STUDIES IN THE MASS SPECTROMETRY OF ORGANIC COMPOUNDS

A thesis presented for the degree of

DOCTOR OF PHILOSOPHY

to

The University of Glasgow

by

Donald H. Robertson

The Chemistry Department

July 1973

ACKNOWLEDGEMENTS

The author wishes to extend his thanks to Dr. R.I. Reed, who showed great patience while providing sound direction in the approach to the problem ; both in mathematics and in use of the English language.

He is likewise indebted to his laboratory colleagues, especially Dr. L. Fraser Monteiro, Dr. M. de L. Fraser Monteiro, Dr. William Laurie and Miss Maria da Silva, all of whom proved to be the ideal colleagues.

On the American scene, he would like to thank Dr. Charles Merritt, Jr. for encouraging the study programme; to John Holz and John Cavagnaro, who helped to ensure an agile mind; Richard Graham, who was responsible for the significant programming and to Chan, who remained faithful on both sides of the Atlantic.

TABLE OF CONTENTS

Abstract of Ph.D. Thesis by D.H. Robertson entitled "Studies in the Mass Spectrometry of Organic Compounds"

A general historical introduction begins the thesis; there follows a survey of the various types of mass spectrometers and their component parts.

In order to place in perspective the main work of the thesis, which is the application of data processing to mass spectrometry, a brief introduction to this field is included.

Next a review of some of the earlier efforts by R.I. Reed are discussed and the necessary background in probability and set theory is presented prior to treatment of the data processing techniques per se.

Since one of the major problems of mass spectrometry is that of the vast amounts of information which must be processed in order to achieve meaningful answers, the aim of this work has been the creation and application of methods for reducing this information to manageable proportions.

The first of the three methods utilizes a principle borrowed from the discipline of information theory which is commonly called the Khinchine entropy function. This principle allows creation of a single-valued fraction for each chemical compound under investigation. The utility of this function is discussed with illustrations of its uniqueness

The second principle comes from the field of mathematics and is referred to as the divergence function. It allows comparisons to be effected between spectra thereby expressing their similarity or dissimilarity. In practice, it has been used to advantage to resolve those cases in use of the Khinchine entropy which involve two values that are not sufficiently different to be diagnostic.

The third and most widely applicable principle involves coding of mass spectra in octally coded binary format; hence it has been called octal coding.

Use of a code based on binary representation is most appropriate for data processing with a digital computer. The

code has the special advantages of being capable of creation on-line in real-time during operation of the mass spectrometer; it also displays insensitivity to intensity factors and the presence of impurities.

Examples are shown from each principle in tabular form in the body of the thesis and supplementary data is presented in the appendix. Also included there are copies of the computer programmes and related publications by the author.

We do not have a simple event A causally connected with a simple event B, but the whole background of the system in which the events occur is included in the concept, and is a vital part of it.

Bridgman

INTRODUCTION

Mass spectroscopy has a close analogy in optical spectroscopy; separation of charged particles in an ion beam compares favourably with the frequency analysis of a light beam. The earliest instruments for ion beam analysis utilised photographic methods of recording phenomena and as such were truly spectroscopic in nature. With the development of electronic means of detection to replace the photographic ones, the more commonly used term, mass spectrometry, came into use. Although photoplate detection is employed in numerous instruments today, especially for recording high resolution spectra, the use of mass spectrometry as a descriptive term is still applied, even if for the purist, there is slight misuse of terminology.

The earliest appearance of work that may be considered related to development of mass spectrometry was in 1886 when Goldstein (1) reported the discovery of positive rays in a low pressure discharge tube. Subsequently W. Wien (2) proved conclusively that the rays were positive by observation of their deflexion in magnetic and electric fields.

In the meantime, J.J. Thomson (2a,3,4) had shown

the existence of electrons, thereby providing coherance to a theory of cathode tube behaviour. His book, (6) " Rays of Positive Electricity ", published in 1913, summarized the researches in which he used what was to be the forerunner of all mass spectrometers, the so-called parabola apparatus; with it he pioneered the mass analysis of ion beams.

Such an analyser had first been used in 1901 by Kaufman (5) as a means of studying cathode rays. In his apparatus a beam of ions of varying mass and energy passes through uniform parallel electric and magnetic fields, after which traversal the ions describe a family of parabolas upon striking a fluorescent screen or photographic plate. Each parabola corresponds to a given m/q ratio; the lengths of the parabolas are a function of energy spread in the incident ions. It was with this apparatus that Thomson proved the existence of two stable isotopes of neon.

The true founder of modern mass spectrometry was F.W. Aston, originally a student of Thomson at Cambridge. The arrangement of electrical and magnetic fields which he employed in his first design of a mass spectroscope, demonstrated a velocity focusing property. The first instrument built by Aston was capable of a resolution of 1 part in 130; later refinements in design made possible the attainment of a resolution of 1 part in 2000. Optically speaking Aston used his deflecting

plates as prisms. An important discovery made with this instrumentation was that of the isotopes of chlorine.

During the same period A. J. Dempster (8) was developing another version of the mass spectrograph which provided for separation of charged particles by varying the accelerating potential of the ions prior to their entry into a 180 degree uniform magnetic field. Such a field possesses directional focusing of the ion beam; as opposed to velocity focusing where a monoenergetic source of ions is not important, such a requirement exists here. Since detection in Dempster's instrument was electrical, it constituted historically the first example of a mass spectrometer. This design was best suited for problems of intensity measurements such as those concerned with ions produced by electron bombardment and therefore represents the prototype of the most common present-day commercial analytical instrumentation. By analogy with optics, the magnetic field is being used here as a lens.

From this point onward development of the deflexion-type instruments involved variations on a theme already established. Thus it was that Bainbridge (9) obviated the need for a monoenergetic ion source by employing a Wien filter between the ion source and the entry point of the ion beam into the magnetic field. This filter consisted of crossed co-terminous electric and magnetic fields which to-

gether provided monoenergetic ion beams to the main magnetic analyser. Addition of the Wien velocity filter to the basic Dempster design allowed a greater variety of ions to be analysed; a not unimportant result from these extended capabilities was the first experimental proof of the Einstein mass-energy relationship.

Single Focusing.

Barber (10) and Stephens (11) showed that operation of a mass spectroscope with a 180 degree magnetic sector was only a special case of focusing action for any wedged-shaped magnetic field. Nevertheless, it was not until 1940 that the tradition of using 180 degree fields was broken by A. Nier (12,13) who designed a 60 degree angular deflexion instrument; since then it has served as a prototype for many commercial units. Owing to the great effectiveness of a small sector (smaller than 180 degrees) the resulting instrument can be smaller. Perhaps of the most importance is the almost complete freedom of source and detector from mass discrimination influence from the field of the magnetic analyser. For purposes of electrical detection the fact that all masses are brought to focus on the same point is of considerable importance.

Double Focusing

In the early thirties the idea of combining velocity

and directional focusing appeared. Hence, the double-focusing instruments. Such a combination was suggested by Bartly and Dempster in 1929 (14); however, their design had a limited mass range and no commercial exploitation followed.

A certain amount of investigation into the focusing properties of magnetic and electric fields (ion optics) was a necessary prerequisite to practical development of the double-focusing instrument. Those whose investigations in the field of ion optics substantially aided this development include Hughes and Rojansky (15), Barber (16), Henneberg (17), Stephens (11), Bruche and Scherzer (18), Smythe (19), Herzog (20) and Herzog and Mattauch (21). Especially Herzog is important in this list for his studies in the derivation of general focusing equations for radial electric and/or homogeneous magnetic fields. These equations have made it readily possible to design instrumentation with a minimum of velocity focusing and a maximum of directional focusing.

In very close succession double-focusing instruments were developed by Dempster (22), Bainbridge and Jordan (23), Mattauch and Herzog (24) and Mattauch (25). During ensuing years numerous refinements and modifications have been introduced which improved the performance and range of usefulness of these instruments.

For purposes of study, double-focusing instrumentation may be classified under four headings:

1. Trochoidal Instruments (26)

Development of this design arose from the fact that 360 degree homogeneous fields cannot be used for mass analysis; the focal point in this geometry is not mass dependent, albeit perfect double-focusing is achieved. If, however, an electrostatic field is imposed perpendicular to the magnetic field, the ions are made to follow a trochoidal path. This combination of fields results in mass dispersion and perfect double-focusing in the plane normal to the magnetic field.

2. Dempster and Jordan-Bainbridge Design (23)

The basic design created by Bainbridge and Jordan (based on the original design of Dempster) utilises an electrostatic field with focusing properties corresponding to those in a semicircular magnetic field i.e. image and object images are zero. The mass scale in this spectrograph is linear and the lines produced by the ions on the photographic detection plate are sharply focused over a wide region of mass. It is double-focusing for one mass only.

3. Mattauch and Herzog Instruments (24,25)

Owing to its ability to detect all masses simultaneously by using photographic plates, this design has proven very popular as a commercial instrument. The mass scale in this instrument is quadratic, its focusing properties first order for all masses.

4. Nier-Johnson Design (27)

Peak matching techniques available in commercial versions of this instrument enable one to do precise mass measurements. A combination of 90 degree electrostatic analyser and 60 degree magnetic analyser in tandem results in first order angular focusing, second order angular focusing and first order energy focusing.

Other Types of Mass Spectrometer

Although the vast majority of mass spectral data useful for qualitative and quantitative analyses has been obtained on magnetic deflexion instruments, there are several distinct types of mass spectrometer which are in common use owing to some special characteristic which satisfies the need of the experiment.

The magnetic deflexion instruments which have already been described are static mass spectrometers; that is, there is no time-dependent factor which is fundamental to mass analysis. Essentially all other mass spectrometers may be classified as dynamic instruments in which a time-dependent parameter is essential to analysis of mass. Within this classification are four subdivisions which are convenient for further classification. These subdivisions are:
1. energy balance 2. time-of-flight 3. path stability and 4. characteristic frequency generator spectrometers.

Many of the instruments which Blauth discusses in his book on dynamic mass spectrometers (28) have mostly academic

interest since their use is not widespread and their spectra are not usually considered for standard compound identification. A brief description of the four major classifications of dynamic spectrometers and mention of the most widely used representatives of each class will suffice.

Some of the requirements of mass spectrometry research which could not be satisfactorily met by static instruments include very rapid analysis of component variation in fast reactions, residual gas analysis, precise determination of large masses as well as the additional factor of compactness and light weight which makes dynamic units so useful for space research studies.

1. energy balance - the mass spectrum is obtained by varying mass-dependent resonance conditions; separation of ions is based on energy analysis; ion energies are obtained by energy exchange with a radio frequency field.
2. time-of-flight - ions of varying mass are caused to leave the source at uniform energy or momentum, under which condition their traversal of a field free distance effects a mass dependent separation based on time of flight or time of traversal.
3. path stability - the path of ions is controlled by radio frequency fields such that selection is possible, based on velocity, phase or mass. The range of " stability " conditions can be made very narrow, thus effecting a filtering action on

the ions. If stability is made contigent upon m/q, mass analysis is possible.

4. characteristic frequency generator - oscillation frequencies characteristic of specific ion masses are measured and presented as a frequency spectrum.

Probably the most common of these dynamic types is the Time-of-Flight mass spectrometer which was pioneered by the Bendix Corporation (29,30). It was originally designed for study of fast reactions for which it is well adapted owing to the capability of rapid scanning which allows 10,000 or more spectra to be generated per second. Numerous advances have been made to the basic instrument, mostly owing to the efforts of Damoth (31). Currently the instrument is finding application in the technique of combined gas chromatography - mass spectrometry (32,33,34) where its rapid scanning capability is useful for monitoring the continuously changing composition of effluent gas from a chromatographic column.

Another important commercial dynamic mass spectrometer is based on the quadrupole design (35), which is an example of the path stability category. Features which are considered favourable in this type of instrument are fast scanning capability, high transmission rate for the ions and lack of dependence upon energy distribution of the ion beam. A more complete descriptive term is quadrupole mass filter.

Other applications of radio-frequency techniques have

been made for the purpose of mass separation by Bennett (36), and Redhead (37). The omegatron(38,39), chronotron (40,41) and mass synchrometer (42-46) are examples of instrumentation in which a combination of radio frequency and magnetic fields is utilised for mass separation.

Ion Sources

Roboz (47) cites six characteristics which are important in design of an ion source:

1. energy of the ions it produces
2. sensitivity or per cent of ions produced based on amount of neutral material provided.
3. nature of the ionic species produced.
4. background and memory effect.
5. mass discrimination.
6. ion current stability and noise.

The electron bombardment source or electron impact source is the most common design. The performance characteristics of this source have made it well suited to use with numerous types of analysers. A standard setting of 70 eV has been used in gathering the majority of recorded spectra. This source is widely used in analytical applications owing to the uniformity of the spectra which it produces; it is the pattern of the ionization and dissociation by electron bombardment which has been the basis for mass spectral identification of organic compounds. That is,

the " finger print " pattern thus produced serves to characterise spectra based on the functional groups they contain.

Historically, most present day sources (especially for analysis of organic compounds) of this type are based upon the design proposed by Nier (48). Other references to this important design are to be found in work of Bleakney (49) and Barnard (50).

Within the ion gun of such an electron bombardment source there is some degree of velocity focusing, thus allowing its use with instruments which exhibit only directional focusing properties. Sample handling and operation are easy. Depending upon specific application there are several other ion production techniques which are rather commonly employed in mass spectrometric investigations.

1. thermal ionization source (51) - a highly selective source with ionization efficiencies approaching 100% for samples within the appropriate volatility range; sample volatility is the limiting factor. It produces a spectrum composed almost exclusively of singly charged ions which show a very narrow energy spread (0.2eV). Additionally, electrical noise is minimal allowing use of electron multipliers. The chief application is to the analysis of the elements and as such the technique has low applicability to organic mass spectrometry.

2. vacuum breakdown (spark source) (52,53) whereby a potential is built up between a pair of electrodes of—

the material to be analysed; its principal merits are high sensitivity and applicability to all elements. Formation of multicharged species is a disadvantage for most applications.

3. ion bombardment source (54,55) - if the surface of a solid sample is bombarded with positive ions, sputtering or vaporisation of the surface occurs, producing positive and negative ions diagnostic of the sample. Utility of the technique is subject to doubt for analytical applications but it is useful for the study of molecular energy states.

4. field ionization (56,57) - in this technique ions are formed as the result of an intense electrostatic field set up at the end of a fine metal tip. Energy spread of the ions thus formed is narrow. Thus, single focusing analysis is satisfactory. The principal analytical feature is the fact that essentially only parent ions are formed. Recent developments suggest highly useful applications to organic structure determination. Commercial sources are now available, a fact suggesting even more extensive application to organic chemistry.

5. photoionization (58-61) - this process provides for ionization to be achieved by electromagnetic radiation of short wavelength, thereby allowing more efficient extraction of ions from the ion source. In addition, investigation of the fine structure of ionic species is possible. The principal objection to this method, namely, lack of sufficiently monochromatic radiation throughout a range of energies has been somewhat overcome owing to introduction of the laser as an ionization device.

6. ion-molecule reactions and chemical ionization. (62,63) The former, exemplified by the equation $X^+ + YH = XH^+ + Y$, which represents an abstraction of H, is the only significant ion-molecule reaction in ordinary mass spectrometry. By its use, mechanisms of higher order have been studied, also the nature of reaction intermediates and a wide range of simple organic molecules, including reactions which produce substantial amounts of negative ions. Study has been motivated by requirements in many fields including physics of the ionosphere and radiation-induced chemistry. Insomuch as peak intensities resulting from ion-molecule reactions are proportional to sample pressure it is easy to detect ions formed in this manner by observing if their intensity varies with sample concentration (pressure). A very special case of the ion-molecule reaction which is rapidly developing importance in mass spectrometric research is chemical ionization (64). In this technique, reactant ions are produced in the source (to date, methane is the most common origin of the reactant ions) from a pure material which is introduced for that purpose; these ions react with molecules of the substance the spectrum of which one chooses to obtain. There are differences between chemical ionization spectra and those obtained from electron bombardment and field ionization. As such, these spectra are an aid to molecular structure determinations and compound type analysis of organic mixtures.

Analysis by Mass Spectrometry

The widespread application of mass spectrometry today is a direct result of development work that provided highly reliable instruments capable of careful control of analysis parameters. In turn, the development itself was due in great part to the acceleration of electronics research concurrent with World War II. Initial applications were in the petroleum industry where much of the early work was carried out.

There are several assumptions upon which the use of a mass spectrometer for analytical purposes is founded:

1. each component of a mixture acts as if it were alone; i.e. ion intensities are linearly additive.
2. each pure component has a unique spectrum
3. ion beam intensities for individual components are proportional to partial pressures (concentrations) of the components.
4. mass spectral cracking patterns are reproducible under constantly maintained experimental conditions.

At first, sample inlet systems were available for handling only gaseous material. Subsequently the development of heated inlet systems (65,66) allowed the range of compounds which could be analysed to be extended, although it is obviously possible to encounter thermal decomposition among unstable (heat labile) compounds. The use of a sintered glass disc covered with gallium (65,66), which is a liquid over a range of 30-1983 degrees C, was a standard heated ~~inlet~~ design

for readily volatilised liquids during a long period of popularity.

The development and improvement of air locks (67,68), whereby the ion source may be isolated from the atmosphere by a suitable arrangement of valves, have made possible the direct insertion of a sample of low volatility. The availability of these solid inlet systems on a commercial basis has stimulated a vast area of investigation in the field of organic mass spectrometry.

Although some elaborate methods have been developed for analysis of complex mixtures, methods involving solution of simultaneous equations and matrix algebra methods (69,70), the analysis of multi-component mixtures is still not an easy task. More recently there have been some elegant methods and techniques applied to mixture analysis (71). Of these methods, probably the most important is common availability of facilities for computer processing of data.

The most significant milestone in mixture analysis has been the in line combination of gas chromatography and mass spectrometry which has had an enormous success in the analysis of multicomponent mixtures contained in small samples. Coupling the two techniques involves a combination of principles of operation which are mutually contradictory. The gas chromatograph must be operated at high pressures relative to the vacuum required in the source of the mass spectrometer. Initially, sample splitting devices were used to void to atmos-

phere all but a tiny amount of sample, thereby maintaining ion source pressure low enough for operation. More sophisticated devices in the form of molecular separators were developed by such people as Ryhage (72), Watson and Biemann (73), and Lipsky (74). They preferentially diffuse or otherwise exclude the carrier gas molecules (low molecular weight) thereby effecting a many-fold enrichment of the sample molecules. It is well known that the technique of gas chromatography can separate (assuming proper selection of parameters) complex mixtures, providing in the ideal case, mono-component elution from the chromatographic column. As each component is eluted it passes through a molecular separator of some type whereby the contaminating carrier gas is removed. The major consideration of a mass spectrometer for operation with a gas chromatograph is a rapid scanning capability owing to multiplicity and rapidity of peak elution as well as for scanning during elution of a "single " peak to determine its purity.

Theory

Interpretation of mass spectra per se has been principally based upon correlation studies whereby the nature of a compound is deduced from similarities of its spectrum with spectra of known compounds. As such, mass spectrometry has been eminently useful for a wide variety of applications. These correlations are based upon the assumption that a given molecule will fragment under controlled conditions in a reproducible manner. A deficiency in any integrated study of mass spectra is the lack

of a completely successful theory to describe the phenomena that are observed.

Historically, two theoretical approaches have been presented: both have their own validity and usefulness, neither is entirely satisfactory. The first of these is the quasi-equilibrium theory developed by Rosenstock et al. (80,81). The basic assumptions of this theory are:

1. the initial ionization process involves vertical Franck-Condon transitions within the parent molecule-ion; the latter is generally in an electronically excited state.

2. most parent molecule-ions will have low symmetry, an odd electron and a multiplicity of low-lying states, the energies of which constitute a continuum.

3. a time delay occurs before dissociation of the parent molecule-ion during which most excess electronic energy becomes randomly distributed as vibrational energy over the entire molecule.

4. rates of dissociation are related to the probabilities of random energy distribution.

5. rearrangements of the parent molecule-ion and of the fragment ions therefrom occur in a manner similar to that described in 4. i.e. random distribution.

6. decomposition of the parent molecule-ion is based on the initial energy of the ion.

Criticisms of this theory have been based on a) lack of

its applicability to large molecules b) discontinuities in its quantitative aspects and c) failure to describe the system accurately in the vicinity of the ionization threshold.

The second theoretical approach for which McLafferty (77) is chiefly responsible utilizes physical-organic theory and therein assumes analogies with organic solution chemistry; it is basically an empirical interpretation. The specific proposal is that "the unimolecular degradation reactions of the energetic ions demonstrated in the mass spectrum are similar to, and controlled through much the same energy effect as ordinary chemical reactions (78)." The means of explanation invoked are the same as those used by organic chemists to explain and predict chemical reactivities. Among these are inductive and steric effects, resonance and ion stability (79). There have been several strong advocates of this general approach, most notable of whom are Djerassi et al. (80) and Biemann (81).

The school of Djerassi has been the chief proponent of charge localization as an explanation or "rationalization" of many fragmentation patterns. It would seem that such use of charge localization as well as the assumption implicit in the McLafferty school that the shape of the parent ion as a precursor to fragmentation of the parent molecule-ion is the same as that of the original molecule are both suspect (82). However, both approaches have had high levels of success and

certainly have not been rigorously challenged to date.

The oldest technique in use for study of mass spectrometric reaction mechanisms is that of isotope labelling which allows identification of specific ions, whereby a proposed mechanism may be supported or discounted. Isotopic effects manifest themselves most markedly in the cases of carbon, silicon, sulphur, chlorine and bromine; the most commonly used isotopes are ^{2}H, ^{13}C, ^{15}N and ^{17}O. Certain problems such as specific but incomplete or non-specific deuteration make interpretation of ^{2}H substituted compounds subject to great care.

Types of Ion

The ions which may be formed in the source of a mass spectrometer are qualitatively assigned to the following categories:

1. parent or neutral ion
2. fragment ions
3. meta-stable ions
4. rearrangement ions
5. ions produced by intermolecular or ion-molecule reactions
6. multiply-charged ions
7. ions formed with excess kinetic energy
8. negative ions

Although all types of ions would be taken into consideration in a rigorous study of a mass spectrum, it is the first four categories which find most active use in conven-

tional correlation studies. Two categories of these, meta-stable and negative ions, deserve additional comments based upon the prominence of research pertaining to them. Study of the former has been accepted practice for some time. In support of arguments favouring controversial fragmentation explanations, renewed interest has been generated in meta-stable ion study. Jennings (83) has shown that it is possible to obtain a pure spectrum free from other ions in the normal spectrum. The expansion of double-focusing mass spectrometry as an analytical tool has provided further incentive for understanding these ions, owing to the possibility of observing meta-stable ions in three distinct regions of the instrument:

a) between the source exit plate and the electrostatic sector (a field-free region)

b) between the monitor and the start of the magnetic sector

c) in the region below the midpoint of the magnet toward the collector

Because of the delayed dissociation characteristic of a meta-stable ion (i.e., it dissociates outside the ion source) there is some doubt that they should be necessarily related to the normal fragmentation which occurs in the source. Nevertheless, meta-stable ion investigations have been enormously helpful in organic chemistry studies, as exemplified especially by Beynon et al. (84).

Although negative ions are generally agreed, based on investigation in the discipline of organic chemistry, to have no wide range applications, their study does promise much in explanation of ionization phenomena which occur in the source of the mass spectrometer. Extended interest in negative ions has been of recent vintage; reports have however appeared embracing topics of negative ion research such as ion-molecule reactions, double-charged ions, kinetic energy distributions and electron affinity studies. There is application of negative ions to organic chemistry in the study of hydroxylated steroids and alkaloids as well as polyfluorinated and other halogenated compounds which unlike their positive ion spectra display a molecular ion when studied via negative ions. In 1970 a book appeared which was devoted entirely to the subject of negative ions (85).

Data Processing

Manual processing of mass spectra, from whatever origin, is a long and detailed task. As interest in mass spectrometry spreads and increasing numbers of spectra become available, the need for high-speed automatic processing of these data becomes very desirable; in many cases, essential. The mass spectral information explosion in recent years has produced spectra of so many different compounds that m/q versus intensity values as a means of recording data have become unwieldy and impractical. The introduction of rapid

scanning capability for most types of instrument makes it possible to amass thousands of spectra in a few minutes or in some cases in a few seconds.

One can now send the mass spectral data signal directly to a computer for on-line processing. This on-line capability leads to the generation of dramatically increasing amounts of raw data which must be reduced and compared with tabulated spectra of known compounds for purpose of identification. There is a serious problem in determining which method of known data storage is most useful in terms of the ease with which comparison can be made with similar data for an unknown compound. Maximum benefit would accrue from a system which would be universally acceptable to all workers in the field. Great potential exists for data banks and spectral searching techniques using modern electronic equipment; however, these techniques will remain only partially successful as long as the standardization of mass spectral data is not achieved on a universal basis. Good examples of the effectiveness of high speed data processing can be found in a tabulating system for the high resolution of organic compounds and fast-scanning high resolutions studies (86,87). The main substance of this work deals with a mathematical approach to the processing of mass spectral data which was realised for the purpose of providing better use of data files through simplification of spectral representation in the file itself.

Analysis of Structure

The proliferation of mass spectrometric investigations has produced a massive amount of data, the classification of which has been recognized as an important prerequisite to its effective utilization. Techniques for storage of these data have proceded from simple tabular listings of m/q versus intensity to more sophisticated storage media such as magnetic tape and other common peripheral computer storage devices i.e. discs or drums. However, storage on these latter devices has been essentially in the same format as that utilized for tabular representation on punch cards or other hard-copy media. A perusal of the recent literature will suggest that a wide range of approaches have been tried for purposes of establishing a practical method of storage and search of such data files (88, 89, 90, 91, 92, 93 and 94).

The traditional method has been that of acquiring a large amount of data in numeric form(tables of m/q versus intensities). It was rationalized that if cracking or fragmentation patterns were obtained for a sufficient number of known compounds it would follow that any unknown could be identified. Although this approach works reasonably well and there have been numerous tabulations of this type of data, two serious defects mitigate against its adoption as a universal system of identification.

These are:

1) lack of reproducibility in the spectrum of the same

compound run on the same instrument and

2) different characteristics of spectra from instruments of different design. In neither case has there been enough investigation done to determine firmly how important these differences actually are. When one proceeds to the actual task of matching spectra, there are two queries to be made:

a) how precisely does one go about matching spectra and

b) what criteria should be established to define a good match.

The answer to the first question has been selection of varying numbers of peaks from the spectra of the known and unknown compounds which are subsequently compared mass by mass. Investigations by Kelley, Ridley et al. (95) have been carried out in this area.

However, the question of what constututes a good match is much more difficult to define in a qualitative and quantitative way. It has been the assumption of the author that one should start at the beginning, unhampered by prior assumptions and attempt to develop a statistical basis for spectral matching which would answer this question. Along the lines already suggested in this section (i.e. mathematical approach) the investigations which are subsequently presented may be categorized within the general area of information theory.

BIBLIOGRAPHY

1. E. Goldstein, Berl. Ber., 39, 691 (1996)

2. W. Wien, Ann. Physik, 65, 446 (1898)

2a. J. J. Thomson, Phil. Mag. V, 44, 293 (1897)

3. J. J. Thomson, Nature, 90, 645 and 663 (1913)

4. J. J. Thomson, Nature, 91, 333 (1913)

5. W. Kaufman, Phys. Z., 4, 55 (1903)

6. J. J. Thomson, "Rays of Positive Electricity", Longmans, Green and Co., London (1913)

7. F. W. Aston, Phil. Mag., 38, 707 (1919)

8. A. J. Dempster, Phys. Rev., 11, 316 (1918)

9. K. T. Bainbridge, Phys. Rev., 43, 103 and 1056 (1933); J. Franklin Inst., 215, 509 (1933)

10. N. F. Barber, Proc. Leeds Phil. Lit. Soc. Sci. Sec., 2, 427 (1933)

11. W. E. Stephens, Phys. Rev., 45, 513 (1934)

12. A. O. Nier, Rev. Sci. Instr., 11, 212 (1940)

13. A. O. Nier, Rev. Sci. Instr., 18, 398 (1947)

14. W. Bartky and A. J. Dempster, Phys. Rev., 33, 1019 (1929)

15. A. L. Hughes and V. Rojansky, Phys. Rev., 34, 284 (1929)

16. N. F. Barber, Proc. Leeds Phil. Lit. Soc., 2, 427 (1933)

17. W. Henneberg, Ann. Phys., Lpz., 19, 335 (1934)

18. E. Brueche and O. Scherzer, Geometrische Electronenoptik, Julius Springer, Berlin (1934)

19. W. R. Smythe, Phys. Rev., 45, 299 (1934)

20. R. F. K. Herzog, Z. Phys., 89, 447 (1934)

21. R. Herzog and J. Mattauch, Ann. Phys., Lpz., 19, 345 (1934)

22. A. J. Dempster, Nature, London, 135, 542 (1935)

23. K. T. Bainbridge and E. B. Jordan, Phys. Rev., 50, 282 (1936)

24. J. Mattauch and R. F. K. Herzog, Z. Phys., 89, 786 (1934)

25. J. Mattauch, Phys. Rev., 50, 617, 1089 (1936)

26. W. Bleakney and J. A. Hipple, Phys. Rev., 53, 521 (1938)

27. E. B. Johnson and A. O. Nier, Phys. Rev., 91, 10 (1953)

28. Eric W. Blauth, "Dynamic Mass Spectrometers", Elsevier Publishing Company, Amsterdam (1968)

29. A. E. Cameron and D. F. Eggers, Rev. Sci. Instr., 19, 2621 (1948)

30. W. B. Wiley and I. H. McLaren, Ibid., 26, 1150 (1955)

31. D. C. Damoth in "Advances in Analytical Chemistry and Instrumentation," Vol. 4, C. N. Reilley, Ed., Wiley, New York pg. 371 (1965)
Also in "Mass Spectrometry," R. I. Reed, Ed., Academic Press, New York pg. 61 (1965)

32. R. S. Gohlke, Anal. Chem. 31, 535 (1959)

33. J. A. Vollnum. I. Omura, J. Seibl, K. Grob and W. Simon, Helv. Chim. Acta, 49, 1768 (1966)

34. S. Stenhagen-Stallberg and E. Stenhagen in "Topics in Mass Spectrometry," A. L. Burlingame, Ed., Wiley Inter-Science (1967) Chapter 5

35. W. Paul and H. Steinwedel, Z. Naturforsch., 152, 143 (1958)

36. W. H. Bennett, J. Appl. Phys., 21, 143 (1950)

37. P. A. Redhead, Canad. J. Phys., 30, 1 (1952)

38. J. A. Hipple, H. Somer and H. A. Thomas, Phys. Rev., 76, 1877 (1949)

39. Ibid., 78, 332 (1950); Ibid., 82, 697 (1951)

40. S. A. Goudsmit, Phys. Rev., 74, 622 (1968)

41. E. E. Hays, P. I. Richards and S. A. Goudsmit, Phys. Rev., 84, 824 (1951)

42. L. G. Smith, Rev. Sci. Instr., 22, 115 (1951)

43. L. G. Smith Phys. Rev., 81, 295 (1951)

44. L. G. Smith, Ibid., 85, 767 (1952)

45. L. G. Smith and C. C. Damm, Phys. Rev., 90, 324 (1953)

46. Ibid., 91, 481 (1953)

47. J. Roboz, "Introduction to Mass Spectrometry - Instrumentation and Techniques," Interscience Publishers, New York, (1968)

48. A. O. Nier, Rev. Sci, Instr., 11, 212 (1940); 18, 398 (1947)

49. W. Bleakney, Phys. Rev., 34, 157 (1929)

50. G. P. Barnard, "Modern Mass Spectrometry", The Institute of Physics, London, p. 60 (1953)

51. A. J. Dempster, Phys. Rev., 11, 316 (1918)

52. A. J. Dempster, Proc. Am. Phil. Soc., 75, 755 (1935)

53. A. J. Dempster, Rev. Sci. Instr., 7, 46 (1936)

54. R. F. Herzog and F. P. Brehbock, Phys. Rev., 76, 855 (1949)

55. G. K. Werner, Advan. Electron. Electron Phys., 7, 239 (1955)

56. E. W. Muller and K. Bahadur, Phys. Rev., 102, 624 (1956)

57. R. Gomer and M. G. Inghram, J. Am. Chem. Soc., 77, 500 (1955)

58. A. Terenin and B. Popov, Phys. Z. Sowjetunion, 2, 299 (1932)

59. F. P. Lossing and I. Tanaka, J. Chem. Phys., 25, 1031 (1956)

60. R. F. Herzog and F. F. Marmo, J. Chem. Phys., 27, 1202 (1957)

61. H. Hurzeler, M. G. Inghram and J. D. Morrison, J. Chem. Phys., 27, 313 (1957)

62. D. P. Stevenson, J. Phys. Chem., 61, 1453 (1957)

63. V. Cermak and Z. Herman, Nucleonics, 19, 106 (1961)

64. F. H. Field, Accounts Chem. Resh., 1, 42 (1968)

65. M. J. O'Neal and T. P. Wier, Anal. Chem., 23, 830 (1951)

66. M. J. O'Neal, "Applied Mass Spectrometry, " The Institute of Petroleum, London, p. 27 (1954)

67. P. de Mayo and R. I. Reed, Chem. and Ind., 1481 (1956)

68. J. H. Beynon, R. A. Saunders and A. E. Williams, Appl. Spectroscopy, 17, 63 (1963)

69. P. D. Crout, Am Inst. Elec. Engrs. Trans., 60, 1235 (1941)

70. E. C. Daigle and H. A. Young, Anal. Chem., 24, 1190 (1952)

71. D. H. Smith, Anal. Chem., 44, 536 (1972)

72. R. Ryhage, Anal. Chem., 36, 1135 (1964)

73. J. T. Watson and K. Biemann, Anal. Chem., 36, 1135 (1964)

74. S. R. Lipsky, C. G. Horvath and W. J. McMurray, Ibid., 38, 1586 (1966)

75. H. M. Rosenstock, M. B. Wallenstein, A. L. Wahrhaftig and H. Eyring, Proc. Natl. Acad. Sci., U. S. 38, 667 (1952)

76. A. L. Wahrhaftig in "Advances in Mass Spectrometry", J. D. Waldron, Ed. Pg. 274, Pergamon Press, London (1959)

77. F. W. McLafferty in "Mass Spectrometry of Organic Ions", Ed. F. W. McLafferty, Academic Press, New York, Chapt. 7, Pg. (1963)

78. F. W. McLafferty, "Mass Spectrometry" in F. C. Nachod and W. D. Phillips, Eds. "Determination of Organic Structure by Physical Methods", Academic Press, New York (1962)

79. F. W. McLafferty in "Advances in Mass Spectrometry", J. D. Waldron, Ed. Pg. 355 (1959)

80. H. Budzikiewicz, C. Djerassi and D. H. Williams, "Mass Spectrometry of Organic Compounds", Holden-Day, Inc., San Francisco, pg. 9 (1967)

81. K. Biemann, "Mass Spectrometry", McGraw-Hill, New York (1962)

82. R. I. Reed and D. H. Robertson, Applied Spectroscopy, 24, No. 2, pg. 175 (1970)

83. K. R. Jennings, J. Chem. Phys., 43, 4176 (1965)

84. J. H. Beynon, "Mass Spectrometry and Its Applications to Organic Chemistry", Elsevier Publishing Co. New York (1960)

85. C. E. Melton, "Principles of Mass Spectrometry and Negative Ions", Marcel Dekker, Inc. New York (1970)

86. W. J. McMurray, B. N. Greene and S. R. Lipsky, Anal. Chem., 38, 1194 (1966)

87. D. Henneberg and K. Casper, Z. Anal. Chem., 38, 1194 (1966)

88. R. A. Hites and K. Biemann, Anal. Chem., 39, 905 (1967)

89. G. Petterson and R. Ryhage, Anal. Chem., 39, 790 (1967)

90. N. Barber and W. A. Wolstenhome, Sci. J., pg. 76 (Sept. 1967)

91. S. Abrahamson, Sci. Pool., (LKB House Journal) 14, 29 (1967)

92. L. K. Crawford and J. D. Morrison, Anal. Chem., 40, 1464, 1469 (1948)

93. T. L. Isenhour, P. C. Jurs and C. N. Reilley, Amer. Chem. Soc., 15th National Meeting, New York, Sept. 1969

94. R. G. Ridley in "Biochemical Applications of Mass Spectrometry", G. R. Waller, Ed., John Wiley, New York Chapter 6 (1972)

95. B. A. Knock, I. C. Smith, D. E. Wright, R. G. Ridley, Anal. Chem., 42, 1516 (1970)

Mathematics is a science in which we never know what we are talking about nor whether what we say is true.

Bertrand Russell

PROBABILITY AND THE GENERAL TREATMENT OF SET THEORY

Modern mathematical analyses have been applied to the interpretation of some very important problems in communication theory. As such the broad classification into which these studies might be placed is that of cybernetics.

Louis de Broglie (1) defined cybernetics as the science "of trigger effects," that is, of small scale actions involving negligible quantities of energies but unleashing appreciably larger phenomena.

It is not necessarily planned to equate the analysis of mass spectra with an exercise in cybernetics. However, de Broglie's definition readily lends itself to the approach to analysis which has been adapted for this presentation, i.e., creation of single-valued diagnostic functions which are capable of representing a large amount of data within their outward simplicity.

One of the principal branches of cybernetics can thus be considered that of information and communication. In turn, information theory itself may be defined as a branch of probability theory, concerned with the likelihood of the transmission of messages within a specified range of accuracy, considering the fact that the transmission is subject to various types of transmission and reception failure.

The basic problem in information theory is that of finding methods of coding information which make for the

most efficient use of a channel. Great sophistication has been achieved within the confines of artificial intelligence (2) which is the study of learning machines. The major feature of such a machine (usually a computer used in special mode) is its ability to improve performance based on its own experience in solving a class of problems. It is concerned thus with inductive ability and the evaluation of prior results.

Technically, the ergodic theorem * of information theory is obtained from two of the most well known theorems in probability, i.e., the individual ergodic theorem and the martingale convergence theorem. The process x (t), which represents a stochastic process is called a martingale if the expectation is finite for all t. That is, the expected value of what is to take place next (t_{n+1}) is the value of what has just occurred (t_n).

The cornerstone of information theory is the concept of redundancy, wherein a measure of the information content is expressed as entropy. For purposes of quantifying this sometimes elusive principal, the relative entropy is expressed as a ratio of what the value is to what it may be

*A random process generating x(t) is ergodic if and only if the probability associated with every stationary subensemble is either 0 or 1. Here lies the basis for a binary coding classification which has special importance in reference to computer processing of data.

at a maximum value. In this case 1-relative entropy = redundancy.

The ultimate classification has been achieved by Louis Fein (3) who objects seriously to use of the term artificial intelligence on the grounds that there is no comparable term for this field when applied to animate processors (the field which corresponds to natural intelligence has been dubbed, of course, psychology). By analogy with biota and biology he proposes nöeta and nöology as more appropriate terms for description of cognitive functions of inanimate structures such as computer, i.e., computer nöology is that branch of nöology which deals with inanimate intelligence functions.

For purposes of codifying mass spectra in a manner which renders them easier to interpret, especially when used with small-medium sized laboratory systems, two concepts, namely the Khinchin entropy function and divergence have been developed both of which treat of probabilities. Basic to both classifications is the concept of set. There thus follows a brief non-rigorous presentation of probability and its relation to the above functions.

The application of probability in the classical sense involves a frequency of "events" per se, with no consideration being assigned to existing conditions which may influence the way in which the events take place.

In approaching probability as utilized in these pages,

a consideration of types of probability will be worthwhile owing to the multiplicity of terms which are commonly used when the subject is under discussion.

Physical probability is synonomous with material probability and intrinsic probability in which propensity for an event to occur is equated with chance. There is no "a priori" consideration about the nature of the system being studied and/or its influence on the probability values.

Logical probability treats of a credible event or occurence, i.e., the likelihood of the event taking place is strengthened by the fact that the event is a logical or credible one.

A purely mathematical probability (tautological) is generally specified by definition for the circumstances or "model" with which it deals.

It is when an attempt is made to define the somewhat more difficult concept of personal or psychological probability that semantic difficulty may be encountered. This latter classification is subjective rather than objective, intuitive rather than deductive in the formal logical sense.

Application of personal probability is made in a formal manner in the discipline currently referred to as Bayesian (4) statistics. In shorthand notation, a joint probability, connected with the intersection of two events is written $P(A \cap B)$ or $P(AB) = P(A) \cdot P(B)$. Conditional probability is the

probability which is associated with the occurrence of a particular event, given the fact that another event has already occurred. This notation appears as follows:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

In the special case where the occurrence of B precludes the occurrence of A, $P(A|B) = 0$. If A and B are independent, the occurrence of B has no effect on the likelihood of A occurring:

$$P(A|B) = P(A)$$

When the two events, A and B, are dependent, joint probability is expressed by the formula:

$$P(AB) = P(A) \cdot P(B|A)$$

or

$$P(AB) = P(B) \cdot P(A|B)$$

Bayes' theorem expresses the general condition for $A_1, A_2 \ldots\ldots A_n$ being mutually exclusive and exhaustive. This formula is also called the probability of causes, a verbal statement of this theorem follows:

$$P(\text{Hypothesis}|\text{Datum}) = \frac{P(\text{Datum}|\text{Hypothesis})P(\text{Hypothesis})}{P(\text{Datum})}$$

The theory of probability deals with certain objects which possess a mathematical structure. Insomuch as the approach to analysis of mass spectral data which has been utilized for this study deals with set theory, it is considered feasible to provide a non-rigorous development of probability from set theory.

The concept of a set is frequently referred to as one of the most basic in all mathematics. This concept, namely a well-defined collection of objects, is widely applicable owing to the high prevalence of such collections of objects in the real world. Contrary to certain popular thought, mathematics is not solely concerned with numbers, thus, it is that non-numeric objects (such as sets) can be manipulated by specific rules which " in toto " form an algebra of sets that in turn leads to the subject of logic. The latter gives rise to one way of formulating certain concepts of probability. Thus, it is possible to define relations and functions in terms of sets, use of both of which is quite essential for the translation into mathematics of real problems.

The specific branch of probability with which one is concerned is information theory. Both the Khinchin entropy function and the divergence calculation, which have already been mentioned, convey a certain amount of information about the mass spectrum which they represent. Both are treated in later sections wherein examples of the information distribution are given.

Notes on Set Theory.

The basic principles of set theory are few in number. A statement of these principles is necessary as use has been made of them in the discussions on mass spectral interpretation.

A set or class may be defined as a collection of objects called the elements of the set. Specifying the elements in a set defines that set. Set theory is the study of the relations of sets to one another and to their subsets; the application of set theory provides a means of defining classes of objects in a very precise manner and of establishing those relationships which exist between various groups of objects, i.e., the ions of a mass spectrum.

All sets, regardless of what other elements they may contain, have as a member the null or empty set which is designated by the symbol $\emptyset$; this is quite different from the set which contains zero as an element.

Two sets are equal when they contain exactly the same elements or members; the order in which the elements are arranged in the two sets undergoing comparison is not important. Subset may be formally defined as any set which is contained in a given set. Any set may be a subset of itself; however, the special case of proper subset is defined as a subset which contains fewer members than the parent set. The number of possible subsets in a set is

given by 2^n, where n is equal to the number of elements in the parent set.

The universal set (the set containing all possible subsets or elements in a given category) is different for each specific set of conditions which we describe.

The complement of a given set Z, designated as Z', contains all those elements in the universal set of which Z is a member which are indeed not members of the set Z itself. The complement of the universal set U is seen to be the null set $\emptyset$ and is designated as U'. Any two sets under consideration are said to be disjoint if they do not have any elements in common.

Two further categories of sets may be defined in the following simple terminology:

1) equivalent sets - sets which contain the same number of elements although not necessarily the same ones.
2) ordered set - a set in which the elements thereof are arranged in a serial relationship based on some pre-defined rule.

These latter two categories are particularly suitable to application in the treatment of mass spectral data.

SPECIFIC BIBLIOGRAPHY

1. L. de Broglie, " New Perspectives in Physics ", Basic Books, Inc., New York , p. 52 (1962)

2. T. L. Isenhour and P. C. Jurs, Anal. Chem., 43, 20A (1971)

3. L. Fein, " Computers and Information Science ", J.T. Tou and R.H. Wilcox, Editors, Spartan Books, Inc., Washington, D.C. (1964)

4. P. Meyer, " Introductory Probability and Statistical Applications ", Addison Wesley, Reading , Mass. (1965)

GENERAL BIBLIOGRAPHY

1. A. M. Arthurs, " Probability Theory ", Routledge and Kegan Paul, Ltd. ; Dover Publications, New York (1967)

2. E. Kamke, " Theory of Sets ", Dover Publications, New York (1950)

3. A. Mizrahi and M. Sullivan, " Finite Mathematics with Applications ", John Wiley and Sons, New York (1973)

4. P. Suppes, " Axiomatic Set Theory ", Dover Publications, New York (1972)

5. J.A.Green, " Sets and Groups ", Routledge and Kegan Paul, Ltd., Dover Publications, New York (1965)

Whatever the instrument, or the compound being studied, a basic knowledge of all parameters influencing the mass spectrum is important to successful interpretation of the information (that) mass spectrometry provides.

F. W. Karasek

The Approach to Data Processing

The analysis of mass spectral data has never been an easy task, even with the advent of sophisticated data reduction and analysis techniques. Much of the difficulty in obtaining new insights lies in a reluctance to look at spectral information other than the traditional ones of mass and intensity. Certain diagnostic features of a mass spectrum such as doubly-charged ions, metastable transitions and isotope ratios provide means of distinguishing one spectrum from another. High resolution provides the determination of exact masses which can in turn lead to absolute identification of a compound.

In every instance, it is necessary to compare the analytical data for an unknown material with a catalogue of standard reference values; a satisfactory match of data leads to identification, either tentative or positive.

Particularly with development of computer technology for commonplace use in the research laboratory, the potential for automated data processing has been increased many-fold and the capability now exists for looking at these data with an eye

on the elucidation of relationships which may be too complex or obtuse to be readily deduced without extensive data reduction and processing capability.

The problem of looking at these data may be arbitrarily assigned two subdivisions:

a) relationships concerning data in a spectrum per se
b) comparison for identification purposes with a file of known data

Techniques exist for searching mass spectral files whereby all peaks in the known spectrum are compared for nearness of fit, utilising the individual peak intensities. In a typical case, tolerances are assigned for the "fit" criterion and from the catalogue of known components, the computer is made to select all those compounds which satisfy this criterion. A suitable analogy is the superposition of two nearly-identical planar figures; a percentage of non-congruence is allowed in the computer programme, thus resulting in rejection of all figures which do not coincide with the model figure within these tolerance levels.

The selection of a probability model for the experiment must consider two factors:

1) choice of a set to represent the possible outcomes
2) allocating probabilities to these possible outcomes

If conditions are held constant in the mass spectrometer there is probability for each pattern of ions (the mass spectrum) produced that the ions under consideration came from a specific arrangement of atoms in the sample molecule. In the practical

approach to the problem one searches for a probability model (mathematically defined) which most accurately approximates the real system.

For defining such a model, one needs to develop an expression. To formulate this expression, one must measure something. The availability of large amounts of tabulated mass spectral data in various formats has provided a good starting point to check out various theories of probability for large amounts of data.

Specifically, one must consider the choice of a probability model which will effect this superimposition in a mathematical way that is amenable to programming on a mini computer system.

Once a sample has been introduced into a mass spectrometer, a sequence of steps follows which involves:

1. measurement of the spectra (1 or more)
2. processing of the data in some meaningful way
3. interpretation of these data with the idea of making an identification of the sample.

In this three step summary of what happens, a great deal is assumed. Possibly the most important assumption is that the spectra which are recorded represent those of pure compounds.

There has already been extensive development of techniques which in one or more ways aid the recording and interpretation of mass spectra.

Combined gas chromatography/mass spectrometry has proven a very powerful technique for analysis of mixtures of compounds

which are readily volatilised. Owing to its applicability to a wide range of problems, this technique has been highly developed through commercial exploitation. Of course purification of mixtures is the most significant contribution made by the gas chromatograph.

Specially designed inlet systems currently make possible the recording of spectra of compounds of widely varying vapour pressure by means of ion source temperatures covering a range from -100° to +500°C.

In quite another area, at least for data processing, high resolution mass spectrometry allows determination of the empirical formulae of molecules and their fragments by means of precise mass determination.

Sophisticated computer applications have been made in high resolution mass spectrometry. Most of the successful work (for example, production of element maps) would be hopelessly time consuming without computer assistence.

There are many less genuine requirements for high resolution spectra than for the standard low resolution spectra which have been traditionally the basis of mass spectral data libraries, representing all the fragments given by a compound under electron bombardment.

Because of such an expansion in capability of mass spectrometry, much interest has been shown lately in methods of data handling, both with regard to construction of a data file and to development of search algorithms for identification of a compound

within a file.(1-15)

A logical and ultimate outcome of detailed investigations in these areas will be the establishment of computer routines which allow "ab initio" identification (3,4) of a compound on the basis of its mass spectrum alone i.e. without the restriction that the compound be in the data file. However, this area is too ill-defined at the present to merit discussion.

In the general case, data manipulation in an analytical mass spectrometry laboratory has been traditionally a matter of careful consideration; advent of the computer for use on a routine basis in processing these data has allowed great increase in the output of such laboratories.

However, the general problem of data reduction and processing continues to persist in the area of mass spectrometry. To date, there have been no clear-cut indications that any particular method of data-handling is superior. Especially intractable is the problem of searching a data file or a data bank of mass spectra for the purpose of comparing these data with an unknown spectrum. What work has been reported in general utilised iterative processes which require the use of large computer systems in order to effect the calculations in a practically finite period of time. There is no doubt that large scale processing, such as pattern classifier and artificial intelligence calculations, should continue, owing to the high likelihood that worthwhile relationships will emerge that could never be visualised from traditional data processing techniques.

Also required is some attention to the simpler techniques which demand less powerful computing capability. In terms of rendering a practical service to present day mass spectrometry, three approaches to the classification of mass spectral data which have both a theoretical and practical interest have been investigated as an outcome of the earlier work of Reed (16,17) and Reed and Robertson (18).

1) the Khinchin entropy function
2) the divergence function
3) octal coding

Application of rudimentary set theory was originally made by R.I. Reed for the purpose of giving clues to the solution of the mass spectral data processing problem within the confines of calculations which could be performed on a small laboratory computer system. For this purpose, analogies have been taken from the field of information theory, which theory may be considered as a branch of probability theory that is concerned with the likelihood of transmission of messages when the information comprising the messages are subject to varying probabilities of transmission failure. In mass spectrometry. the analogy is that of the absolute qualitative and quantitative relationship among ions formed in the ion source being distorted by the analyser and detector stages of mass spectrometer operation.

In a more advanced treatment it is reasonable to assume that the source of the spectra under consideration would become less important, especially the relative intensities of the diag-

nostic ions. Optimum use of a data bank would arise from a diagnostic system wherein simply the presence or absence of a peak in the spectrum would serve, in combination with presence or absence of other possible peaks in a spectrum, to define uniquely the compound being analysed. More will be said of this approach under the title of octal coding.

GENERAL BIBLIOGRAPHY

1. S. Abrahamsson, S. Stallberg-Stenhagen, and E. Stenhagen, Biochem. J., 92, 2 (1964)

2. B. Petersson and R. Ryhage, Arkiv. Kemi, 26, 293 (1967)

3. L.R. Crawford and J. D. Morrison, Anal. Chem. 40, 1464 (1968)

4. L.R. Crawford and J. D. Morrison, Anal. Chem. 43, 1790 (1971)

5. R.A. Hites and K. Biemann, Advan. Mass Spectrom., 4 , 37 (1968)

6. B.A. Knock, I.C.Smith,D.W.Wright,R.G.Ridley, and W.Kelly, Anal. Chem., 42, 1516 (1970)

7. F.E. Lytle and T.L. Brazie, Anal. Chem. 42, 1532 (1970)

8 S.L. Grotch, Anal. Chem., 43, 1362 (1971)

9 Ibid. 42, 1214 (1970)

10. Ibid. 45, 2 (1973)

11. L.E.Wangen, W.S.Woodward and T.L.Isenhour, Anal.Chem. 43, 1605, (1971)

12. H.S.Herz, R.A.Hites and K.Biemann, Anal. Chem., 43, 681 (1971)

13. R.G.Ridley, in " Biochemical Applications of Mass Spectrometry ", G.R.Waller, Ed., Wiley-Interscience, New York, Chap. 6, (1972)

14. D.H.Robertson and R.I.Reed, Proc. of 19th Annual Conference on Mass Spectrometry and Allied Topics, Atlanta,Ga., p.68 (1971)

15. D.H.Robertson, J.Cavagnaro, J.Holz and C. Merritt,Jr., Proc. of the 20th Annual Conference on Mass Spectrometry and Allied Topics, Dallas, Texas, p.359 (1972)

16. R.I.Reed, Rev. Port. Quim., 10, 129 (1968)

17. R.I.Reed, Rev. Port. Quim., 12, 16 (1970)

18. R.I.Reed and D.H.Robertson in " Recent Topics in Mass Spectrometry ", R.I.Reed,Ed., Gordon and Breach, New York (1971)

Mathematicians have never been in complete agreement on their science, though it is said to be the science of self-evident verities - absolute, indisputable and definitive. They have always been in controversy over the developing aspects of mathematics, and they have always considered their own age to be a period of crisis.

Henri Lebesgue

Prior to a specific treatment of the three major diagnostic functions with which the author has experimented for facilitating mass spectral file coding and search, a preliminary treatment of the antecedent investigations of R.I. Reed should be cited. (1,2)

There has already been a firm basis established in the use of mathematical analysis in the fields of communication theory and data reduction. A considerable amount of expertise has been developed in application of this kind of information to practical problems of transmission and receipt of "messages". R. I. Reed has been concerned with the possibility of developing straightforward methods of manipulating mass spectral data. The most extensive of these investigations, based on elementary set theory, has been referred to as naive analysis. (1,2) In this application, "naive" is taken to mean "non-rigorous" in the mathematical sense.

The problem of identification of an unknown by comparison with a data bank of known spectra has received considerable attention. The specific technique employed for the retrieval generally depends on whether one is working with high or low resolution data. A promising beginning has been made with respect to low resolution spectra (3);the situation with high resolution is considerably more difficult to define and will not be considered in this presentation. Most recently, the principle of artificial intelligence (4) has been applied to these low resolution analyses. In the latter, the computer is provided

with a programme which allows it to estimate the agreement of the unknown spectrum when it is compared against a so-called training set of spectra; in essence the computer is thus able to decide to which classification the unknown belongs. Inherent herein is the concept of set; that is, the concept of belonging.

Insomuch as modern mathematical techniques are highly dependent upon the use of computers for the purpose of reducing and processing gross quantities of otherwise unwieldy data, it has been considered desirable to develop methods of mass spectral analysis and/or codification which would incorporate the binary principle. Appropriate use of set theory allows this to be realised. It has already been shown that a simple basis for probability may be derived from set theory and as such give foundation to entropy, divergence and octal codification concepts as presented subsequently.

In its simplest form, naive analysis allows one to consider the ions produced in a mass spectrometer as subsets or (in the case of an individual ion) as a member of the universal set of all possible ions which may arise from the fragmentation of an organic molecule. With this basic premise, naive theory has been used successively to examine alkanes, alkenes and cycloalkanes. As the work has progressed it has been found that the simple criterion which was assumed for alkanes does not completely satisfy the latter two classifications for hydrocarbons. The problem of criteria upon which to base a mass spectral analysis has not been solved with complete satisfaction.

Since one can freely define what constitutes the membership of a set, the potential inherent in the use of naive set analysis is great. In the simplest situation which can be imagined, one could select a subset from the universal set which would embody all the ions necessary to provide a unique diagnostic for a given class of compound, i.e., set A is composed of all the ions which are required to identify unambiguously the class, hydrocarbons. The set A will likewise be subject to subdivision into further subsets, which represent, in the case of the above example such classifications as alkanes, alkenes, alkynes and cycloalkanes and cycloalkenes. It is further possible to indicate points of branching within the molecule by set manipulation. (1)

In retrospective consideration of the effectiveness of the Khinchin entropy function, the divergence calculation and octal coding in simplifying the identification of mass spectra of diverse types of compounds, especially with regard to the extreme simplicity of their calculation, the approach as presented in the formal treatment for hydrocarbons (see Appendix A) becomes cumbersome. Although well grounded in mathematics, the practical applications of this latter approach vis-à-vis computer processing of data, seems to be limited.

Reference to the reprint in Appendix A will show that the initial concept of naive analysis (through means of probability and set theory) is more theoretical in attitude than the practical utility of the techniques suggests is advisable for immediate application to in-house production problems.

The current method of mass spectrum interpretation is semi-intuitive and as such tends to be time-consuming as well as uncertain in results. Especially in consideration of developing new methods which rely heavily upon the use of computers, a systematic approach is desirable. It was with such an idea in mind that the functions presented in this section were developed.

It seems reasonable to say that the average man who is confronted with a massive data reduction task in the field of mass spectrometry has less use for philosophy than for results from his researches.

It is recognised that only a small portion of mass spectral information, most particularly in the case of a high resolution spectrum, is utilised effectively, in great part because of the time-consuming nature of data acquisition and reduction. Widespread availability of high-speed computer systems for processing analytical data such as those from mass spectrometers has encouraged development of novel and basic ways of manipulating them so that greater utility may be made of the data which is available.

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

Figure 1

Khinchin Entropy Function

The first of these three approaches which concern themselves with probabilities is the so-called Khinchin entropy function.

Let us consider an event the probability of which, based upon information available, is designated as p. The desired goal is the application of a basic numerical definition which represents the amount of information that it conveys about the event of interest. Good (5), in his treatise makes two demands on such a definition:

1) it should be a decreasing function of p and
2) the amount of information provided by two or more independent events should be the sum of their separate amounts. The events are the individual ions produced in the mass spectrometer.

These conditions are satisfied by functions of the type, -log p or -ln p. From the standpoint of information theory, the expression in figure 1 is defined as the entropy of the experiment; that is, the probability that the information transmitted by the sender is received by the receiver. In slightly different terms, it is possible to consider entropy as a representation of likelihood that a "message" will be transmitted, within certain specified ranges of accuracy, when the information is subject to certain probabilities of failure in transmission, owing to such factors as distortion and contamination with noise, either random or

periodic. This function-η is the Khinchin entropy function (6) just mentioned and named after the Russian mathematician of the same name. The equation defining the function appears in the work of C.E. Shannon on communication theory (7); 'i' is the index of message units or number of ions in the spectrum under consideration. The p_i values represent a ratio A/B where A is the individual intensity of a single ion and B is the sum of all ion intensities for a given compound.

Entropy of the experiment, as defined by Shannon (7) is without dimension and may be considered as analogous with entropy as it is customarily defined in statistical mechanics. More specifically it is the statistical mechanical equivalent of ordinary entropy divided by the Boltzman constant. Thus is being expressed a set of relative weight factors supporting the probability of the occurence of a specific collection of events.

Within the framework of consideration for this discussion the events are the ions appearing in a given mass spectrum. For these calculations the probabilities represent the ratios of individual ion currents to total ion currents. This is generally equivalent to relative intensity of each ion divided by the total ion current as a sum of individual ion intensities.

In this equation, as mentioned above, 'i' is the index of units corresponding to the individual ions and 'p' is the individual probability for each ion in the spectrum. If only one event has a probability of unity and all others are zero, the value of the entire function is zero; if all probabilities are equal,

Relative Intensities	$A/B=p_i$	$\ln p_i$	$p_i \ln p_i$
38	.1180	-2.1369	-.25215
100	.3106	-1.1694	-.36316
89	.27639	-1.2859	-.35541
68	.21111	-1.5550	-.32827
23	.07142	-2.6390	-.18847
4	.01242	-4.3882	-.05450
322 = B			-1.54196 = Σ

A = individual relative intensities for each ion
B = total of relative intensities for all ions
$-\eta$ = total of individual Khinchin functions

Khinchin function = -1.54196 (ln) or -0.66976 (log)
0.43429 x ln = $\log_{10}$

Subsequent tables of Khinchin functions are in base$_{10}$ Ratios among values are independent of base, however.

Table 2

Detailed Calculation of Khinchin Function for 1-hexene

COMPOUND	KHINCHIN FUNCTION (K)
n-butane	0.9266
2-methylpropene	0.8121
t-2-butene	1.0418
3-methyl-1,2-butadiene	1.2160
1,3,5-hexatriene	1.3396
1,5-hexadiyne	1.6053
3-heptyne	1.3791

Table 3

the function will have a maximum value. There lies in the continuum between zero and maximum value, a collection of values which has been shown to have certain diagnostic nature for a variety of organic compounds, most especially alkanes, alkenes and alkynes

A detailed example of a calculation is given in table 2, indicating how the individual probabilities vary from ion to ion. Probability of Khinchin function for the molecule as a whole is the sum of the individual values. This function, represented by the sum for all the ions, reflects a net probability for a particular collection of ion-current values to occur. In the case of figure 2 the sum is -1.54196. This sum for all the ions of a spectrum will be henceforth designated as K. In this regard, K values are diagnostic for specific ranges of probability which correlate with group classification of compounds. For these calculations, individual probabilities represent the ratios of individual ion currents to total ion current for each ion. In other words, relative intensity of each ion divided by total ion current.

Table 3 lists a group of hydrocarbons whose Khinchin functions serve to identify the compounds when comparison is made between and among members of the alkane, alkene and alkyne families of hydrocarbons.

These compounds have been selected to show the typical variation in the entropy value which is expected for the variation in degree of unsaturation in the molecule. In a

search of a file of precalculated Khinchin values, a matching index may be used to establish the correspondence of the value for an unknown compound with the library value.

When the mass spectra are converted to their corresponding entropy functions in this way, a data file can be constructed consisting of these members.

There is contained in table A a listing of K values for a large classification of organic compounds. The distribution of values is in general such that the K value (Khinchin entropy function) may be used for diagnostic purposes. Such a single-valued function is highly desirable when considering computer search of an unknown. Utilising this concept, it is only necessary to calculate a Khinchin function for an unknown compound when its "tabulated" spectrum is presented and to compare the single value thus obtained with a library file of values which have already been determined for a collection of compounds. Computer search routines under these conditions are extremely fast; in addition, a minimum amount of library storage is required; both of which are highly useful for small mini-computer systems, which currently entertain especial popularity in analytical chemistry laboratories.

In actual practice,it has been possible to utilise this function for libraries up to approximately 200 compounds and still retain unique diagnostic characteristics. When expansion is effected to extensive data libraries (e.g. The Atlas of Mass Spectral Data) the non-uniformity of recorded spectral data in

TABLE A

Compound	Khinchin Function
methane	0.4921
methane	0.4128
methane	0.5183
methane	0.4274
ethane	0.6861
ethane	0.7457
ethane	0.6729
ethane	0.7294
ethene	0.7111
ethene	0.6389
ethene	0.7579
1,2, trans-dideuterothene	0.8253
1,1, trans-dideuteroethene	0.8344
tri-D-ethene	0.8655
ethylene-D4	0.7869
tetra-deuteroethene	0.6935
ethane	0.7055
ethane	0.7418
ethane	0.7549
propane	1.0130
propane	1.0009
propane	0.9767

TABLE A cont.

Compound	Khinchin Function
propyne	0.6883
propane	0.9362
propane	0.9549
propane	0.8466
propane	0.9605
2-butyne	1.0127
1,4, dideuterobutane	1.0713
allene(propadiene)	0.6856
allene(Propadiene)	0.8663
propene	0.8837
propene	0.9626
propene	0.8539
propene	0.9707
propene	0.9574
propene-2D	0.9258
propene-1,1-D2	1.0738
propene 3,3,3,-D3	1.0942
1,2-butadiene	1.0840
1,3-butadiene	1.0106
1,3-butadiene	0.9959
1,2-butadiene	1.0036
1,3-butadiene	1.0686
trans-2-butene	1.0418

TABLE A cont.

Compound	Khinchin Function
trans-2-butene	0.9366
cis-2-butene	1.0132
2-methylpropene	1.0393
n-butane	0.9803
2-methylpropane	0.8038
n-butane	0.9460
n-butane	0.9266
n-butane	0.9769
n-butane	0.9313
n-butane	0.9520
2-methylpropane	0.8121
isobutane	0.8864
isobutane	0.8953
ethanamide	1.0808
isopropylamide	0.4278
n-propylamine	0.7511
2-deuterobutane	1.0435
n-butane-2D	0.9926
2-D-2-methylpropane	0.8626
2-D-2-methylpropane	0.9771
1-deuterobutane	1.1228
n-butane-1,1,1-D3	1.1438
ethanethiol	1.1245
n-butane-1,1,1,3,3-D5	1.2042

TABLE A cont.

Compound	Khinchin Function
n-butane-1,1,1,2,2,3,3-D7	1.2043
1,2-butadiene	1.0833
1-butene	1.0149
1-butene	0.9955
1-butene	1.0560
1-butene	1.0206
1-butene	1.0411
cic-2-butene	1.0607
1-butene	0.9836
trans-2-butene	1.0643
2-butene	1.0280
1-butene	1.0459
2-methylpropene	0.9913
cis-2-butene	1.0646
3-penten-1-yne	1.1208
2-methyl-1-buten-3-yne	1.1373
1,2-pentadiene	1.2707
2-pentyne	1.2092
1-pentyne	1.0633
2,3-pentadiene	1.2512
2-pentene	1.0864
cis-2-pentene	1.0734
1-pentene	1.0634

TABLE A cont.

Compound	Khinchin Function
3-methyl-1-butene	0.9944
cis-2-pentene	1.0922
2-methyl-2-butene	1.0002
trans-2-pentene	1.0902
1-pentene	0.9843
2-methyl-1-butene	0.9971
3-methyl-1-butene	1.0761
2-methyl-2-butene	1.1049
2-methylbutane	1.2251
cyclopentadiene	0.9651
trans-2-pentene-4-yne	1.1270
1,3-cyclopentadiene	1.0487
cyclopentene	1.0821
3-methyl-1-butyne	1.2092
2-pentyne	1.1812
2-methyl-1,3-butadiene	1.1923
2-methyl-1,3-butadiene	1.1829
1-cis-3-pentadiene	1.2319
2-pentyne	1.2637
cyclopentene	1.0719
cyclopentene	1.0751
1-pentyne	1.2075
3-methyl-1,2-butadiene	1.2160
spiropentane	1.1418

TABLE A cont.

Compound	Khinchin Function
1,4-pentadiene	1.2137
2-methyl-1,3-butadiene	1.1298
1-trans-3-pentadiene	1.2327
1-pentyne	1.2084
n-hexane	1.0510
n-hexane	1.0470
n-hexane	1.0940
2,2-dimethylbutane	1.0858
2,2-dimethylbutane	1.0317
2-methylpentane	0.9916
2-methylpentane	1.0553
2,3-dimethylbutane	0.9462
2,3-dimethylbutane	0.8689
3-methylpentane	0.9814
3-methylpentane	1.0344
1,5-heptadiene-3-yne	1.2449
1-hexyne	1.2353
2-hexyne	1.2923
1-hexyne	1.2280
3,3-dimethyl-1-butyne	1.0637
3-hexyne	1.3308
3-hexyne	1.2994
2-hexyne	1.3019

TABLE A cont.

Compound	Khinchin Function
2,3-dimethyl-1,3-butadiene	1.2647
2-methyl-1,3-pentadiene	1.2866
cis-2-hexene	1.2165
2-methyl-2-pentene	1.0875
2,3-dimethyl-2-butene	1.1051
4-methyl-cis-2-pentene	1.1051
cis-3-hexene	1.2064
3,3-dimethyl-1-butene	1.0769
4-methyl-1-trans-2-pentene	1.0931
3-methyl-cis-2-pentene	1.1834
2,3-dimethyl-1-butene	1.0518
2-methyl-1-pentene	1.1818
3-methyl-trans-2-pentene	1.1578
3-methyl-1-pentene	1.1664
2-ethyl-1-butene	1.1902
2-ethyl-1-butene	1.2014
4-methyl-1-pentene	1.0763
1-hexene	1.1661
trans-2-hexene	1.1624
trans-3-hexene	1.1760
1-hexene	1.1749
1,5-hexadiene	1.1193
2,4-hexadiyne	1.1390
1,5-hexadiyne	1.1605

TABLE A cont.

Compound	Khinchin Function
1,3,5-hexatriene	1.3396
1,5-hexadiene	1.0917
1,6-heptadiyne	1.1126
1,6-heptadiyne	1.1378
4-methyl-2-hexyne	1.3154
3-heptyne	1.3791
2-methyl-1-hexene	1.1166
2-methyl-1,5-hexadiene	1.2457
5-methyl-1-hexyne	1.3109
1,6-heptadiene	1.2136
5-methyl-2-hexyne	1.3607
2,4-dimethyl-1,3-pentadiene	1.3190
1-heptyne	1.3195
5-methyl-1-hexyne	1.3109
2-heptyne	1.3523
3-heptyne	1.3791
1-heptyne	1.3195
trans-3-heptene	1.2181
1-heptene	1.2032
4-methyl-1-hexene	1.0833
4,4-dimethyl-cis-2-pentene	1.2016
3-methyl-2-ethyl-1-butene	1.2520
2,3,3-trimethyl-1-butene	1.1390

TABLE A Cont.

Compound	Khinchin Function
2,4-dimethyl-1-pentene	1.1600
5-methyl-trans-2-hexene	1.2043
2-methyl-2-heptene	1.2008
n-nonane	0.6661
3,3-diethylpentane	0.6993
2,4-dimethyl-3-ethylpentane	0.9187
2,2,5-trimethylhexane	0.7167
2,2,4-trimethylhexane	0.7190
4,4-dimethylheptane	0.7257
3,3-dimethylheptane	0.7190
2,2-dimethylheptane	0.7934
3-ethylpentane	0.7524
2-methyloctane	0.7903
3-methyloctane	0.7560
4-methyloctane	0.7615
2,2,5,5-tetramethyl-trans-3-hexene	1.3587
2,2,5,5-tetramethyl-cis-3-hexene	1.3480
chloroethane	0.9782
1,1,1-trichloroethane	1.0573
1,1,2-trichloroethane	1.2098
5-nonanone	1.1749
2-thiapropane	1.0787

TABLE A Cont.

Compound	Khinchin Function
2-thiabutane	1.1774
2-thiapentane	1.2079
3-thiapentane	1.2357
3-methyl-2-thiabutane	1.2321
1-pentanol	1.1396
1-pentanol	1.1348
2-pentanol	0.8982
3-pentanol	0.9245
2-methyl-1-butanol	1.1770
3-methyl-1-butanol	1.1714
3-methyl-1-butanol	1.2397
3-methyl-1-butanol	1.2604
2-methyl-2-butanol	1.1983
2-methyl-2-butanol	1.0904
3-methyl-2-butanol	0.8831
2,5-hexanediol	0.9850
2-hexanelhiol	1.1472
n-butyl formate	1.1817
isobutyl formate	1.2257
sec. butyl formate	1.1980
n-propyl acetate	0.9387
ethyl propanoate	0.9462
ethyl propanoate	0.9442

TABLE A Cont.

Compound	Khinchin Function
methyl butanoate	1.1541
methyl isobutyrate	1.1533
2-hexanone	0.9632
4-methyl-2-pentanone	1.0064
2-pentanone	0.8906
3-methyl-2-butanone	0.7244
2-butanone	0.7579
1,2-dichloroethane	1.0669
3-pentanone	0.8664
cis-1,2-dichloroethane	1.0702
2-propanone	0.7892
chloromethane	0.7665
3-hexanone	1.0635
t-1,2-dichloroethene	1.0832

these libraries results in inconsistencies among Khinchin values which render them less than ideally useful for the unique identification of every compound in the library. Although not universally applicable, this approach is eminently practical with limited libraries, especially those which have been generated under the same instrumental conditions as data for the unknown compounds. A file of single valued functions, one for each compound in the library, can thus be stored in the minimum configuration of core storage for most currently produced mini-computer systems. Search algorithms are extremely simple for the comparison of an unknown with the members of the library; in total, the codification and search process is essentially trivial in nature and can be easily executed on an inexpensive mini-machine.

Summary

In this exposition, a development of set theory and probability, namely the Khinchin entropy function, is shown to provide diagnostic values when applied to classification of mass spectral data.

Because of great variation in mass spectral data from one laboratory to another, a set of values may not be definitive in every case. Such variation is well demonstrated with Khinchin values for the list of compounds in table A, the data for which came from the standard A.P.I. collection of spectra.

Current techniques in instrumentation and data processing by mini-computer allow calculation of the Khinchin function from mass and intensity data which may be acquired on-line in real-time.

In this way, individual laboratories can "construct" data files of Khinchin functions as they obtain the spectra for pure compounds. Especially with small files of compounds for specific applications, a high degree of success may be anticipated.

It was the lack of sufficient uniqueness when applied to large mass spectral data files from various sources, i.e. the standard A.P.I. data, which directed investigations to more "unique" classification methods. The next section deals with one of these investigations for developing a unique classification; namely, the divergence function. It has been used successfully to resolve ambiguities arising from those cases where Khinchin function values for a pair of compounds were too close to provide for unique identification.

NOTE:

Appendix A contains an extensive listing of Khinchin values which were made possible by the generosity of S. Grotch at the Jet Propulsion Laboratories in Pasadena, California, U.S.A. The library used is essentially the Atlas of Mass Spectral Data to which has been added a collection of spectra from the laboratory of K. Biemann at Massachusetts Institute of Technology, Cambridge, Mass. U.S.A. It was this opportunity to calculate Khinchin functions for a library of 6680 compounds that confirmed their lack of universal uniqueness for a multi-source data file.

Owing to space restrictions, only a portion of the total listing is given here.

$$J(1,2) = N_1 \sum_i^n (p_{1i} - p_i) \log_e \frac{p_{1i}}{p_i}$$

Figure 1

$$+ \quad N_2 \sum_i^n (p_{2i} - p_i) \log_e \frac{p_{2i}}{p_i}$$

$$p_i = \frac{p_{1i} + p_{2i}}{2}$$

Divergence

Also associated with the searching of a data file for establishing nearness of match between any two spectra is the calculation which is referred to as divergence. This calculation allows one to compare two spectra with a single value resulting from the calculation which indicates by its magnitude how different it is from the "reference" compound. In the comparison of two spectra, one is always chosen as the "reference" spectrum. Thus, a measure of the divergence between two spectra is calculated, using the equation in figure 1 which has found application in information theory. In this equation, J(1,2) is the divergence (8,9) between an unknown spectrum and a spectrum from a data file of known compounds; N_1 and N_2 are the total ion currents of the mass spectra in question (the general assumption is made that the sum of individual ion intensities in a normalised mass spectral tabulation is approximately equivalent to their ion currents). P_{1i} and P_{2i} represent the probabilities of the i'th ion in samples 1 and 2; P_i is defined as $\frac{1}{2}$ the sum of P_{1i} and P_{2i}. The equation comes from a general statistical approach (8) to the comparison of populations. Two mass spectra are considered as two independent random samples N_1 and N_2 .

In table 1 the divergence values resulting from comparing in seven different cases, two spectra of the same compound chosen at random from the A.P.I. files are shown. That is, a total of four methane spectra have been calculated and a total of six ethane spectra. In each instance, if the two spectra were identical, the value of J would be zero. Hence, the small values of J here reflect the slight differences present in spectra of the same compound

Compound Name	Average Divergence
methane (1)	0.06772
methane (2)	0.27666
ethane (1)	0.01339
ethane (2)	0.31431
ethane (3)	0.36150
propane	0.57211
cyclopropane	0.03265

Table 1

Compound Pair	Average Divergence
2-methyl-1-ene vs. 3-methyl-1-ene	0.95724
cyclohexane vs. hex-1-ene	1.11170
2,3-dimethylpentane vs. 2,4-dimethylpentane	1.46830
n-heptane vs. 2,4-dimethylpentane	2.61200
2-thiapentane vs. 3-thiapentane	10.23720
1-butanol vs. 2-butanol	17.44280

Table 2

because of operating conditions in the mass spectrometer on which the spectra have been obtained.

Table 2 is concerned with comparison of a group of pairs of compounds which are different in structure. The trend within the first four pairs which are hydrocarbons is for a higher value of J to reflect a greater difference in structure. The last two pair show much increased values for position isomerism in hetero-atom systems.

All compounds listed in table 3 in the comparison compound column have been compared with hexane as a "template" (i.e. the fully saturated straight chain C-6 compound). Again, the J values are indicative of the degree of difference from the compounds with which the comparison is made.

It has been found convenient to refer the calculation of divergence of a given compound in the aliphatic hydrocarbon series to that of the normal alkane of the same carbon number. Thus, in a library file of divergence values, a group of subsets is established, corresponding to the values for compounds having the same carbon number. This greatly reduces the number of values to be searched and correspondingly the time to execute the search.

In table 4 , a detailed accounting for a typical divergence calculation is given. It would seem likely from a preliminary standpoint that a consideration of these peak by peak divergence values would enable one to locate points of structural differences between any two compounds being compared.

Comparison Compound	Divergence (J)
n-hex-1-ene	4.2269
2-methylpent-2-ene	7.7805
cyclohexane	10.0238
3-hexyne	11.8302
2-methylpentene	12.5054
1-hexyne	18.7163
2-hexyne	25.8092

Table 3

Comparison of Divergence Calculations,
Relative to n-hexane as reference compound.

m/q	Intensity (A)	Intensity (B)	Individual Divergence	Accumulative Divergence
27	37.9	16.3	95.87	95.87
39	13.0	10.3	3.71	99.59
41	28.4	36.0	7.54	107.13
42	12.4	33.0	96.33	203.46
43	100.0	100.00	0.21	203.67

Table 4

Detailed divergence calculations of the pair of compounds A vs. B
A = n-butane; B= 2-methylpropane

However, time did not favour such an investigation. It was thus necessary to focus attention on the utility of divergence calculations for establishing single-valued diagnostic functions. Two tables follow which list a number of compounds for which this function has been calculated. Table A represents values which were calculated from all the peaks in a spectrum, whereas table B was calculated for the five most intense peaks only. In both cases, a range of values is obtained which are generally diagnostic for the chemical compounds which they represent. Probably the most obvious feature is that the magnitude of divergence values is greater for values of table A than those of table B. This is a function of the number of mass spectral peaks used in the calculation and should not be considered diagnostic when intercomparing tables A and B.

In the framework of implementing this calculation and the Khinchin entropy function calculation on a small computer system, the general aim of the techniques discussed in this work, it can be seen that the divergence is considerably more time consuming in so much as a full study of a library file of mass spectra implies a calculation for each member of the library against the "reference" compound or "template".

The individual calculation of divergence, although much more laborious when performed manually, does not involve appreciably more central processor time on a mini system.

Very significant use has been made of this calculation, however, to resolve situations such as those illustrated in table 5 ; namely, the case where the Khinchin values are not sufficiently

Divergence Calculations, Utilising All Peaks in Spectrum
n-heptane is the reference compound

Table A

Compound	Divergence
2-heptyne	202.205
3-heptyne	471.685
t-1,3-dimethylpentane	93.042
2-methylhexane	22.224
2-methylhexane	26.797
3-ethyl-1-pentene	32.501
ethylcyclopentene	218.361
2,3-diethylpent-2-ene	305.315
1-t-2-dimethylcyclopentene	89.364
3-methyl-t-hex-3-ene	80.311
1-c-2-dimethylcyclopentene	92.786
2,3-dimethylpentane	67.343
2,2,3-triethylbutane	69.342
2,4-dimethylpentane	5.205
2-methylhexane	22.224
2-methylhexane	33.889
2,2-dimethylpentane	54.189
2,4-dimethylpentane	6.443
n-heptane	30.319
2,2-dimethylpentane	57.838
3-methylhexane	0.142
3-methylhexane	0.545

Divergence Calculations, Utilising All Peaks in Spectrum

n-heptane is the reference compound (cont.)

Table A

Compound	Divergence
2,2,3-trimethylbutane	62.578
2,3-dimethylpentane	69.791
3-ethylpentane	50.664
3-ethylpentane	59.973
3,3-dimethylpentane	36.095
3,3-dimethylpentane	26.483
2-methyl-1-hexene	25.153
3-methyl-cis-3-hexexe	233.464
cycloheptane	55.551
1-cis-3-dimethylcyclopentane	55.575
2-methyl-2-hexene	48.450
3,4-dimethyl-cis-2-pentene	399.831
3,4-dimethyl-1-pentene	177.750
2-methyl-trans-3-hexene	89.794
1,1,2,2-tetramethylcyclopropane	155.913
3,4-dimethyl-trans-2-pentene	387.142
5-methyl-1-hexene	33.866
3,3-dimethyl-1-pentene	54.007
3-ethyl-2-pentene	246.568
4-methyl-trans-2-hexene	56.115
2,3-dimethyl-1-pentene	57.180
3-methyl-1-hexene	92.779

Divergence Calculations, Utilising All Peaks in Spectrum

n-heptane is the reference compound (Cont.)

Table A

Compound	Divergence
3-methyl-cis-2-hexene	174.482
4,4-dimethyl-1-pentene	27.504
2,4-dimethyl-2-pentene	130.658
4,4-dimethyl-trans-2-pentene	346.985
cycloheptane	60.419
trans-2-heptene	94.951
1,1-dimethylcyclopentane	194.856
2,3,3-trimethyl-1-butene	326.876
2,4-dimethyl-1-pentene	57.336
5-methyl-trans-2-hexene	131.262
4-methyl-1-hexene	11.664
4,4-dimethyl-cis-2-pentene	102.513
3-methyl-2-ethyl-1-butene	243.065
1-heptyne	52.260
1-heptyne	52.260
trans-3-heptyne	37.417
1-heptene	29.661
2,4-dimethyl-1,3-pentadiene	424.824
5-methyl-1-hexyne	49.939
1,6-heptadiene	335.286
5-methyl-2-hexyne	373.344

Divergence Calculations, Utilising All Peaks in Spectrum
n-hexane is the reference compound

Table A

Compound	Divergence
cyclohexane	10.0238
2,4-hexadiyne	26.2104
1,5-hexadiyne	29.8075
n-hex-1-ene	4.2269
2-methylpent-1-ene	12.5054
2-hexyne	25.8092
3-hexyne	11.8302
2-methylpent-2-ene	7.7805
hex-1-yne	18.7163

Divergence Calculations, Utilising All Peaks in Spectrum

n-pentane is the reference compound

Table A

Compound	Divergence
2-methylbutane	17.8871
2,2-dimethylpropane	62.2172
2-methylpentane	18.9821
n-pentane	5.8080

Divergence Calculations, Utilising All Peaks in Spectrum
n-butane is the reference compound
Table A

Compound	Divergence Value
2-methylpropane	40.6926
n-butane	15.1075
n-butane	16.8479
n-butane	15.0465
n-butane	24.9250
n-butane	14.5552
2-methylpropane	43.4982
2-methylpropane	23.5398
2-methylpropane	7.4067
2-deuterobutane	76.5472
n-butane-2D	235.1620
2-D-2-methylpropane	126.7840
2-D-2-methylpropane	134.5390
1-D-butane	72.3092

Divergence Calculations, Five Most Intense Peaks in Spectrum

Reference compound is n-butane

Table B

Compound	Divergence
n-butane	0.0000
2-methylpropane	1.1100
n-butane	0.1157
n-butane	0.5059
n-butane	0.2706
n-butane	0.0129
n-butane	0.4718
isobutane	0.1798
isobutane	0.1561
isobutane	0.9622
1-buten-3-yne	1.9753
1-buten-3-yne	1.9041
1,3-butadiene	2.6054
1,3-butadiene	2.7042
1,3-butadiene	2.2764
1,2-butadiene	0.0602
1,2-butadiene	0.1713
1,2-butadiene	0.1868
1-butyne	1.0467
2-butyne	0.2946
cyclobutene	2.0796
isobutene	0.5676

Divergence Calculations, Five Most Intense Peaks in Spectrum

Reference compound is n-butane

Table B

Compound	Divergence
1-butene	0.0629
1-butene	0.1327
1-butene	0.1279
1-butene	0.0679
1-butene	0.3907
1-butene	0.4413
trans-2-butene	0.0424
trans-2-butene	0.0488
trans-2-butene	0.0488
trans-2-butene	0.5493
cis-2-butene	0.5942
cis-2-butene	0.0561
cis-2-butene	0.0542
isobutene	0.5877
2-butene	0.3968
2-methylpropene	0.7732
1-deuterobutane	2.2662
2-deutero-2-methylpropane	1.1874
2-deuterobutane	0.1586
n-butane-2-D	1.1791
1,4-dideuterobutane	0.2897
n-butane-1,1,1-D3	1.9439

Divergence Calculations, Five Most Intense Peaks in Spectrum

Reference compound is n-pentane

Table B

Compound	Divergence
isopentane	1.4173
isopentane	1.2925
neopentane	1.5213
1,3-cyclopentadiene	0.2119
3-penten-1-yne	0.2612
2-methyl-1-buten-3-yne	0.7823
cyclopentadiene	0.3690
cyclopentane	1.3410
cyclopentane	1.3163
trans-2-penten-4-yne	0.3219
1,4-pentadiene	2.4509
1-cis-3-pentadiene	2.4987
1-trans-3-pentadiene	2.8911
2,3-pentadiene	0.70281
1,2-pentadiene	4.2457
3-methyl-1-butyne	0.2565
1-pentyne	2.7697
1-pentyne	1.3645
cyclopentene	0.5487
cyclopentene	0.5742
spiropentane	2.7470

Divergence Calculations, Five Most Intense Peaks in Spectrum

Reference Compound is n-pentane (cont.)

Table B

Compound	Divergence
2-methyl-1,3-butadiene	3.0803
2-methyl-1,3-butadiene	3.7703
2-pentyne	0.5428
2-pentyne	0.5428
2-pentyne	0.9036
methylcyclobutane	2.2872
ethylcyclopropane	0.6368
ethylcyclopropane	0.1982
1-pentene	0.0586
1-pentene	0.0377
2-methyl-2-butene	0.6841
3-methyl-1-butene	1.4291
2-methyl-1-butene	0.6635
3-methyl-1-butene	1.3904
cis-1,2-dimethylcyclopropane	0.3959
trans-1,2-dimethylcyclopropane	0.2249
2-methyl-2-butene	0.7536
cis-2-pentene	0.3258
cis-2-pentene	0.2332
trans-2-pentene	0.3007
2-pentene	0.5437
2-methyl-1-butene	0.7532

Divergence Calculations, Five Most Intense Peaks in Spectrum

Reference compound is n-propane

Table B

Compound	Divergence
propyne	1.3389
propyne	1.2555
propadiene	1.1617
allene (propadiene)	1.4852
propene	1.1209
propene	1.1066
propene	0.7896
propene-2-D	0.2613
propylene	1.0855
cyclopropane	1.1653
cyclopropane	1.4013
propane	0.0253
propane	0.0880
propane	0.0631

Compound	Entropy	Divergence
2,4-hexadiene	0.473	10.97
3-methyl-1,3-pentadiene	0.474	26.19
2-ethyl-1,3-butadiene	0.497	7.79
1,3-hexadiene	0.498	9.17

Table 5

Use of Divergence to Resolve Non-unique Khinchin Functions

In each case the divergence value is calculated using n-hexane as the reference compound.

different to be considered diagnostic.

In the scheme of computer operations, a subroutine is invoked at the point where a search of the library file for a unique Khinchin function fails. There is then performed a divergence calculation on the data just used in calculation of the Khinchin values. When dealing with a limited library for specific use, such as those already described with respect to Khinchin functions, a high degree of success was reached in resolving "ties" which resulted from Khinchin calculations.

Work on more advanced forms of the divergence calculation, i.e. intercomparison among more than two spectra is in progress in the group of R.I. Reed at Glasgow University .

GENERAL BIBLIOGRAPHY

1. R.I. Reed, Rev. Port. Quim., 10, 129 (1968)

2. R.I. Reed, Rev. Port. Quim., 12, 16 (1970)

3. " Catalogue of Mass Spectral Data ", serial # 44, Thermodynamics Research Centre Data Project, Thermodynamics Research Centre, Texas A. and M., University College Station, Texas

4. P.C. Jurs, B.R.Kowalski and T.L.Isenhour, Anal. Chem., 41, 21 (1969)

5. L.J.Good, " Probability and the Weighing of Evidence ", Charles Griffin, London (1950)

6. A.I.Khinchin, " Mathematical Foundations of Information Theory ", Dover Publications, Inc. , New York (1957)

7. C.E.Shannon, " Bell Syst. Tech. J. ", 27, 379-423; 623-656 (1948)

8. S. Kullback, " Information Theory and Statistics ", Dover Publications, Inc., New York (1968)

9. G.H.Hardy, J.E.Littlewood and G. Polya, " Inequalities ", 2d Edition, Cambridge University Press (1967)

It is always easier to explain the presence of a particular peak in a mass spectrum than to predict the main peaks to be expected in the spectrum of a particular compound and it is only when this latter problem is attempted that the naivety of our present level of understanding is fully realised.

John Beynon

Octal Coding

The initial concept of this dissertation was based on mathematical manipulation of mass spectral data with little specific consideration of the nature of the compounds whose spectra were being studied. Mathematical manipulation is a more purely academic exercise; consideration of the compound type being studied stands a somewhat better chance of being immediately useful, especially in applications which utilise on-line real-time data acquisition techniques. It was with this more practical approach in mind that a special codification scheme was created; it has been named octal coding.

A basic problem in the processing of any type of spectroscopic data by computer is that of data file management. This involves coding of data for storage in the file and technique of searching the file for purposes of finding a match among file data for an unknown spectrum. Thus, there has been conceived and tested a codification procedure for use with low resolution mass spectral data banks which allows significant compression of the library file through selective binary coding of characteristic peaks and use of variable-length logical records.

The searching of mass spectral libraries, when these libraries consist of a full tabulation of mass and intensity pairs for every ion produced by each compound, is time- and

core storage-consuming in operation. Because not every laboratory using mass spectrometry as a diagnostic tool has access to a large computer system and because the so-called mini computer systems are becoming increasingly more available, the consideration of methods whereby use of core storage is reduced and the time of library search for component identification is likewise reduced is most important.

The concepts of Khinchin entropy function and divergence which have already been discussed were investigated with the aim in mind of providing a unique diagnostic value for each member of a spectral library. File searches for such single valued functions (i.e. Khinchin and divergence) satisfy the requirement for small core storage and/or peripheral storage requirements as well as reducing file search time.

However, in the real world of data acquisition from such common devices as a tandem gas chromatograph/mass spectrometer, as well as from conventional mass spectrometer operation with rapid scanning capability, the production of mass spectra reaches a rate which renders processing all mass and intensity data so generated impractical, if not indeed impossible.

It will be remembered that the Khinchin entropy function and the divergence calculation require, in the first stage, all values of mass versus intensity, although the final result is a single-valued function in each case.

In the case of octal coding, the values required may be extracted readily during on-line data acquisition, thereby leading

directly to on-line creation of a library without a requirement for storage of all mass and intensity data.

In actual practice a library would be created by coding the mass spectra for a group of compounds expected in a certain analysis. Subsequently, during mass spectrometric analysis of unknown compounds, codification of the spectrum would occur under conditions much more likely to reproduce those utilised in creating the original library of known spectra.

In the general case, identification of unknown compounds by spectroscopic means assumes the existence of a file of data for a large number of known components and the ability to search the contents of the file in a manner which facilitates component identification. Although the general problems of data file management and the design of information retrieval algorithms are common to all areas of spectroscopy, the field of mass spectrometry offers perhaps one of the greatest challenges in terms of the complexity and quantity of data. As such, much attention has been given to the problem of data manipulation in this field.

The traditional approach has dealt with the basic tables of m/q versus intensity values that constitute digitised mass spectra. These tables have been managed in numerous ways.(1-16)In general, identification of an unknown mass spectrum involves testing the unknown against a known spectrum to observe their similarity. A matching index is calculated whereby a numerical value may be assigned to similarity or dissimilarity of the two spectra as dictated by the criteria of the test. By comparing the unknown

to each known spectrum in the library file, computing a matching index for each examination and selecting the compounds in the file whose matching indices indicate a high probability that they are the same as the unknown compound, it is possible to achieve identification.

Because of time and space requirements in computer processing of mass spectral data when they are represented by the conventional tabulation of mass and intensity for all ions, considerable attention has been given recently to condensed codification.

In a system developed by Petterson and Ryhage (17) a preliminary filtering was performed, based on molecular weight, followed by search for the six highest peaks from the total mass spectrum. An alternate approach searched for the highest peaks between certain mass numbers, i.e. six peaks for molecular weight up to 200 and 10 peaks for molecular weights above 200.

The searching procedure developed by Hites and Biemann (18) utilised an abbreviated spectrum which retained the two most intense peaks in consecutive regions of 14 mass units. Ratios are calculated for these selected masses using respective intensities from unknown and known compounds; those ratios of 1 or more were weighted according to an intensity scale.

Crawford and Morrison (19) based a searching system on the six strongest peaks in 3200 A.S.T.M. spectra; normalisation techniques were employed.

Still another extensive study (8000 standard spectra) was performed by Knock, Smith, Ridley and Kelly (20). Varying num-

bers of peaks were used, both in normal and decreasing intensity arrangements; provision was also made to account for possible mass discrimination effects. Counting similar isomers, the retrieval record was 97 per cent.

An early exercise by Abrahamsson, Haggstrom and Stenhagen (21) utilised all m/q and intensity values for each spectrum; it was possible to select the search mode used by the computer, i.e. five strongest intensities, five strongest weighted intensities or the most intense masses in selected m/q intervals of the spectrum. Spectra chosen by any of these search techniques or modes were compared with the unknown.

The techniques cited above treat data which is represented alphanumerically or numerically in decimal format (base 10) i.e., input for m/q and intensity value is in base 10; the compound name is in alphanumeric format. Insomuch as a digital computer operates in the binary representation or base 2, there is much wasted time involved in the sequence: decimal input - conversion to binary for internal processing and back to decimal for output formatting. Such a procedure is wasteful of time and core memory of the computer.

A novel approach to the problem of coding and searching of library files, which is closely related to machine level digital computer operations, has been provided by Grotch (22,23) In this approach, mass spectral intensity data is systematically quantitised to binary levels related to the presence or absence of a peak for each m/q value within a specified range (typically

12-200 amu). Extensive statistical testing, using each binary spectrum in the library file as an unknown and comparing it to all other members of the file, was carried out to show that a binary pattern representation was indeed a unique representation of the data. A mismatch criterion was developed, using Boolean functions for "exclusive or" and "and", for which a value of 0 is an index of a perfect match. Uniform logical record lengths for each binary mass spectrum were used to construct the library file.

Most recently (23), Grotch has extended the concept of one bit coding to an "abbreviated" mass spectrum, wherein the mass position of the single most intense peak in each 14 amu grouping is encoded. Elimination of encoding words for peak heights results in greatly increased search speeds as well as providing for the obvious reduction in amount of computer storage required. Results are comparable to those found with the original one bit encoding procedure.

The variability in the values of relative ion abundances with variation in mass spectrometer design and operation produces considerable uncertainty in the reliability of a diagnostic based on measurement of spectral intensity factors, and strongly points out the efficiency of using a coding technique which does not depend upon the reproducibility of an intensity factor.

Along these lines, a codification procedure has been recently conceived and tested for use with low resolution mass spectral data banks; such codification was based on the octal or base 8 numerical system. Extremely rapid rates of library search result

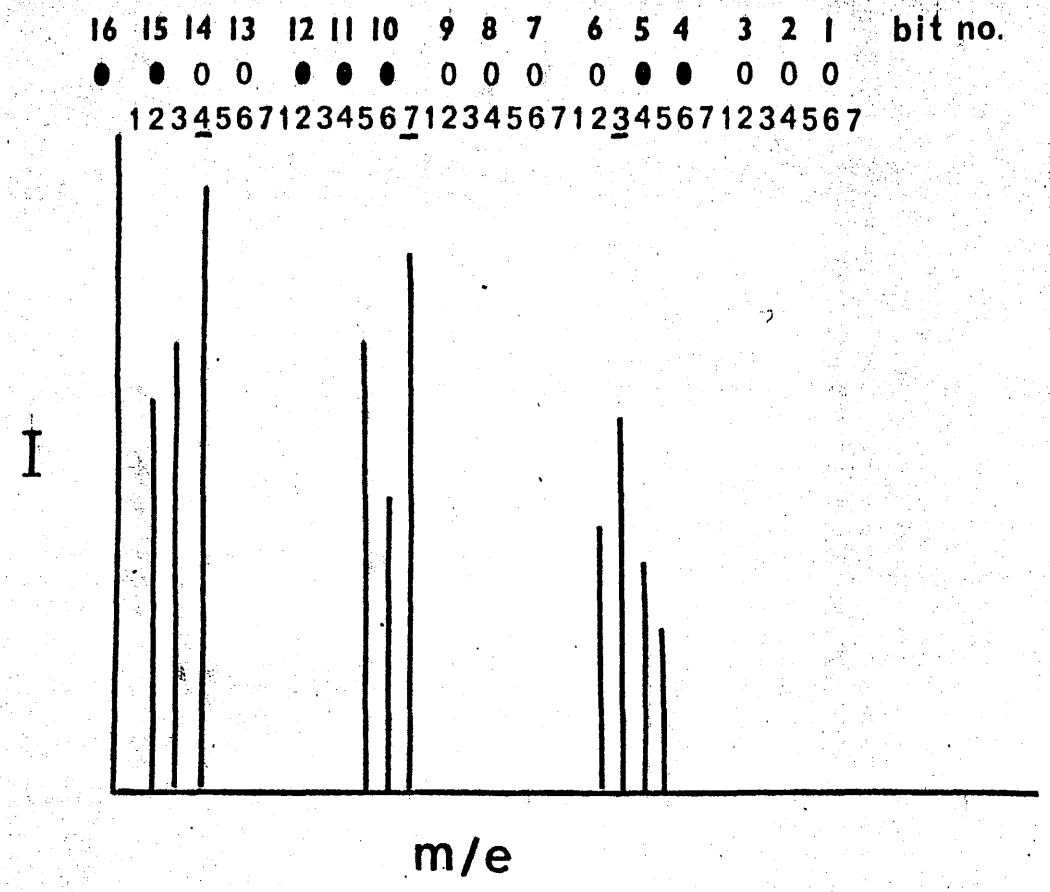

16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 bit no.
● ● 0 0 ● ● ● 0 0 0 0 ● ● 0 0 0
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7
I
m/e

Figure 1

from use of octal coding for the following reasons:

1. A binary code whereby the octal codification is represented allows highly efficient use of core storage in a digital computer

2. The computation of matching indices by logical operations is more suited to basic computer design than the standard computations, excepting the work of Grotch, already described.

In actuality a compression of data occurs in two instances: a) by use of initial coding in octally coded binary to obviate conversion from decimal code and b) by use of variable word length based on the number of peaks appearing in the spectrum. Figure 1 indicates the procedure for encoding a mass spectrum. Selective binary coding of characteristic peaks is accomplished by arbitrarily dividing the mass range of interest into multiple groups of seven. The number corresponding to the spectrum peak having the highest intensity is then encoded as a three bit binary number. Thus, the fourth peak is encoded in the first grouping, the seventh peak in the second and so on; zero is used to denote the absence of a peak within the grouping, thereby giving a total of eight possible values, hence the name octal coding.

Representation of an octal number within the computer requires three bits; thus, in a 16 bit machine such as the Hewlett-Packard 2116B used first in setting up this system, five octal characters can be stored in each computer word, with one bit left over. Thereby a single computer word is capable of storing information which covers a range of 35 atomic mass units (m/q units). Compounds requiring a greater range of masses to be encoded require an additional number of computer words. As many

Mass ranges	23-29	30-36	37-43	44-50
m/q to be encoded	24	32	43	0
Position of m/q in octet	2	3	7	0
Binary code	010	011	111	0
Octal code	2	3	7	0

Table 1

are used as are needed to encode the spectrum. The last word is then designated by setting a flag in the 16th bit; that is, bit 16 is "on".

If consideration is given to the m/q values which occur most often in the spectra of organic compounds, a series of octal ranges beginning with the group of 7 masses, 23-29, serves to provide greater diagnostic character for this method of coding. Subsequent mass ranges would be 30-36, 37-43, etc. (see Table 1). The first octade, containing masses 12-14-15-16-17-18-19 was included in the original codification procedure but when it was learned that no additional information was gleaned from using these m/q values, the entire octade was dropped from consideration.

The efficiency of a coding procedure is reduced by the need to use "0" for coding, i.e. coding which leads to a large number of zeros. If one codes in octades (or some larger sized grouping) there is more likelihood of a peak appearing and thus, according to the basic principles of information theory, greater entropy, as reflected by more efficient transmission of information. In general, the encoding and retrieval of data from a number system such as this is predicated primarily on the principle of simple manipulation of the numbers. As suggested by Smith's (24) use of compound classifier, selecting the mass ranges so that certain ions fall characteristically in particular octades, it is possible to develop qualitative information content as well, that may relate to the functional group structure of the molecule. However, sufficient computer power was not available at the time

```
2970 1097 164 P-ALPHA,ALPHA-TRIMETHYLPHENETHYL    52777216 16171224 06101000 00000000 00000000 00000000
2971 4969  94 2,3-DITHIABUTANE (DIMETHYL DISUL    53021427 16200000 00000000 00000000 00000000 00000000
2972 4947  94 2,3-DITHIABUTANE (DIMETHYL DISUL    53121427 16200000 00000000 00000000 00000000 00000000
2973 1703 220 1,2-DIBROMOCHLOROETHANE             53140400 36161072 72354004 00004100 000C0C00 00000C00
2974 1855 252 TRIMETHYL TRIMESATE                 53261514 36615606 16161620 30203050 610C0000 00000000

2975  603 132 1,1,1-TRIFLUORO-3-CHLORO-PROPANE    53261656 72466165 17720720 00000000 00000000 00C00000
2976 2553 164 2,3-DIMETHYL-5-ISOBUTYLPYRAZINE     53313126 27272727 16162C00 00000C00 00C00000 0000C000
2977 4973 112 2,3-DIMETHYLTHIOPHENE               53321236 71516100 00000000 00000C00 00C00000 00000000
2978 4975 112 2,5-DIMETHYLTHIOPHENE               53321256 11515100 0C000000 C0000000 000C0C00 00000C00
2979 5395 112 2,5-DIMETHYLTHIOPHENE               53321256 17515100 00000000 00000000 000C0000 00000000

2980 5174 126 2,3,4-TRIMETHYLTHIOPHENE            53321256 76515161 00000000 00000000 00000000 00000000
2981 5440 124 (1-THIAETHYL)-BENZENE (METHYL PH    53321517 16173041 07000000 00000000 00000000 00000000
2982 6388 192 1,2,3,4,5,6,7,8-OCTAHYDRO-9-THIA    53321616 16545231 26162617 20200000 00C00C00 00000000
2983 6395 262 8,9-DIMETHYL-11-THIABENZO(B)-FLU    53321654 43254434 54271766 55443726 15212000 00000000
2984 6397 264 9,10-DIMETHYL-6,7-DIHYDRO-11-THI    53321654 43254435 54471726 55443726 35416000 00000000

2985 6391 248 8-METHYL-11-THIABENZO(B)FLUORENE    53321654 43524347 54271716 55443716 20000000 00000000
2986 6329 188 1,2,3,4-TETRAHYDRODIBENZOTHIOPHE    53321656 17237271 56451275 10000000 00000C00 00000000
2987 6327 190 2-ETHYL-5,7-DIMETHYLBENZO(B)THIO    53321656 22237271 76451617 10000000 00000C00 00000000
2988 6322 176 2-METHYL-7-ETHYLBENZO(B)THIOPHEN    53321656 23137271 16161710 00000000 00000000 00000000
2989 6325 176 2,5,7-TRIMETHYLBENZO(B)THIOPHENE    53321656 23137271 76161710 00000000 00000000 00000000

2990 6319 162 2,5-DIMETHYLBENZO(B)THIOPHENE (2    53321656 26132271 16171000 00000000 00000000 00000000
2991 6320 162 2,7-DIMETHYLBENZO(B)THIOPHENE (2    53321656 26132271 16171700 00000000 00000000 00000000
2992 6318 162 7-ETHYLBENZO(B)THIOPHENE (7-ETHY    53321656 34142217 16171000 00000000 00000000 00000000
2993 6354 136 1-THIA-(2,3-DIHYDROINDENE) (1-TH    53321656 46132617 16300000 00000000 00000000 00000000
2994 4630  98 2-METHYLTHIOPHENE                   53323151 40510000 00000000 00000000 00C00000 00000000

2995 4631  98 3-METHYLTHIOPHENE                   53323151 47510000 00000000 00000000 00000C00 00000000
2996 5385  98 2-METHYLTHIOPHENE                   53323151 47510000 00000000 00000000 00000000 00000000
2997 5386  98 3-METHYLTHIOPHENE                   53323151 47510000 00000000 000C0000 00000000 00000C00
2998 5330 112 3-ETHYLTHIOPHENE                    53323156 61516100 00000000 00000000 00000000 00000000
2999 4972 112 2-ETHYLTHIOPHENE                    53323156 71516100 00000000 00000000 00C00000 00000000

3000 5411 154 TRICYCLO(3.3.1.1(3,7))-2-THIADEC    53323237 16565761 50610000 00000000 00000000 00000000
// UNLOAD SYS002
/&
02 001  0CODE33   08 45 40  08 56 51          SLG 8 1
```

Figure 2

of this study to test this principle on a statistically significant level with an extensive library.

It was possible recently, through the cooperation of Dr. Grotch at the Jet Propulsion Laboratories in Pasadena, California, U.S.A. to encode a library of 6880 compounds (basically the so-called Wiley tape of mass spectral data) in the octal format. A sample printout is shown in figure 2. It should be pointed out that a 16 bit computer word enables 64 thousand compounds to be coded uniquely ($2^{16} = 65,536$). Using the 6880 compound library, there was printed out a listing of all compounds with the same octal coding. In more than 90% of the cases compounds with identical codes represented duplicates in the library or isomers which are not readily differentiated by the mass spectrometer.

A highly significant feature of octal coding, as already cited, is that it may be easily achieved from digitised mass and intensity data acquired on-line in real-time from the detector of a mass spectrometer. A normal sequence of data processing would involve the following on-line operations.

1. Conversion of the analogue mass spectrometer output to digital format, i.e. mass and intensity data.

2. Condensation of the data to octal format by means of a routine which determines the most intense peak in each octal grouping and provides the binary equivalent as the "spectrum" for which the data library is searched.

3. When binary "1" is sensed in the 16th bit, the number of words to encode the "spectrum" is known; thus, it is not

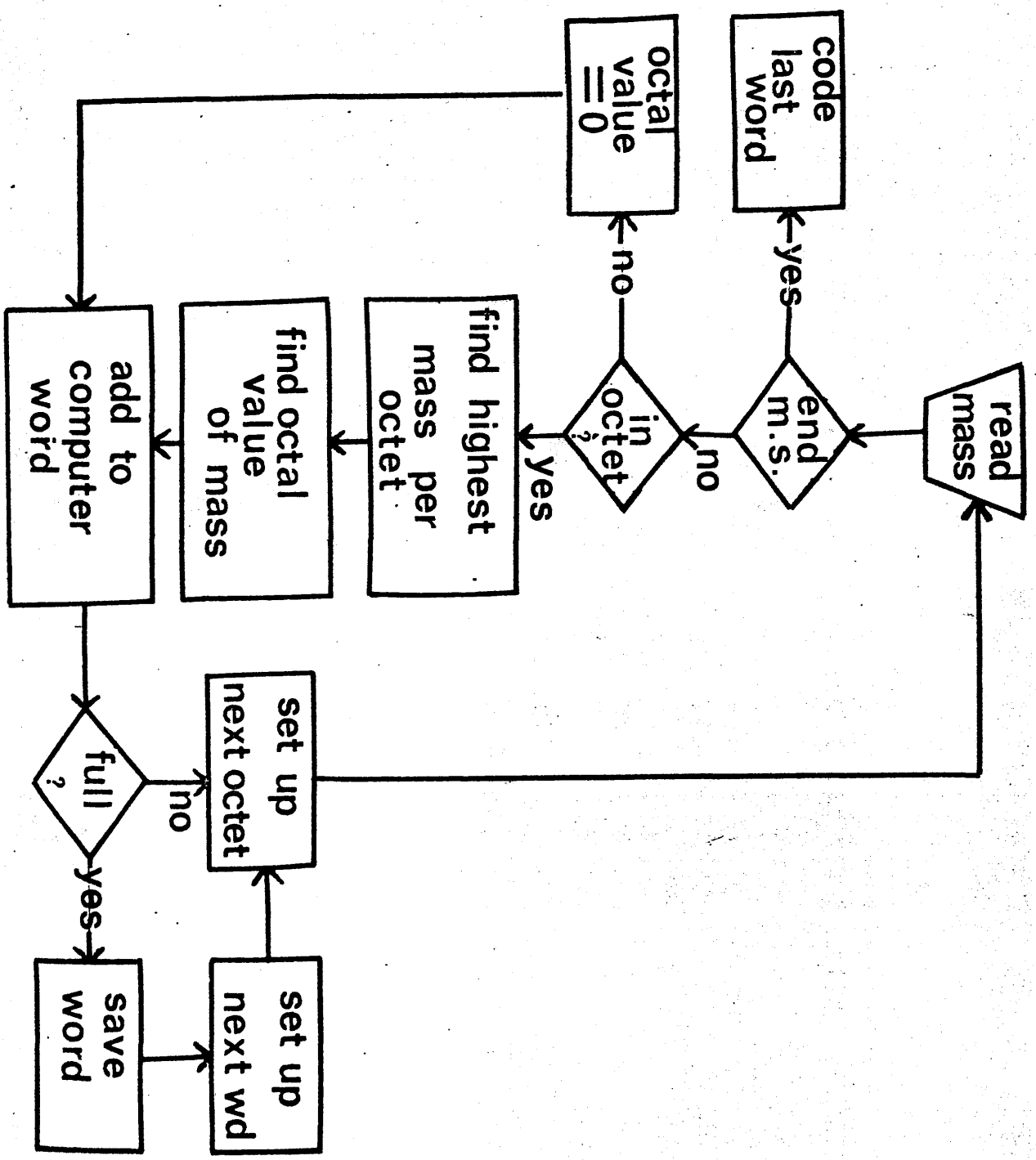

read mass
end m.s.
yes
code last word
no
in octet ?
yes
find highest mass per octet
find octal value of mass
add to computer word
octal value = 0
no
set up next octet
full ?
yes
save word
set up next wd
no

Figure 3

necessary to search the entire library but only the subset or sub-library collection of spectra which require that number of words for coding.

A flow chart for the routine to effect octal coding is found in figure 3.

While it is useful to devise various coding schemes, particularly those which may be readily manipulated with Boolean functions, the practical advantage to the analysis process is limited unless the library file is properly organised for subsequent searching. With this selective coding technique, it is possible to divide the total file into a series of sub-files based on the number of computer words necessary to selectively code the spectrum to its highest observed m/q. This particular organisational form appears to be of special utility as suggested earlier in fully automated gas chromatographic/mass spectrometric analysis systems, since the highest observed m/q is a quantity readily extracted during the data reduction process. In actual use, such a file organisation implies pre-filtering of the data tables, since only those sub-files having the same number of words as the unknown must be searched. As has been indicated, the sub-files, as constructed in this work, use the 16th bit of the word to signify the end of the logical record. The word immediately following the end of the logical record thereby contains an integer pointer to a separate file containing the alphanumeric characters of the compound name. The division into sub-files of variable logical record length and creation of a name file were designed to make maximum use of random access mass

Number of words to code in octal	Name of compound	Octal Code
1	methanol	100372
2	acetaldehyde	000371 100710
3	2-butanone	000370 000717 100071
4	diacetyl	000370 000710 000000 100010
5	1,4-dimethyl-benzene	000350 000371 000616 000160 100700

Table 2 - Subsets of octally coded spectra

storage devices such as discs, which have recently been reduced in price to the point that they may be included in a modest computer system.

As an example, the comparison of the octal code of compounds with spectra of varying complexity is given in table 2. This table also serves as an example of the subset files for classifying octally-coded library files. Appendix A contains a more extensive list of examples of octally-coded compounds.

If uniform record lengths are utilised, in the case where all m/q values are coded for each spectrum, the number of records utilised in the library will be that number required to code the compound whose spectrum displays the greatest number of m/q values, i.e. the greatest number of ions. For a library in which the maximum molecular weight occurs at m/q 250, this will be sixteen, i.e. 16 words of 16 bits each will allow encoding 16x16 or 256 m/q values.

In the case of an octal codification procedure wherein variable record length is employed, a single word suffices for coding the spectrum of methanol, whereas 5 words are needed for coding the spectrum of 1,4-dimethylbenzene.

Since one computer word is capable of representing 35 mass positions, five computer words are needed to encode to a mass of 175. Since the average molecular weight of 6652 compounds contained in the Atlas of Mass Spectral Data (the so-called Wiley file) is 167, it is seen that five computer words is the average number required for the total collection of data. For each unknown being searched for in the library, a matching

index is calculated; the five best matches, in decreasing order of goodness of match, are printed. This feature is expected to be most useful in future cases when much expanded library files are being searched and the possibility exists for the same matching index to be calculated for more than one compound.

Although it is possible to code approximately 64K compounds with 16 binary bits, one must expect, because of non-uniform production of spectra, that identification will not in every case be that of a unique compound.

In the direction of minimising such non-unique cases, it has been noted that octal coding provides a dampening effect on variations in spectral characteristics because of its relative insensitivity to errors in digitisation. For example, errors which may occur in the initial codification of an unknown spectrum, caused for example by a spurious signal or the additive effect of impurities upon peak intensities, have significant influence upon correct identification of the compound in question. Likewise, in the case where the spectrum of an unknown compound is incorrectly coded because of variation in relative intensity values caused by mass spectrometer instabilities, it is shown that correct identification may result even in cases where a coding error occurs in each word of the spectrum. Examples of this behaviour pattern are depicted in table 3 below.

Compound - 2-pentanol

Correct code	Incorrect code
052725 106200	052726 106100

STRAWBERRY
WHOLESOMENESS LIBRARY

TTY:<NAMLIB.DA

1
2
34 HYDROGEN SULFIDE
32 METHANOL
2
5
48 METHYL MERCAPTAN
56 PROPENAL
44 ACETALDEHYDE
46 FORMIC ACID
46 ETHANOL
3
23
58 ACETONE
72 BUTYRALDEHYDE
70 CROTONALDEHYDE
58 PROPIONALDEHYDE
74 PROPANOIC ACID
60 ACETIC ACID
74 2-METHYL-1-PROPANOL
88 1-PENTANOL
88 3-METHYL-1-BUTANOL
88 2-METHYL 1-BUTANOL
88 2-METHYL-2-BUTANOL
88 3-METHYL-2-BUTANOL
88 2-PENTANOL
74 2-BUTANOL
74 1-BUTANOL
60 1-PROPANOL
74 METHYL ACETATE
102 N-PROPYL ACETATE
116 N-BUTYL ACETATE
102 3-HEXANOL
116 AMYL FORMATE
76 1,1-DIMETHOXYMETHANE
72 2-BUTANONE
4
20
86 3-METHYL-2-BUTANONE
86 DIACETYL
86 2-PENTANONE
88 2-METHYL PROPANOIC ACID
102 PENTANOIC ACID
88 BUTANOIC ACID
88 ETHYL ACETATE
102 N-HEXANOL
102 2-HEXANOL
98 2-HEXENAL
86 1-PENTENE-3-OL
116 I-AMYL FORMATE
130 I-AMYL ACETATE
144 N-HEXYL ACETATE
90 1,1-DIMETHOXYETHANE
88 ACETOINE
96 FURFURAL
94 DIMETHYL DISULFIDE
98 CYCLOHEXANONE
84 CYCLOPENTANONE
5
32
120 PHENYL METHYL KETONE

114 2-HEPTANONE
100 HEXALDEHYDE
116 HEXANOIC ACID
102 2-METHYL BUTANOIC ACID
108 BENZYL ALCOHOL
102 METHYL ISOBUTYRATE
102 METHYL BUTYRATE
102 ETHYL PROPIONATE
144 ETHYL-N-CAPROATE
130 ISOPROPYL BUTYRATE
116 ETHYL BUTYRATE
130 ETHYL N-VALERATE
130 ETHYL-3-METHYL BUTYRATE
116 2-HEPTANOL
130 2-ETHYLHEXANOL
136 BENZALDEHYDE
142 TRANS-2-HEXENYL ACETATE
104 1-METHOXY-1-ETHYOXYETHANE
118 1,1-DIETHOXYETHANE
132 1-ETHOXY-1-PROPOXYETHANE
172 1ETHOXY-1HEX-3-ENOXYETHANE
146 1,1-DIETHOXY BUTANE
160 1,1-DIETHOXY PENTANE
110 METHYLFURFURAL
110 2-ACETYLFURAN
116 ETHYLISOBUTYRATE
116 METHYL-2-METHYLBUTYRATE
198 CIS-3-HEXENYL-N-HEXANOATE
106 1,3-DIMETHYLBENZENE
106 1,2-DIMETHYLBENZENE
106 1,4-DIMETHYLBENZENE
6
8
172 HEXYL BUTYRATE
154 LINALOOL
122 PHENETHYL ALCOHOL
154 ALPHA-TERPINEOL
132 1,1-DIETHOXYPROPANE
174 1,1-DIETHOXY HEXANE
130 ETHYL ACETOACETATE
130 ETHYL-2-METHYLBUTYRATE
7
13
158 ETHYL HEPTANOATE
156 ISOPROPYL HEXANOATE
154 BORNEOL
154 I-FENCHYL ALCOHOL
142 1-METHYLNAPHTHALENE
142 2-METHYLNAPHTHALENE
150 BENZYL ACETATE
160 1-ETHOXY-1-PENTOXYETHANE
174 1-ETHOXY-1-HEXOXYETHANE
158 METHYL-N-OCTANOATE
150 ETHYL BENZOATE
144 METHYL HEPTANOATE
156 METHYL OCTANOATE
8
7
228 ETHYL LAURATE
172 ETHYL CAPRYLATE
200 ETHYL CAPRATE
162 METHYL T-CINNAMATE
176 ETHYL-C-CINNAMATE
166 ETHYL SALICYLATE
172 METHYL NONANOATE

9
3
198 2-HEXENYL-HEXANOATE
186 METHYL DECANOATE
198 TETRADECANE
10
2
230 1,1-DIHEXOXYETHANE
214 METHYL-N-DODECANOATE
11
1
226 HEXADECANE
12
0
13
0
14
0
15
0
585

FIGURE 4

```
000353
000720
100000
000652
000371
000616
000160
100620
000052
000716
000273
000020
100300
000052
000715
000273
000020
100300
000372
000727
000104
000020
100300
000372
000725
000373
000537
100203
000052
000715
000372
000040
100302
000072
000725
000372
000030
100203
000372
000527
000362
000730
100202
000371
000527
000173
000727
100302
000053
000525
000160
000506
100200
000072
000727
000162
000536
100070
000070
000371
000606
000000
100700
000053
000717
000431
```

FIGURE 5

In general, a single mistake in coding may occur in each computer word used to represent the compound without incorrect identification resulting.

The coding system has proven practical in the processing of data for projects which involve a small library of octally coded spectra of compounds known from previous experience to be commonly present in certain types of samples. In the current hardware configuration these compounds are determined by a combined gas chromatography/mass spectrometry system. Figure 4 indicates the structure of a data file for compounds determined in the course of protocol development for studies in the wholesomeness of strawberries; organisation is based on the sub-file category already presented, i.e. on the increasing number of words required to represent the spectrum.

The code on line one of this figure identifies the library which it contains; the format for the remainder of the library is as follows: The value on the next line is the number of computer words needed to encode the spectrum and the number on the third line is the number of compounds in the subfile. The names of the compounds in the subfile are printed immediately below in each case with the molecular weight appearing immediately to the left of the name.

In figure 5, appear the actual octal codes for some members of the library already listed in figure 4. The numbers on the two lines prior to each octal representation have the same meaning as in figure 4.

Figure 6 shows a reproduction of the printout indicating

SCAN NO.: 10 START TIME: 626. SEC

COMPOUND NAME	MATCHING INDEX
TRIDECANE	9
N-TETRADECANE	25
1-TETRADECENE	43

Figure 6

Printout for a typical computer identification

TTY:<NAMLIB.DA

Baff Baff Wholesomeness Library
7/6/72

1
4
ETHENE
ACETYLENE
ETHANE
METHANE
2
6
PROPANE
PROPENE
1-BUTENE
METHYL ACETYLENE
2-METHYLPROPANE
1-BUTENE
3
7
2-BUTANONE
2-METHYLBUTANE
N-PENTANE
1-PENTYNE
ACETONE
1-PENETENE
N-BUTANE
4
6
N-HEXANE
2-METHYLPENTANE
1-HEXENE
1-HEPTENE
1-HEPTYNE
1-HEXYNE
5
10
1 OCTENE
2-METHYLHEPTANE
N-HEPTANE
2-METHYLHEXANE
CIS-1,2-DICHLOROETHYLENE
PHENYL METHYL KETONE
N-OCTANE
1-DECYNE
1-OCTYNE
1-NONYNE
6
4
2-METHYLOCTANE
1-DECENE
N-NONANE
1-DODECYNE
7
3
N-UNDECANE
1-UNDECENE
N-DECANE
8
3
N-DODECANE
1-TRIDECENE
1-DODECENE
9
3
N-TETRADECANE
1-TETRADECENE
TRIDECANE
10
3
1-PENTADECENE
1-HEXADECENE
N-PENTADECANE
11
2
N-HEXADECANE
N-HEPTADECANE
12
1
N-OCTADECANE
13
1
N-EICOSANE
14
1
N-HENEICOSANE
15
2
N-DOCOSANE
N-TRICOSANE
585

Figure 7

the identification of the component found for mass spectrum no. 10 in a typical chromatographic analysis. The octal code subfile which was searched was for a compound represented by 9 computer words. Only three compounds were found in the file and the best match was given by tridecane (see figure 7) which is a name file of compounds for another special library - this time for beef wholesomeness studies. The formatting here is slightly different from that utilised in figure 4. In particular, the molecular weight is not shown. In each case the first number represents the number of computer words required to encode the spectrum in octal format, the second number indicates how many compounds requiring that number of words will be found in the subfile. In this particular case (as indicated by the arrow in figure 7) the subfile requiring 9 words to encode contains 3 compounds:

n-tetradecane, 1-tetradecene and tridecane

In summary, octal coding provides for utilisation of analogue mass spectra in a binary format which facilitates the file search for identification purposes. The overall size of the data files to be searched may be substantially reduced and the need for searching the entire file is avoided by using variable record lengths to provide subfile classifications.

In actual practice this method of coding and searching of data files for the identification of compounds in gas chromatography/mass spectrometry has greatly expedited the time required to complete the analysis and has as well reduced the amount of time a skilled spectroscopist must devote to routine interpretation.

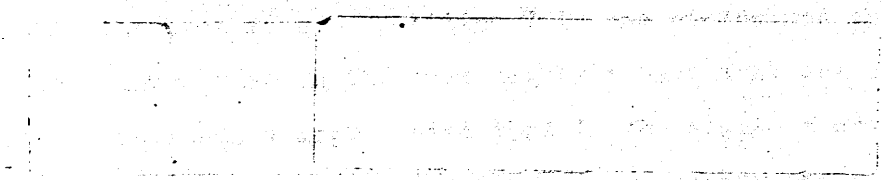
START MASS	TRANSITION			
	.01	1.	5.	10.
20	17.5	13.0	5.9	8.1
21	17.5	13.3	6.3	8.6
22	17.5	13.3	6.4	8.7
23	17.4	12.6	5.8	7.9
24	17.2	12.2	5.7	7.7
25	17.0	12.0	5.5	7.4
26	16.	12.5	5.9	8.0

Table 4

Number of ones coded
(average per spectrum)

Although the assignment of mass for the octal groupings utilised in the this work is based upon a start mass of 23 for the first "octade", it has not been established with complete certainty that this mass distribution is the best one to use. It would seem axiomatic that the application of specific mass spectral information to the codification process would result in more highly diagnostic binary patterns which may be recognised by the computer and correlated with like patterns in the similarly coded library. For example, a practicing mass spectroscopist would feel that starting mass for the groups or octades would affect the success of the search algorithms in differentiating among spectra, insomuch as there exists a series of ions which are routinely used as diagnostics for specific functional groups. The need to consider statistical data in light of the basic character of the discipline producing them is illustrated in table 4.

Since the number of binary bits or "ones" coded in the library represents its information content, such representation is frequently employed. Considering a range of starting masses for the first group from 20-26, it can be seen for the intensity found at each transition level (threshold level) the number of ones coded remains essentially constant. The information content of the spectrum as an entity remains constant. Such a conclusion is obvious; but also misleading because it imputes to the code the ability to select the most diagnostic mass in each group of seven when it is actually selecting the most intense mass in each group of seven amu. A look at some statistical values based on



30	31	53	54	91	92
225	226	281	282	309	310

Table 5

m/q Values at Window Boundaries, Octal Code

One mass in seven is coded

a 6880 compound library illustrates the point as suggested in table 5.

Each pair of masses in this table (selected from the statistical distribution of ones when the starting mass is 23) represents respectively the last mass or position 7 in one window and the first or position 1 mass in the next higher adjacent window or group. It can be readily appreciated that if the scale is moved left or right to accomodate any starting mass that any one of these mass values would be moved into an adjacent window where it might or might not be coded as the most intense peak, depening of course upon the intensity of the other peaks in the window under consideration. In other words, for this group of adjacent mass pairs or indeed for any adjacent mass pair, a shift of starting mass used in the code may eliminate the boundaries which allow each mass to be coded in a separate window. Appearance of adjacent masses is in many instances a highly diagnostic feature. Commonly known examples would be the P and P-1 peaks and the isotope distribution peaks.

In terms of practicality as regards computer power needed to do the job and ease with which the chosen technique provides for on-line real-time processing of mass spectral data, it certainly need not be a unique compound. In most practical cases, the investigator is in possession of some ancillary information about the compound or compounds he expects to identify; so the fact that more than one answer is given by a search algorithm does not mitigate against its utility.

There has been, in the practical applications of this work, a higher degree of success when working with special restricted libraries of compounds expected to appear in particular sample types; these libraries contain spectra coded from the mass spectrometer on which the "unknown is being run. Therefore greater internal self-consistency of data is achieved. When one goes to larger libraries, the spectra in which come from numerous sources, the internal self-consistency factor is reduced, often dramatically.

If progress in purification of data libraries can be encouraged at a rate near that at which codification and search techniques are proliferating at the present time, there is indeed promise for major advance in the general area of mass spectral data processing.

GENERAL BIBLIOGRAPHY

1. S. Abrahamsson, S. Stallberg-Stenhagen, and E. Stenhagen, Biochem. J., 92, 2 (1964)

2. B. Petersson and R. Ryhage, Arkiv. Kemi, 26, 293 (1967)

3. L.R. Crawford and J. D. Morrison, Anal. Chem. 40, 1464 (1968)

4. L.R. Crawford and J. D. Morrison, Anal. Chem. 43, 1790 (1971)

5. R.A. Hites and K. Biemann, Advan. Mass Spectrom., 4 , 37 (1968)

6. B.A. Knock, I.C.Smith,D.W.Wright,R.G.Ridley, and W.Kelly, Anal. Chem., 42, 1516 (1970)

7. F.E. Lytle and T.L. Brazie, Anal. Chem. 42, 1532 (1970)

8 S.L. Grotch, Anal. Chem., 43, 1362 (1971)

9 Ibid. 42, 1214 (1970)

10. Ibid. 45, 2 (1973)

11. L.E.Wangen, W.S.Woodward and T.L.Isenhour, Anal.Chem. 43, 1605, (1971)

12. H.S.Herz, R.A.Hites and K.Biemann, Anal. Chem., 43, 681 (1971)

13. R.G.Ridley, in " Biochemical Applications of Mass Spectrometry ", G.R.Waller, Ed., Wiley-Interscience, New York, Chap. 6, (1972)

14. D.H.Robertson and R.I.Reed, Proc. of 19th Annual Conference on Mass Spectrometry and Allied Topics, Atlanta,Ga., p.68 (1971)

15. D.H.Robertson, J.Cavagnaro, J.Holz and C. Merritt,Jr., Proc. of the 20th Annual Conference on Mass Spectrometry and Allied Topics, Dallas, Texas, p.359 (1972)

16. R.I.Reed, Rev. Port. Quim., 10, 129 (1968)

17. B. Pettersson and R. Ryhage, Arkiv. Kemi, 26, 293 (1967)

18. R.A.Hites and K. Biemann, " Advances in Mass Spectrometry ", 4, 37 (1968)

19. L.R. Crawford and J.D. Morrison, Anal. Chem., 40, 1464 (1968)

20. B.A.Knock, I.C.Smith,D.E.Wright, R.G.Ridley and W.Kelly, Anal. Chem. 42, 1516 (1970)

21. S. Abrahamsson, G. Haggstrom and E. Stenhagen, 14th Annual Conf. Mass Spect. and Allied Topics, pp 522-7 (1966)

22. S.L.Grotch, Anal. Chem., 42, 1214 (1970)

23. S.L.Grotch, Anal. Chem., 43, 1362 (1971)

24. D.H.Smith, Anal. Chem., 44, 536 (1972)

APPENDIX A

COMPOUND	OCTAL CODE
methane	150000
ethane	147300
n-propane	147475
methanol	140600
ethanol	141606
1-propanol	106000
n-butane	047274 105000
n-pentane	040274 174050
1-butanol	007672 173071
1-pentanol	000673 172030
2-propanol	041276 160000
2-butanol	000676 146060
1-butene	047272 143000
cis-2-butene	047272 173000
iso-butene	047272 143000
t-2-butene	040272 103000
1-pentene	040273 172730
t-2-pentene	040473 172730
c-2-pentene	047273 172730
benzene	047170 150374

COMPOUND	OCTAL CODE
n-hexane	040474 074040 150000
n-heptane	040474 074040 140500
n-octane	040474 074040 140005
1-hexanol	041674 073720 130000
1-heptanol	071402 003737 120300
1-octanol	071472 072727 137203
2-pentanol	000276 002060 170000
2-hexanol	071206 004020 160500
2-heptanol	041276 072030 120307
3-pentanol	000672 076000 170000
3-hexanol	071614 006060 140500
3-heptanol	041672 076727 160307
cyclohexanol	071405 004747 110500

COMPOUND	OCTAL CODE
cyclopentyl-methanol	041272 074717 110500
o-methyl-cyclohexanol	071202 004717 140105
m-methyl-cyclohexanol	071202 004747 110103
p-methyl-cyclohexanol	071402 004737 110105
2,5-dimethyl-pyridine	007271 041575 164050
2,6-dimethyl-pyridine	007273 051615 104050
3,4-dimethyl-pyridine	007273 051615 104050
3,4-dimethyl-pyridine	007272 051515 104050
3,5-dimethyl-pyridine	007272 051575 104050
c-1,2-di-chloroethene	027130 027130 107100
t-1,2-di-chloroethene	027130 017130 107100
methyl-benzene	040276 050573 153000
1,2-dimethyl-benzene	040272 050573 103740
1,3-dimethyl-benzene	040272 050573 103740
1,4-dimethyl-benzene	040272 050573 103740

COMPOUND	OCTAL CODE
n-nonane	000474 074040 040400 105000
n-decane	000474 074740 040404 100050
1-nonanol	071402 003727 020300 103000
2-octanol	071276 072720 030203 107000
3-octanol	071404 006747 020603 106000
3-nonanol	071404 006727 040206 103000
2,3-dimethyl-pyridine	007271 051615 004050 150000
2,4-dimethyl-pyridine	000051 061515 014451 150000
1-hydroxy-2,3-dimethyl-pyridine	007272 052573 053750 161000
1-hydroxy-2,3-dimethyl-pyridine	007272 052573 053750 161000
1-hydroxy-2,5-dimethyl-pyridine	000272 052573 013750 161000
1-hydroxy-2,6-dimethyl-pyridine	007272 057573 053750 161000
1-hydroxy-3,4-dimethyl-pyridine	000272 052573 003750 160000
1-hydroxy-3,5-dimethyl-pyridine	007375 051614 005050 160000
o-chloro-toluene	037276 057361 053451 173000
p-chloro-toluene	077276 047361 053422 173010
m-chloro-toluene	047276 057361 053431 173000
1,2-dicyano-benzene	007350 047471 071610 105000

COMPOUND	OCTAL CODE
n-undecane	040474
	074740
	040304
	040000
	150000
n-dodecane	000474
	074747
	040404
	004040
	100500
n-tridecane	040474
	074740
	040404
	004040
	140005
1-decanol	071472
	072737
	027273
	005010
	170000
1-undecanol	071404
	002720
	020203
	007000
	130000

Qualitative Analysis of Gas Chromatographic Eluates by Means of Vapor Phase Pyrolysis

II. Classification by Set Theory

Charles Merritt, Jr., and D. H. Robertson

Pioneering Research Laboratory, U.S. Army Natick Laboratories, Natick, Mass. 01760

Development of a simple diagnostic function which is easy to calculate and manipulate in a data bank or data storage file has been achieved. The function, taken from the discipline of information theory, is designated as the Khinchine entropy function. Although originally conceived for the special data file problems associated with mass spectrometry, the function is highly useful in differentiating structures on the basis of their pyrolysis patterns alone.

THE AUTHORS of a recent book (*1*) on the subject of identification techniques in gas chromatography have predicted that pyrolysis gas chromatography (pyrograms) will become a standard method of identification of gas chromatographic peaks. Their optimism is not unfounded, since prior reports (*2–4*) demonstrate both that pyrograms are uniquely representative of the parent compound, and may be obtained in a reproducible manner by appropriate control of operating conditions. The complexity of the pyrograms as well as a great similarity in their characteristics has tended to discourage the use of the method as a means of identification. One approach toward simplification of the problem is to employ a combination of retention volume data and "small molecule pyrograms" as described in Part I of this paper (*5*). This part describes a second method which utilizes a mathematical approach to aid in the interpretation of the pyrograms.

In prior work, patterns relating component abundance and peak number or Kovats index (*6*) have been established in the manner of mass spectra (*2, 4*) to aid in the comparison of pyrograms of known compounds with those of unknowns. Likewise, correlations of pyrograms with the structure of molecules (*3, 7, 8*) have helped to improve the interpretation of pyrograms. As with mass spectrometry, however, a method of data processing which provides for classification of the data in a systematic manner is an important prerequisite to its efficient use. In the fields of mass spectrometry and gas chromatography, in particular, a need has been generated for data file searches to accomplish the selection of the correct compound from a collection of many similar data. The manner of this search and the processing of data from the resultant selection has been the subject of several different investigations (*9–11*).

Recently set theory classification has been used for purposes of solving problems of data reduction and processing of mass spectra (*12–14*) especially with the aim of deriving the type of calculation which may be easily performed on a small laboratory computer system.

In this study, a development of set theory—namely, the Khinchine entropy function—is shown to provide diagnostic values when applied to quantitative distribution of pyrolysis products resulting from gas chromatographic separation of these products. With these distribution data, from each pyrolyzed compound, a function may be calculated which shows unique diagnostic characteristics.

Because of the great variation in results from pyrolysis studies, particularly those due to the wide choice of temperature and other operating parameters employed, a set of values obtained in one laboratory is not necessarily equivalent to that obtained in another. However, if the acquisition of data is accomplished with a view toward standardization of operating conditions among laboratories, the utility of statistically derived diagnostic values in providing easily calculated and easily searched files offers great promise when a compound is characterized by intensity values.

The Khinchine entropy function is developed from the specific branch of probability which is concerned with information theory (*15*). In statistical terms, the sum of all the independent probabilities must be equal to the certainty that an event will occur, *i.e.*,

$$\sum_{i}^{n} p_i = 1 \qquad (1)$$

This is achieved in relation to a pyrogram by expressing the abundance of a single chromatographic component as a fraction of the total quantitative abundance of peaks eluted from the gas chromatographic column.

Let us consider an event the probability of which, based upon information available at the time of consideration, is

(1) D. A. Leathard and B. C. Shurlock, "Identification Techniques in Gas Chromatography," John Wiley and Sons, Ltd., London, 1970, p 123.
(2) C. A. M. G. Cramers and A. I. M. Keulemans, *J. Gas Chromatogr.*, **5**, 58 (1967).
(3) E. J. Levy and D. G. Paul, *ibid.*, **5**, 136 (1967).
(4) D. L. Fanter, J. W. Walker, and C. J. Wolf, ANAL. CHEM., **40**, 2168 (1968).
(5) C. Merritt, Jr., and C. DiPietro, *ibid.*, **44**, 57 (1972); "Advances in Chromatography 1971," A. Zlatkis, Ed., Chromatography Symposium, University of Houston, Houston, Texas, 1971, p 174.
(6) E. sz. Kovats, in "Advances in Chromatography," Vol. I., J. C. Giddings and R. Keller, Ed., Marcel Dekker, New York, N.Y., 1965, p 229.
(7) W. D. Dencker and C. J. Wolf, *J. Chromatogr. Sci.*, **8**, 534 (1970).
(8) L. P. Turner and W. R. Barr, *ibid.*, **9**, 176 (1971).
(9) B. Knock, D. Wright, W. Kelly, and R. G. Redley, 17th Annual Conference on Mass Spectrometry and Allied Topics, Dallas, Texas, 1969, p 398.
(10) P. C. Jurs, B. R. Kowalski, T. L. Isenhour, and C. N. Reilley, ANAL. CHEM., **42**, 1387 (1970).
(11) S. L. Grotch, *ibid.*, p 1214.
(12) R. I. Reed and D. H. Robertson, in "Recent Topics in Mass Spectrometry," R. I. Reed, Ed., Gordon and Breach, New York, N.Y., 1971, p 301.
(13) R. I. Reed and D. H. Robertson, in "Recent Developments in Mass Spectrometry," K. Ogata and T. Hayakawa, Ed., University Park Press, Baltimore. Md., 1970.
(14) D. H. Robertson and R. I. Reed, Proceedings of 19th Annual Conference on Mass Spectrometry and Allied Topics, American Society for Mass Spectrometry, Atlanta, Ga., May 1971.
(15) C. E. Shannon, *Bell Syst. Tech. J.*, **27**, 379–423; 623–656 (1948).

Reprinted from ANALYTICAL CHEMISTRY, Vol. 44, Page 60, January 1972

Table I. Detailed Calculation of Khinchine Entropy Function for 2-Methyl-1-Pentene[a]

Intensity values from Pyrogram, A	$A/B = p_i$	$\ln p_i$	$p_i \ln p_i$
60.16	0.23593	−1.44392	0.341
15.14	0.05937	−2.83022	0.168
25.49	0.09996	−2.31264	0.231
100.00	0.39252	−0.93395	0.367
0.8	0.00314	−5.80914	0.003
24.7	0.09686	−2.34341	0.227
24.7	0.09686	−2.34341	0.227
3.2	0.01250	−4.42285	0.055
0.8	0.00314	−5.80914	0.003
254.99			Σ 1.622

Total = B

$\eta = (1.622)(0.43429) = 0.7044$

$0.43429 \ln = \log_{10}$

[a] Data from Reference *17*.

Table II. Khinchine Entropy Values for Selected Alkanes[a]

Compound	Khinchine
n-Hexane	0.572
2-Methylpentane	0.632
3-Methylpentane	0.664
Methylcyclopentane	0.285
2,2-Dimethylbutane	0.475
n-Nonane	0.666
3,3-Diethylpentane	0.699
2,4-Dimethyl,3-ethylpentane	0.919
2,2,5-Trimethylhexane	0.717
2,2,4-Trimethylhexane	0.775
4,4-Dimethylheptane	0.726
3,3-Dimethylheptane	0.719
2,2-Dimethylheptane	0.793
3-Ethylpentane	0.752
2-Methyloctane	0.790
3-Methyloctane	0.756
4-Methyloctane	0.761
n-Decane	0.762
n-Heptane	0.634
n-Octane	0.681
2,2,3-Trimethylpentane	0.796
2,2,4-Trimethylpentane	0.594
2,3,4-Trimethylpentane	0.784
2,3,3-Trimethylpentane	0.841

[a] Data from Reference *4*.

designated as p. In this case the event, p, is the probability of an individual component appearing in the pyrogram. The desired goal is the application of a basic numerical definition that represents the amount of information which it conveys about the event. Good (*16*) makes two demands on such a definition: It should be a decreasing function of p_i, and the amount of information provided by two or more independent events should be the sum of the separate amounts. These conditions are satisfied by functions of the type $-\ln p$. According to the principles of information theory we are following here, the expression

$$\eta = -\sum_{i}^{n} p_i \ln p_i \qquad (2)$$

is defined as the entropy of the experiment; or the Khinchine entropy function. The p_i values represent a fraction A/B where A is the individual intensity of each component being considered and B is the total of all the intensities. Entropy of the experiment is without dimensions and may be considered analogous with entropy as it is customarily defined in statistical mechanics—*i.e.*, the property of the system which is related to probability of state. More specifically, it is the statistical mechanical equivalent of ordinary entropy divided by the Boltzmann constant. A set of relative weight factors is thus expressed which represents the probability of the occurrence of a specific collection of events. Again, the events are the relative intensities of the chromatographic peaks resulting from chromatography of pyrolysis products from pure compounds. When the data showing quantitative distribution of components separated by the chromatographic process are subjected to treatment as a probability distribution (the Khinchine entropy function calculation), the resulting values fall into categories which are diagnostic for the material, the chromatographic profile of which was used for the calculation.

An example of the calculation is presented in Table I. Individual values of $-p_i \ln p_i$ are calculated and summed; the total is the Khinchine function for the compound in question. The data are taken from pyrograms reported by Cramers (*17*) and the value of η is computed by means of an appropriate program written for a Hewlett-Packard Model 2116B computer (The calculation is trivial and may, if desired, be performed on a programmable desk calculator.).

The utility of the entropy function for evaluation of pyrolysis patterns is illustrated in the following examples. The data

Table III. Repeatability of Khinchine Entropy Function for 1-Hexene[a]

(1)	0.668
(2)	0.672
(3)	0.667
(4)	0.668
(5)	0.670
(6)	0.672
(7)	0.672
(8)	0.674
Av.	0.671

$\sigma = 0.003$

[a] Data from Reference *4*.

Table IV. Khinchine Entropy Function Values for Selected Normal Alkenes[a]

Compound	Khinchine
1-Hexene	0.670
2-Hexene[b]	0.694
3-Hexene[b]	0.560
1-Heptene	0.688
trans-2-Heptene	0.691
trans-3-Heptene	0.825
1-Octene	0.697
trans-2-Octene	0.759
trans-3-Octene	0.887
trans-4-Octene	0.829
1-Nonene	0.810
trans-2-Nonene	0.724
trans-3-Nonene	0.765
trans-4-Nonene	0.751
1-Decene	0.802
trans-2-Decene	0.765
trans-3-Decene	0.794
trans-4-Decene	0.787
trans-5-Decene	0.927

[a] Data from Reference *4*.
[b] Compounds of unknown geometric structure. More precise values are seen in Table X.

(16) L. J. Good, "Probability and the Weighing of Evidence," Charles Griffin, London, 1950.
(17) C. A. Cramers, in "Gas Chromatography, 1968," C. L. A. Harbourn, Ed., Institute of Petroleum, London, 1969, p 395.

Table V. Khinchine Entropy Function Values for Selected Branched Alkenes[a]

Compound	Khinchine
Cyclohexene	0.568
3-Methyl-1-pentene	0.758
4-Methyl-2-pentene	0.692
4-Methyl-1-pentene	0.579
3-Methyl-2-pentene	0.351
2-Methyl-2-pentene	0.601
3-Methylcyclopentene	0.843
2-Ethyl-1-butene	0.661
2,3-Dimethyl-1-butene	0.666
3,3-Dimethyl-1-butene	0.648
3,3,5-Trimethyl-1-hexene	0.888
3,5-Dimethyl-3-heptene	0.786
2,3-Dimethyl-2-heptene	0.889
2,6-Dimethyl-3-heptene	0.885
2,2-Dimethyl-3-heptene	0.841
2-Methyl-3-octene	0.792
2-Methyl-2-octene	0.737
2-Methyl-1-octene	0.768

[a] Data from Reference *4*.

Table VI. Khinchine Entropy Function Values for Selected Alkynes[a]

Compound	Khinchine
1-Hexyne	0.822
2-Hexyne	0.615
3-Hexyne	0.511
4-Methyl-1-pentyne	0.684
4-Methyl-2-pentyne	0.698
1-Nonyne	0.914
2-Nonyne	0.954
3-Nonyne	0.884
4-Nonyne	0.919
7-Methyl-3-octyne	0.865

[a] Data from Reference *4*.

Table VII. Khinchine Entropy Function Values for Selected Alkadienes[a]

Compound	Khinchine
1,3-Hexadiene	0.498
1,4-Hexadiene	0.853
1,5-Hexadiene	0.766
2,4-Hexadiene	0.473
1,3-Cyclohexadiene	0.688
2-Methyl-1,3-pentadiene	0.603
3-Methyl-1,3-pentadiene	0.474
4-Methyl-1,3-pentadiene	0.289
2-Methyl-1,4-pentadiene	0.538
2,3-Dimethyl-1,3-butadiene	0.629
3-Ethyl-1,3-butadiene	0.497
2,7-Nonadiene	0.678

[a] Data from Reference *4*.

from which the computations are made are primarily selected from the work of Fanter, Walker, and Wolf (*4*) because of the availability of pyrograms for a large number of compounds having a variety of structural relationships and because of the apparent close control of the parameters by which they were obtained.

Table II shows the values for the Khinchine entropy function for a variety of alkanes. In general, a wide range of values is obtained reflecting the differences in structure. An indication of the reproducibility of the values may be deduced from the repeatability data shown in Table III. Using this criterion, one may assert that the compounds listed in Table II may be distinguished from one another. The entropy function values for the series of normal and branched alkenes, and for alkynes and alkadienes are given in Tables IV through VII, respectively. Again, the values are found to cover a wide range. It may be observed that occasionally there is a correspondence between values for an alkane, or an alkene and one of the other functional classes, but within a given class the values differ sufficiently to provide identification of most of the individual members.

Table VIII. Portion of a Khinchine Function Search File for Hydrocarbon Compounds[a]

Compound	Khinchine
Methylcyclopentane	0.285
3-Methyl-2-pentene	0.351
2,4-Hexadiene	0.473
3-Methyl-1,3-pentadiene	0.474
2,2-Dimethylbutane	0.475
3-Ethyl-1,3-butadiene	0.497
1,3-Hexadiene	0.498
3-Hexyne	0.511
2-Methyl-1,4-pentadiene	0.538
cis-3-Hexene	0.545
trans-3-Hexene	0.551
3-Hexene	0.560
Cyclohexene	0.568
n-Hexane	0.572
4-Methyl-1-pentene	0.579
trans-2,*trans*-4-Hexadiene	0.587
2,2,4-Trimethylpentane	0.594
2-Methyl-2-pentene	0.601
2-Methyl-1,3-pentadiene	0.603
2-Hexyne	0.615
2,3-Dimethyl-1,3-butadiene	0.629
2-Methylpentane	0.632
n-Heptane	0.634
3,3-Dimethyl-1-butene	0.648
2-Ethyl-1-butene	0.660
3-Methylpentane	0.664
n-Nonane	0.666
2,3-Dimethyl-1-butene	0.666
1-Hexene	0.670
2,7-Nonadiene	0.678
n-Octane	0.681
4-Methyl-1-pentyne	0.684
1,3-Cyclohexadiene	0.688
1-Heptene	0.688
4-Methyl-2-pentene	0.692
2-Hexene	0.694
4-Methyl-2-pentyne	0.698
3,3-Diethylpentane	0.699

[a] Data from Reference *4*.

It is expected that the Khinchine function values will be utilized in a computer search file as a single valued diagnostic number to identify unknown compounds. A representation of a portion of such a search file covering the Khinchine values from ~0.2 to ~0.7 is presented in Table VIII. Within the limits of precision established from the reproducibility data (Table III), in most cases a single compound will be identified by the number. In six cases a pair of compounds would be selected, but the structure of the compounds is sufficiently different in these instances to permit the identification to be made from the correlation with component composition of the pyrogram.

It is most impressive to observe the wide divergence of values which represent compounds of very close similarity in structure. This may be observed for certain pairs or groups such as the 2-methyl- and 3-methyl-pentanes or, the 2-, 3-, and 4-methyl octanes (Table II); the 2 methyl-, 1-, 2-, or 3-

Table IX. Pyrogram Patterns for *cis*- and *trans*-2-Hexene[a]

Compound	Retention number[b]						
	1	2	3	4	6	8	10
trans-2-Hexene	70	100	24	78	9	69	9
cis-2-Hexene	71	100	17	82	10	71	10

[a] Data from Reference *4*.
[b] Adapted from retention index interval *4*.

Table X. Khinchine Entropy Function Values for Selected Cis/Trans Hexenes and Hexadienes[a]

Compound	Khinchine
trans-2-Hexene	0.734
cis-2-Hexene	0.727
trans-3-Hexene	0.551
cis-3-Hexene	0.545
trans-2-*trans*-4-Hexadiene	0.587
cis-2-*trans*-4-Hexadiene	0.631
cis-2-*cis*-4-Hexadiene	0.607

[a] Data from reference *4*.

Table XI. Khinchine Entropy Function Values for Selected Nucleosides, Nucleotides, and Related Compounds[a]

Compound	Khinchine
Adenine	0.859
Guanine	0.752
D-Ribose	0.794
Uracil	0.928
Cyclosine	0.856
Thymine	0.949
Adenosine	0.862
Guanosine	0.849
Cytidine	0.969
Uridine	0.904
Thymidine	0.911
5'-AMP	0.880
5'-GMP	0.871
5'-CMP	0.876
5'-UMP	0.932
Adenylyl (3' → 5') adenosine	0.887
Uridylyl (3' → 5') adenosine	0.929
Cytidylyl (3' → 5') adenosine	0.864
Cytidylyl (3' → 5') uridine	0.923
Uridylyl (3' → 5') cytosine	0.867
Guanylyl (3' → 5') adenosine	0.861
Adenylyl (3' → 5') guanosine	0.850
Adenylyl (3' → 5') uridine	0.791

[a] Data from Reference *8*.

octenes (Table V); and the 1,3-, 1,4-, and 1,5-hexadienes (Table VII).

The capability to discern geometric as well as positional isomerism is likewise impressive. The cis/trans hexenes, for example, are seen to have very similar pyrogram patterns (Table IX) and according to Fanter, Walker, and Wolf (*4*) cannot be identified. In Table X, however, the value of the Khinchine function is seen to distinguish easily between the two isomers. Further examples of the identification of cis/trans isomers are also seen in Table X.

The Khinchine entropy function may also be applied to pyrograms resulting from the pyrolysis of solids. Values calculated from data obtained for some purine bases, nucleosides, nucleotides, and related compounds are given in Table XI. As with the pyrograms obtained from vapor phase pyrolysis, the values are unique and likewise diagnostic.

Table XII. Khinchine Functions for Data from Different Laboratories (Calculated from Pyrolysis Patterns)

Compound	[a]	[b]	[c]
n-Hexane	0.432	0.399	0.441
Hexene-1	0.670	0.656	
Hexene-2	0.695	0.759	
Hexene-3	0.560	0.604	
2,3-Dimethylbutene-1	0.666	0.669	
3,3-Dimethylbutene-1	0.6476	0.679	
Heptene-1	0.583		0.536
Octene-1	0.604		0.586

[a] Reference *4*.
[b] References *2*, *17*.
[c] Reference *3*.

Obviously the value of η as a diagnostic value cannot by its calculation from the original data improve their reproducibility. An enhancement is achieved which enables closer distinctions between compounds with very similar pyrogram patterns. Most data reported herein are based on hydrocarbon pyrograms. It is questionable that, without modification, this diagnostic function will be equally useful for study of compounds of other functionality. Preliminary investigation with other special functions suggests that their use in conjunction with the Khinchine function will provide a very powerful diagnostic tool.

Entropy function values derived from pyrograms obtained in different laboratories may be expected to vary because of the difference in temperature and other operating conditions employed in the pyrolysis device. The extent of this variation is seen in Table XII. If data are to be compared from one laboratory to another, reliable, reproducible, standardized pyrolysis procedures must be employed. As the values of pyrolysis patterns, expressed in terms such as the Khinchine entropy function, become better appreciated for the identification of GC eluates, the incentive to develop the required standardization should appear. The efficacy of a single valued diagnostic criterion to employ in search of data banks will then not only be valid for analyses conducted within one laboratory, but for many laboratories.

The approach as presently conceived suffers one serious limitation—namely, the inability to cope with any but pure components. In addition to the obvious solution of improving GC separations, it is expected that further consideration of information theory may provide for deconvolution and subsequent identification of partially resolved eluates.

Received for review July 16, 1971. Accepted September 3, 1971.

Reprinted from

RECENT DEVELOPMENTS IN MASS SPECTROSCOPY

Proceedings of the International Conference
on Mass Spectroscopy, Kyoto

edited by
KOREICHI OGATA
and TERUO HAYAKAWA

UNIVERSITY OF TOKYO PRESS
1970

Naive Analysis of Structure☆

R. I. Reed and D. H. Robertson

Department of Chemistry, University of Glasgow, Scotland, U. K.

Modern mathematical analysis has been used to advantage in the area of communication theory and in processing and reducing large amounts of data from various fields of analysis.

In this presentation, rudimentary set theory has been applied to interpretation of the mass spectra of alkanes; its utility for alkenes and alkynes has also been demonstrated. The ions produced in a mass spectrometer may be considered as subsets or (in the case of an individual ion, as a single member) of the universal set of all possible ions which may arise from the fragmentation of organic molecules. For the sake of simplicity, initial studies have dealt with saturated hydrocarbons and for the purpose of introducing the concept, only the simple case of the saturated hydrocarbons is considered.

Modern data processing is of course greatly dependent upon the use of computers; because most computers operate in a binary code, it was desirable to develop a method of analysing mass spectra which would incorporate the binary principle. For this purpose an index has been developed which is "1" if the ion under consideration has diagnostic value in the spectrum and "0" if it does not. In addition, the principles of set theory are sufficiently well developed to allow a much fuller application to mass spectral analyses than has been presented in these pages. A systematic development of these concepts relative to mass spectrometry is not possible within the scope of this article.

Therefore it is intended to introduce two approaches to naive analysis of structure: one dealing with the analysis of existing spectra; the other with synthesizing spectra from a known structure.

Before we actually apply set theory to the analysis of existing spectra, certain assumptions must be made, relevant to alkane analysis.

1. Fragmentation of a parent ion is favored at points of chain-branching with preferred elimination of the largest branch.
2. Ions of even mass often arise in association with such fissions, due to concomitant hydrogen rearrangements.
3. The parent molecular ion of a branched-chain alkane is less abundant than that of the isomeric straight-chain alkane.

It is intended to investigate, within the confines of set theory, how far these rules can be used to determine the overall structure of an unknown alkane. Within this framework, it is necessary to cite three problems which must also be considered:

1. Isomeric ions may be formed, in which case the observed ion-current for the given mass will be the sum of the individual isomeric ions.
2. Since the use of set theory implies a collection of all possible ions which can be formed from a given structure, regardless of whether or not they are actually produced in the mass spectrometer, suitable operations must be employed to remove undesirable or unacceptable structures.
3. Complications may arise owing to the occurence of sequential reactions; however, allowance for them cannot be made without some assumptions about the structure which is sought.

☆ Presented by D. H. Robertson.

The treatment of saturated hydrocarbons is based on the observation that the major differences among their spectra occur in the branched-chain compound of a given carbon number. Therefore, one is provided with a means of picking out the branching points in an isomeric alkane by comparing its principal ions of the form C_nH_{2n+1} with its counterpart in the straight-chain compound. A similar comparison is made in the case of alkenes and alkynes where ions of the form C_nH_{2n} and C_nH_{2n-2} respectively, have diagnostic significance. Based on the assumption that total ion current is the same for each isomer, intensities of ions formed by fragmentation at a branch will be greater than the same empirical ion from the straight-chain reference standard.

A binary oriented ratio (already mentioned) is employed which works out to be "1" or "0", depending upon whether the ion does or does not have diagnostic value. Ratios with a value of "1" are found to have a symmetrical distribution around the center of the molecule; this center is n/2 for even-numbered carbon compounds and between (n-1)/2 and (n+1)/2 for odd-numbered compounds.

The ionization process for the molecule XY will be considered as:

$$XY + e \longrightarrow X^+ \quad Y^{\cdot} + 2e \quad \text{or} \quad X^{\cdot} \quad Y^+ + 2e$$

It is assumed that bond dissociation energies of the carbon-carbon bonds in neutral alkane molecules occur in decreasing order: primary, secondary, tertiary and thus the concentration of X and Y radicals would be greater for the branched-chain structure, assuming concentrations of straight- and branched- chain species to have been the same before ionization. Abundances of the branch-chain structure are represented by $[X_E^+]$, those of the straight-chain reference compound by $[X_K^+]$. Based on bond energies we may say.

$$\frac{[X_E^+]}{[X_K^+]} \geq 1$$

The same expression may be written for Y:

$$\frac{[Y_E^+]}{[Y_K^+]} \geq 1$$

Thus, ion abundances ratios which are greater than unity will appear in pairs, and the sum of their masses will equal the mass of the parent molecular ion. For purpose of developing a notation system, two further points must be considered:

1. Ionization cross section depends to a great extent upon constitution of a molecule rather than upon its structure.
2. Total ion current represents a reasonable approximation of the ion cross section.

As the ion cross section and total ion current in the isomeric alkanes would be the same, for the purposes of the argument the observed values are treated as if they were mathematically exact. We therefore, in the case where differences are observed, scale the ion currents of the branched-chain alkanes up or down to bring them into line with those of the n-alkanes; in other words, the branched-chain molecules are normalised to the straight-chain molecule having the same number of carbons.

For a general analysis, the series of alkyl ions and their associated abundances have been considered the most useful. Not only can they be assigned in pairs by the relationship $P^{+\cdot} = X^+ + Y^+$, but they are usually the most important of all the ions associated with a given carbon number.

The following notation is hereby adopted:

$$C_nH_{2n+2}^{+\cdot} = P^{+\cdot} = (C_\ell H_{2\ell+1} + C_mH_{2m+1}) = \ell + m$$

which in the shorthand nomenclature becomes

$$\frac{[C_\ell H^+_{2\ell+1};\text{BRANCHED}]}{(\Sigma;\text{BRANCHED})} \times \frac{(\Sigma;\text{STRAIGHT})}{[C_\ell H^+_{2\ell+1};\text{STRAIGHT}]} = r_\ell = \frac{\ell_E}{\ell_K} \times \frac{\Sigma K}{\Sigma E}$$

Finally, we adopt the index q_ℓ which is assigned the value of unity if r_ℓ is equal to or greater than one and a value of zero if r_ℓ is less than one.

A condensed form of the nomenclature for 3-methylheptane appears in the following figure where the straight-chain reference compound, ℓ_k is n-hexane and the branched-chain hydrocarbon is represented by ℓ_E.

Table 1

C_nH_{2n+1}	ℓ	ℓ_E	ℓ_K	r_ℓ	q_ℓ
15	1	3.90	3.00	1.08	1
29	2	42.50	34.50	1.02	1
43	3	100.00	100.00	0.83	0
57	4	67.30	34.20	1.64	1
71	5	3.05	28.30	0.09	0
85	6	48.60	29.50	1.37	1
99	7	0.76	0.07	9.03	1
114	(8)	2.99	6.74	0.37	(1)
Σ		485.26	403.71		

$$r_\ell = \frac{\ell_E}{\ell_K} \times \frac{\Sigma_K}{\Sigma_E} = \frac{\ell_E}{\ell_K} \times \frac{403.71}{485.26}$$

Distribution of q values is now further developed. If molecular species of the general form $C_nH_{2n+1}^{+\cdot}$ are arranged equidistantly along an axis they should group in pairs:

$\ell = 1$ and n-1
$\ell = 2$ and n-2 etc.

The ℓ values which are equidistant from the center of the molecule should have the same q values. In the case of 3-methylheptane, for instance:

$\ell = 2, \quad q_2 = 1$

$\ell = 6, \quad q_6 = 1$

both of which are equidistant from the center of the molecule at carbon number four. In the determination of the structure of the molecule, one derives $\ell + m$, where $1 < \ell$, $m < n/2$ or $(n-1)/2$. In the case of 3-methylpentane, there are three such pairs: 4, 4; 2, 6; 1, 7 each pair of which represents q_ℓ values of unity or "1."

The critical operation in this approach to mass spectral analysis appears to be the selection of the reference compound (i.e. the n-hexane in the example just given) with which the "unknown" is compared. Although reasonable success has been achieved by using reference compounds corresponding to the straight-chain compound of the same carbon number, it is not completely appreciated how close in structure the "unknown" and reference compounds must be to allow a satisfactory analysis.

Application of set theory also allows one to generate the possible structure of ion groupings from an unknown. These groups in the case of the alkanes are of the form $\cdot CH_3$, $\cdot C_2H_5$ etc. and shall be represented by the arabic numerals 1, 2 etc. If a radical appears more than once in a single molecule, it is given a separate notation for each time it appears.

For example, in 2, 2', 3-trimethylpentane, there are four different methyl groups which may be uniquely identified as 1, 1', 1'', 1'''. This collection of symbols represents subsets of the complete or universal set represented by P or the appropriate carbon number; this number would be '8' of course in the case of the octanes. In addition, a null or empty set containing no elements, must be considered. Each symbol (1, 1', 1'', 1''', etc.) represents a proper subset of the universal set and there are also two improper subsets, the null set and the universal set itself. With the exception of the null set, each set, or subset contains elements which are identified with the number of carbons contained in the radical under consideration.

Thus, trimethylpentane is represented by the universal set 8 as well as by at least one subset 4; these sets contain 8 and 4 elements respectively. In a system of finite order, an important relationship exists between sets and subsets.

$$A \smile B \smile \cdots\cdots \smile N = \sum_{A}^{N} A - \sum_{AB}^{MN} {}^{n}C_2(A \frown B)$$

$$+ \sum_{ABC}^{LMN} {}^{n}C_3(A \frown B \frown C) \cdots\cdots (-1)^{n-1}[A \frown B \cdots\cdots \frown N]$$

Capital letters represent sets or subsets, $\smile$ represents union of sets and $\frown$ represents the intersection of sets.

Again, referring to 2, 2', 3-trimethylheptane in which the methyl group occurs four times, we depict in Fig. 1 a set representation of this compound:

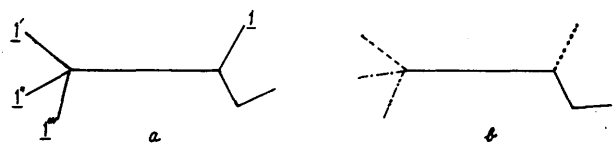


Fig. 1

automatically, four distinct heptyl groups are generated, represented by 7, 7', 7'', and 7'''. In the figure on the right, the solid line represents the portion of the structure which is common to each of the heptyl groups; i.e. a butyl chain. 7 represents the total structure less the methyl indicated by the dotted line; 7', the total structure less the methyl group indicated by the dashed line etc.

In order to make further analysis easier, we introduce the concept of set intersection. Since it is clear that none of the subsets 1, 1', 1'', 1''' can intersect with each other, we may say that the intersection of 1 with 1' for instance, is the null set.

However, this condition does not apply to members of the subset which is represented by the numeral 7 because the intersection of 7 with 7' will include all of the molecule which is common to both.

Dealing with the methyl radicals can be simplified by considering the property of sets which states that the intersection of any set with the null set is itself the null set. Since all intersections of the subsets 1, 1', etc., taken two at a time, yield only the null-set, any intersection of three or more subsets will be equivalent to intersection of an individual set with the null set.

The situation with the heptyl radicals is different; taken three at a time, the intersection of 7, 7', and 7'' yields a subset with five carbon atoms. An illustration of this application follows:

Consider a set A which is associated with a family of subsets, B_1, B_2, etc., having the following properties:

a. A is a union of sets, B_1, B_2, etc.

b. for a given pair of subsets, say B_1 and B_2, either $B_1 = B_2$ or the intersection of B_1 and $B_2 = 0$.

Such a family of sets is called a partition of A. In octane, for instance, the subsets 4 and 4' are a partition of P for 4 intersection 4' equals 0 and one of these is branched at the point of attachment.

Since subsets 1 and 2 were observed in the actual analysis, we may write:

$$\underline{1} \cup \underline{2} \cup \underline{4} = \underline{1} + \underline{2} + \underline{4} - \sum_{1,2}^{2,4} {}^{n}C_2(1{:}2) + \underline{1}{:}\underline{2}{:}\underline{4} = 7 - \phi - 1{:}4 - 2{:}4 + \phi$$

and since $1{:}2 = \phi \quad \therefore 4 = 7 - \underline{1}{:}\underline{4} - 2{:}4$ or $3 = \underline{1}{:}\underline{4} + \underline{2}{:}\underline{4}$

The intersection of 1 with 4 contains one element only; therefore the intersection of 2 with 4 contains two elements and the structure for the butyl group appears. Other possible subsets may be deduced in a similar way.

Detailed investigation of several aspects of set theory are at present in progress at the University of Glasgow; further development will be published as they are available.

Some considerations of the naive analysis of structure

R. I. REED and D. H. ROBERTSON

Department of Chemistry, University of Glasgow
Glasgow W2, Scotland, U.K.

It is always easier to explain the presence of a particular peak in a mass spectrum than to predict the main peaks to be expected in the spectrum of a particular compound and it is only when this latter problem is attempted that the naïveté of our present level of understanding is fully realised.

JOHN BEYNON

The use of mathematical analysis is well established in the fields of communication theory and data reduction; computer programmes are available for handling this kind of information and the expertise whereby the technique may be applied to practical problems has reached a highly developed state. The authors of this presentation have been concerned with the possibility of applying one or more of these mathematical methods to the analysis of mass spectra. Using fuzzy set theory as a starting point, a straightforward method is developed whereby mass spectral data can be manipulated to provide diagnostic criteria for the various ions in the spectrum under consideration.

The problem of identification of an unknown by comparison with a data bank of mass spectra has received considerable attention. The specific technique employed for the retrieval depends whether one is working with high or low resolution data. A promising beginning has been made with respect to low resolution spectra[1]; the situation with high resolution is con-

siderably more difficult to define. Most recently the principle of artificial intelligence has been applied to these analyses[2]. In the latter case the computer is provided with a programme which allows it to estimate the agreement of the unknown spectrum when it is compared against a so-called training set of spectra; the computer is thus able to decide to which classification the unknown belongs.

Inasmuch as modern mathematical techniques are highly dependent upon the use of computers for the purpose of reducing and processing gross quantities of otherwise unwieldy data, it has been considered desirable to develop a method of mass spectral analysis which would incorporate the binary principle. Appropriate use of set theory allows this to be realised. Some of the investigations, especially those involving cycloalkanes, suggest that strict adherence to a binary classification may not be feasible.

In its simplest form, naive analysis allows one to consider the ions produced in a mass spectrometer as subsets or (in the case of an individual ion) as a member of the universal set of all possible ions which may arise from the fragmentation of an organic molecule. With this basic premise we have used naive set theory successively to examine alkanes, alkenes and cycloalkanes. As the work has progressed it has been found that the simple criterion which was assumed for alkanes does not completely satisfy the latter two classifications of hydrocarbons. The problem of a criterion or criteria upon which a mass spectral analysis is based has not been solved with complete satisfaction; however, the development of these applications will be presented as they have occurred in the laboratory. How they may be used at present in spectrum analysis will be indicated and areas for future investigation will be suggested, whereby application of the technique may be extended to further classes of organic compounds.

The source of data in each case is the A.P.I.[3] file of mass spectra, the use of which assumes the comparison of an unknown with one or more reference compounds. The comparison is made ion for ion between unknown and reference compound. In the ideal case, the structure of the unknown may be deduced by determining whether or not it has a positive correlation with an ion in the reference compound. (It is implicitly assumed that each ion which appears in the spectrum of the reference compound has a particular relationship to the structure of the molecule from which it comes.) There is variable success in "predicting" an unknown structure depending on which reference compound is used and what criterion (criteria) is (are) employed. The reference compound which has been selected for the initial analyses is

the straight-chain alkane containing the same number of carbons as the "unknown" structure with which it is compared. Although some experimentation has been made with the use of different compounds, the results are far from definitive. A somewhat more detailed discussion of this problem is treated in the section on alkenes.

Inasmuch as we can freely define what constitutes the membership of a set, the potential inherent in the use of naive set analysis is great. The universal set may be looked upon as containing all the possible ions resulting from fragmentation in the mass spectrometer of all the known organic compounds (we tacitly assume that the treatment is restricted to the realm of organic chemistry). In the simplest situation which can be imagined, one could select a subset from the universal set which would embody all the ions necessary to provide a unique diagnostic for a given class of compound i.e. set A is composed of all the ions which are required to identify unambiguously the class, hydrocarbons. The set A will likewise be subject to subdivision into further subsets, which represent, in the case of the above example, such classifications as alkanes, alkenes, alkynes, cycloalkanes and cycloalkenes.

There are some pitfalls which one may fall into when making the translation from naive set theory to more rigorous considerations. Among these are the axiom of choice, Zorn's lemma and the well-ordering theorem. However, at the present point of development there is need for little more than an awareness of these pitfalls. As the set theory approach is formalised in future work mere awareness will not be enough.

ALKANES

Before the actual manipulation of data is begun, it is necessary to provide a series of assumptions which in this initial case are applicable to alkanes in particular. Later in this article we shall discuss their application to the alkenes and cycloalkanes.

These initial assumptions (relative to alkanes) are:

1. fragmentation of a parent ion is favoured at points of chain branching with preferred elimination of the largest branch,

2. ions of even mass often arise in association with such fissions, due to concomitant hydrogen rearrangements, and

3. the parent molecular ion of a branched chain alkane is less abundant than that of the isomeric straight-chained alkane.

It is also necessary to consider the fact that isomeric ions may be formed, in which case the observed ion currents (as obtained in these calculations by the addition of the individual normalised ion intensity values taken directly from the A.P.I. data sheets) for any given mass will be represented by the sum of the individual isomeric ions.

In so for as the use of set theory implies a collection of all possible ions which can be conceived as being formed from a given structure, regardless of whether or not they are actually produced in the mass spectrometer, a criterion must be adopted for rejecting unacceptable ions i.e. those which do not actually appear in the list of m/q* ratios in the A.P.I. tables of spectra. In addition there is also the possibility of sequential reactions occurring; however, it is obviously not possible to make any allowances for this phenomenon without first assuming some information about the structure which one is attempting to determine. In the case of an actual analysis, structural information about the unknown would of course be absent. As the rationale behind the technique is to provide a means of comparison between reference compound and unknown which will allow identification of the latter, such a possibility shall not be considered in this treatment. Thus it is that the ions which result in a given spectrum are used for analysis without scrupulous consideration of their origin.

The approach to saturated hydrocarbons (alkanes) is based on the observation that the major differences which occur among their respective spectra will be found in the branched chain compound of a given carbon number. It is in this manner that a means of locating the branching point in an isomeric alkane is obtained i.e. by comparing its principal ions of the general formula $C_nH_{2n+1}^+$ with the comparable ions in the straight chain compound.

Based on the assumption that total ion current is the same for each isomer, intensities of ions formed by fragmentation at a branch will be greater than the same empirical ion from the straight chain reference standard. Analogous behaviour has been observed with regard to the alkenes.

A binary oriented ratio is employed which works out to be "1" if the ion has a diagnostic value relative to the reference compound and "0" if it does not. In terms of our set theory treatment, it is observed that ratios with a value of "1" are found to have a symmetrical distribution around the centre of the molecule; this centre (again in terms of set theory and for the $n-1$ ion, where n = number of carbons) is at $n/2$ for ions from odd numbered

* The use of m/q follows the suggestion of the American Institute of Physics.

carbon compounds. It should be noted in regard to the f values, which are subsequently developed, that their distribution is likewise symmetrical around the centre of the molecule. (Please see note in Table 1.)

The ionization process for the molecule XY is considered as:

$$XY + e \rightarrow X^+ + \dot{Y} + 2e \quad \text{or} \quad \dot{X} + Y^+ + 2e$$

Table 1

C_nH_{2n+1}	l	I_E	I_K	r_l	f_l
15	1	3.90	3.00	1.08	1
29	2	42.50	34.50	1.02	1
43	3	100.00	100.00	0.83	0
57	4	67.30	34.20	1.64	1
71	5	3.05	28.30	0.09	0
85	6	48.60	29.50	1.37	1
99	7	0.76	0.07	9.03	1
114	(8)	2.99	6.74	0.37	(1)
Σ	·	485.26	403.71	·	·

$$r_l = \frac{I_E}{I_K} \cdot \frac{\Sigma_K}{\Sigma_E} = \frac{I_E}{I_K} \cdot \frac{403.71}{485.26}$$

K = the known reference standard; in this case it is n-octane.
E = the unknown compound whose structure is sought; here, it is 3-methylheptane.
Σ = Total ion current taken as the sum of normalised individual ion intensities.
I_E = an ion intensity (unknown).
I_K = an ion intensity (reference compound).
Values for f_l which are diagnostic for the ions are distributed symmetrically around the value for C_4, which represents both C_4 fragments arising from bond cleavage between C_4 and C_5 (n-octane) and/or C_3–C_4 (3-methylheptane).

It is assumed that bond dissociation energies of the carbon–carbon bonds in neutral alkane molecules occur in decreasing order: primary > secondary > tertiary and thus the concentration of X and/or Y radicals would be greater for the branched chain structure, assuming the concentration of straight- and branched chain species to have been the same before ionization. Abundance of the branched-chain ion is represented by $[X_E^+]$, that of the straight chain reference compound by $[X_K^+]$. Based on bond energies, we may say that the ratio of branched- to straight-chained structures is equal to or greater than "1", i.e. $X_E^+/X_K^+ \geq 1$. The same type of expression may be written for the "Y" ion, i.e. $Y_E^+/Y_K^+ \geq 1$. Thus, in general, ion

abundance ratios which are greater than unity will appear in pairs, and the sum of their masses will of course equal the mass of the parent molecular ion (see Table 1).

For the purposes of this investigation it shall be assumed that ionization cross section depends to a greater extent upon constitution of a molecule than upon its structure and that total ion current represents a reasonable approximation of the ion cross section. Ötvos and Stevenson[4] suggest that there is simply a proportionality between cross section and total ion current. It will be pointed out in the discussion of cycloalkane analysis that for the purpose of obtaining diagnostic values, the original approximation is satisfactory.

Ideally, the ion cross section (σ) and total ion current (Σ) for the isomeric hydrocarbons would be the same. For the purpose of the argument the observed values are treated as if they were mathematically exact, although in actual fact σ is only approximately so for different isomers. Therefore, in the case where differences are observed, the ion currents of the branched-chain alkanes are scaled down to bring them into line with those of the *n*-alkanes; in other words, the branched chain molecules are normalized to the straight chain molecule having the same number of carbons.

For a general analysis, the series of alkyl ions and their associated abundances have been considered the most useful. Not only can they be assigned in pairs by the relationship $P = X^+ + Y^+$ but they are usually the most important of all the ions associated with a given carbon number.

The following simple notation is hereby adopted which underlies the connexion between the ion formed from the parent:

$$C_nH_{2n+2}^{+\cdot} = P^{+\cdot} = (C_lH_{2l+1} + C_mH_{2m+1})^{+\cdot} = l + m$$

The shorthand nomenclature for normalizing to the straight chain compound is as follows:

$$\left(\frac{l_E}{l_K}\right)\frac{\Sigma_K}{\Sigma_E} = \frac{[C_lH_{2l+1}^+ \text{ (branched)}]}{\Sigma \text{ (branched)}} \times \frac{\Sigma \text{ (straight chain)}}{[C_lH_{2l+1}^+ \text{ (straight chain)}]} = r_l$$

Finally, we adopt one additional index f_l which is assigned the value of "1" if r_l is equal to or greater than "1" and a value of "0" if r_l is less than "1".

A condensed form of the nomenclature, using 3-methylheptane as an example is given in Table 1. Again, only the "*l*" ion is represented. The reference compound, denoted by subscript *K*, is *n*-octane.

Distribution of the f_l values is now further developed. If ionic species of the general form $C_nH_{2n+1}^+$ are arranged equidistantly along an axis, they should group in pairs e.g.

$$l = 1 \quad \text{and} \quad l = n - 1$$

$$l = 2 \quad \text{and} \quad l = n - 2$$

based on the fact that a compound of carbon number n produces an ion of carbon number l and an ion of carbon number $n - l$. The values of l which are equidistant from the centre of the molecule should have the same f_l values. In the case of 3-methylheptane,

$$l = 2, \quad f_2 = 1$$

$$l = 6, \quad f_6 = 1$$

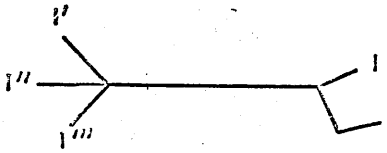


Figure 1

both of which are equidistant from the centre of the molecule which is at carbon 4. (See note in Figure 1.) In the determination of the structure of the molecule, one derives $l + m$, where $1 < l$, $m < n/2$ or $(n - 1)/2$ and $m = n - l$. Again referring to 3-methylheptane, there are three such pairs:

4,4; 2,6 and 1,7

The use of set theory also allows one to generate the possible structure of ion groupings from an unknown. These groups are of the form $\cdot CH_3$, $\cdot C_2H_5$, etc. and are represented by the arabic numerals *1*, *2* etc., corresponding to the carbon number. If a radical appears more than once in a single molecule, it must be given a separate notation each time it appears. For example, in 2,2′ 3-trimethylpentane, there are four different methyl groups (as indicated in Figure 1) which may be uniquely identified as 1, 1′, 1″, 1‴. This collection of symbols represents subsets of the complete or universal set which is itself represented by *P* or the appropriate carbon number; this number would of course be *8* in the case of the octanes. In addition, a null or empty set which contains no elements must be considered.

Each symbol (1, 1′, 1″, 1‴, 2, 3, etc.) represents a proper subset of the universal set; there are also two improper subsets, the null set and the universal set itself. With the exception of the null-set, each set or subset contains elements which are identified with the number of carbons contained in the radical under consideration. Thus, trimethylpentane is represented by the universal set *8* as well as by at least one subset *4*; these sets contain *8* and *4* elements respectively.

$$A \cup B \cup \cdots \cup N = \sum_{A}^{N} A - \sum {}^{n}C_{2}\,(A \cap B)$$

$$+ \sum_{ABC}^{LMN} {}^{n}C_{3}\,(A \cap B \cap C \cdots$$

$$(-1)^{n-1}\,[A \cap B \cdots \cap N]$$

Figure 2

In a system of finite order, an important relationship exists between sets and their subsets. Capital letters represent sets or subsets, $\cup$ represents the union of sets and $\cap$ represents the intersection of sets. Again referring to 2,2′,3-trimethylpentane in which the methyl group occurs four times, we depict in Figure 3 a set representation of the compound. Automatically, 4 distinct heptyl groups can now be generated; these are represented by 7, 7′, 7″ and 7‴. (Since there are four distinct methyl groups, there will be an equal number of heptyl groups i.e. groups of the form $P—CH_3$, where P denotes the parent. In Figure 3, the solid line represents the portion of the structure which is common to each of the heptyl groups; i.e., a butyl chain. The numeral *7* represents the total structure less the methyl indicated by the dotted line; 7′, the total structure less the methyl group indicated by the dashed line, etc.

In order to make further analysis easier, we introduce the concept of set intersection. Since it is clear that none of the subsets 1, 1′, 1″, 1‴ can inter-

Figure 3

sect with each other, we may say that the intersection of 1 with 1′, for instance, is the null set. However, this condition does not apply to members of the subset represented by the numeral 7 because the intersection of 7 with 7′ will include that part of the molecule which is common to both.

Consideration of the methyl radicals can be simplified by considering the property of sets which states that the intersection of any set with the null set is itself null. It is clear that the subsets 1, 1′, 1″ and 1‴ cannot intersect one with the other, so that $1 : 1' = \emptyset$ for all intersections of methyl groups; $\sum_{1}^{1'''} {}^{n}C_2 (1 \cap 1') = \emptyset$.

The situation with the heptyl (C7) radicals is different; taken three at a time, the intersection of 7, 7′ and 7″ yields a subset with 5 carbon atoms, as seen in Figure 4. An illustration of this application follows:

$7 \cap 7' \cap 7''$

Figure 4

Consider a set A which is associated with a family of subsets $b-1$, $b-2$, etc., having the following properties:

1. A is a union of sets, for instance, b and c.

2. For any given pair of subsets b and c, either b equals c or $b \cap c$ equals $\emptyset$. Such a family of sets is called a partition of A. In octane, for instance, the subsets 4 and 4′ are the partition of P since $4 \cap 4'$ equals $\emptyset$.

Inasmuch as subsets 1 and 2 were observed in the actual analysis, we may write:

$$1 \cup 2 \cup 4 = 1 + 2 + 4 - \sum_{1,2}^{2,4} {}^{n}C_2 (1{:}2)\ 1{:}2{:}4$$

$$= 7 - \emptyset - 1{:}4 - 2{:}4 + \emptyset$$

Since $1{:}2 = \emptyset \quad 4 = 7 - 1{:}4 - 2{:}4$

or $3 = 1{:}4\ 2{:}4$

Figure 5

The intersection of 1 with 4 contains one element only, therefore 2 ∩ 4 contains two elements and the structure for the butyl or C_4 group emerges. Other possible subsets may be deduced in a similar way.

ALKENES

When olefines are considered, two features of the molecule become important from the standpoint of naive analysis: the points of chain branching and the location of the double bond. The following rules may be used with both alkanes and/or alkenes:

1. the parent molecular ion is biggest for the straight chain isomer
2. loss of a C_1 fragment is likely only for a CH_3 side chain
3. fragmentation is in general most likely at highly branched carbons
4. C_3 and C_4 fragments are always prominent in alkane spectra.

Owing principally to the presence of the unsaturation, ions from an alkene are more stable than and occur with greater relative intensities than analogous ions from alkane structures. It is also observed that the abundance of the parent peak in an alkane, relative to its fragment peaks, is greater than in the case of saturated hydrocarbons. It is unfortunate for the purposes of establishing rules for alkenes, that the latter trend is found to reverse itself above C_6; this may be related to the observation that low mass unsaturated ions are more stable than the saturated ions of comparable mass (i.e. comparable ions from paraffins). Generally speaking, one may say that the preferred point of fragmentation in the olefine molecule is allylic to the double bond. In addition there exists a greater tendency for rearrangement to occur in the mass spectrometric analysis of unsaturated hydrocarbons with the net result that their mass spectra are less sensitive to variations in molecular structure. Given these considerations, a series of olefine spectra will be studied (olefines which are structually similar) with attention to the 41 and 55 peaks and the two series:

a) 41, 55, 69, 83, etc.

b) 42, 56, 70, 84, etc.

The series of ions of the form C_nH_{2n-1} (a, above) predominates but with a decrease in magnitude with increasing molecular weight; there is at the same time an increase in prominence of the molecular ion.

Table 2

Compound (A.P.I. no.)	PAIR	a	b
99	1,5	+, +	−, +
	2,4	+, +	−, +
	3,3	0.73, 0.73	0.53, 0.53
100	1,5	+, +	+, +
	2,4	+, +	+, +
	3,3	+, +	+, +
101	1,5	+, +	+, +
	2,4	+, +	+, +
	3,3	+, +	+, +
102	1,5	+, +	+, +
	2,4	0.86, +	0.77, −
	3,3	+, +	+, +
103	1,5	+, +	+, +
	2,4	−, +	−, −
	3,3	+, +	+, +
104	1,5	+, +	+, +
	2,4	0.88, +	0.78, +
	3,3	+, +	+, +
276	1,5	+, +	+, +
	2,4	0.85, +	0.77, +
	3,3	+, +	+, +
278	1,5	+, +	0.84, +
	2,4	−, +	−, −
	3,3	+, +	+, +
279	1,5	+, +	+, +
	2,4	−, +	−, −
	3,3	+, +	+, +
399	1,5	+, +	+, +
	2,4	0.99, +	0.88, +
	3,3	+, +	+, +
524	1,5	+, +	+, +
	2,4	−, +	−, −
	3,3	+, +	+, +

Table 2 (cont.)

Compound (A.P.I. no.)	PAIR	a	b
525	1,5	+, +	+, +
	2,4	−, +	−, −
	3,3	+, +	+, +

99 – trans-hex-2-ene
100 – trans-hex-3-ene
101 – 3-methylpent-1-ene
102 – 4-methylpent-1-ene
103 – 2-methylpent-2-ene
104 – 3-methyl-cis-pent-2-ene
276 – 2-methylpent-1-ene
278 – 4-methyl-trans-pent-2-ene
279 – 4-methyl-cis-pent-2-ene
399 – cis-hex-3-ene
524 – 2,3-dimethylbut-1-ene
525 – 2,3-dimethylbut-2-ene

Column *a* represents data for n-hexane; column *b* data for hex-1-ene
+ indicates that a correlation exists; − that a correlation is absent; where decimal fractions are given, the intention is to suggest that values of the ratios which are less than 1 (see the "1" or "0" criterion in text) may have diagnostic value, especially when they are near 1.00 in value

It must be noted that the even mass peaks may be confused with those formed by double fragmentation in paraffins unless a metastable transition is observed. The hydrogen rearrangements which are so very prevalent in alkenes result in migration of radical sites along the chain; therefore the location of the double bond has very little influence upon the spectrum except in those instances where the double bond has undergone tetrasubstitution.

Another basic problem in the use of the naive analysis of spectra lies in the selection of the reference compound (see treatment of alkanes) with which an unknown is compared for purposes of getting diagnostic ions. Thus it is that the question again arises: How close in structure to the unknown compound must the reference compound be in order to provide such ions as occur in the observed spectra?

As a means of comparing reference compounds, the diagnostic pairs re-

sulting from a C_6 hydrocarbon (see alkane data) are listed in Table 2, namely: 1,5; 2,4 and 3,3. Each C_6 monoalkene has been compared for these pairs using (a) *n*-hexane and (b) hex-1-ene as reference compounds. Based on our earlier comments as regards the general lack of influence which saturation has in alkene spectra, it would not be supposed that any significant difference in diagnostic value of the various C_nH_{2n-1} ions would manifest itself when the reference compound is changed. The data in Table 2 indicate that such a difference does exist (specifically with respect to the 2,4 pair of ions) depending upon which of the two reference compounds is used. The 2,4 pair corresponds to ions at masses 27 and 55 respectively, both of which would be expected in alkene spectra since they are of the form C_nH_{2n-1}.

Study of the twelve C_6 alkenes in the table indicates that if the molecule is symmetrical with respect to the distribution of side chains or is straight chained (consider compounds 99, 101, 104, and 399) the same diagnostic value is obtained for the 2,4 ion pair from both reference compounds. However, in the case of unsymmetrical molecules (see compounds 102, 103, 278, 279, 524 and 525) one fails to obtain diagnostic value for the 2,4 ion pair when hex-1-ene is used as the reference compound; this failure occurs specifically with the C_4 member of the pair. One is thus led to postulate that in the treatment by naive analysis the unsaturated function assumes a degree of importance in establishing the diagnostic value of certain ions. It is further observed that those compounds with a low intensity in their mass spectra for the C_5 ion at mass 55 are the same compounds which fail to give diagnostic value for the C_4 ion when the reference compound is changed from *n*-hexane to hex-1-ene.

In attempting to apply set theory to analysis of mass spectra, there is an awareness of the shortcomings in general applicability of this concept to all types of compounds. It was felt initially that hydrocarbons offered the simplest group of spectra with which to test the theory. Particularly after the observation has been made that the olefine spectra, when subjected to set theory analysis, are sensitive to structural variation, it seems possible that the simplicity of structural elements (C and H) conceals much more than was expected from the early investigations concerning alkanes.

CYCLOALKANES

In the preceding treatment of alkanes and alkenes, the basis of analysis was the occurrence of preferred fragmentation at a point of chain branching, to

yield ions of the formula $C_nH_{2n+1}^+$ and $C_nH_{2n-1}^+$ respectively. Use of these ions has been shown to be satisfactory for generation of a structural formula, at least under those conditions imposed upon the analysis in this early stage of development.

When dealing with alkanes, chain branching constituted the only possible difference in structure between the reference straight chain hydrocarbon and the unknown alkane being considered. In the case of the alkenes, position and extent of unsaturation becomes a second factor in comparison with the reference compound and hence the need for more exhaustive investigation of the effect upon analysis of a change in reference compound.

As only the simplest case of an alkyl substituted fully saturated ring system is studied (in the initial approach to cyclic hydrocarbons) the types of ions which must be considered for an analysis are still only two: a) the alkyl ions which arise from the alkyl chains in the original cyclic structure and b) an alkenyl series, resulting from any sequence containing an alicyclic ring. In the earlier work, if r_i was greater than or equal to unity $f = 1$, otherwise $f = 0$.

In the present series, (inasmuch as the reference compound of choice is still the straight chain alkane of the same carbon number as the cyclic-molecule being analysed) one is comparing the abundance of the corresponding alkenyl ions from a cyclo- and a normal alkane.

The alkenyl ion from a cycloalkane is formed by simple fission; however, the distribution of positive charge between the resultant moieties is not equal. Other considerations being the same, the cyclic portion of the molecule will yield the preponderance of the fragment ions, if only because of its lower ionization potential. Formation of an alkenyl ion from the reference alkane should involve loss of two hydrogen atoms from the alkyl ion, which is assumed to be formed first and thus the process does not lend itself to straightforward thermochemical arguments.

Even if both ions have the same ionic formula, it cannot be assumed "a priori" that there is any value in the ratio of their abundances.

Thus it is that we alter our test for diagnostic requirement of a ratio value for any given ionic species; the "0" or "1" criterion is abandoned and the occurrence of a local maximum in the ratio is taken as the new criterion. However, the general conclusion that I_E increases at a branch point is unimpaired in that in a survey of ratios a local maximum may be considered in the same manner i.e. a local maximum rather than the "0" versus "1" evaluation is observed to reveal a branching point.

Ionization cross section values of Ötvos and Stevenson[4] were used in the calculations with alkanes and alkenes, in which cases they were assumed to approximate the values for total ion current obtained by addition of the intensity values for the ions appearing in the A.P.I. spectra. The availability of a distinct set of ionization cross section values (cf. Mann[5]) prompted a comparison of results using both sets. This comparison showed a difference of the order of 3%; but, as the positions of the maxima were not observed to change when Mann's values were substituted; investigations were continued with those of Ötvos and Stevenson.

The present method of calculation is thus as follows. Separate the ions of the mass spectrum into their various series $C_nH_{2n+1}^+$, $C_nH_{2n}^+$, etc. Let l_E represent the observed ion current for $C_nH_{2n+1}^+$ in the unknown and l_K that of the corresponding ion in the known; Σ_E is the total ion current of E and Σ_K the same value for K. The ionization cross-section for E is represented by σ_E and similarly for K. Then the corrected ratio values of the ion currents for the ion $C_nH_{2n}^+$ are $l_E \cdot (\sigma_E/\Sigma_E)$ and $l_K \cdot (\sigma_K/\Sigma_K)$. The comparison of these ion currents is more difficult than formerly, as both σ and Σ differ for the two compounds. To obtain a reasonable assessment, we must adjust the value of one (say E) to allow for these differences. Let $l_E \cdot \sigma_E/\Sigma_E = h$ and correct it to bring it to the same conditions as obtain for K. In correcting for the differences of ionization cross section we replace h by $h \cdot (\sigma_E/\sigma_K)$; by making a further correction for the total ion current, the expression becomes

$$h \cdot \frac{\sigma_E}{\sigma_K} \cdot \frac{\Sigma_K}{\Sigma_E}$$

The corrected value of the ion current for E becomes

$$l_E \cdot \frac{\sigma_E}{\Sigma_E} \cdot \frac{\sigma_E}{\sigma_K} \cdot \frac{\Sigma_K}{\Sigma_E}$$

and that for the ratio between the unknown and known ions

$$l_E \cdot \frac{\sigma_E \cdot \sigma_E \cdot \Sigma_K}{\sigma_K \cdot \Sigma_E \cdot \Sigma_E} \Big/ \frac{l_K \cdot \sigma_K}{\Sigma_K}$$

Hence, by regrouping

$$\left(\frac{\sigma_E}{\sigma_K \Sigma_E \Sigma_E}\right) \cdot l_E \cdot \sigma_E \Sigma_K / l_K \cdot \sigma_K \left(\frac{1}{\Sigma_K}\right)$$

and rearranging

$$\left(\frac{\sigma_E \cdot \Sigma_K}{\Sigma_{E'} \cdot \sigma_K}\right) I_{E'} \cdot \sigma_E \Sigma_K / I^{K'} \sigma_K \Sigma_E$$

Accordingly, as was assumed in the previous communication[6], it is supposed that

$$\frac{\sigma_E \cdot \Sigma_K}{\Sigma_E \cdot \sigma_K} \approx 1$$

and that the wanderings from this simple theory are due to variations in ion-collection efficiencies and similar small perturbations whose origins are not fully understood.

It has been assumed that the error introduced by assuming a strict proportionality of the total ion currents of C_nH_{2n} and C_nH_{2n+2} with their ionization cross sections is small. Within these limits analyses have been carried out upon a series of cycloalkanes of the following general types: cyclohexyleicosanes, cyclic systems with a terminal ring, cyclic systems with multiple substitution and compounds having more than one cyclic group.

For the purposes of illustrating the method of data manipulation for this category of compound, the case of cyclohexyleicosanes is discussed. They are very appropriate because a series of position isomers of such compounds is available. The calculation is carried out in two stages. The first comparison is made between the alkyl ions of unknown and reference for the purposes of locating any branchpoints in the alkyl chains. Secondly, a comparable investigation is carried out for the alkenyl ions. The discussion of deductions for the isomer 5-cyclohexyleicosane will serve as an example.

The mass spectrum yields the molecular formula, $C_{26}H_{52}$. A series of abundant alkyl ions are observed at $m/q = 281$ (the most abundant) and corresponding to formula $C_{20}^+H_{41}$; $m/q = 155$ ($C_{11}^+H_{23}$); $m/q = 211$ ($C_{15}^+H_{31}$) and $m/q = 309$ ($C_{22}^+H_{45}$). The alkenyl ion sequence shows the presence of $C_6^+H_{11}$, $C_{11}^+H_{21}$, $C_{17}^+H_{33}$, and $C_{20}^+H_{39}$ ($m/q = 83$, 153, 237, and 279). Excepting $C_{20}^+H_{39}$ and $C_{22}^+H_{45}$ these ions form a self-consistent elemental set which, by the methods used in the treatment of alkanes, yields a structure of the type C_4H_9—$CH(C_6H_{11})$—$C_{15}H_{30}$.

Three points must now be taken into consideration: 1) the nature of the C_6H_{11} ring system; 2) the possibility of chain branching in the $C_{20}H_{41}$ alkyl group and 3) the origin of the two ions $C_{20}H_{39}$ and $C_{22}H_{45}$. In reference to point 1), the cyclic group may be either cyclohexyl, an isomer of the methyl-

Table 3

	n-Hexacosane.	Position of substitution (for cyclohexyleicosanes)					
		(−2)	(−3)	(−4)	(−5)	(−7)	(−9)
A.P.I. No.	669	1263	1264	1265	1487	1036	1037
Σ	578.02	846.69	937.37	876.21	1084.49	1318.06	1372.64
				Alkyl ions			
l							
3	14.800	0.399	0.407	0.451	0.351	0.303	0.277
4	17.300	0.244	0.271	0.307	0.241	0.247	0.248
5	9.913	0.197	0.217	0.258	0.240	0.214	0.215
6	6.834	0.180	0.207	0.248	0.179	0.215	0.211
7	1.510	0.170	0.219	0.259	0.218	0.319	0.298
8	0.986	0.176	0.214	0.264	0.211	0.303	0.328
9	0.723	0.165	0.216	0.258	0.209	0.302	0.316
10	0.562	0.215	0.201	0.262	0.199	0.272	0.280
11	0.441	0.146	0.200	0.256	0.212	0.222	0.287
12	0.369	0.139	0.195	0.239	0.187	0.288	0.279
13	0.318	0.135	0.183	0.231	0.187	0.278	0.235
14	0.282	0.130	0.170	0.233	0.176	0.252	0.216
15	0.246	0.121	0.165	0.204	0.189	0.190	0.211
16	0.220	0.107	0.142	0.203	0.159	0.139	0.134
17	0.201	0.079	0.140	0.156	0.091	0.077	0.085
18	0.185	0.093	0.238	0.083	0.058	0.048	0.074
19	0.173	0.020	0.040	0.035	0.025	0.015	0.030
20	0.159	**2.126**	**2.676**	**3.175**	**3.476**	**1.913**	**2.129**
21	0.144	0.024		0.028	0.015	0.006	0.003
22	0.125	0.017	0.300	0.016	0.138	0.004	0.003
23	0.083			0.266	0.013	0.011	
24	0.050		0.487	0.027			
25	0.005	0.694					
26	0.005	1.667	0.375	0.900		0.177	0.256
				Alkenyl ions			
3	6.055	0.820	0.820	0.903	0.530	0.531	0.499
4	5.675	1.129	1.099	1.193	0.877	0.766	0.744
5	2.855	1.244	1.155	0.989	0.790	0.609	0.551
6	2.059	**4.480**	**3.232**	**3.317**	**2.199**	1.636	1.509
7	1.251	1.372	1.173	1.650	1.492	1.417	1.331

Table 3 (cont.)

	n-Hexacosane	Position of substitution (for cyclohexyleicosanes)					
		(−2)	(−3)	(−4)	(−5)	(−7)	(−9)
A.P.I. No.	669	1263	1264	1265	1487	1036	1037
Σ	578.02	846.69	937.37	876.21	1084.49	1318.06	1372.64
				Alkenyl ions			
8	0.512	**7.448**	1.464	1.405	1.575	1.790	1.656
9	0.213	1.175	9.156	1.787	1.338	2.389	2.276
10	0.078	1.330	2.281	**19.084**	2.021	2.681	3.455
11	0.042	1.530	2.496	3.291	**28.085**	3.161	4.122
12	0.028	1.753	2.717	3.103	3.083	3.819	3.881
13	0.017	2.073	0.353	3.804	3.238	**32.298**	4.361
14	0.012	2.241	3.646	4.167	3.243	5.556	5.125
15	0.007	2.871	5.268	5.429	5.154	6.139	**58.659**
16	0.005	3.472	6.000	7.057	5.699	5.398	5.897
17	0.003	4.419	9.167	10.000	6.964	5.294	8.286
18	0.003	3.954	7.917	7.111	5.536	3.971	97.429
19	0.002	2.069	6.563	6.338	4.324	5.556	5.106
20	0.002	4.138	8.750	5.667	8.378	116.444	2.553

$\sigma_E = 160.16$ in all instances, for the compounds are isomers.
$\sigma_K = 162.16$ for *n*-hexacosane, the reference compound.

$$\frac{x\text{-cyclohexyleicosane}}{n\text{-hexacosane}} = \frac{I_x}{I_{669}} \times \frac{\sigma_E}{\sigma_K} \times \frac{\Sigma_K}{\Sigma_E}$$

where x = position of substitution
I_x has two values, one for the alkyl ion series, one for the alkenyl ion series.
Ratios in bold type are considered the most reliable.
Values for mass 26 relate to the parent molecular ion and as such are not members of the alkyl ion sequences.

cyclopentyl group or (less likely) a substituted cyclobutyl group. The latter is unlikely on the basis of observed fragmentation patterns from cyclobutyl-containing systems. The present form of naive analysis is not able to distinguish reliably between the first two of these possibilities.

The second point, namely the possibility of chain branching, is considered to be remote because any such occurrences would imply a terminal cyclic group. If one assumes this fragmentation to provide all the necessary alkyl ions it would require two cleavages in the parent molecular ion, which phenomenon is excluded by our analytical principles.

In regard to the third point, the origins of $C_{20}^{+}H_{39}$ and $C_{22}^{+}H_{45}$ must be considered. The latter may readily be assumed to arise from cleavage within the ring, which would yield the observed $R = C_{20}H_{41}$. The complementary ion does not belong to either ion sequence here studied and thus is an example of the utility of the present status of naive analysis.

Despite the failure to provide complete explanation for all ions occurring in the course of a particular analysis, the method would appear to have advantages, especially its applicability in the technique of using a reference compound which may be analysed in admixture with the unknown. Techniques for subsequent recovery of the mass spectrum under these conditions are known. Thus, there is the possibility of overcoming problems due to short term instrumental variations and thereby allowing extension of the analysis to any mass spectrometer with the requisite minimum resolution.

A more detailed treatment of the cycloalkane systems has been published[6]. An initial effort, dealing with alkanes, has also appeared[7].

Worthwhile progress is currently being made with respect to a more complete treatment of systems with double bonds and those which include hetero-atoms i.e. atoms other than C and H. Although far from complete, these germinal ideas may serve as a beginning for a treatment of mass spectral data which will in time assume appreciable significance.

NOTES ON SET THEORY

The basic principles of set theory are few in number. A statement of these principles is necessary as use has been made of them in the discussion on mass spectral interpretation.

A set or class may be defined as a collection of objects called elements of the set. Specifying the elements in a set defines that set. Set theory is the study of the relations of sets to one another and to their subsets; the application of set theory provides a means of defining classes of objects in a very precise manner and of establishing those relationships which exist between various groups of objects i.e. the ions of a mass spectrum.

All sets, regardless of what other elements they may contain, have as a

member the null or empty set which is designated by the symbol Ø. This is quite different from the set which contains zero as an element.

Two sets are equal when they contain exactly the same elements or members; the order in which the elements are arranged in the two sets undergoing comparison is not important. Subset may be formally defined as any set which is contained in a given set. Any set may be a subset of itself; however, the special case of proper subset is defined as a subset which contains fewer members than the parent set. The number of possible subsets in a set is given by 2^n, where n is equal to the number of elements in the parent set.

The universal set (the set containing all possible subsets or elements in a given category) is different for each specific collection of conditions which we describe.

The complement of a given set Z, designated as Z', contains all those elements in the universal set U, of which Z is a member, which are *not* indeed members of the set Z itself. The complement of the universal set U is seen to be the null set Ø and is designated as U'. Any two sets under consideration are said to be disjoint if they do not have any elements in common.

Two further categories of sets may be defined in the following simple terminology:

1) equivalent sets—sets which contain the same number of elements although not necessarily the same ones.

2) ordered set—a set in which the elements thereof are arranged in a serial relationship based on some pre-defined rule.

These latter two categories are particularly suitable to application in the treatment of mass spectral data.

References

1. Kn[illegible], B., Wright, D., Kelly, W., and Ridley, R.G., *17th Annual Conference on Mass Spectrometry and Allied Topics*, A.S.T.M.—E-14, Dallas, Texas, 18–23 May 1969.
2. Jurs, P.C., Kowalski, B.R., and Isenhour, T.L., *Anal. Chem.*, **41**, 21 (1969).
3. "Catalogue of Mass Spectral Data", Serial No. 44, Thermodynamics Research Centre Data Project, Thermodynamics Research Centre, Texas A and M, University College Station, Texas.
4. Ötvos, J.W., and Stevenson, D.P., *J. Am. Chem. Soc.*, **78**, 546 (1956).
5. Mann, J.B., *J. Chem. Phys.*, 46, 1650 (1967).
6. Reed, R.I., *Rev. Port Quim.*, 10, 129 (1968).
7. *ibid.*, 12, 16 (1970).

The following pages contain representative pages of computer printouts for the Khinchine function, abstracted from a total of 6880 values which were determined. The excessive bulk involved in reproduction of the entire printout scheme made it a practical expedient that selected values be shown to illustrate the range of Khinchine values which were obtained.

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1	6476	278	154	20	38	109**********		0.0000	2,5-DIMETHYL-7-N-OCTYLBICYCLO(4.4.0)DECANE
2	6512	1318	1325	94	91	57	0.00	0.0	N-TETRANONACONTANE
3	6050	2	1	0	2	2	99.99	0.0	HYDROGEN
4	6064	2	1	0	2	2	99.99	0.0	HYDROGEN
5	5800	346	238	25	46	55**********		0.0000	1,7-DICYCLOPENTYL-4-(2'-CYCLOHEXYLETHYL)-HEPTANE
6	4924	4	2	0	0	4	101.09	0.0866	DEUTERIUM
7	4923	2	2	0	2	2	102.09	0.1446	HYDROGEN
8	1763	226	2	16	18	105	102.99	0.1900	B-PHENYLETHYL ETHER
9	139	86	2	4	3	86	103.99	0.2352	1-CHLORO-1-BUTEN-3-YNE
10	12	32	4	0	0	32	103.73	0.2433	OXYGEN
11	8	28	6	1	0	28	103.18	0.2615	CARBON MONOXIDE
12	6	28	4	0	0	28	104.75	0.2977	NITROGEN
13	6066	28	3	0	0	28	105.90	0.3407	NITROGEN
14	1769	228	2	3	1	83	109.33	0.4210	PENTACHLOROACETONE
15	1	20	3	0	0	20	110.29	0.4653	NEON
16	6051	20	3	0	0	20	110.55	0.4715	NEON
17	1440	192	2	4	4	157	113.32	0.5225	1,1,3,3-TETRACHLORO-2-METHYLPROPENE
18	6063	32	6	0	0	32	109.12	0.5350	OXYGEN
19	4627	28	7	1	0	28	108.55	0.5410	CARBON MONOXIDE
20	6057	40	4	0	0	40	113.26	0.5445	ARGON
21	6052	40	6	0	0	40	113.40	0.5514	ARGON
22	22	40	5	0	0	40	114.99	0.5850	ARGON
23	1946	340	2	23	32	325	116.93	0.5969	2,2-BIS(4-HYDROXY-3-TERT-BUTYL) PROPANE
24	6065	40	4	0	0	40	116.23	0.6140	ARGON
25	1026	160	3	12	16	160	115.44	0.6423	ISOPROPYL ISOPROPENYLBENZENE
26	5689	30	6	0	0	30	111.92	0.6515	NITROGEN OXIDE (NITRIC OXIDE)
27	1848	248	2	17	28	233	121.17	0.6686	2,6-DI-TERT-BUTYL-4-ISOPROPYLPHENOL
28	4	18	5	0	2	18	122.77	0.7807	WATER
29	6053	44	13	1	0	44	115.41	0.8231	CARBON DIOXIDE
30	348	110	2	3	4	75	139.51	0.8599	2,3-DICHLOROPROPENE
31	1323	182	6	11	15	167	130.82	0.9081	X-CHLORODIETHYLTOLUENE
32	6067	44	10	1	0	44	118.67	0.9705	CARBON DIOXIDE
33	1556	202	3	10	12	167	133.56	1.0076	AR,AR-DI(CHLOROMETHYL)-AR-ETHYLBENZENE
34	4628	44	6	1	0	44	124.42	1.0351	CARBON DIOXIDE
35	4565	27	6	1	1	27	125.89	1.0712	HYDROGEN CYANIDE
36	13	32	9	0	0	28	130.89	1.1083	AIR
37	4704	27	10	1	1	27	127.13	1.1360	HYDROGEN CYANIDE
38	1010	158	6	7	10	43	126.36	1.1449	2-PROPENE-1,1-DIOL DIACETATE
39	3	26	9	2	2	26	130.33	1.1627	ACETYLENE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
40	1575	206	4	14	22	191	135.60	1.1962	AR,AR-DI-TERT-BUTYLPHENOL
41	159	88	27	1	0	69	121.38	1.2092	CARBON TETRAFLUORIDE
42	605	132	3	10	12	132	199.82	1.2320	2,4-DIMETHYL-STYRENE
43	4561	17	5	0	3	17	190.09	1.2902	AMMONIA
44	4872	88	12	1	0	69	129.90	1.2992	TETRAFLUOROMETHANE
45	27	44	8	1	0	44	131.69	1.3007	CARBON DICXIDE
46	66	65	34	2	1	65	121.71	1.3511	1,1-DIFLUOROETHYLENE-2-D.
47	4543	26	8	2	2	26	136.99	1.3625	ETHYNE (ACETYLENE)
48	3250	104	22	0	0	85	125.16	1.3649	SILICON TETRAFLUORIDE 251CE
49	4531	16	6	1	4	16	188.48	1.3756	METHANE
50	18	36	6	0	1	36	151.14	1.3966	HYDROCHLORIC ACID
51	4564	36	4	0	1	36	154.79	1.3984	HYDROGEN CHLORIDE
52	2	16	8	1	4	16	193.28	1.4078	METHANE
53	1018	160	6	6	9	81	145.49	1.4144	3-BROMOCYCLOHEXENE
54	4566	46	5	0	0	30	168.99	1.5547	NITROGEN DIOXICE
55	173	88	5	4	8	45	184.58	1.5989	2-VINYLOXYETHANOL
56	3177	102	66	8	6	43	130.24	1.6062	PHENYLACETYLENE
57	2334	102	66	4	6	43	130.24	1.6062	ACETIC ANHYDRIDE
58	2875	156	72	12	12	156	123.95	1.6128	2-6 DIMETHYL NAPHTHALENE
59	1019	160	4	12	16	145	222.11	1.6154	AR-ISOPROPENYL-1,2,4-TRIMETHYLBENZENE
60	4926	20	7	1	0	20	208.13	1.6168	TETRADEUTEROMETHANE
61	4581	16	7	1	4	16	215.33	1.6494	METHANE
62	4567	44	10	0	0	44	160.99	1.6598	DINITROGEN OXIDE (NITROUS OXIDE)
63	4563	76	10	1	0	76	154.09	1.6671	CARBON DISULFIDE
64	160	88	10	4	5	53	216.97	1.6763	CHLOROPRENE
65	260	102	31	4	6	43	136.54	1.6885	ACETIC ANHYDRIDE
66	4568	64	11	0	0	64	174.69	1.7170	SULFUR DICXIDE
67	4472	16	8	1	4	16	217.56	1.7217	METHANE
68	1952	356	7	28	20	356	147.69	1.7254	1,2,3-TRIPHENYLNAPHTHALENE
69	15	34	9	1	3	34	218.35	1.7390	FLUOROMETHANE
70	5690	34	9	0	3	34	180.50	1.7519	PHOSPHINE
71	1557	202	6	10	12	153	166.99	1.7534	AR,ALPHA-DICHLORO-AR-ISOPROPYLTOLUENE
72	47	58	4	3	6	58	232.70	1.7654	VINYL METHYL ETHER
73	4316	262	10	15	18	205	153.99	1.7676	DIHYDROHETERO-PEUCENIN
74	450	120	7	8	8	91	168.97	1.7696	PHENYL-ACETALDEHYDE
75	287	104	30	1	0	69	143.55	1.7883	CHLOROTRIFLUOROMETHANE
76	2508	117	41	5	11	44	136.19	1.7992	ALANINE ETHYL ESTER
77	4925	34	7	0	2	34	198.69	1.8040	HYDROGEN SULFIDE
78	2591	73	23	4	11	30	136.82	1.8097	N-BUTYLAMINE
79	4562	60	9	1	0	60	189.19	1.8122	CARBON OXYSULFIDE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
200	2148	62	16	2	6	31	193.25	2.4305	ETHYLENE GLYCOL
201	3611	132	32	6	12	45	202.99	2.4310	PARALDEHYDE
202	3855	148	37	10	12	105	188.43	2.4317	ISOBUTYROPHENONE 877CE
203	4313	260	11	15	16	205	286.98	2.4346	PEUCENIN
204	238	100	37	2	0	81	277.25	2.4364	TETRAFLUOROETHYLENE
205	802	146	74	6	10	43	155.89	2.4373	1,2-ETHANE DIACETATE
206	4847	58	32	3	6	43	171.26	2.4453	2-PROPANONE (ACETONE)
207	1486	198	41	13	10	105	183.67	2.4467	PHENYL BENZOATE
208	2502	143	57	7	13	70	161.98	2.4560	PROLINE ETHYL ESTER
209	31	46	27	2	3	28	177.89	2.4629	FLUOROETHYLENE
210	3027	96	13	5	4	96	284.44	2.4662	FURFURAL
211	4334	499	14	24	33	178	192.69	2.4671	DIMETHYL PUROMYCIN
212	5253	86	48	4	6	43	184.59	2.4683	2,3-BUTANEDIONE (BIACETYL)
213	6879	95	27	2	0	69	232.21	2.4704	TRIFLUOROACETONITRILE (TRIFLUOROMETHYLCYANIDE)
214	54	60	18	2	4	31	247.86	2.4714	METHYL FORMATE
215	4314	274	11	16	18	231	292.98	2.4721	PEUCENIN-7-METHYL ETHER
216	145	86	38	1	1	51	160.22	2.4815	CHLORO-DIFLUORO-METHANE
217	65	64	16	2	2	64	242.17	2.4843	1,2-DIFLUOROETHYLENE
218	6855	104	29	1	0	69	192.80	2.4910	CHLOROTRIFLUOROMETHANE
219	1834	242	81	16	18	121	153.79	2.4976	1,2-BIS(P-METHOXYPHENYL) ETHANE
220	4473	30	17	2	6	28	222.98	2.4999	ETHANE
221	3598	166	69	3	0	69	196.20	2.5004	ACETONE,HEXAFLUORO- 618CE
222	6756	105	43	1	0	105	277.99	2.5022	CYANOGEN BROMIDE
223	1988	148	37	10	12	105	191.37	2.5073	ISOBUTYROPHENONE
224	2595	73	23	4	11	44	166.75	2.5151	SEC-BUTYLAMINE
225	4769	72	23	4	8	43	182.40	2.5176	2-BUTANONE (METHYL ETHYL KETONE)
226	2495	61	20	2	7	30	175.43	2.5188	ETHANOLAMINE
227	1723	222	25	9	9	139	187.94	2.5222	(TRICHLOROETHYL) TOLUENE
228	4705	41	19	2	3	41	220.88	2.5276	ETHANENITRILE
229	6113	120	35	1	0	85	189.97	2.5291	DIFLUORODICHLOROMETHANE
230	4771	46	16	1	2	29	268.40	2.5331	METHANOIC ACID (FORMIC ACID)
231	3509	98	21	1	0	63	239.63	2.5346	CARBONYL CHLORIDE 524CE
232	45	58	36	3	6	43	195.16	2.5350	ACETONE
233	20	37	19	1	0	37	334.64	2.5381	FLUOROMETHANE-D3
234	108	74	40	4	10	45	169.77	2.5445	METHYL-N-PROPYL ETHER
235	5926	73	53	4	11	30	157.48	2.5451	1-AMINO-2-METHYLPROPANE (ISO-BUTYLAMINE)
236	4559	50	14	1	3	50	256.79	2.5463	CHLOROMETHANE (METHYL CHLORIDE)
237	1411	190	66	8	14	87	251.83	2.5467	DIETHYLENE GLYCOL DIACETATE
238	4669	388	39	7	0	69	196.52	2.5511	HEXADECAFLUOROHEPTANE
239	5111	60	26	2	4	43	310.32	2.5581	ETHANOIC ACID (ACETIC ACID)

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
240	34	52	28	4	4	52	241.32	2.5595	1-BUTEN-3-YNE
241	3548	90	38	3	6	31	208.23	2.5636	SYM-TRIOXANE 563CE
242	2254	90	38	3	6	31	208.23	2.5636	SYM-TRIOXANE
243	4900	72	25	4	8	43	191.20	2.5661	NOR-BUTANONE
244	52	59	38	3	9	30	165.61	2.5677	N-PROPYLAMINE
245	4348	264	10	15	20	249	225.59	2.5729	VULGARIN
246	32	48	19	2	5	47	198.09	2.5772	FLUOROETHANE
247	4549	48	19	1	4	47	294.74	2.5836	METHANETHIOL (METHYL MERCAPTAN)
248	58	60	31	3	8	45	171.53	2.5856	ISOPROPYL ALCOHOL
249	1970	82	23	5	6	43	180.55	2.5858	METHYL ALLYL KETONE
250	2002	101	69	6	15	30	156.44	2.5868	N HEXYL AMINE
251	3129	101	69	6	15	30	156.44	2.5868	N-ETHYL-NORM-BUTYLAMINE 130CE
252	5122	86	46	5	10	43	173.91	2.5903	2-PENTANONE (METHYL-NOR-PROPYL KETONE)
253	1560	204	49	6	5	204	221.33	2.5921	IODOBENZENE
254	5115	72	34	4	8	43	188.92	2.5944	2-BUTANONE (METHYL ETHYL KETONE)
255	5207	338	30	6	0	69	176.72	2.5952	TETRADECAFLUORO-2-METHYLPENTANE
256	155	87	74	5	13	30	153.84	2.5961	N-AMYLAMINE
257	5635	29	16	2	3	29	273.85	2.5982	MONODEUTEROETHENE (MONODEUTEROETHYLENE)
258	1288	178	54	10	10	105	182.86	2.6006	PHENACYL ACETATE
259	90	72	41	4	8	43	184.51	2.6015	METHYL ETHYL KETONE
260	4256	112	62	6	12	30	155.00	2.6036	OMEGA-AMINOCAPRONITRILE
261	103	74	27	3	6	43	195.21	2.6066	METHYL ACETATE
262	6118	58	31	3	6	43	201.61	2.6071	2-PROPANONE (ACETONE)
263	5923	50	24	1	3	50	269.87	2.6101	CHLOROMETHANE (METHYL CHLORIDE)
264	3061	42	27	2	2	14	287.54	2.6117	KETENE 34CE
265	2136	42	27	2	2	14	287.54	2.6117	KETENE
266	2953	141	69	9	19	98	155.55	2.6124	N-N BUTYL PIPERIDINE
267	11	32	26	2	0	32	279.18	2.6130	ETHYLENE-D4
268	4833	200	42	4	0	100	305.38	2.6178	OCTAFLUOROCYCLOBUTANE
269	6505	372	66	27	48	133	154.28	2.6196	1-MESITYLOCTADECANE (1,3,5-TRIMETHYL-2-N-OCTADECYLBENZENE)
270	334	108	50	4	9	45	162.35	2.6212	3-CHLORO-2-BUTANOL
271	4556	58	26	3	6	43	209.09	2.6218	2-PROPANONE (ACETONE)
272	523	127	51	2	0	92	159.42	2.6242	DICHLOROFLUORO-ACETONITRILE
273	5925	59	49	3	9	30	165.80	2.6288	1- AMINO-NOR-PROPANE (NOR-PROPYLAMINE)
274	5114	60	29	3	8	45	183.74	2.6296	2-PROPANOL (ISOPROPYL ALCOHOL)
275	851	148	16	11	16	133	219.84	2.6307	TERT-BUTYLTOLUENE
276	4894	52	20	4	4	52	245.03	2.6325	1-BUTEN-3-YNE
277	5606	40	28	3	4	40	298.18	2.6341	PROPYNE (METHYLACETYLENE)
278	16	35	16	1	1	33	279.08	2.6342	METHANE-D3-OL
279	186	90	34	3	6	45	203.84	2.6352	METHOXY ACETIC ACID

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
280	4134	184	99	12	8	28	157.24	2.6384	2-PHENYLBENZOPHENONE
281	4756	60	21	3	8	45	181.84	2.6397	2-PROPANOL (ISOPROPYL ALCOHOL)
282	335	108	72	4	9	45	154.37	2.6399	2-CHLOROPROPYL METHYL ETHER
283	3217	156	62	2	5	156	311.23	2.6433	ETHYL IODIDE
284	657	134	8	6	14	57	277.25	2.6434	TRIMETHYL ORTHOPROPIONATE
285	1067	162	58	6	10	45	189.06	2.6463	METHOXYACETIC ANHYDRIDE
286	314	106	22	2	3	27	355.58	2.6487	BROMOETHYLENE
287	51	58	32	4	10	43	221.84	2.6544	2-METHYLPROPANE
288	1879	268	30	1	0	191	245.50	2.6546	TRIBROMOFLUOROMETHANE
289	2165	70	13	5	10	42	218.53	2.6557	METHYLCYCLOBUTANE
290	209	94	29	2	3	59	207.15	2.6583	METHYL CHLOROFORMATE
291	2182	74	61	3	10	44	173.78	2.6635	PROPYLENEDIAMINE
292	3366	72	61	3	8	44	173.78	2.6635	PROPYLENEDIAMINE
293	125	80	60	2	5	31	158.91	2.6656	2-CHLORO-ETHANOL
294	4272	238	32	18	22	147	169.91	2.6676	1,3,4,5-TETRAMETHYL-2-PHENETHYL-BENZENE
295	6343	95	43	2	0	69	234.96	2.6719	TRIFLUOROETHANENITRILE (TRIFLUOROACETONITRILE)
296	4848	72	42	4	8	43	193.26	2.6733	2-BUTANONE (METHYL ETHYL KETONE)
297	5232	46	30	2	6	45	324.79	2.6786	DIMETHYL ETHER
298	4547	60	24	3	8	45	188.29	2.6792	2-PROPANOL (ISOPROPYL ALCOHOL)
299	184	88	49	5	12	45	173.44	2.6798	N-BUTYL METHYL ETHER
300	5623	130	46	6	10	57	246.76	2.6800	PROPANOIC ANHYDRIDE
301	17	35	9	2	1	32	342.14	2.6814	ETHANE-D5
302	4535	58	35	4	10	43	225.22	2.6854	2-METHYLPROPANE (ISOBUTANE)
303	61	62	31	2	3	62	257.43	2.6858	CHLOROETHYLENE
304	4793	102	23	6	14	45	219.03	2.6871	DIISOPROPYL ETHER
305	4670	250	32	5	0	131	209.58	2.6882	DECAFLUORCCYCLOPENTANE
306	4349	262	10	15	18	247	243.99	2.6886	ARGLANINE
307	4858	88	30	4	8	43	196.62	2.6910	ETHYL ACETATE
308	6345	67	17	1	3	15	283.18	2.7000	METHYL-N,N-DIFLUOROAMINE
309	4797	88	23	4	8	43	193.85	2.7027	ETHYL ACETATE
310	5113	60	32	3	8	31	182.47	2.7036	1-PROPANOL (NOR-PROPYL ALCOHOL)
311	357	112	45	1	2	112	283.24	2.7042	BROMOFLUOROMETHANE
312	3965	110	18	6	6	110	204.45	2.7044	PYROCATECHOL DOW
313	4295	401	19	29	23	255	195.99	2.7077	1,3-DIPHENYL-3-ETHYL-4-OXO-2-SPIRO-(9-FLUORENYL)-AZETIDINE
314	4320	260	11	15	16	217	334.98	2.7079	DIHYDRO-DESOXYKARENIN
315	5025	62	39	3	7	47	190.41	2.7106	2-FLUOROPROPANE
316	141	86	55	4	6	43	171.36	2.7153	2,3-BUTANE-DIONE
317	5637	30	16	2	2	30	277.50	2.7184	1,CIS-2-DIDEUTEROETHENE (1, CIS-2-DIDEUTEROETHYLENE)
318	5605	40	26	3	4	40	317.93	2.7209	PROPADIENE (ALLENE)
319	3353	61	36	2	7	30	195.86	2.7213	ETHANOLAMINE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
440	4150	124	86	7	8	44	255.82	2.8942	2,6-DIMETHYL-1,4-PYRONE
441	6854	62	29	2	3	27	295.38	2.8951	CHLOROETHENE (CHLOROETHYLENE)
442	6342	86	41	1	1	51	194.98	2.8952	CHLORODIFLUOROMETHANE (FREON-22)
443	4836	60	36	3	8	31	192.83	2.8981	1-PROPANOL (NOR-PROPYL ALCOHOL)
444	6087	132	46	10	12	66	174.77	2.9001	3A,4,7,7A-TETRAHYDRO-4,7-METHANOINDENE (DICYCLOPENTADIENE)
445	3291	98	36	4	2	26	266.73	2.9016	MALEIC ANHYDRIDE 297CE
446	2303	98	35	4	2	26	266.73	2.9016	MALEIC ANHYDRIDE
447	5233	62	39	2	6	29	360.94	2.9018	DIMETHYL PEROXIDE
448	6112	136	38	1	0	101	250.12	2.9030	FLUOROTRICHLOROMETHANE
449	4979	68	47	4	4	39	266.27	2.9051	FURAN
450	3580	129	42	8	19	44	176.65	2.9072	HEPTANE,2-AMINO-5-METHYL- 598CE
451	2012	129	42	8	19	44	176.65	2.9072	2 AMINO 5 METHYL N HEPTANE
452	1081	162	9	8	18	74	413.15	2.9076	TRIMETHYLORTHOVALERATE
453	4765	58	18	3	6	29	327.46	2.9091	PROPANAL (PROPIONALDEHYDE)
454	4781	120	50	9	12	91	190.56	2.9093	NOR-PROPYLBENZENE
455	6848	94	23	3	1	94	339.68	2.9104	1,1,1-TRIFLUOROPROPYNE
456	3644	138	60	8	10	43	177.39	2.9113	ALLYLIDENE ACETATE
457	3852	260	117	19	32	105	180.71	2.9171	TRIDECANE,2-PHENYL- 874CE
458	2695	260	117	19	32	105	180.71	2.9171	2 PHENYLTRIDECANE
459	46	58	42	2	2	18	205.78	2.9189	GLYOXAL
460	6402	102	36	5	10	43	198.90	2.9192	ISOPROPYL ACETATE
461	1015	160	41	7	6	125	298.89	2.9195	AR,AR-DICHLOROTOLUENE
462	6104	78	36	3	7	43	216.47	2.9199	2-CHLOROPROPANE
463	2596	130	33	8	18	57	199.89	2.9211	DIISOBUTYLETHER
464	4614	128	65	9	20	57	218.97	2.9227	2,2,4,4-TETRAMETHYLPENTANE
465	218	96	31	2	2	61	347.21	2.9244	CIS-1,2-DICHLOROETHYLENE
466	4301	298	10	19	26	298	427.97	2.9248	DIHYDRO-ANTIRHINE
467	4775	78	35	6	6	78	207.61	2.9253	BENZENE
468	200	92	49	3	5	43	181.07	2.9260	CHLOROACETONE
469	4585	58	37	4	10	43	254.61	2.9278	2-METHYLPROPANE (ISOBUTANE)
470	4896	60	29	3	8	31	199.17	2.9308	1-PROPANOL (NOR-PROPYL ALCOHOL)
471	4709	53	25	3	3	26	382.12	2.9310	PROPENENITRILE
472	2143	58	38	3	6	28	245.48	2.9328	TRIMETHYLENE OXIDE
473	5128	88	46	5	12	45	195.44	2.9336	3-METHYL-2-BUTANOL
474	180	88	59	5	12	45	188.74	2.9367	3-METHYL-2-BUTANOL
475	4772	60	15	2	4	60	444.44	2.9371	ETHANOIC ACID (ACETIC ACID)
476	848	148	70	5	9	69	201.46	2.9374	BROMOCYCLOPENTANE
477	3158	74	42	3	3	39	229.77	2.9409	PROPARGYL CHLORIDE 157CE
478	2181	74	42	3	3	39	229.77	2.9409	PROPARGYL CHLORIDE
479	5204	200	26	4	0	131	285.00	2.9422	OCTAFLUORO-1-BUTENE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
760	5572	174	23	10	22	45	389.16	3.1765	1,1-DI-SEC-BUTOXYETHANE (DI-SEC-BUTYL ACETAL)
761	215	96	34	2	2	61	376.14	3.1777	TRANS-1,2-DICHLOROETHYLENE
762	1123	166	87	10	14	110	183.63	3.1778	P-BUTOXYPHENOL
763	2123	240	100	16	16	105	330.47	3.1784	2 METHYLBENZYL 2 METHYLBENZOATE
764	1084	164	54	1	0	85	234.61	3.1787	DIFLUOROCHLOROBROMOMETHANE
765	1501	198	50	5	11	71	311.50	3.1812	2-IODOPENTANE
766	1661	214	56	4	8	135	404.17	3.1815	1,3-DIBROMOBUTANE
767	106	74	36	3	6	73	276.23	3.1837	1,3-DIOXOLANE
768	5234	43	28	2	5	42	369.86	3.1838	ETHYLENIMINE
769	1545	202	60	11	22	71	280.39	3.1840	BUTYL BUTYRYL LACTATE
770	4583	44	28	3	8	29	326.64	3.1856	PROPANE
771	4755	60	29	3	8	31	228.16	3.1861	1-PROPANOL (NOR-PROPYL ALCOHOL)
772	3623	128	42	7	12	55	271.98	3.1861	N-BUTYL ACRYLATE
773	2045	142	42	8	14	55	271.98	3.1861	N BUTYL ACRYLATE
774	862	150	74	5	11	43	324.66	3.1876	3-BROMO-PENTANE
775	2928	148	74	9	8	105	295.64	3.1899	1-PHENYL-1, 2-PROPANEDIONE
776	611	132	26	9	8	132	396.77	3.1900	ARYL-VINYL-BENZALDEHYDE
777	2806	0	46	0	0	73	194.24	3.1906	MUCO INOSITOLS
778	2750	218	62	16	26	105	187.10	3.1908	2- PHENYL DECANE
779	3666	128	65	9	20	57	244.54	3.1910	2,2,5-TRIMETHYL HEXANE
780	801	146	54	8	18	57	209.60	3.1923	TERT-BUTYL SULFIDE
781	233	98	58	4	2	54	329.56	3.1933	MALEIC ANHYDRIDE
782	4646	78	42	6	6	78	228.44	3.1938	BENZENE
783	1259	176	71	7	13	97	256.19	3.1939	1-METHYL-3-BROMOCYCLOHEXANE
784	643	134	44	6	14	59	277.07	3.1947	BIS-(2-METHOXYETHYL) ETHER
785	404	116	42	5	8	43	216.96	3.1951	LEVULINIC ACID
786	1941	334	94	21	18	149	223.29	3.1963	2-(4-PHENOXYPHENOXY)ETHYL BENZOATE
787	1273	178	81	12	18	94	196.41	3.1974	HEXYL PHENYL ETHER
788	3632	120	35	3	5	41	264.64	3.1975	1-BROMO-1-PROPENE
789	489	122	55	8	10	94	225.22	3.1986	PHENETOLE
790	5132	100	42	6	12	43	255.34	3.1998	2-HEXANONE
791	67	66	42	5	6	66	273.26	3.1999	CYCLOPENTADIENE
792	352	112	59	6	5	112	254.45	3.2010	CHLORO-BENZENE
793	3214	148	75	9	8	105	296.07	3.2014	1-PHENYL-1, 2-PROPANEDIONE
794	463	122	87	8	10	107	204.43	3.2015	P-ETHYL PHENOL
795	2023	100	52	5	8	29	322.03	3.2017	VINYL PROPRIONATE
796	699	138	59	8	10	110	236.53	3.2017	P-ETHOXYPHENOL
797	3076	100	52	5	8	29	322.03	3.2017	PROPIONATE, VINYL- 48CE
798	4902	59	47	4	9	44	250.56	3.2018	2-DEUTERO-2-METHYLPROPANE
799	4587	70	47	5	10	42	266.47	3.2022	CYCLOPENTANE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
800	6875	198	53	4	1	29	255.73	3.2033	PERFLUOROBUTYRALDEHYDE
801	660	136	75	8	8	91	216.21	3.2046	PHENYLACETIC ACID
802	3502	70	22	4	6	41	416.26	3.2057	2-BUTENAL(CROTONALDEHYDE)
803	2161	70	22	4	6	41	416.26	3.2057	2-BUTENAL
804	5074	152	32	1	0	117	406.81	3.2061	TETRACHLOROMETHANE (CARBON TETRACHLORIDE)
805	4523	114	64	8	18	57	295.71	3.2067	2,2,3-TRIMETHYLPENTANE
806	4765	120	55	9	12	105	211.06	3.2070	1-METHYL-4-ETHYLBENZENE
807	541	128	60	7	12	55	280.67	3.2070	N-BUTYL ACRYLATE
808	4167	116	53	5	12	72	303.31	3.2071	TETRAMETHYL UREA
809	4643	42	25	3	6	42	418.39	3.2087	CYCLOPROPANE
810	302	104	53	5	12	59	205.60	3.2099	3-METHOXY-2-BUTANOL
811	4605	128	60	9	20	57	265.50	3.2113	2,2,4-TRIMETHYLHEXANE
812	5077	98	48	2	4	63	296.68	3.2114	1,1-DICHLOROETHANE
813	246	100	52	4	4	28	241.64	3.2116	SUCCINIC ANHYDRIDE
814	2255	90	74	3	6	45	319.84	3.2120	LACTIC ACID
815	3163	90	74	3	6	45	319.84	3.2120	LACTIC ACID
816	1618	210	33	11	14	69	217.06	3.2131	ALLYLIDENE DICROTONATE
817	3070	170	77	3	7	43	332.34	3.2132	ISOPROPYL IODIDE 42CE
818	1325	182	73	13	10	105	295.20	3.2136	BENZOPHENONE
819	5088	79	67	5	5	79	310.91	3.2139	PYRIDINE
820	1089	164	91	10	12	135	198.56	3.2141	P-METHOXY PROPIOPHENONE
821	3532	136	29	4	9	57	341.97	3.2143	2-BROMOBUTANE
822	5596	45	28	2	7	30	249.55	3.2148	AMINOETHANE (ETHYLAMINE)
823	4584	58	35	4	10	43	311.45	3.2160	NOR-BUTANE
824	1547	202	51	10	18	57	192.45	3.2160	2-METHYL-1,3-PROPANEDIOL DIPROPIONATE
825	5701	146	74	8	18	57	207.93	3.2163	2,2,4,4-TETRAMETHYL-3-THIAPENTANE
826	925	152	65	12	8	152	208.86	3.2185	ACENAPHTHALENE
827	784	146	61	6	4	146	294.22	3.2201	P-DICHLOROBENZENE
828	3307	68	41	3	4	68	367.51	3.2222	IMIDAZOLE 316CE
829	2158	68	41	3	4	68	367.51	3.2222	IMIDAZOLE
830	1229	174	48	9	18	57	233.23	3.2225	DI-ISOBUTYL CARBONATE
831	4474	44	29	3	8	29	324.83	3.2227	PROPANE
832	3597	57	46	2	3	18	487.52	3.2234	GLYCOLONITRILE 617CE
833	5914	84	59	6	0	84	217.94	3.2246	HEXADEUTEROBENZENE
834	3565	184	28	4	9	57	351.06	3.2251	SEC-BUTYL IODIDE 580CE
835	839	148	12	10	12	133	521.71	3.2260	2-METHYL-2,3-DIHYDRO-1,4-BENZOPYRAN
836	1340	184	55	4	9	57	332.82	3.2275	1-IODOBUTANE
837	158	88	40	3	4	29	355.17	3.2276	CYCLIC ETHYLENE CARBONATE
838	83	72	29	3	4	43	328.48	3.2286	PYRUVALDEHYDE
839	280	102	65	6	14	45	211.97	3.2289	4-METHYL-2-PENTANOL

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
840	1650	212	89	14	12	170	177.42	3.2291	4-BIPHENYLYL ACETATE
841	234	98	62	2	4	63	254.03	3.2297	1,1-DICHLORCETHANE
842	1514	200	94	3	6	121	375.15	3.2318	1,2-DIBROMOPROPANE
843	649	134	100	9	10	57	191.88	3.2327	2,3-DIHYDRO-2-METHYLBENZOFURAN
844	3568	184	28	4	9	57	351.95	3.2354	ISOBUTYL IODIDE 584CE
845	858	150	51	3	0	131	445.24	3.2384	PERFLUOROPROPENE
846	5545	342	186	25	42	117	179.17	3.2387	1-NOR-HEXADECYL-(2,3-DIHYDROINDENE) (1-NOR-HEXADECYLINDAN)
847	2242	88	54	4	8	45	305.89	3.2389	3-HYDROXY-2-BUTANONE
848	3161	88	54	4	8	45	305.89	3.2389	3-HYDROXY-2-BUTANONE
849	2507	133	44	5	11	60	253.37	3.2394	SERINE ETHYL ESTER
850	427	118	47	0	0	83	283.83	3.2397	SULFURYL-CHLORO-FLUORIDE
851	5147	252	185	19	24	167	179.82	3.2400	1,1-DIPHENYLHEPTANE
852	4604	128	56	9	20	57	311.56	3.2404	2,2,3-TRIMETHYLHEXANE
853	5019	45	28	3	7	30	330.76	3.2409	2-DEUTEROPROPANE
854	6722	142	104	10	22	57	302.61	3.2412	2,2,5,5-TETRAMETHYLHEXANE
855	324	108	41	2	5	108	416.83	3.2418	BROMOETHANE
856	265	102	75	5	10	57	236.15	3.2428	PIVALIC ACID
857	6223	136	62	8	8	105	309.83	3.2430	METHYL BENZOATE
858	2359	106	40	2	3	27	353.38	3.2436	VINYL BROMIDE
859	1063	162	73	10	10	117	210.41	3.2439	(AR-VINYLPHENYL) ACETIC ACID
860	5361	190	124	14	22	105	195.23	3.2442	2-PHENYLOCTANE
861	2531	88	46	4	8	45	325.24	3.2447	ACETOIN
862	2600	178	40	11	14	105	304.25	3.2449	N-BUTYL BENZOATE
863	4853	116	57	6	12	43	229.98	3.2452	4-HYDROXY-4-METHYL-2-PENTANONE (DIACETONE ALCOHOL)
864	6734	86	61	6	14	43	339.18	3.2461	2,3-DIMETHYLBUTANE
865	6098	64	39	2	5	64	484.38	3.2494	CHLOROETHANE
866	1336	184	18	8	9	105	448.70	3.2512	4-BROMO-O-XYLENE
867	3736	122	51	4	10	45	227.38	3.2513	DIETHYLENE GLYCOL 757CE
868	2369	106	52	4	10	45	227.38	3.2513	DIETHYLENE GLYCOL
869	2445	114	65	6	10	43	208.21	3.2527	4-METHYL-2,3-PENTANEDIONE
870	3729	114	65	6	10	43	208.21	3.2527	4-METHYL-2,3-PENTANEDIONE
871	793	146	42	7	14	43	275.17	3.2536	DI-N-PROPYL CARBONATE
872	3172	98	64	6	10	43	210.91	3.2548	5-HEXEN-2-ONE
873	1740	224	81	13	17	133	192.04	3.2551	A-CHLOROISOBUTYRO-2,4-DIMETHYLPHENONE
874	4719	128	52	9	20	57	308.84	3.2562	2,2-DIMETHYL-3-ETHYLPENTANE
875	205	92	64	4	9	43	299.80	3.2569	1-CHLORO-2-METHYLPROPANE
876	4477	72	56	5	12	43	327.97	3.2574	NOR-PENTANE
877	4216	212	27	10	9	128	364.98	3.2581	1,2-DIACETOXY-3-FLUOROBENZENE
878	4721	78	45	6	6	78	230.59	3.2583	BENZENE
879	347	110	75	6	6	95	229.28	3.2584	2-FURYL METHYL KETONE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
880	4475	58	42	4	10	43	311.46	3.2586	NOR-BUTANE
881	4541	56	31	4	8	41	297.08	3.2588	2-METHYLPROPENE
882	4888	78	35	6	6	78	239.64	3.2588	BENZENE
883	3290	96	48	2	2	61	316.84	3.2591	VINYLIDINE CHLORIDE 296CE
884	2290	96	45	2	2	61	316.84	3.2591	VINYLIDINE CHLORIDE
885	5270	88	52	5	12	45	255.97	3.2593	ETHYL ISOPROPYL ETHER
886	4620	86	31	6	14	57	421.07	3.2594	3-METHYLPENTANE
887	1974	98	64	6	10	43	211.09	3.2615	5 HEXEN 2 ONE
888	2406	110	59	2	6	79	274.48	3.2631	DIMETHYL SULFITE
889	5242	76	59	3	8	45	261.50	3.2633	2-METHOXY-1-ETHANOL (ETHYLENE GLYCOL MONOMETHYL ETHER)
890	4849	86	59	5	10	43	240.41	3.2639	2-PENTANONE (METHYL-NOR-PROPYL-KETONE)
891	2828	174	68	8	14	43	205.81	3.2641	BUTYLENE DIACETATE
892	6081	66	39	5	6	66	287.40	3.2650	CYCLOPENTADIENE
893	147	86	84	4	6	18	209.49	3.2658	SUCCINALDEHYDE
894	42	56	37	4	8	41	302.03	3.2669	2-METHYL PROPENE
895	3543	184	30	4	9	57	390.53	3.2674	N-BUTYL IODIDE 558CE
896	6608	190	85	14	22	105	195.74	3.2675	2-PHENYLOCTANE
897	254	100	53	2	3	31	233.55	3.2677	2,2,2-TRIFLUORCETHANOL
898	4606	128	64	9	20	57	270.39	3.2678	2,2,5-TRIMETHYLHEXANE
899	1101	164	57	6	13	43	327.07	3.2681	2-BROMOHEXANE
900	131	84	57	5	8	43	366.52	3.2684	METHYL-CYCLOPROPYL KETONE
901	4623	70	27	5	10	42	373.31	3.2687	1-PENTENE
902	5927	73	51	4	11	44	216.25	3.2698	1-AMINO-1-METHYLPROPANE (SEC-BUTYLAMINE)
903	552	130	51	7	14	57	281.92	3.2698	N-BUTYL-PROPIONATE
904	1165	170	100	6	12	45	226.99	3.2719	BIS-(2-CHLORO-1-METHYL) ETHYL ETHER
905	3534	142	26	8	14	41	376.63	3.2725	ISO-BUTYL METHACRYLATE
906	2065	142	26	8	14	41	376.63	3.2725	ISOBUTYL METHACRYLATE
907	4984	67	40	4	5	67	385.03	3.2732	PYRROLE
908	1745	224	84	13	17	133	182.43	3.2735	P-ETHYL-ALPHA-CHLOROISOBUTYROPHENONE
909	1261	176	72	7	13	97	264.92	3.2749	1-METHYL-4-BROMOCYCLOHEXANE
910	5278	100	49	5	8	43	223.19	3.2761	2-PROPENYL ACETATE (ALLYL ACETATE)
911	5020	46	34	3	6	31	329.56	3.2763	2,2-DIDEUTEROPROPANE
912	2594	73	21	4	11	58	271.76	3.2766	ETHYLDIMETHYLAMINE
913	5134	100	47	6	12	43	269.91	3.2767	3-METHYL-2-PENTANONE
914	64	64	35	2	5	64	461.10	3.2771	CHLORDETHANE
915	373	114	85	3	5	65	192.73	3.2771	1-CHLORO-2,2-DIFLUOROPROPANE
916	2188	75	58	3	9	44	253.44	3.2780	2-AMINO-1-PROPANOL
917	3364	75	58	3	9	44	253.44	3.2780	2-AMINO-L-PROPANOL
918	4538	56	33	4	8	41	290.55	3.2781	1-BUTENE
919	4361	562	30	36	66	479	241.29	3.2781	OXOHEXAKIS-(VINYLACETATO)-TETRABERYLLIUM(II)

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
920	1338	184	53	4	9	57	299.16	3.2793	1-IODO-2-METHYLPROPANE
921	3008	127	97	8	17	84	204.43	3.2807	N-N-W-DIMETHYL CYCLOHEXYLAMINE
922	3052	102	56	6	14	43	228.78	3.2826	DI-N-PROPYL ETHER 24CE
923	2331	102	56	6	14	43	228.78	3.2826	DI-N-PROPYL ETHER
924	3133	129	77	8	19	58	198.98	3.2829	TERT. OCTYL AMINE
925	3628	170	51	2	0	85	280.14	3.2829	ETHANE,1,2-DICHLORO-1,1,2,2-TETRAFLUORO- 648CE
926	2007	129	77	8	19	58	198.98	3.2829	T OCTYL AMINE
927	5165	130	34	1	1	51	247.68	3.2831	DIFLUOROBROMOMETHANE
928	3028	72	36	4	8	43	409.34	3.2832	ISOBUTYRALDEHYDE
929	250	100	41	5	8	55	223.38	3.2838	ETHYL ACRYLATE
930	1458	194	61	11	14	101	252.12	3.2842	2-PHENOXYETHYL PROPIONATE
931	426	118	52	6	14	45	271.74	3.2843	2-ISOPROPOXY-1-PROPANOL
932	3365	75	55	3	9	18	231.21	3.2844	2-PROPANOL,1-AMINO 374CE
933	6867	80	42	2	2	45	299.34	3.2850	1-CHLORO-1-FLUOROETHENE
934	168	88	43	4	8	57	241.61	3.2850	3-BUTENE-1,2-DIOL
935	6398	74	35	3	6	31	381.58	3.2854	ETHYL FORMATE
936	837	148	101	10	12	105	224.15	3.2859	N-BUTYROPHENONE
937	5116	72	37	4	8	43	409.46	3.2860	2-METHYLPROPANAL (ISOBUTYRALDEHYDE)
938	4619	86	32	6	14	43	312.05	3.2871	2-METHYLPENTANE
939	39	56	38	4	8	41	293.81	3.2872	1-BUTENE
940	395	116	63	7	16	55	296.92	3.2892	4-HEPTANOL
941	2006	129	69	8	19	58	198.77	3.2896	2 AMINO 5 METHYL HEPTANE
942	3132	129	69	8	19	58	198.77	3.2896	2-AMINO-2 METHYL-N-HEPTANE
943	3041	150	82	6	11	43	211.54	3.2907	BUTANE, THREO-3-CHLORO-2-ACETOXY- 13CE
944	730	140	70	8	9	91	199.55	3.2912	(2-CHLOROETHYL) BENZENE
945	4542	54	28	4	6	54	455.20	3.2914	1,3-BUTADIENE
946	6771	74	52	2	7	73	334.31	3.2914	DIMETHOXYBORANE
947	188	90	38	3	6	45	454.73	3.2916	DIMETHYL CARBONATE
948	133	84	61	5	8	43	414.48	3.2932	3-METHYL-3-BUTENE-2-ONE
949	2526	157	62	10	23	30	193.38	3.2943	1-DECYLAMINE
950	5202	162	23	4	0	93	420.90	3.2945	HEXAFLUORO-2-BUTYNE
951	850	148	49	3	4	81	222.90	3.2947	1,1,2-TRIFLUORO-2-CHLOROETHYL METHYL ETHER
952	1677	214	82	14	14	141	217.97	3.2954	ETHYL-1-NAPHTHALENEACETATE
953	264	102	50	5	10	57	315.64	3.2972	ETHYL PROPIONATE
954	1151	168	42	9	9	139	242.55	3.2985	4-CHLOROPROPIOPHENONE
955	1982	128	52	8	16	43	276.42	3.2986	OCTANONE 2
956	95	72	42	5	12	43	333.48	3.2991	N-PENTANE
957	555	130	71	6	10	57	335.14	3.2993	1-ACETOXY-2-BUTANONE
958	3069	59	42	3	9	44	231.52	3.2993	ISOPROPYLAMINE 41CE
959	5236	56	35	2	4	28	426.13	3.2994	DIAZOETHANE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
960	4625	70	28	5	10	55	301.86	3.3002	3-METHYL-1-BUTENE
961	3757	102	60	5	10	59	287.19	3.3003	2-METHYL-2-HYDROXY-BUTANE-3-ONE 778CE
962	2341	102	60	5	10	59	287.19	3.3003	2-METHYL-2-HYDROXY-BUTANE-3-ONE
963	4758	74	29	4	10	45	245.28	3.3004	2-BUTANOL (SEC-BUTYL ALCOHOL)
964	5595	45	30	2	7	44	347.44	3.3008	DIMETHYLAMINE
965	1788	232	43	8	9	232	234.48	3.3010	2-IODO-1,4-DIMETHYLBENZENE
966	37	54	31	4	6	39	465.50	3.3020	1,3-BUTADIENE
967	2714	246	65	18	30	119	200.21	3.3023	2 PHENYL,2 METHYL UNDECANE
968	1782	232	47	8	9	232	238.45	3.3025	2-IODO-1,3-DIMETHYLBENZENE
969	4839	74	46	4	10	45	237.98	3.3029	2-BUTANOL (SEC-BUTYL ALCOHOL)
970	3231	72	44	3	4	42	391.42	3.3030	PROPIOLACTONE,BETA 230CE
971	2176	72	44	3	4	42	391.42	3.3030	B-PROPIOLACTONE
972	4831	64	29	2	2	64	383.95	3.3034	1,1-DIFLUOROETHENE
973	1372	186	70	2	4	107	413.07	3.3038	1,1-DIBROMOETHANE
974	97	72	52	5	12	43	457.28	3.3042	ISOPENTANE
975	3622	128	53	8	16	43	276.60	3.3043	OCTANONE-2
976	4165	136	57	9	12	94	195.90	3.3056	PHENYL-N-PROPYL ETHER
977	740	142	80	4	8	93	325.62	3.3058	BIS-(2-CHLOROETHYL) ETHER
978	315	106	53	7	6	77	436.97	3.3067	BENZALDEHYDE
979	5030	142	50	10	22	57	292.06	3.3070	2,2,4-TRIMETHYLHEPTANE
980	195	90	46	4	10	31	326.68	3.3075	2-ETHOXYETHANOL
981	6843	162	45	4	0	93	387.16	3.3091	PERFLUORO-2-BUTYNE
982	4624	70	27	5	10	55	334.00	3.3098	2-METHYL-1-BUTENE
983	627	134	75	6	14	59	260.83	3.3100	2-(2-HYDROXY PROPOXY)-1-PROPANOL
984	1996	59	42	3	9	44	231.86	3.3100	ISOPROPYL AMINE
985	6627	302	101	22	38	105	190.14	3.3104	2-PHENYLHEXADECANE
986	2166	70	24	4	6	41	370.96	3.3113	DIMETHYLKETENE
987	4985	74	56	3	6	46	246.74	3.3119	THIACYCLOBUTANE
988	6103	76	36	3	5	41	310.51	3.3128	1-CHLORO-1-PROPENE
989	154	86	46	4	6	55	238.39	3.3147	METHYL ACRYLATE
990	96	72	42	4	8	41	395.95	3.3163	1,2-EPOXY-2-METHYL-PROPANE
991	2139	0	36	3	4	55	281.50	3.3165	PROPARGYL ALCOHOL
992	2797	198	104	10	14	69	258.11	3.3168	ETHYLENE DIMETHACRYLATE
993	229	98	50	2	1	98	249.72	3.3169	1-CHLORO-2,2-DIFLUOROETHYLENE
994	1892	276	105	10	19	57	195.47	3.3176	CHLORAL DI-SEC-BUTYL ACETAL
995	681	136	78	4	9	57	261.17	3.3181	2-BROMO-2-METHYLPROPANE
996	5246	122	67	3	7	43	301.71	3.3185	2-BROMOPROPANE
997	5613	89	64	3	7	43	432.83	3.3190	1-NITROPROPANE
998	466	122	82	4	7	43	238.92	3.3190	2-CHLOROETHYL-ACETATE
999	4864	116	44	6	12	43	253.91	3.3198	NOR-BUTYL ACETATE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1000	3153	56	37	3	4	55	281.60	3.3199	PROPARGYL ALCOHOL 152CE
1001	5241	60	41	3	8	45	301.94	3.3201	METHYL ETHYL ETHER
1002	1066	162	70	6	11	83	245.48	3.3204	BROMOCYCLOHEXANE
1003	4653	70	54	5	10	42	271.13	3.3206	CYCLOPENTANE
1004	4626	70	27	5	10	55	330.03	3.3208	2-METHYL-2-BUTENE
1005	3762	100	40	5	8	69	288.00	3.3208	TRANS-METHYL CROTONATE
1006	3631	100	40	5	8	69	288.00	3.3208	TRANS-METHYL CROTONATE
1007	2019	100	40	5	8	69	288.00	3.3208	TRANS METHYL CROTONATE
1008	2314	100	40	5	8	69	288.00	3.3208	TRANS-METHYL CROTONATE
1009	44	58	50	3	6	29	422.44	3.3215	PROPIONALDEHYDE
1010	819	148	77	3	1	113	258.51	3.3216	1,1,3,3-TETRAFLUORO-3-CHLORO-PROPENE
1011	2138	56	20	3	4	55	305.30	3.3223	PROPARGYL ALCOHOL
1012	3152	56	20	3	4	55	305.30	3.3223	PROPARGYL ALCOHOL 152CE
1013	2997	146	41	8	18	45	340.62	3.3225	1,1-DIPROPOXYETHANE
1014	3235	92	72	3	5	29	350.66	3.3226	PROPIONYL CHLORIDE 235CE
1015	2268	92	72	3	5	29	350.66	3.3226	PROPIONYL CHLORIDE
1016	2127	266	113	18	18	118	302.21	3.3230	2 5 DIMETHYLBENZYL 2 4 DIMETHYLBENZOATE
1017	417	118	45	6	14	45	282.57	3.3234	1,1-DIETHOXYETHANE
1018	4990	150	88	6	14	43	234.35	3.3240	2,5-DIMETHYL-3,4-DITHIAHEXANE
1019	797	146	75	10	7	146	187.35	3.3240	2-FLUORONAPHTHALENE
1020	4611	128	57	9	20	57	260.27	3.3256	3,3-DIETHYLPENTANE
1021	4851	100	55	6	12	43	260.60	3.3257	4-METHYL-2-PENTANONE
1022	4760	74	28	4	10	59	293.40	3.3265	2-METHYL-2-PROPANOL (TERT-BUTYL ALCOHOL)
1023	36	54	29	4	6	54	331.75	3.3289	1,2-BUTADIENE
1024	408	116	40	6	12	43	271.22	3.3291	2,2,3-TRIMETHYL-1,3-DIOXALANE
1025	6720	85	51	5	11	42	401.42	3.3300	1-METHYLPYRROLIDINE
1026	403	116	59	7	16	87	316.33	3.3301	3-ETHYL-3-PENTANOL
1027	5928	73	49	4	11	58	233.52	3.3303	1-AMINO-1,1-DIMETHYLETHANE (TERT-BUTYLAMINE)
1028	582	132	61	6	12	45	214.52	3.3311	2-(2-VINYL-OXY-ETHOXY)-ETHANOL
1029	63	64	65	2	5	31	197.49	3.3316	2-FLUOROETHANOL
1030	5922	60	32	2	4	43	418.62	3.3323	ETHANOIC ACID (ACETIC ACID)
1031	4325	178	17	11	14	178	279.99	3.3323	METHYL EUGENOL
1032	2660	322	137	23	46	43	489.09	3.3323	11-TRICOSENE
1033	629	134	102	6	14	59	286.66	3.3336	BIS-(1-METHYL-2-HYDROXYETHYL) ETHER
1034	1266	176	58	9	20	57	270.61	3.3342	DI-N-BUTOXY METHANOL
1035	2440	114	51	2	4	79	303.64	3.3343	BIS(CHLOROMETHYL)ETHER
1036	3396	114	51	2	4	79	303.64	3.3343	BIS(CHLOROMETHYL)ETHER
1037	6114	122	41	3	7	43	280.94	3.3356	1-BROMOPROPANE
1038	1973	84	27	5	8	69	418.98	3.3374	TRANS METHYL PROPENYL KETONE
1039	6594	358	106	26	46	105	191.62	3.3389	2-PHENYLEICOSANE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1040	849	148	68	11	16	133	227.02	3.3393	AR-ETHYL-1,2,4-TRIMETHYLBENZENE
1041	2128	266	120	18	18	118	325.31	3.3395	2 5 DIMETHYLBENZYL 2 5 DIMETHYLBENZOATE
1042	639	134	93	10	14	105	200.87	3.3402	O-N-PROPYLTOLUENE
1043	4617	72	29	5	12	43	477.24	3.3403	2-METHYLBUTANE (ISOPENTANE)
1044	3361	76	50	3	5	41	296.39	3.3424	2-CHLOROPROPENE
1045	1300	180	60	5	9	43	235.40	3.3427	1-BROMOISOPROPYL ACETATE
1046	3984	59	30	3	9	58	298.50	3.3428	TRIMETHYLAMINE
1047	6615	246	100	18	30	105	193.44	3.3433	2-PHENYLDODECANE
1048	680	136	63	8	8	105	257.80	3.3435	A-HYDROXYACETOPHENONE
1049	2326	101	80	6	15	58	246.86	3.3449	N-ETHYL-N-BUTYL AMINE
1050	3130	102	80	5	14	58	246.86	3.3449	PROPYLAMINE,3 DIMETHYLAMINO- 131CE
1051	4637	100	56	7	16	43	314.22	3.3451	3-ETHYLPENTANE
1052	3873	74	29	4	10	45	250.87	3.3460	SEC-BUTANOL
1053	2185	74	29	4	10	45	250.87	3.3460	SEC-BUTANOL
1054	6401	90	50	4	6	43	211.43	3.3462	ETHYL-1,1-D2 ACETATE (1,1-DIDEUTEROETHYL ACETATE)
1055	3660	130	52	6	10	55	369.44	3.3469	ACRYLATE,2-METHOXYETHYL- 678CE
1056	3879	114	31	6	10	43	257.02	3.3470	METHALLYL ACETATE
1057	2028	114	31	6	10	43	257.02	3.3470	METHALLYL ACETATE
1058	1453	194	61	6	11	43	255.57	3.3473	2-BROMO-1-METHYLPROPYL ACETATE
1059	5223	182	137	14	14	91	204.71	3.3501	1,2-DIPHENYLETHANE (BIBENZYL)
1060	6605	162	83	12	18	105	204.14	3.3507	2-PHENYLHEXANE
1061	3519	120	23	5	12	45	342.05	3.3516	2-/2-METHOXYETHOXY/-ETHANOL 534CE
1062	6487	142	56	10	22	57	275.46	3.3529	4-N-PROPYLHEPTANE
1063	3473	72	28	3	4	27	424.28	3.3543	ACRYLIC ACID
1064	309	106	53	8	10	91	224.49	3.3551	ETHYL BENZENE
1065	6055	56	36	4	8	41	306.82	3.3554	1-BUTENE
1066	4782	120	51	9	12	105	225.20	3.3558	ISOPROPYLBENZENE (CUMENE)
1067	175	88	38	4	8	28	344.95	3.3561	P-DIOXANE
1068	4641	100	60	7	16	43	307.03	3.3572	3,3-DIMETHYLPENTANE
1069	2536	89	43	4	11	58	233.93	3.3572	2-METHYL-2-AMINO-1-PROPANOL
1070	2172	72	27	3	4	72	427.25	3.3578	ACRYLIC ACID
1071	3382	94	53	5	6	41	344.22	3.3582	GLUTARONITRILE
1072	795	146	58	10	7	146	190.87	3.3587	1-FLUORONAPHTHALENE
1073	2493	159	77	8	17	86	254.50	3.3600	NORLEUCINE ETHYL ESTER
1074	2246	88	47	4	8	87	341.78	3.3606	1,3-DIOXANE
1075	4887	56	34	4	8	28	416.69	3.3606	CYCLOBUTANE
1076	2598	101	38	5	11	59	251.33	3.3613	3-METHYLBUTYLAMIDE
1077	3229	136	69	4	9	57	325.88	3.3624	TERT-BUTYL BROMIDE
1078	5612	57	42	3	7	42	301.38	3.3624	N-METHYLETHYLENIMINE
1079	3280	78	59	4	2	51	376.79	3.3628	VINYLIDINE CYANIDE 284CE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1080	5138	114	50	7	14	57	325.96	3.3634	3-HEPTANONE
1081	1444	194	93	10	10	163	199.36	3.3635	DIMETHYL PHTHALATE
1082	3478	112	41	6	5	112	316.85	3.3654	CHLORCBENZENE
1083	2423	112	41	6	5	112	316.85	3.3654	CHLOROBENZENE
1084	5080	84	55	6	0	84	230.79	3.3658	HEXADEUTEROBENZENE
1085	1750	226	78	16	18	105	231.02	3.3676	A-PHENYLETHYL ETHER
1086	430	118	63	8	6	118	239.90	3.3684	BENZOFURAN
1087	2335	102	62	8	6	102	222.82	3.3686	PHENYLACETYLENE
1088	2096	164	57	10	12	91	221.56	3.3699	EHTYLPHENYL ACETATE
1089	3710	164	57	10	12	91	221.56	3.3699	ETHYLPHENYL ACETATE
1090	2316	100	41	5	8	43	335.79	3.3703	2,3-PENTANEDIONE
1091	5634	128	62	9	20	57	326.87	3.3709	2,2-DIMETHYL-3-ETHYLPENTANE
1092	6059	56	37	4	8	41	309.79	3.3711	1-BUTENE
1093	6111	112	50	6	5	112	278.35	3.3714	CHLORCBENZENE
1094	5585	104	36	5	12	31	413.13	3.3717	DIETHOXYMETHANE (DIETHYL FORMAL, ETHYLAL)
1095	107	74	61	4	10	45	228.31	3.3720	SEC-BUTYL ALCOHOL
1096	3530	136	32	4	9	57	362.38	3.3724	PROPANE,1-BROMO-2-METHYL- 545CE
1097	3273	66	57	5	6	66	352.72	3.3729	1,3-CYCLOPENTADIENE 277CE
1098	2155	66	55	5	6	66	352.72	3.3729	1,3-CYCLOPENTADIENE
1099	746	142	66	2	1	142	333.79	3.3746	BROMO-1,2-DIFLUOROETHYLENE
1100	1757	226	46	7	15	57	256.86	3.3748	1-IODOHEPTANE
1101	6870	168	66	2	1	83	291.43	3.3757	1,1-DIFLUORO-1,2,2-TRICHLOROETHANE
1102	2130	266	101	18	18	119	250.30	3.3765	2 4 DIMETHYLBENZYL 2 5 DIMETHYLBENZOATE
1103	4727	120	85	9	12	91	209.21	3.3771	NOR-PROPYLBENZENE
1104	2802	180	45	6	12	73	215.96	3.3791	L-INOSITOLS
1105	2785	194	49	12	18	75	214.18	3.3792	1,1-DIETHOXY-2-PHENYLETHANE
1106	3395	114	32	2	1	29	286.96	3.3795	CHLORODIFLUORO ACETALDEHYDE
1107	353	112	53	3	6	76	285.62	3.3795	1,3-DICHLORO-PROPANE
1108	1762	226	91	16	18	170	195.35	3.3801	4-BIPHENYLYL-N-BUTYL ETHER
1109	3665	128	45	8	4	128	215.74	3.3805	1,2-DICYANOBENZENE
1110	3176	102	63	4	6	102	223.22	3.3815	ACETIC ANHYDRIDE
1111	971	156	61	3	3	77	201.56	3.3817	3-BROMO-3,3-DIFLUORO-1-PROPENE
1112	1572	206	49	1	1	129	311.61	3.3825	DIBROMOCHLOROMETHANE
1113	1815	236	76	14	20	87	209.84	3.3826	ETHYL-P-TERT-BUTYLPHENOXYACETATE
1114	5121	88	47	4	8	43	296.24	3.3847	2-METHYLPROPANOIC ACID (ISOBUTYRIC ACID)
1115	207	92	42	4	9	57	328.37	3.3849	2-CHLORO-2-METHYLPROPANE
1116	2715	246	64	18	30	119	204.95	3.3851	2 PHENYL 2,5,7,7, TETRA METHYL OCTANE
1117	464	122	82	8	10	91	280.14	3.3872	PHENETHYL-ALCOHOL
1118	2364	106	66	7	6	77	467.98	3.3875	BENZALDEHYDE
1119	3443	106	66	7	6	77	467.98	3.3875	BENZALDEHYDE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1120	3590	0	45	0	0	57	278.33	3.3878	1-PENTENE-2-AL,4,4-DIMETHYL- 610CE
1121	3023	0	45	0	0	57	278.33	3.3878	2-AL-4, 4, DIMETHYL-1PENTENE
1122	202	92	68	3	5	29	359.34	3.3902	PROPIONYL CHLORIDE
1123	3362	76	42	3	5	41	305.99	3.3904	TRANS-1-CHLOROPROPENE
1124	213	94	58	6	6	94	218.91	3.3905	PHENOL
1125	121	78	45	3	7	42	281.01	3.3906	1-CHLORO-PROPANE
1126	1287	178	55	11	14	105	312.32	3.3906	N-BUTYL BENZOATE
1127	579	132	44	2	3	97	312.90	3.3908	1,1,1-TRICHLORO-ETHANE
1128	2086	158	33	9	18	43	421.71	3.3912	2 METHYLBUTYL ISOBUTYRATE
1129	953	156	58	3	6	77	265.07	3.3916	1-CHLORO-2-BROMOPROPANE
1130	738	142	52	11	10	142	292.86	3.3917	1-METHYLNAPHTHALENE
1131	4794	90	22	4	10	59	371.55	3.3925	1,1-DIMETHOXYETHANE (DIMETHYL ACETAL)
1132	4638	100	66	7	16	57	387.65	3.3927	2,2-DIMETHYLPENTANE
1133	3552	121	39	7	7	103	259.62	3.3932	BENSAMIDE C6H5CONH2
1134	124	80	45	2	2	80	324.91	3.3933	1-FLUORO-2-CHLORO-ETHYLENE
1135	3086	94	41	1	3	93	482.53	3.3935	METHYLENE BROMIDE 83CE
1136	4615	128	56	9	20	43	332.55	3.3940	2,3,3,4-TETRAMETHYLPENTANE
1137	2115	196	24	12	20	82	368.82	3.3952	CIS 3 HEXENYL TRANS 2 HEXENOATE
1138	5009	79	42	6	5	79	228.81	3.3955	DEUTEROBENZENE
1139	5310	128	63	9	20	57	299.54	3.3963	2,2-DIMETHYLHEPTANE
1140	499	124	89	7	8	95	209.37	3.3974	2-FURYL-ETHYL-KETONE
1141	4540	56	34	4	8	41	328.00	3.3977	TRANS-2-BUTENE
1142	1787	232	46	8	9	232	229.18	3.3982	4-IODO-M-XYLENE
1143	461	121	85	7	4	121	201.79	3.3983	P-FLUORO-BENZONITRILE
1144	1282	178	69	7	15	57	259.91	3.3988	3-BROMCHEPTANE
1145	1485	198	49	14	14	92	290.09	3.3997	BENZYL ETHER
1146	2807	0	47	0	0	73	217.68	3.4005	SCYLLO INOSITOLS
1147	170	88	38	4	8	73	391.47	3.4012	2-METHYL-1,3-DIOXOLANE
1148	6214	246	127	18	30	105	202.92	3.4013	2-PHENYLDODECANE
1149	3355	60	36	3	5	59	310.09	3.4017	2-FLUOROPROPENE
1150	3503	71	25	3	5	28	367.73	3.4019	2-HYDROXYPROPANENITRILE(LACTONITRILE)
1151	2168	71	25	3	5	28	367.73	3.4019	2-HYDROXYPROPANENITRILE
1152	770	143	48	2	3	97	349.89	3.4022	1,1-DICHLORO-1-NITROETHANE
1153	5620	116	70	6	12	43	236.30	3.4024	4-HYDROXY-4-METHYL-2-PENTANONE (DIACETONE ALCOHOL)
1154	84	72	44	3	4	72	396.46	3.4026	ACRYLIC ACID
1155	2033	116	55	6	12	43	335.49	3.4045	T BUTYL ACETATE
1156	3122	116	55	6	12	43	335.49	3.4045	T-BUTYL ACETATE 123CE
1157	3767	86	42	4	6	45	220.09	3.4059	2,3-BUTANEDIOL
1158	2257	90	42	4	10	45	220.09	3.4059	2,3-BUTANEDIOL
1159	6629	330	109	24	42	105	194.75	3.4061	2-PHENYLOCTADECANE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1160	162	88	45	4	8	57	332.93	3.4061	METHYL PROPIONATE
1161	1703	220	63	2	3	143	307.16	3.4062	1,2-DIBROMOCHLOROETHANE
1162	235	98	68	6	10	69	236.94	3.4065	4-METHYL-1-PENTYNOL-1
1163	527	128	77	1	2	49	413.02	3.4066	CHLORO-BROMO-METHANE
1164	4275	374	20	28	22	374	369.18	3.4072	2,6-DI-ALPHA-NAPHTHYLIDINE-CYCLOHEXANONE
1165	964	156	81	4	6	29	211.27	3.4078	ETHYL DICHLOROACETATE
1166	2518	115	98	6	13	58	228.07	3.4082	DIACETONEAMINE
1167	868	150	72	9	10	105	297.39	3.4083	ETHYL-BENZOATE
1168	548	130	45	2	1	130	483.07	3.4086	TRICHLOROETHYLENE
1169	6581	190	92	14	22	105	202.82	3.4088	2-PHENYLOCTANE
1170	38	56	35	4	8	41	334.58	3.4089	2-BUTENE
1171	4642	100	63	7	16	57	420.67	3.4090	2,2,3-TRIMETHYLBUTANE
1172	5022	78	58	3	7	42	297.06	3.4093	1-CHLOROPROPANE
1173	2018	100	53	5	8	55	301.63	3.4095	ETHYL ACRYLATE
1174	60	62	43	3	7	29	292.45	3.4098	1-FLUOROPROPANE
1175	189	90	28	3	6	31	282.74	3.4098	ETHYLENE GLYCOL MONOFORMATE
1176	261	102	72	5	10	71	261.92	3.4101	TETRAHYDROFURFURYL ALCOHOL
1177	2260	90	69	3	3	27	319.11	3.4104	ACRYLYL CHLORIDE
1178	206	92	56	4	9	56	298.92	3.4112	1-CHLOROBUTANE
1179	4248	108	50	6	8	54	323.55	3.4118	ALPHA-BETA-DIMETHYL-SUCCINONITRILE
1180	5276	88	59	5	12	45	232.25	3.4133	METHYL ISOBUTYL ETHER
1181	6869	100	59	2	3	65	229.73	3.4137	1,1-DIFLUORO-1-CHLOROETHANE
1182	3664	128	47	8	4	128	220.64	3.4153	1,4-DICYANOBENZENE
1183	169	88	48	4	8	45	309.27	3.4160	ISOPROPYL FORMATE
1184	5277	100	64	5	8	43	240.87	3.4161	2,4-PENTANEDIONE (ACETYLACETONE)
1185	6097	214	34	4	1	45	455.12	3.4166	HEPTAFLUOROBUTANOIC ACID
1186	992	157	106	5	4	157	412.74	3.4173	3-BROMOPYRIDINE
1187	602	132	59	7	16	59	233.41	3.4176	1-ISOPROPOXY-2-METHYL-2-PROPANOL
1188	4921	88	40	4	8	28	304.41	3.4180	1,4-DIOXANE
1189	4354	502	23	18	18	430	288.49	3.4184	OXOHEXAKIS-(2,2-D2-PROPIONATO)-TETRABERYLLIUM(II)
1190	6612	218	100	16	26	105	198.82	3.4187	2-PHENYLDECANE
1191	2532	134	65	9	10	43	279.04	3.4190	PHENYLACETONE
1192	5235	45	37	2	7	30	263.33	3.4207	AMINOETHANE (ETHYLAMINE)
1193	183	88	49	5	12	57	315.58	3.4209	2,2-DIMETHYL-1-PROPANOL
1194	212	94	56	2	3	50	264.43	3.4210	CHLOROACETIC ACID
1195	6881	283	71	5	0	69	253.30	3.4211	PERFLUOROBUTYLMETHYLENEIMINE
1196	240	100	75	6	12	43	276.45	3.4212	METHYL ISOBUTYL KETONE
1197	2126	266	126	18	18	118	302.61	3.4224	2 4 DIMETHYLBENZYL 3 5 DIMETHYLBENZOATE
1198	3658	130	57	6	10	43	315.99	3.4230	TETRAHYDROFURFURYL ACETATE
1199	174	88	49	5	12	45	233.48	3.4232	2-PENTANOL

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1200	4595	56	37	4	8	41	324.03	3.4233	2-METHYLPROPENE
1201	3780	160	64	3	3	43	189.31	3.4237	1,1,1-TRICHLORC-PROPANONE
1202	2214	84	49	2	4	28	330.16	3.4237	3-AMINC-2,1,4-TRIAZOLE
1203	1631	212	76	6	13	43	286.07	3.4239	1-IODOHEXANE
1204	3077	78	58	2	3	43	271.61	3.4246	ACETYLCHLCRIDE 58CE
1205	2194	78	58	2	3	43	271.61	3.4246	ACETYL CHLORIDE
1206	135	84	57	5	8	69	233.17	3.4248	METHYLBUTYNOL
1207	1431	192	48	9	20	59	228.08	3.4248	TRIPROPYLENE GLYCOL'
1208	4526	114	69	8	18	43	341.36	3.4250	2,3,4-TRIMETHYLPENTANE
1209	2092	164	68	10	12	133	277.44	3.4251	3 5 DIMETHYL METHYLBENZOATE
1210	4621	86	30	6	14	43	517.23	3.4252	2,2-DIMETHYLBUTANE
1211	6620	274	106	20	34	105	195.70	3.4254	2-PHENYLTETRADECANE
1212	2269	92	28	7	8	91	315.01	3.4258	BICYCLC(2.2.1)-2,5-HEPTADIENE
1213	3479	116	45	6	12	43	235.64	3.4268	2-BUTYLACETATE
1214	2034	116	45	6	12	43	235.64	3.4268	2 BUTYLACETATE
1215	2190	76	54	3	8	45	284.40	3.4268	PROPYLENE GLYCOL
1216	3159	76	54	3	8	45	284.40	3.4268	PROPYLENE GLYCOL
1217	3733	116	53	6	12	43	243.85	3.4270	DIACETONE ALCOHOL
1218	2463	116	53	6	12	43	243.85	3.4270	DIACETONE ALCOHOL
1219	286	104	55	5	12	45	357.24	3.4275	1-ETHOXY-2-PROPANOL
1220	3626	142	59	4	8	93	415.08	3.4275	ETHER,BIS-/2-CHLOROETHYL/- 646CE
1221	6820	136	35	8	8	105	321.45	3.4279	METHYL BENZOATE
1222	4509	54	31	4	6	54	335.73	3.4281	2-BUTYNE
1223	704	138	84	8	10	107	257.04	3.4284	1-PHENYL-1,2-DIHYDROXYETHANE
1224	6753	172	48	1	2	93	484.65	3.4291	METHYLENE BROMIDE
1225	512	126	90	7	7	91	217.10	3.4292	ALPHA-CHLORO-TOLUENE
1226	942	154	90	12	10	154	338.67	3.4296	ACENAPHTHENE
1227	685	136	53	8	8	121	263.13	3.4299	P-HYDROXYACETOPHENONE
1228	4777	106	51	8	10	91	234.97	3.4300	ETHYLBENZENE
1229	6404	106	48	5	6	46	230.53	3.4301	ISOPROPYL-1-D ACETATE-D3 (1-DEUTEROISOPROPYL TRIDEUTEROACETA
1230	460	120	40	6	4	90	362.13	3.4301	BENZOFURAZAN
1231	1989	186	67	6	0	186	308.87	3.4304	HEXAFLUORCBENZENE
1232	276	102	37	6	14	59	319.17	3.4314	ETHYL BUTYL ETHER
1233	3293	111	77	5	5	43	307.45	3.4317	VINYL ACETATE,ALPHA-CYANO- 300CE
1234	4980	103	61	7	5	103	237.73	3.4319	BENZONITRILE (CYANOBENZENE)
1235	163	88	52	4	8	60	260.57	3.4323	N-BUTYRIC ACID
1236	3510	100	26	5	8	28	400.91	3.4336	GAMMA-VALEROLACTONE
1237	185	89	45	3	4	49	358.73	3.4338	3-CHLOROPROPIONITRILE
1238	2145	60	32	2	8	60	506.10	3.4349	SYM-DIMETHYLHYDRAZINE
1239	847	148	15	10	12	117	543.68	3.4350	METHYL-AR-VINYLBENYL ETHER

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1240	3042	108	80	4	9	45	220.07	3.4377	2-BUTANOL,ETHRYO-3-CHLORO- 14CE
1241	2385	108	80	4	9	45	220.07	3.4377	ETHYRO-3-CHLORO 2-BUTANOL
1242	388	114	23	6	10	55	399.02	3.4381	6-HYDROXYHEXANCIC ACID LACTONE
1243	3893	136	35	4	9	57	312.39	3.4381	SEC-BROMOBUTANE
1244	176	88	51	5	12	59	278.55	3.4394	3-PENTANOL
1245	4236	70	60	4	6	43	407.03	3.4404	DIVINYL ETHER
1246	3356	60	41	3	5	59	301.79	3.4408	ALLYL FLUORIDE
1247	2250	88	60	4	8	45	320.94	3.4409	3-HYDROXY-2-BUTANONE
1248	4242	94	76	5	6	94	230.47	3.4410	4-AMINOPYRIDINE
1249	945	154	48	9	11	91	226.07	3.4424	(3-CHLOROPROPYL) BENZENE
1250	2017	98	65	5	6	55	369.19	3.4426	VINYL ACRYLATE
1251	6740	78	64	2	3	43	272.15	3.4429	ACETYL CHLORIDE
1252	1497	198	47	5	11	43	360.66	3.4431	1-IODOPENTANE
1253	5262	136	65	4	9	57	347.54	3.4433	2-BROMOBUTANE
1254	1851	250	70	14	18	149	195.08	3.4436	DI-N-PROPYL PHTHALATE
1255	6850	132	37	3	1	113	393.86	3.4437	1,1,1,3,3-PENTAFLUOROPROPENE
1256	4522	114	64	8	18	43	330.94	3.4441	3-METHYL-3-ETHYLPENTANE
1257	693	136	49	7	4	136	412.12	3.4446	O-PHENYLENE CYCLIC CARBONATE
1258	6003	67	42	4	5	67	437.96	3.4462	PYRROLE
1259	201	92	64	3	8	61	323.70	3.4467	GLYCEROL
1260	5427	166	77	12	22	97	374.02	3.4469	1,1-DICYCLOPENTYLETHANE
1261	755	142	56	3	4	63	380.34	3.4469	2-CHLOROETHYL CHLOROFORMATE
1262	3706	114	42	7	14	43	359.32	3.4474	2,4 DIMETHYL PENTANAL
1263	306	104	70	5	12	45	270.41	3.4479	2,4-PENTANEDIOL
1264	5283	118	52	5	10	29	383.44	3.4480	ETHYL CARBONATE
1265	5584	90	38	4	10	59	365.45	3.4486	1,1-DIMETHOXYETHANE (DIMETHYL ACETAL)
1266	3363	76	46	3	5	41	314.42	3.4487	CIS-L-CHLOROPROPENE
1267	2420	111	75	5	5	43	308.35	3.4497	A CYANOVINYL ACETATE
1268	1144	168	85	10	16	43	249.54	3.4497	METHYL ISOPROPENYL KETONE DIMER
1269	4746	84	64	6	12	56	330.24	[illegible]	ETHYLCYCLOBUTANE
1270	3554	122	134	2	3	43	307.65	3.4508	ACETYL BROMIDE 569CE
1271	2293	96	56	5	4	96	443.79	3.4519	2-FURALDEHYDE
1272	2219	84	56	4	4	42	265.76	3.4521	CYCLOBUTANE-1,3-DIONE
1273	3286	82	67	4	2	42	265.76	3.4521	DIKETENE 290CE
1274	5018	45	29	3	7	29	457.08	3.4522	1-DEUTEROPROPANE
1275	5023	78	60	3	7	43	276.46	3.4524	2-CHLOROPROPANE
1276	4482	86	59	6	14	57	445.50	3.4529	3-METHYLPENTANE
1277	1335	184	47	2	2	186	434.52	3.4532	1,2-DIBROMOETHYLENE
1278	3330	350	66	7	0	131	270.45	3.4535	1-HEPTENE,PERFLUORO- 339CE
1279	4592	56	38	4	8	41	327.60	3.4538	1-BUTENE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1280	5258	122	65	4	7	43	249.13	3.4539	2-CHLOROETHYL ACETATE
1281	4640	100	72	7	16	43	374.33	3.4541	2,4-DIMETHYLPENTANE
1282	6088	160	94	12	16	80	214.91	3.4542	2,5-DIMETHYL-(3A,4,7,7A-TETRAHYDRO-4,7-METHANOINDENE)(DI(MET
1283	4330	414	15	29	50	91	540.95	3.4547	BETA-SITOSTEROL
1284	3420	128	28	1	2	49	389.65	3.4549	METHYLENE CHLOROBROMIDE 433CE
1285	5308	59	39	4	9	44	316.01	3.4550	2-DEUTEROBUTANE
1286	2217	84	24	5	8	55	418.91	3.4556	CYCLOPENTANONE
1287	774	144	73	8	16	43	288.14	3.4558	TETRAHYDRO-2,5-DIMETHYLPYRAN-2-METHANOL
1288	701	138	78	3	7	45	212.02	3.4558	1-BROMO-2-PROPANOL
1289	5301	130	79	8	18	57	274.96	3.4564	DI-NOR-BUTYL ETHER
1290	6399	76	42	3	4	33	361.31	3.4583	ETHYL-1,1-D2-FORMATE (1,1-DIDEUTEROETHYL FORMATE)
1291	4947	94	51	2	6	94	380.72	3.4588	2,3-DITHIABUTANE (DIMETHYL DISULFIDE)
1292	4865	116	45	6	12	57	346.32	3.4591	NOR-PROPYL-NOR-PROPANOATE
1293	1747	226	67	14	10	105	302.07	3.4591	BENZOIC ANHYDRIDE
1294	2386	108	84	4	9	45	221.49	3.4599	THREO 3-CHLORO 2-BUTANOL
1295	3043	108	84	4	9	45	221.49	3.4599	2-BUTANOL, THREO-3-CHLORO- 15CE
1296	3171	96	57	5	4	96	444.25	3.4601	2-FURALDEHYDE
1297	1202	172	59	5	10	93	328.83	3.4604	BIS(2-CHLOROETHOXY) METHANE
1298	2216	84	32	2	4	28	401.96	3.4609	4-AMINO-1,2,4-TRIAZOLE
1299	3155	60	33	2	8	60	508.44	3.4614	HYDRAZINE,SYM-DIMETHYL 154CE
1300	1483	198	54	5	11	43	345.02	3.4631	3-METHYL-1-IODOBUTANE
1301	1972	84	25	5	8	41	449.33	3.4633	CIS METHYL PROPENYL KETONE
1302	4488	100	59	7	16	43	336.28	3.4639	3-ETHYLPENTANE
1303	4577	84	65	6	12	41	356.78	3.4639	3,3-DIMETHYL-1-BUTENE
1304	601	132	95	9	8	131	284.05	3.4647	2-METHYL-BENZOFURAN
1305	333	108	47	6	4	108	366.24	3.4657	QUINONE
1306	732	140	82	7	5	105	274.68	3.4674	BENZOYL CHLORIDE
1307	5929	99	68	6	13	56	237.20	3.4685	CYCLOHEXYLAMINE
1308	578	132	61	2	0	132	328.89	3.4690	1,1-DICHLORO-2,2-DIFLUORO-ETHYLENE
1309	3550	196	22	12	20	41	505.14	3.4692	LINALYL ACETATE 565CE
1310	5024	79	70	3	6	44	260.08	3.4704	2-CHLORO-2-DEUTEROPROPANE
1311	5250	72	47	4	8	44	511.10	3.4712	ETHYL ETHENYL ETHER (ETHYL VINYL ETHER)
1312	2941	145	82	10	11	144	307.56	3.4712	2,5-DIMETHYL INDOLE
1313	4492	100	59	7	16	43	328.89	3.4717	3,3-DIMETHYLPENTANE
1314	4579	98	66	7	14	57	319.76	3.4721	4,4-DIMETHYL-1-PENTENE
1315	5563	118	46	6	14	45	311.39	3.4721	1,1-DIETHOXYETHANE (DIETHYL ACETAL, ACETAL)
1316	4478	72	55	5	12	43	489.05	3.4729	2-METHYLBUTANE (ISOPENTANE)
1317	4720	128	56	9	20	43	309.42	3.4732	2,4-DIMETHYL-3-ETHYLPENTANE
1318	4499	56	37	4	8	41	319.99	3.4733	2-METHYLPROPENE
1319	6662	358	110	26	46	105	201.85	3.4740	2-PHENYLEICOSANE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1320	3459	122	62	8	10	107	405.88	3.4743	ALPHA-METHYL BENZYL ALCOHOL
1321	2125	266	129	18	18	119	284.50	3.4744	2 4 DIMETHYLBENZYL 2 4 DIMETHYLBENZOATE
1322	6096	102	46	5	10	57	282.59	3.4750	2,2-DIMETHYLPROPANOIC ACID
1323	3215	150	98	6	14	45	266.60	3.4754	TRIETHYLENE GLYCOL
1324	2922	150	98	6	14	45	266.60	3.4754	TRIETHYLENE GLYCOL
1325	3226	64	39	2	5	27	546.63	3.4755	ETHYL CHLORIDE 225CE
1326	2151	64	39	2	5	27	546.63	3.4755	ETHYL CHLORIDE
1327	3531	136	28	4	9	57	387.03	3.4771	BUTANE,1-BROMO- 546CE
1328	43	58	42	3	6	28	471.62	3.4781	PROPYLENE OXIDE
1329	3422	154	30	2	0	85	398.24	3.4782	ETHANE,CHLOROPENTAFLUORO- 435CE
1330	3533	142	27	8	14	41	379.28	3.4784	N-BUTYL METHACRYLATE
1331	2066	142	27	8	14	41	379.28	3.4784	N BUTYL METHACRYLATE
1332	4539	56	34	4	8	41	337.74	3.4787	CIS-2-BUTENE
1333	5166	164	47	1	0	85	282.63	3.4789	DIFLUOROCHLOROBROMOMETHANE
1334	2177	72	38	3	4	43	289.57	3.4791	PYRUVALDEHYDE
1335	3157	72	38	3	4	43	289.57	3.4791	PYRUVALDEHYDE
1336	1449	194	66	11	14	120	243.53	3.4798	N-BUTYL SALICYLATE
1337	6005	79	46	5	5	79	361.65	3.4812	PYRIDINE
1338	996	158	56	10	22	71	345.64	3.4812	PENTYL ETHER
1339	293	104	81	4	8	45	207.22	3.4813	2-HYDROXY PROPIONATE
1340	625	134	85	10	14	105	218.22	3.4828	SEC-BUTYLBENZENE
1341	1493	198	69	8	16	91	282.00	3.4835	4,4-DICHLOROBUTYL BUTYL ETHER
1342	5587	118	46	6	14	45	320.54	3.4837	1,1-DIETHOXYETHANE (DIETHYL ACETAL, ACETAL)
1343	50	58	40	3	6	57	315.04	3.4838	ALLYL ALCOHOL
1344	3526	130	23	6	10	29	458.84	3.4840	ALLYLETHYL CARBONATE
1345	1279	178	70	11	14	107	314.93	3.4855	P-ETHYLPHENYL PROPIONATE
1346	111	76	54	3	5	41	327.27	3.4859	1-CHLORO-PROPENE
1347	4496	56	38	4	8	41	322.19	3.4862	1-BUTENE
1348	5931	101	64	6	15	44	256.76	3.4863	N-ISOPROPYL-1-AMINO-2-METHYLETHANE (DIISOPROPYLAMINE)
1349	1054	162	67	1	1	83	295.12	3.4875	BROMODICHLOROMETHANE
1350	3998	154	38	12	10	154	386.91	3.4883	ACENAPTHENE
1351	203	92	47	4	9	56	409.88	3.4886	2-CHLOROBUTANE
1352	4618	86	32	6	14	57	528.52	3.4892	NOR-HEXANE
1353	1980	114	80	7	14	43	250.23	3.4893	5 METHYL 2 HEXANONE
1354	149	86	39	5	10	57	292.05	3.4895	CYCLOPENTANOL
1355	4481	86	62	6	14	43	334.49	3.4910	2-METHYLPENTANE
1356	6309	190	112	14	22	105	214.29	3.4911	2-PHENYLOCTANE
1357	4489	100	67	7	16	57	413.63	3.4919	2,2-DIMETHYLPENTANE
1358	3489	72	53	3	4	44	487.41	3.4919	VINYL FORMATE 502CE
1359	2014	72	53	3	4	44	487.41	3.4919	VINYL FORMATE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1360	6772	102	132	4	11	73	257.79	3.4922	ETHYLDIMETHOXYBORANE
1361	888	150	72	5	11	71	341.07	3.4927	2-BROMO-2-METHYL-BUTANE
1362	4138	214	130	14	14	91	250.70	3.4932	DIBENZYLSULPHIDE
1363	4137	197	94	14	15	91	280.21	3.4943	DIBENZYLAMINE
1364	1360	184	47	4	9	57	275.04	3.4943	2-IODO-2-METHYLPROPANE
1365	5598	59	42	3	9	58	326.13	3.4945	TRIMETHYLAMINE
1366	2321	100	24	5	8	56	431.05	3.4948	Y-VALEROLACTONE
1367	5583	88	38	4	8	73	458.32	3.4949	2-METHYL-1,3-DIOXACYCLOPENTANE (2-METHYL-1,3-DIOXOLANE)
1368	3188	114	82	7	14	43	250.41	3.4960	5-METHYL-2-HEXANONE
1369	4783	120	53	9	12	105	228.83	3.4962	1-METHYL-2-ETHYLBENZENE
1370	747	142	82	3	4	59	296.29	3.4976	METHYL DICHLORACETATE
1371	6733	86	55	6	14	57	423.61	3.5003	3-METHYLPENTANE
1372	5561	90	41	4	10	59	383.96	3.5006	1,1-DIMETHOXYETHANE (DIMETHYL ACETAL)
1373	2823	174	36	10	22	57	459.66	3.5008	1,1-DIBUTOXYETHANE
1374	144	86	47	4	6	86	378.13	3.5023	TRANS-CROTONIC ACID
1375	5617	90	67	4	10	31	347.82	3.5023	2-ETHOXYETHANOL (ETHYLENE GLYCOL MONOETHYL ETHER)
1376	1094	164	71	10	12	122	296.82	3.5023	M-ETHYLPHENYL ACETATE
1377	242	100	76	6	12	57	318.42	3.5026	METHYL-TERT-BUTYL KETONE
1378	2940	145	86	10	11	144	315.16	3.5029	2,6-DIMETHYL INDOLE
1379	3358	80	43	5	6	52	287.53	3.5034	CYANOBUTADIENE
1380	1948	350	81	7	0	69	285.80	3.5035	PERFLUOROMETHYLCYCLOHEXANE
1381	4507	54	33	4	6	39	496.00	3.5039	1,3-BUTADIENE
1382	677	136	45	8	8	121	280.38	3.5053	O-HYDROXYACETOPHENONE
1383	922	152	89	8	8	121	246.72	3.5056	METHYL P-HYDROXYBENZOATE
1384	2037	116	65	6	12	43	243.74	3.5059	ISOBUTYL ACETATE
1385	3039	116	65	6	12	43	243.74	3.5059	ISOBUTYL ACETATE 11CE
1386	2811	180	81	13	8	180	237.48	3.5064	FLUORENONE
1387	6735	92	69	7	8	91	289.56	3.5075	METHYLBENZENE (TOLUENE)
1388	2300	98	28	6	10	41	290.85	3.5075	METHYLPROPYLKETENE
1389	3904	131	59	8	17	46	229.89	3.5080	2-OCTANOL-2-D
1390	905	152	71	2	1	152	301.59	3.5086	IODOACETYLENE
1391	4647	92	71	7	8	91	300.78	3.5094	METHYLBENZENE (TOLUENE)
1392	471	122	55	7	6	105	380.01	3.5095	BENZOIC ACID
1393	289	104	59	8	8	104	288.21	3.5098	STYRENE
1394	5160	129	80	10	7	129	201.40	3.5099	1-DEUTERONAPHTHALENE
1395	4889	92	56	7	8	91	311.86	3.5108	METHYLBENZENE (TOLUENE)
1396	1354	184	58	8	9	105	212.07	3.5111	A-BROMO-P-XYLENE
1397	5139	198	85	14	30	57	292.57	3.5112	2,2,3,3,5,6,6-HEPTAMETHYLHEPTANE
1398	4593	56	36	4	8	41	360.55	3.5119	CIS-2-BUTENE
1399	6729	342	117	25	42	117	193.98	3.5120	(1-N-HEXADECYL INDAN)

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1400	6099	132	36	2	3	97	351.28	3.5121	1,1,1-TRICHLORCETHANE
1401	6094	88	45	4	8	43	304.88	3.5121	2-METHYLPROPANCIC ACID (ISOBUTYRIC ACID)
1402	4191	382	30	28	18	382	264.99	3.5127	1,1-BIS(3-PHENYL ISOINDOLYLIDENE)
1403	3599	198	57	3	0	18	290.69	3.5137	1,3-DICHLCROTETRAFLUOROACETONE
1404	80	70	49	4	6	39	318.60	3.5139	3,4-EPOXY-1-BUTENE
1405	5933	101	71	5	11	43	291.57	3.5141	N-METHYLMORPHOLINE
1406	3841	246	116	18	30	92	322.37	3.5152	1 PHENYL DODECANE (COLGATE) C18H30
1407	2712	246	116	18	30	92	322.37	3.5152	1-PHENYL DODECANE (COLGATE)
1408	284	102	74	6	14	43	237.81	3.5161	N-PROPYL ETHER
1409	94	72	41	4	8	43	361.92	3.5171	2,3-EPOXYBUTANE
1410	1969	84	40	5	8	55	443.28	3.5174	CYCLOPENTANONE
1411	640	134	79	9	10	119	390.78	3.5176	P-METHYLACETOPHENONE
1412	4596	54	29	4	6	54	393.48	3.5180	1,2-BUTADIENE
1413	4789	134	59	10	14	91	289.00	3.5181	ISOBUTYLBENZENE
1414	665	136	52	4	9	57	306.15	3.5185	1-BROMOBUTANE
1415	1225	174	53	9	18	57	305.63	3.5196	DI-N-BUTYL CARBONATE
1416	4613	128	65	9	20	57	380.18	3.5197	2,2,3,4-TETRAMETHYLPENTANE
1417	4784	120	54	9	12	105	235.17	3.5200	1-METHYL-3-ETHYLBENZENE
1418	410	116	59	4	4	54	409.60	3.5206	MALEIC ACID
1419	4745	84	64	6	12	41	320.77	3.5220	1,1,2-TRIMETHYLCYCLOPROPANE
1420	4594	56	34	4	8	41	361.69	3.5222	TRANS-2-BUTENE
1421	444	120	69	9	12	105	230.04	3.5226	ISOPROPYL-BENZENE
1422	1549	202	64	11	22	70	432.39	3.5234	DI-ISOAMYL CARBONATE
1423	3639	116	63	6	16	30	212.73	3.5235	HEXAMETHYLENE DIAMINE 658CE
1424	2458	0	63	0	0	30	212.73	3.5235	HEXAMETHYLENE
1425	3279	78	72	4	2	78	369.04	3.5238	FUMARONITRILE 283CE
1426	4995	84	63	6	12	41	349.38	3.5246	2,3-DIMETHYL-1-BUTENE
1427	5275	88	49	5	12	73	299.21	3.5258	METHYL TERT-BUTYL ETHER
1428	2240	87	70	4	9	44	374.54	3.5259	N,N-DIMETHYL ACETAMIDE
1429	2337	102	68	6	14	45	232.13	3.5259	2-HEXANOL
1430	2354	104	45	4	8	59	349.62	3.5263	2-HYDROXYISOBUTYRIC ACID
1431	3705	132	45	6	12	59	349.62	3.5263	PROPANOIC ACID,2-HYDROXY-2-METHYL- 723CE
1432	3773	102	68	6	14	45	232.15	3.5263	2-HEXANOL 795CE
1433	3758	102	43	6	14	59	275.99	3.5263	2-METHYL PENTANOL-2
1434	435	118	79	6	14	45	294.15	3.5263	2,5-HEXANE-DIOL
1435	3577	184	37	4	9	57	320.24	3.5267	T-BUTYL IODIDE 595CE
1436	329	108	52	7	8	108	339.32	3.5272	ANISOLE
1437	1334	182	36	2	2	83	344.43	3.5272	BIS (DICHLOROMETHYL) ETHER
1438	3854	134	94	9	10	43	293.59	3.5275	2-PROPANONE,PHENYL- 876CE
1439	1986	134	94	9	10	43	293.59	3.5275	PHENYL 2 PROPANONE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1520	3713	120	69	8	8	105	341.80	3.5602	ACETOPHENONE
1521	825	148	57	8	4	104	385.49	3.5605	PHTHALIC ANHYDRIDE
1522	495	124	64	7	8	109	340.32	3.5606	O-METHOXY-PHENOL
1523	1894	278	86	16	22	149	190.82	3.5611	DIBUTYL PHTHALATE
1524	1209	172	88	12	12	144	273.99	3.5612	2-ETHOXYNAPHTHALENE
1525	5260	156	98	4	6	29	233.39	3.5615	ETHYL DICHLOROACETATE
1526	5699	150	71	6	14	43	263.15	3.5620	2,5-DIMETHYL-3,4-DITHIAHEXANE
1527	1100	164	68	6	13	43	351.19	3.5620	3-BROMOHEXANE
1528	2159	68	43	3	4	68	380.97	3.5622	PYRAZOLE
1529	178	88	48	5	12	57	515.80	3.5626	3-METHYL-1-BUTANOL
1530	1083	164	46	2	0	166	474.64	3.5627	TETRACHLOROETHYLENE
1531	4601	112	71	8	16	57	303.73	3.5628	2,4,4-TRIMETHYL-1-PENTENE
1532	5653	202	139	15	22	131	200.78	3.5634	1-NOR-PENTYL-(1,2,3,4-TETRAHYDRONAPHTHALENE)
1533	3542	75	109	2	5	44	535.53	3.5641	CARBAMATE,METHYL 557CE
1534	310	106	76	8	10	91	297.09	3.5642	P-XYLENE
1535	1498	198	65	9	11	91	225.48	3.5649	(3-BROMOPROPYL) BENZENE
1536	4751	96	44	2	2	61	393.14	3.5653	1,CIS-2-DICHLOROETHENE
1537	6121	116	55	6	12	57	286.75	3.5656	METHYL-NOR-PENTANOATE
1538	1864	258	47	2	0	179	389.83	3.5661	1,2-DIBROMOPERFLUOROETHANE
1539	4795	118	26	6	14	44	630.40	3.5662	1,1-DIETHOXYETHANE (DIETHYLACETAL ACETAL)
1540	2323	101	66	6	15	72	377.95	3.5664	DI-N-PROPYL AMINE
1541	3051	101	66	6	15	72	377.95	3.5664	DI-N-PROPYLAMINE 23CE
1542	3605	124	45	6	8	54	272.97	3.5666	BETA,BETA(1)-OXY DIPROPIONITRILE
1543	1368	184	77	12	8	184	219.24	3.5667	DIBENZO-P-DIOXIN
1544	447	120	78	9	12	105	227.06	3.5670	P-ETHYL-TOLUENE
1545	3804	142	76	10	22	57	340.76	3.5672	4-N-PROPYLHEPTANE C10H22
1546	6264	142	75	10	22	57	340.76	3.5672	4-NOR-PROPYLHEPTANE
1547	3763	198	66	5	11	43	347.93	3.5675	N-AMYL-IODIDE 785CE
1548	1064	162	96	11	14	105	262.14	3.5677	N-VALEROPHENONE
1549	753	142	92	11	10	142	297.28	3.5678	2-METHYLNAPHTHALENE
1550	2500	159	72	8	17	86	280.73	3.5685	ISOLEUCINE ETHYL ESTER
1551	912	152	105	9	12	110	206.17	3.5690	P-PROPOXYPHENOL
1552	105	74	43	4	10	56	421.44	3.5693	N-BUTYL ALCOHOL
1553	1638	212	54	14	12	105	287.66	3.5696	BENZYL BENZOATE
1554	5172	260	79	4	10	29	233.37	3.5699	DIETHYLMERCURY
1555	3675	87	43	3	5	42	460.52	3.5701	3 METHYL-2-OXAZOLIDINONE
1556	550	130	53	7	14	43	316.33	3.5713	ISOAMYL-ACETATE
1557	3276	69	40	3	3	43	298.42	3.5714	PYRUVONITRILE 280CE
1558	3689	222	68	10	22	59	263.19	3.5719	TETRAETHYLENE GLYCOL DIMETHYL ETHER C10H22O5
1559	74	70	59	4	6	40	301.70	3.5721	3-BUTYN-1-OL

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1440	2050	130	26	7	14	43	315.03	3.5282	ISOPROPYL BUTYRATE
1441	3433	152	55	2	1	83	286.48	3.5283	1,1-DICHLORO-2,2,2-TRIFLUOROETHANE
1442	1714	220	75	15	24	91	207.37	3.5290	BENZYL OCTYL ETHER
1443	2171	72	31	4	8	42	387.16	3.5294	2-METHOXY PROPENE
1444	3772	72	31	4	8	42	387.16	3.5294	2-METHOXY PROPENE
1445	391	116	49	2	0	116	337.91	3.5297	1-CHLORO-1,2,2-TRIFLUOROETHYLENE
1446	1770	228	97	14	12	91	205.30	3.5303	BENZYL SALICYLATE
1447	3791	142	54	4	8	93	445.82	3.5306	ETHER,BIS-/2-CHLOROETHYL/- 812CE
1448	4815	128	57	9	20	43	361.64	3.5313	4,4-DIMETHYLHEPTANE
1449	5960	96	64	7	12	67	238.57	3.5321	3-ETHYLCYCLOPENTENE
1450	3883	142	33	8	14	43	305.29	3.5322	3-HEXENE-1-YL ACETATE
1451	3894	142	33	8	14	43	305.29	3.5322	3-HEXENE-1-YL ACETATE
1452	2058	142	33	8	14	43	305.29	3.5322	3 HEXENE 1 YL ACETATE
1453	5133	100	51	6	12	43	457.06	3.5328	3-HEXANONE
1454	3387	108	41	6	8	41	333.63	3.5339	ADIPONITRILE
1455	612	132	54	10	12	132	358.11	3.5352	ARYL-ALPHA-DIMETHYL-STYRENE
1456	5119	74	42	4	10	43	468.46	3.5353	2-METHYL-1-PROPANOL (ISOBUTYL ALCOHOL)
1457	439	118	61	6	14	57	332.68	3.5356	2-TERT-BUTOXY-ETHANOL
1458	1030	160	86	7	12	43	199.27	3.5358	1,3-PROPANEDIOL DIACETATE
1459	2189	75	109	2	5	31	470.93	3.5360	METHYL CARBAMATE
1460	4886	34	24	0	0	32	485.05	3.5361	HEXADEUTERODIBORANE
1461	1611	208	85	3	4	81	190.70	3.5364	1,1-DIFLUORO-2-BROMO-2-CHLOROETHYL METHYL ETHER
1462	1756	226	62	7	15	57	302.63	3.5368	2-IODOHEPTANE
1463	2817	178	68	8	18	59	276.46	3.5368	TRIETHYLENE GLYCOL DIMETHYL ETHER
1464	100	74	19	3	6	43	398.78	3.5370	GLYCIDOL
1465	3806	154	95	9	14	97	222.86	3.5374	2-N-PENTYLTHIOPHENE
1466	6516	154	94	9	14	97	222.86	3.5374	2-N-PENTYLTHICPHENE
1467	654	134	82	9	10	105	228.59	3.5382	ALPHA-PHENYLPROPIONALDEHYDE
1468	5188	342	227	25	42	117	196.92	3.5383	1-NOR-HEXADECYL-(2,3-DIHYDROINDENE) (1-NOR-HEXADECYLINDAN)
1469	5002	82	61	6	10	67	294.21	3.5384	3,3-DIMETHYL-1-BUTYNE
1470	3648	178	68	8	18	59	276.54	3.5395	TRIETHYLENE GLYCOL DIMETHYL ETHER 667CE
1471	5630	130	71	7	14	57	297.48	3.5401	ISOBUTYL-NOR-PROPANOATE
1472	4466	102	55	6	14	45	238.50	3.5403	2-HEXANOL
1473	2156	67	48	4	5	67	545.21	3.5410	PYRROLE
1474	3123	130	32	7	14	43	363.22	3.5411	PROPIONATE, T-BUTYL- 124CE
1475	3125	87	52	5	13	30	283.53	3.5416	NORM-AMYLAMINE 126CE
1476	2183	74	52	3	10	30	283.53	3.5416	TRIMETHYLENE DIAMINE
1477	4598	68	51	5	8	67	315.09	3.5416	CYCLOPENTENE
1478	165	88	46	4	8	43	309.34	3.5421	ISOBUTYRIC ACID
1479	6742	98	75	6	15	41	335.43	3.5426	TRIETHYLBORANE

POS	ID	MW	NPK	C	H	BASE	SUM HTS	ENTROPY(BASE2)	
1480	4296	318	24	21	22	69	309.79	3.5427	2,2-DIPHENYL-4,4-DIMETHYL-3-(2-CYANO-ISOPROPYL)-IMINOOXETANE
1481	3977	129	58	8	19	58	325.10	3.5427	3-OCTYLAMINE
1482	2061	142	42	8	14	43	267.60	3.5427	CIS 3 HEXENE 1 YL ACETATE
1483	4560	98	35	2	4	62	412.04	3.5441	1,2-DICHLOROETHANE
1484	2237	87	68	4	9	57	532.61	3.5443	MORPHOLINE
1485	561	130	52	10	10	130	399.62	3.5446	3-METHYL-INDENE
1486	4607	128	58	9	20	43	363.62	3.5453	2,3,3-TRIMETHYLHEXANE
1487	3507	94	29	2	0	94	442.29	3.5454	ACETYLENE,DICHLORO- 522CE
1488	2285	94	29	2	0	94	442.29	3.5454	DICHLOROACETYLENE
1489	4766	72	26	4	8	44	546.51	3.5463	NOR-BUTANAL (NOR-BUTYRALDEHYDE)
1490	4516	114	70	8	18	43	375.43	3.5465	2,3-DIMETHYLHEXANE
1491	308	106	70	8	10	91	295.33	3.5467	M-XYLENE
1492	2897	122	76	7	10	82	413.60	3.5468	2-METHYL ADIPONITRILE
1493	3624	70	45	5	10	55	336.43	3.5469	CIS-PENTENE-2
1494	2009	129	58	8	19	58	322.10	3.5469	3 OCTYLAMINE
1495	6082	80	44	6	8	79	328.85	3.5473	METHYLCYCLOPENTADIENE
1496	5309	60	43	4	8	44	317.31	3.5483	1,4-DIDEUTEROBUTANE
1497	3050	101	66	6	15	44	274.60	3.5485	DI-ISOPROPYLAMINE 22CE
1498	2324	101	66	6	15	44	274.60	3.5485	DI-ISO-PROPYL AMINE
1499	1735	224	27	6	0	155	417.15	3.5486	PERFLUORO-1,3,5-HEXATRIENE
1500	3694	120	39	5	12	45	276.07	3.5498	1,3-DIMETHOXY-2-PROPANOL
1501	5564	132	52	7	16	43	300.76	3.5499	DI-NOR-PROPOXYMETHANE (DI-NOR-PROPYL FORMAL, PROPYLAL)
1502	1626	210	44	6	11	83	315.62	3.5514	IODOCYCLOHEXANE
1503	817	148	60	7	16	59	312.78	3.5521	DIPROPYLENE GLYCOL METHYL ETHER
1504	772	144	111	10	8	144	220.25	3.5522	2-NAPHTHOL
1505	4244	98	69	4	6	27	350.62	3.5522	3,4-DIMETHYL-1,2,5-OXADIAZOLE
1506	733	140	91	8	9	112	229.72	3.5525	P-FLUOROPHENETOLE
1507	4905	84	64	6	12	56	348.93	3.5548	ISOPROPYLCYCLOPROPANE
1508	6403	105	47	5	7	46	249.75	3.5549	ISOPROPYL ACETATE -D3 (ISOPROPYL TRIDEUTEROACETATE)
1509	4493	100	66	7	16	57	459.04	3.5555	2,2,3-TRIMETHYLBUTANE
1510	4786	120	52	9	12	105	267.18	3.5555	1,2,3-TRIMETHYLBENZENE
1511	2912	150	58	6	14	89	271.75	3.5559	1,1 BIS(ETHYL MERCAPTO) ETHANE
1512	3439	184	32	2	2	186	510.59	3.5561	ETHYLENE,CIS,TRANS-DIBROMO- 452CE
1513	866	150	53	6	14	45	226.09	3.5566	TRIETHYLENE GLYCOL
1514	697	137	68	7	4	137	246.80	3.5571	P-CHLOROBENZONITRILE
1515	4194	327	35	23	21	145	245.99	3.5586	3-PHENYLIMINO-2,2-DIMETHYL-4,4-DIPHENYL OXETANE
1516	93	72	56	4	8	42	351.85	3.5587	TETRAHYDROFURANE
1517	1242	174	53	8	8	159	370.14	3.5590	1,2-DICHLORO-4-ETHYLBENZENE
1518	2021	100	70	5	8	55	305.94	3.5595	ETHYL ACRYLATE
1519	1981	120	69	8	8	105	341.80	3.5602	ACETOPHENONE

Programming Language - BASIC

```
10   DIM A[50],W[50],H[50],Q[50],B[50],D[50],F[50]
20   READ P
25   LET B=0
30   FOR V=1 TO P
40   READ A
50   LET B=B+A
60   NEXT V
70   PRINT "SUM="B
80   RESTORE
90   READ P
92   LET W=0
94   LET Q=0
98   LET F=0
100   FOR U=1 TO P
110   READ A
120   LET C=A/B
130   LET D=LOG(C)*.43429
140   LET E=C*D
150   LET F=-100*C*LOG(C)*.43429/1.07874
160   LET W=W+D
170   LET L=W/P
180   LET Q=Q+E
190   LET G=Q/P
200   PRINT C;D;E;F
205   NEXT U
206   PRINT "Q="Q
210   DATA 41
220   DATA 16,2,2,2,8,38,1,1,1,1,1,9,8,114,2,349,1,321,155,7,1,1,24
230   DATA 2,44,368,66,1000,33,1,2,15,62,60,20,77,39,220,502,22,1
250   DATA 8.00000E-02,.27
260   END

READY
```

Khinchin Programme

Programming Language - BASIC

```
1   DIM X[75],Y[75],A[75],B[75],C[75],U[75],V[75]
4   LET T3=0
5   LET S1=0
6   LET S2=0
20   READ N
30   FOR K=1 TO N
50   READ X,Y
60   LET S1=S1+X
70   LET S2=S2+Y
90   NEXT K
95   PRINT "SUMS ARE"S1;S2
100   READ N
110   RESTORE
115   PRINT " DIVERGENCES AND RUNNING DIVERGENCES"
120   READ N
130   FOR I=1 TO N
140   READ X[I],Y[I]
150   LET A[I]=X[I]/S1
160   LET B[I]=Y[I]/S2
170   LET C[I]=(A[I]+B[I])/2
180   LET U[I]=LOG(A[I]/C[I])
190   LET V[I]=LOG(B[I]/C[I])
200   LET J=S1*(A[I]-C[I])*(U[I])+S2*(B[I]-C[I])*(V[I])
205   LET Q=T3/N
215   LET T3=T3+J
216   PRINT X[I];Y[I];J;T3
220   NEXT I
225   PRINT "AVERAGE IS "Q
230   DATA 6
240   DATA 26,8,86,29,171,75,856,754,1000,1000,11,12
250   DATA 6
255   END

READY
```

Divergence Programme