- moves hand with tube to the starting position

Figure 4. Sample Zymate Program.

racy are approximately the same for both techniques. In the robot preparation, the robot prepares the samples and loads them into autosampler vials. The final HPLC determination is identical in both manual and robot prepared assays. It should be noted that since this work was completed, Zymark has announced the availability of an HPLC auto injector module which allows the robot to actually inject each sample it has prepared and monitor the analysis to assure successful completion.

Table I. Comparison of Sample Preparation for Sorbate in Chocolate Syrup

	Robot	Manual
Time (min./sample)	10	20
Accuracy	~ 10%	~ 10%
Precision	~ 2%	~ 2%

In the area of methods development for analytical research, the robot is being used to perform accuracy and precision studies. A researcher can more completely evaluate each phase of methods development which include time dependent operations such as extractions and volume dependent operations such as partitioning or the use of bonded-phase extraction columns. It can serve as the ultimate ruggedness test for each part of a proposed method.

The benefits of the use of robotics in the analytical research laboratory include dependable accuracy and precision. Our early data indicates that accuracy and precision are equal to or greater than manual methods. Since the unit can be programmed for 24 hour operation, a laboratory can become more productive by capitalizing on these extra hours for sample preparation. Since the robot is not a turnkey operation, a great deal of versatility is afforded the researcher. In the methods development phase, multiple extraction and recovery studies using a variety of conditions can be used to optimize the final method. This can result in an accelerated methods development with ruggedness testing included. There are many tangible benefits associated with the use of analytical robotics. In a time when technical manpower is in short supply, a researcher can use a limited laboratory staff to conduct more sophisticated experiments while freeing personnel from simple repetitive tasks. This has the additional benefit of an improved working environment and a resulting increase in laboratory productivity.

In summary, the acquisition of a laboratory robotics unit allows unattended 24 hour a day operation allowing an increase in productivity by permitting scheduling of sample preparation and final instrumental analysis. It also serves to increase accuracy and precision by eliminating human sources of error.

Acknowledgments

The authors wish to thank Millie Sholly for typing the final manuscript and many drafts that are necessary in a project.

Literature Cited

1. Product Brochure, Zymark Corporation, Hopkinton, MA, 1982.
2. Owens, R. D., Eckstein, R. J., Analytical Chemistry 1982, 54, 2347-2351.
3. Dessy, R., Analytical Chemistry 1983, 55, 1100A-1114A.

RECEIVED April 6, 1984