Prediction of Mass Spectra from Structural Information

J. Gasteiger,* W. Hanebeck, and K.-P. Schulz

Organisch-Chemisches Institut, Technische Universität München, D-8046 Garching, Germany

Received January 29, 1992

A system has been developed that simulates mass spectra. The modular design allows gradual incorporation of knowledge on mass spectral reaction types. Statistical analyses of instances of such reaction types provide evaluations for the important processes in the mass spectrometer. Such evaluation procedures for the processes of ionization and of α-cleavages are presented. A second system has been developed that automatically acquires knowledge about mass spectral reactions directly from experimental mass spectra. It provides a detailed scheme of the individual steps of fragmentations and rearrangements of an organic molecule in the mass spectrometer. Illustrative examples for the interpretation of mass spectra are given.

INTRODUCTION

Mass spectroscopy is of major importance in the elucidation of the structure of organic compounds. However, the relationships between structure and mass spectral data are still only poorly understood. The complexity of the relationships between structure and mass spectrum and the vast amount of data to be processed have persuaded chemists and computer scientists quite early on to use computers. The DENDRAL project, initiated in 1965, was an application of computer science to the problem of structure elucidation using mass spectral data. It is still cited as *the* classical example of the application of artificial intelligence techniques to chemical problems.[1] However, it is sad to say that, in the end, the DENDRAL project failed in its major objective of automatic structure elucidation by mass spectral data, and research was discontinued. Therefore, investigations of the relationships between structure and mass spectra by computer techniques suffered severe setbacks.

The two major computer systems for exploiting mass spectral information in the structure elucidation process, STIRS[2] and MassLib,[3] only use identity and similarity searches of mass spectra to derive structural information. The SpecInfo system[4] employs mass spectral data to draw conclusions on substructural features in the unknown compound and uses these substructures in combination with information deduced from other spectroscopic techniques to assign the overall structure. The CHEMICS system[5] also places heavy emphasis on spectral data other than mass spectra, the latter being used only to derive the molecular weight or molecular formula of a compound.

When we embarked on a project to make more use of mass spectral information in structure elucidation, we strongly believed that major advances could only be made by a full modeling of the events occurring in the mass spectrometer. In other words, this requires an understanding and modeling of the fragmentation and rearrangement steps undergone by organic molecules in the mass spectrometer. This would make it possible to predict the mass spectrum of a given structure. Mass spectra simulation will be covered in the first part of this publication. The knowledge required for the prediction of the reactions of a molecule in the mass spectrometer will be extracted from a database of mass spectra. This is the theme of the second part of this work.

A TOOL FOR STRUCTURE ELUCIDATION

However interesting the simulation of mass spectra—from the structure to the spectrum—is from a theoretical point of view, for the practitioner the other direction—from the spectrum to the structure—is more important. Nevertheless, mass spectra simulation can also be a useful tool in the structure elucidation process. Quite often one already has some firm ideas on the potential structure of the compound investigated and can already offer various candidate structures. These structure suggestions may be derived from a knowledge of the reaction that led to the product investigated, or they may have been assigned on the basis of information from other spectroscopic methods. In such a situation, mass spectra are simulated for each of the candidate structures, and these simulated mass spectra are compared with the experimental one. The simulated mass spectrum that shows the highest similarity with the experimental spectrum allows one to infer the most likely structure for the compound investigated (Figure 1).

OVERVIEW OF THE APPROACH

The objective of this work was to develop a computer system that, given the structure of a compound, predicts its mass spectrum.[6] This system, MASSIMO (MAss Spectra SIMulatOr), requires knowledge about the fragmentations and rearrangements occurring in the mass spectrometer in order to derive the reaction steps for the molecule under study. These sequences of reactions and their evaluations will then be translated into a simulated mass spectrum. The knowledge of mass spectral reaction types required for this system was derived in an initial study for one process (α-cleavage) from a manually-compiled database. Subsequently, to speed up the knowledge-acquisition process, a system was developed that automatically generates databases on the various mass spectral reaction types. These databases will then be submitted to machine-learning techniques to automatically derive the knowledge required for mass spectra simulation. The system for knowledge acquisition, FRANZ (Fragmentation and Rearrangement ANalyZer), needs as input the structures and the mass spectra of a series of compounds. The interrelationships of the two systems are shown in Figure 2.

In our approach to simulating mass spectra we could draw from our rich experience in modeling chemical reactions. Over the last 15 years we have developed continually advanced versions of the EROS system (Elaboration of Reactions for Organic Synthesis) for the prediction of the course of organic reactions and the design of organic syntheses.[7] The latest version, EROS 6, was a particularly good starting point for the development of a system for mass spectra simulation.

0095-2338/92/1632-0264$03.00/0

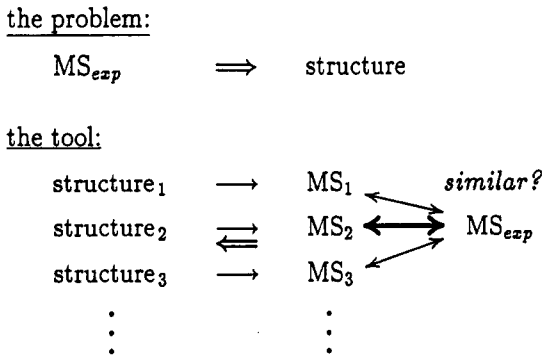


Figure 1. Mass spectra simulation as a tool for structure elucidation.

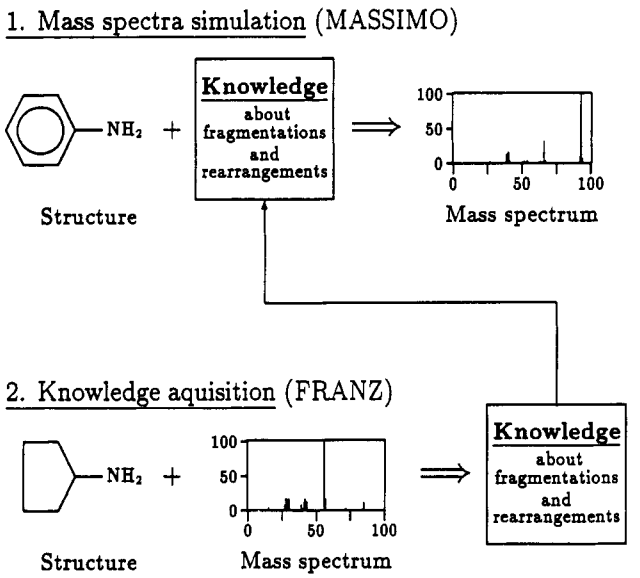


Figure 2. Mass spectra simulation: The objective and the knowledge acquisition through analyses of mass spectra.

EROS 6 distinctly separates the knowledge base on organic reactions from the inference mechanisms that predict the course of chemical reactions.[8] The knowledge base of EROS 6 is twofold: One is procedural in nature and contains a series of empirical methods for the rapid calculation of parameters quantifying electronic and energy effects.[9] The other knowledge base contains details on certain reaction types including functions that quantify chemical reactivity by using the parameters calculated by the empirical procedures.

In a series of studies, data on fundamental gas-phase reactions gathered by ion cyclotron resonance spectroscopy and high-pressure mass spectroscopy had been investigated. We were able to derive linear equations using inductive, polarizability, and hyperconjugation effect parameters calculated by the above empirical procedures. These equations reproduced data on proton affinity of amines, alcohols, and ethers, of thiols and thioethers, and of carbonyl compounds, as well as data on gas-phase acidity of alcohols.[10] This quantitative modeling of data on gas-phase reactions encouraged us to also address the reactions occurring as fragmentations and rearrangements in the mass spectrometer.

SYSTEM DESIGN AND IMPLEMENTATION

The simulation of mass spectra requires the handling of sequences of competing reactions. Thus, it has many similarities with the prediction of the course of "normal" organic reactions where one must make decisions between many alternative reactions and trace a pathway over several reaction steps. Therefore, the overall design of MASSIMO was quite similar to that of EROS 6 (Figure 3).

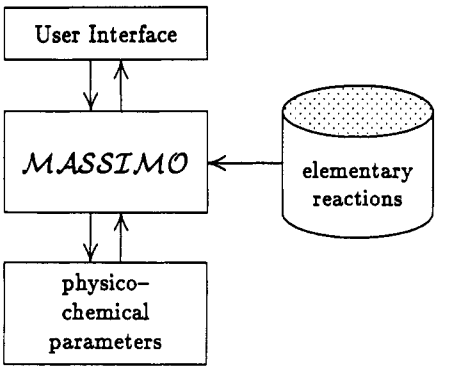


Figure 3. Basic design of the system MASSIMO for mass spectra simulation.

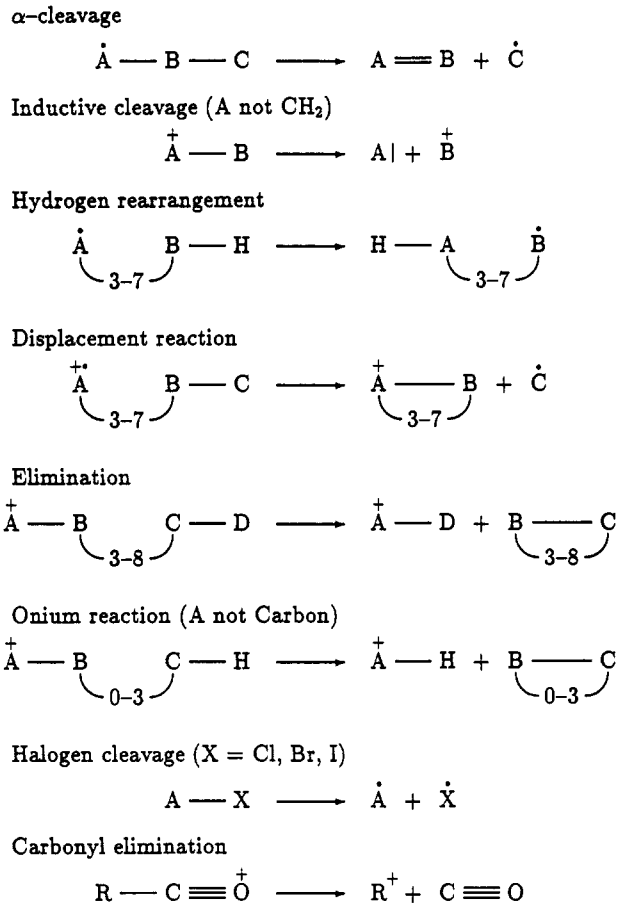


Figure 4. Fundamental reaction steps for the simulation of mass spectra.

The main system performs the basic management work and makes the decisions as to which fragmentations and rearrangements are to be performed. These decisions are based on knowledge about the elementary reaction steps observed in the mass spectrometer. This knowledge is kept in a rule file, separate from the main system. This allows an easy change of the knowledge by editing and also makes it easy to influence the performance of the system without having to do any reprogramming. The rule file on the mass spectral reaction types includes evaluation functions for these fundamental steps using physicochemical parameters on electronic and energy effects that are calculated by those empirical procedures mentioned above.[9] These procedures form the second type of knowledge base, knowledge on how to derive quantitative physicochemical parameters for the species encountered in the mass spectrometer. Clearly, this knowledge base can be extended by other methods including quantum mechanical procedures if desired.

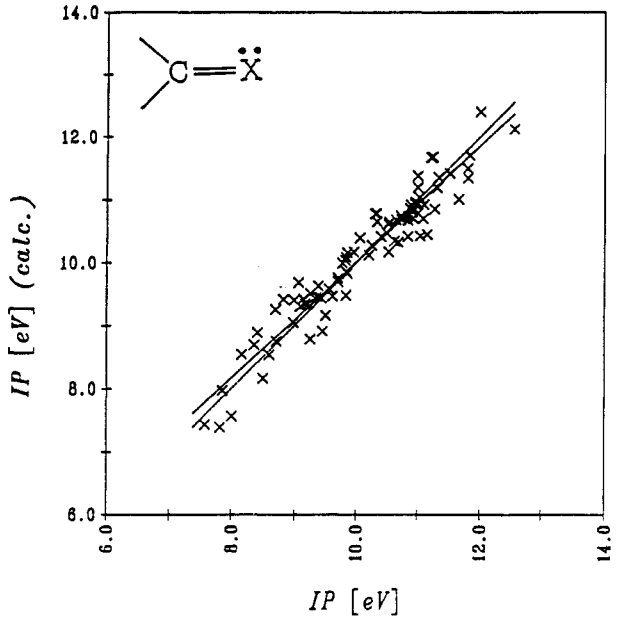


Figure 5. Correlation of experimental ionization potentials with those calculated from eq 1. Number of data points, 86; regression coefficient, r = 0.959; standard deviation = 0.31 eV.

The elementary reaction steps presently contained in MASSIMO are given in Figure 4.

MASSIMO is implemented in standard FORTRAN 77 and consists of about 170 routines with approximately 25 000 lines of code. The rule file consists of separate blocks of information for each elementary reaction written in ASCII-code. A special simple language, VERGIL (Versatile Eros Reactivity functions Generation and Interpretation Language)[11] has been developed that allows a flexible definition of restrictions and evaluations for the individual reaction types of Figure 4.

On program initiation, the rule file is compiled into a binary file that is incorporated into the MASSIMO system.

KNOWLEDGE ACQUISITION

Ionization Potentials. Presently, we are concentrating on simulating electron ionization mass spectra. The primary event in the mass spectrometer is then an ionization of the organic molecule to give a radical cation. Currently, we only allow ionization from a free electron pair and from π-orbitals. Primary ionization from σ-orbitals is not yet included in our system.

If there are several sites for ionization (n- or π-orbitals) in a molecule, we consider several primary molecular ions in our modeling of mass spectra. The ratio of these primary ions is estimated from the values of the ionization potentials at the various sites. Functions for the calculation of ionization potentials have been derived by statistical analyses, using the physicochemical parameters calculated by the empirical procedures.

In essence, the entire range of organic compounds has been divided into five classes:

compounds with lone pair orbitals
carbonyl compounds
π-systems without heteroatoms
π-systems with heteroatoms
aromatic compounds

For each class, a separate equation for the calculation of the ionization potential has been developed by multilinear regression analysis.[12] Figure 5 and eq 1 give the results for carbonyl compounds (aldehydes, ketones, carboxylic acids, esters, and their thio and nitrogen analogs).

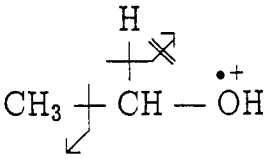


Figure 6. Classification of the potential sites for an α-cleavage in ethanol as observed (–CH_3•) or not observed (–H•).

The statistical analysis of the ionization potential, IP, of 86 carbonyl compounds led to eq 1.

$$IP_{calc} = 7.26 + 0.60\ \chi_{\sigma,min} - 0.16\ \overline{R^+} - 0.30\ \alpha_b - 57.4\ q_{\pi,max} \quad (1)$$

In this equation $\chi_{\sigma,min}$ is the minimum value of the σ-electronegativity in the carbonyl system, R^+ gives the amount of resonance stabilization for the positive charge on the carbon atom when breaking the CO double bond, α_b is the bond polarizability of the CO double bond, and $q_{\pi,max}$ is the maximum value of the π-charge in the carbonyl system.

α-Cleavage. For a given cation or radical cation there are usually several alternative pathways open for further fragmentation or rearrangement, i.e., several of the elementary reactions of Figure 4 will fit a given ion. It must therefore be determined to what extent each of the various reaction channels takes place. Such determinations are based on quantitative evaluations of the various alternatives for reacting. These evaluations are made with mathematical functions, similar to eq 1, derived from experimental observations. However, quantitative data on the various reaction types observed in the mass spectrometer are hard to come by (in the next section we will present an approach that was designed to provide such values). And yet, even in the absence of quantitative data, a mathematical function that somehow calculates how easily a given reaction type takes place should be obtained. How this can be achieved will now be explained.

In the absence of quantitative data on reactivity in the mass spectrometer, one can at least infer from the mass spectrum whether a certain reaction takes place or not in a specific molecular environment. We have analyzed a series of mass spectra in this way to find out where an α-cleavage occurs in the molecular ion and where it does not. The data consisted of monofunctional aliphatic alcohols, ethers, thiols, thioethers, amines, alkyl chlorides, bromides, and iodides. Electron ionization mass spectra, 70 eV, all taken from the literature were analyzed to find out which positions, from among all potential sites for an α-cleavage in the molecular ion, are in fact, submitted to an α-cleavage. Figure 6 shows such a classification in the case of ethanol.

In this way, we ended up with a data set of 144 sites for an α-cleavage with 71 classified as observed (reactive) and 73 as not observed (nonreactive). This data set was analyzed by logistic regression analysis (LoRA).

Logistic regression analysis takes an initial binary classification P (1 or 0, e.g., reactive or not; α-cleavage observed or not) and models it by the following set of equations (eqs 2 and 3)

$$P = \frac{1}{1 + e^{-f}} \quad (2)$$

with

$$f = c_0 + c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \quad (3)$$

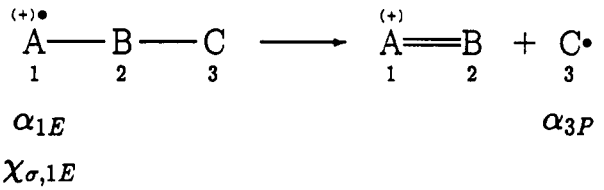


Figure 7. Basic scheme of an α-cleavage indicating the numbering of atoms and the parameters included in eq 4.

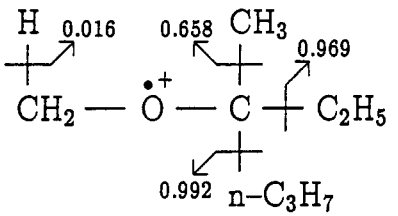


Figure 8. Probabilities for an α-cleavage calculated by eq 4.

x_1, x_2, etc. are variables that influence the classification of an event in one of the two categories. The coefficients c_0, c_1, c_2, etc. are changed in such a way as to make the calculated probability values P as close as possible to the given classification P.

Application of logistic regression analysis to the above data set of α-cleavages led to the following equation (eq 4):

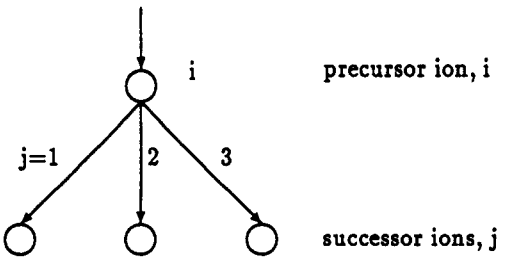
$$f = 74.99 - 0.71\alpha_{1E} - 6.47\chi_{\sigma,1E} + 2.23\alpha_{3P} \quad (4)$$

Figure 7 shows the scheme of an α-cleavage. It is provided to assist in the interpretation of the physicochemical background of eq 4.

The indices of the parameters in eq 4 refer to the numbering of atoms in the scheme of the reaction core of an α-cleavage; the subscripts E and P refer to the educt or product side of this reaction, respectively. To begin with, it should be noted that the higher the value of f, the more the reaction probability P approaches a value of 1, i.e., the more reactive the α-cleavage is. An increase in the polarizability and in the σ-electronegativity on atom 1 of the initial ion both tend to retain the unpaired electron and thus lower the value of f, equivalent to decreasing the reactivity. An increase in the polarizability of atom 3 in the product α_{3P} raises the ease of an α-cleavage as this effect significantly contributes to the stabilization of the radical C• and thus shifts the reaction to the product side. Thus, eq 4, obtained by a statistical method, has a firm physicochemical basis.

Application of eq 4 to the various sites of an ion formally prone to an α-cleavage allows one to assign reaction probabilities to all of these sites. Figure 8 shows the results for an ionized ether.

The values of the reaction probabilities, P, are used to calculate the intensities of peaks in the simulated mass spectrum. A pseudokinetic model based on first-order reaction equations is employed. The intensity of peaks is calculated in a top-down manner starting with the molecular ions. Preliminary estimates for the intensities of the molecular ion $I_{i,\text{old}}$ are calculated from the ionization potentials by a weighting procedure if several sites for ionization are present. Next, the various reaction probabilities, P_{ij}, of each reacting ion are added (Figure 9, eq 5). This allows the calculation of the intensities by eq 6 of all successor ions, j, stemming from this precursor ion, i. The final intensity of an ion, $I_{i,\text{new}}$, can now be calculated from the preliminary intensity, $I_{i,\text{old}}$, by subtracting the intensities of all ions that are derived from this precursor ion (Figure 9, eq 7).



$$N_i = \sum_{j=1}^{n} P_{ij} \quad (5)$$

$$I_j = \left[\frac{P_{ij}}{N_i} \times \left(1 - e^{-N_i t}\right)\right] \times I_{i,old} \quad (6)$$

$$I_{i,new} = I_{i,old} - \sum_j I_j \quad (7)$$

Figure 9. Calculation of signal intensities, I, in the mass spectrum from the reaction probabilities, P.

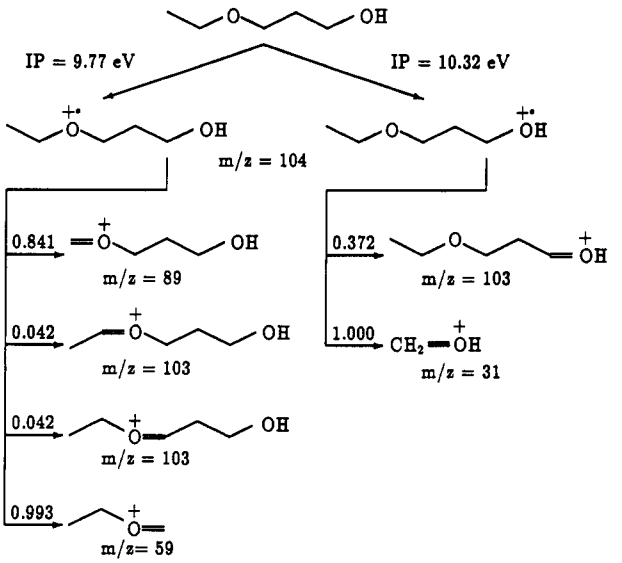


Figure 10. Fragmentation scheme obtained by application of functions for the calculation of ionization potentials and probabilities of α-cleavages (eq 4).

KNOWLEDGE APPLICATION

Mass spectra are not just the result of α-cleavages. We are presently deriving functions for all the other elementary reaction types of Figure 4.

However, even the present state of development already constitutes a valuable tool for peak assignment. The user has control over which elementary types to use in the simulation of a mass spectrum. Thus, the user can get details on the feasibility of certain reactions and can rapidly assign individual peaks in a mass spectrum.

Figure 10 shows the fragmentation scheme that is obtained for 1,3-propanediol monoethyl ether by applying the elementary reaction types of ionization and α-cleavage only.

Two different ionization potentials are calculated for the two sites with free electron pairs. The numbers on top of the

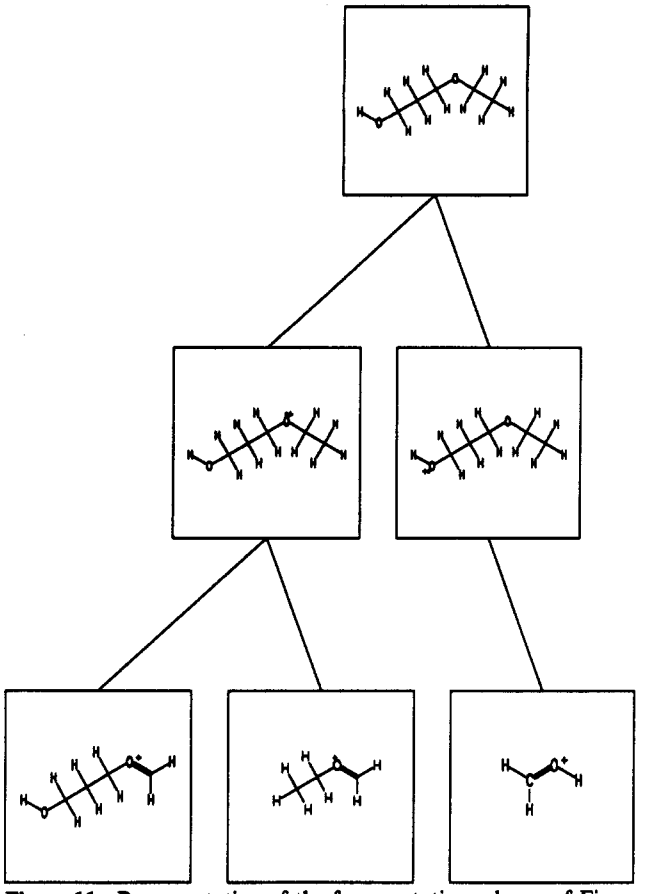


Figure 11. Representation of the fragmentation scheme of Figure 10 by the graphical user interface VEGAS.

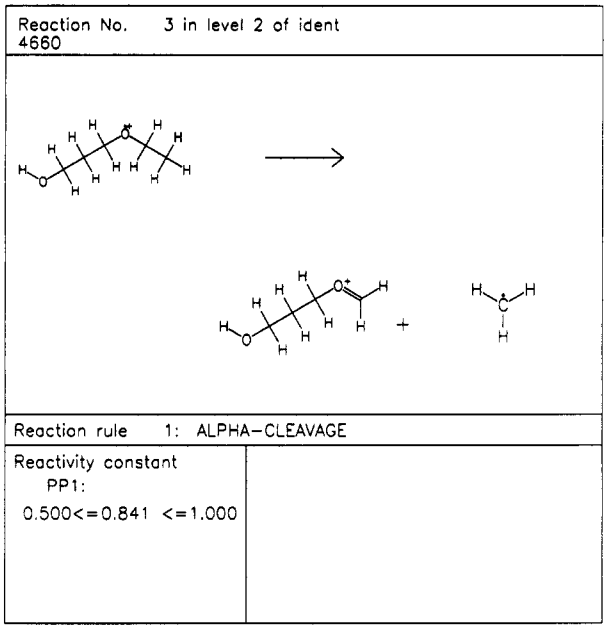


Figure 12. Graphical presentation of an individual fragmentation step by VEGAS.

arrows indicate the reaction probabilities calculated by eqs 2 and 4. Reactions with probabilities less than 0.5 are discarded.

The user is provided with a graphical interface, VEGAS, that gives him an overview of the fragmentation scheme (Figure 11) or lets him extract individual steps in the simulation of mass spectra (Figure 12).

Three α-cleavages are retained from the fragmentation scheme of Figure 10. The intensities calculated for the three ions obtained in these α-cleavages as calculated by eqs 5–7 are compared in Figure 13 with the experimental spectrum.

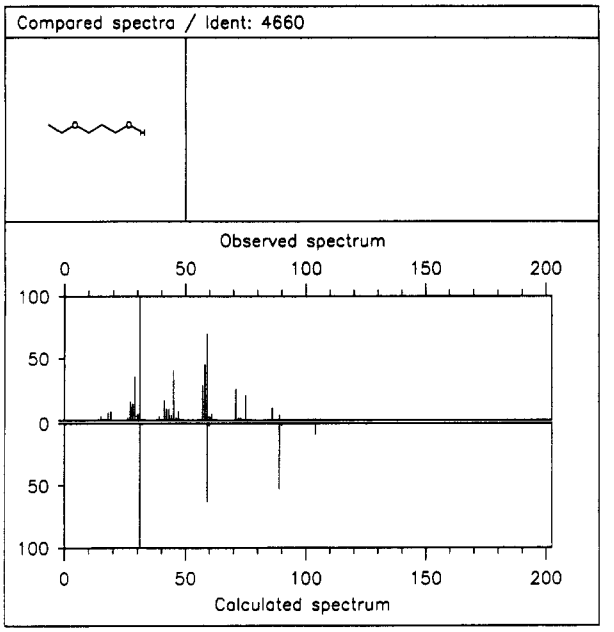


Figure 13. Comparison of the signals obtained from the fragmentation scheme of Figure 10 with the experimental spectrum.

(Obviously, "calculated spectrum" has to be understood with the restrictions imposed onto the system. In this case, they were the following: only show me the α-cleavages of the molecular ions!)

The intensities at m/z 31 and 59 are well reproduced. The intensity of the signal at m/z 89 is higher than in the experimental spectrum. Obviously, this ion rapidly fragments further, by processes other than α-cleavages.

AUTOMATIC KNOWLEDGE ACQUISITION

The construction of the database on α-cleavages was laborious and time-consuming. The experimental mass spectra had to be carefully analyzed by inspection to find out which α-cleavages occur and which ones are not observed. Nearly 2 months of work had to be invested to analyze the appropriate mass spectra, check the assignments with reports in the literature, and supplement the evaluations. In spite of all this work, in the end the database consisted of information gathered from only 70 mass spectra on simple monofunctional compounds.

Clearly, manual analysis cannot fully exploit the rich body of information contained in the huge amount of mass spectra already measured. The problem of making the information embodied in a vast collection of mass spectra available for acquiring knowledge about the reactions occurring in the mass spectrometer can only be solved by automatic methods.

We, therefore, embarked on a project to develop a system that can automatically extract the instances of the various elementary reaction types from a series of mass spectra. The outline of the Fragmentation and Rearrangement ANalyZer (FRANZ) is given in Figure 14.

The system takes the structure of a compound and generates the network of fragmentations and rearrangements for this molecule. Again this is achieved, as in MASSIMO, by elementary reaction types contained in an external file. However in contrast to MASSIMO, no evaluation of the reaction types is performed in FRANZ. In fact, as was already stated, the overall objective of FRANZ was to eventually gain just such evaluation procedures for each individual elementary reaction type. As applied, the elementary reaction types generate in FRANZ all conceivable fragmentation and

rearrangement steps. This network consists of all potential reactions that one can imagine for a structure to occur with a given set of elementary reactions. It is reduced to a network of observed fragmentation and rearrangement reactions by constant cross-checking with the peaks and intensities in the experimental mass spectrum. In addition to extracting only those reaction steps that actually occur in the mass spectrometer, FRANZ also assigns reaction probabilities to these individual steps. This is done in preparation for a statistical analysis of the reaction probabilities to obtain functions, or other forms of knowledge, for an evaluation of the reaction types.

The problem-solving techniques developed here allow another important and useful application of FRANZ: They provide an interpretation of a mass spectrum in order

1. to assign structures to the peaks of a mass spectrum
2. to extract the fragmentation and rearrangement reactions observed for a given structure
3. to assign reaction probabilities to those reaction steps.

The results of such an analysis of a mass spectrum are of considerable interest both to the expert in mass spectroscopy and to the organic chemist as a user.

The major steps in the analysis of a mass spectrum are

1. Generate the network of conceivable fragmentation and rearrangement steps for the given structure by application of the predefined elementary reaction types.
2. Assign structures to the groups of signals in the experimental mass spectrum by comparison with the ions contained in the fragmentation scheme.
3. Determine the amount of each ion.
4. Reconstruct a mass spectrum from the assigned signals.
5. Combine all structures connected through rearrangements to a group of isomers.
6. Calculate the amount of material that passes through each branch of the reaction network. This is done from the bottom (smallest ion) to the top (molecular ion), in a direction opposite to the fragmentation.
7. Reduce the network of reactions to those actually observed.
8. Calculate probabilities for the individual steps in the fragmentation scheme.
9. Assign confidence levels to the probability values.

As already mentioned, the FRANZ system has two major areas of application

1. Elucidation of the sequence of events in the fragmentation of an organic molecule in the mass spectrometer.
2. Construction of databases on the various types of elementary mass spectral processes for knowledge acquisition.

Some examples will be given to illustrate the performance of the system with different types of compounds. They show the merits of the system for the interpretation of a mass spectrum. The investigation of large series of mass spectra for the automatic acquisition of knowledge on mass spectral reaction types is still an ongoing study. No results on larger data sets will therefore be presented here, but the direction of research can already be inferred from the few examples given.

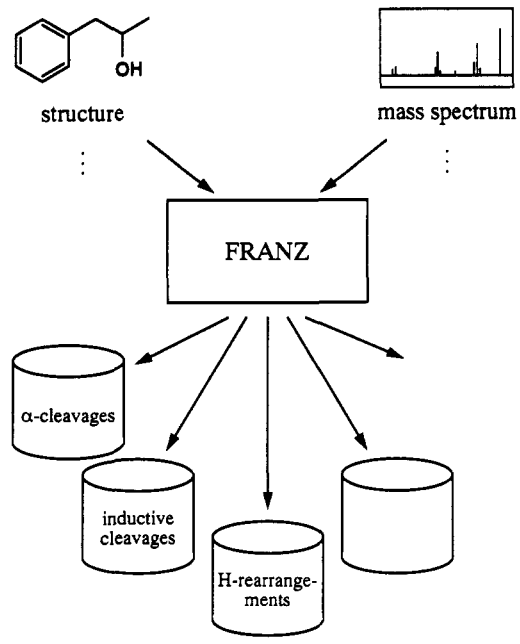


Figure 14. Basic design of the system FRANZ for extracting the reaction steps from a mass spectrum.

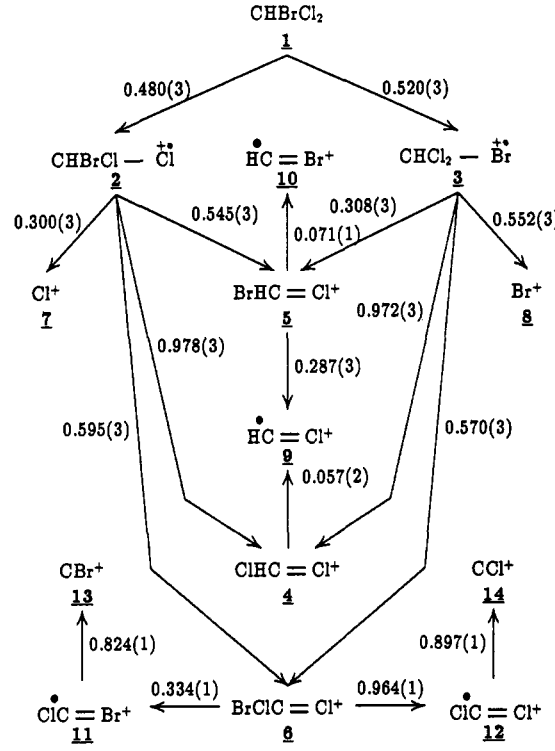


Figure 15. Fragmentation scheme obtained for bromodichloromethane. The numbers give the calculated values of reaction probabilities; the values in parentheses indicate the level of confidence: 1 = very high; 2 = high; 3 = low.

All mass spectra are 70-eV electron ionization spectra taken from the NIH Database. The reaction networks were obtained by application of the elementary processes contained in Figure 4.

Figure 15 shows the fragmentation scheme obtained from an analysis of the mass spectrum of bromodichloromethane. For reasons of clarity of presentation, the neutral molecules and radicals eliminated at the various steps have not been included in the figure.

The ionization at the bromine atom is preferred—particularly if one takes account of the statistical factor—because of the lower ionization potential at the bromine atom as compared to that of the chlorine atom. The

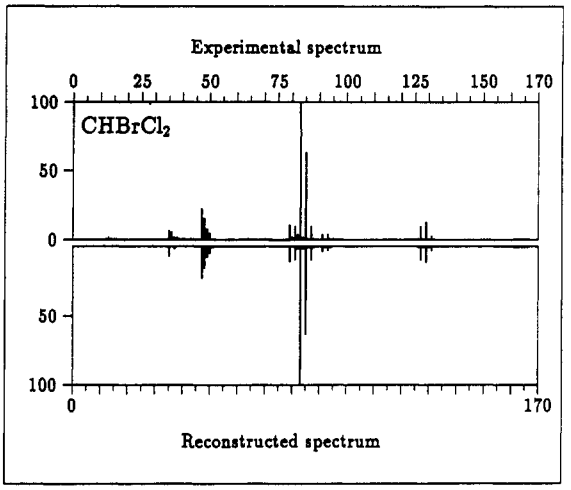


Figure 16. Experimental mass spectrum of bromodichloromethane and mass spectrum reconstructed from the fragmentation scheme of Figure 15. The explanation factor is 93.6%.

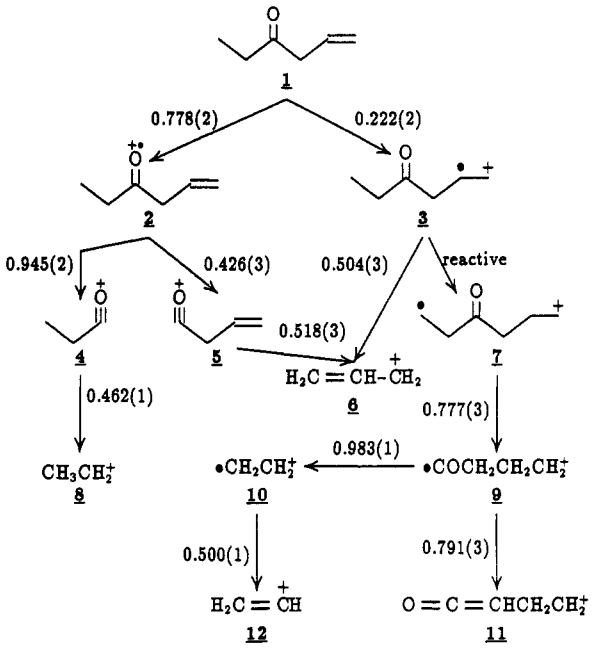


Figure 17. Fragmentation scheme derived for 5-hexen-3-one. The numbers give the calculated reaction probabilities; the values in parentheses indicate the level of confidence: 1 = very high; 2 = high; 3 = low.

quantitative values of the reaction probabilities reflect the more intuitive, qualitative feeling of a mass spectroscopy expert. The mass spectrum reconstructed from the fragmentation scheme in Figure 15 is compared with the experimental mass spectrum of bromodichloromethane in Figure 16. The reconstructed mass spectrum can explain 93.6% of the intensity in the experimental mass spectrum.

In Figure 17 the fragmentation scheme extracted from the mass spectrum of 5-hexen-3-one is given. Again, for clarity of presentation, the neutral molecules and radicals eliminated at the various fragmentation steps have not been included in the scheme.

Both sites of ionization, the free electron pair of the carbonyl group and the isolated CC double bond, are considered with the oxygen of the carbonyl group being the preferred site. The probabilities of the reaction steps leading to the allyl cation, **6**, have a rather low level of confidence because two alternative pathways lead to the allyl cation. Cations **5** and **11** are isomeric, giving the probabilities of the reactions leading to them a low level of confidence. The radical ions

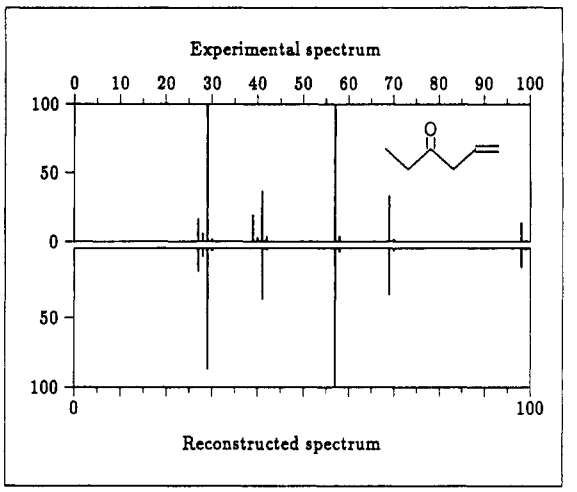


Figure 18. Comparison of the experimental and the reconstructed mass spectrum of 5-hexen-3-one. The explanation factor is 88.7%.

3 and **7** are isomeric to each other and, thus, give rise to the same signal in the spectrum. Thus, there is not enough evidence in the mass spectrum to assign a reaction probability to this step. However, the rearrangement of **3**–**7** is required to occur to explain the formation of the ions **9**–**12** and the associated peaks in the mass spectrum.

The experimental mass spectrum of 5-hexen-3-one is compared in Figure 18 with the one reconstructed from the fragmentation and rearrangement scheme of Figure 17. Most (88.7%) of the intensity in the experimental mass spectrum of 5-hexen-3-one can be explained by the fragmentation scheme of Figure 17. The largest discrepancy shows up with the peak at m/z 39. This peak cannot be interpreted by the above reaction network. The high intensity of this peak points to a high stability of the ion associated with this signal; it probably has the structure of the aromatic cyclopropenyl cation. The elementary reactions contained in Figure 4 do not provide a reaction path that leads from 5-hexen-3-one to cyclopropenyl cation. We are presently investigating which additional elementary processes have to be included in our analysis in order to generate such a pathway.

As a last example, the experimental mass spectrum of 3-chlorocyclopent-2-enone is compared in Figure 19 with the reconstructed mass spectrum that was obtained with the elementary reaction types of Figure 4. The corresponding fragmentation and rearrangement scheme cannot be given here because of its complexity. It contains a network of 895 reaction steps and 501 species that corresponds to 68 ions with different composition. Each ion, each node in the reaction network, and each step and its associated reaction probability have been determined explicitly by FRANZ.

This example stresses the point that the reaction events in the mass spectrometer are indeed quite complex. Nevertheless, the chemical and mathematical models embodied in our approach are able to make suggestions on the individual steps that might occur in the mass spectrometer. Clearly, such a detailed analysis is far beyond any attempt to manually extract and write down the fragmentation and rearrangement scheme.

In this sense, FRANZ provides opportunities for the extraction of knowledge from mass spectra beyond that which is practical for human experts to achieve alone. We strongly believe that this system will revive interest and research in explicitly expressing the events in the mass spectrometer. It gives credence to expressing the structure of the ions by an explicit valence bond structure and to expressing the reactions in the mass spectrometer by explicit shifts of bonds and

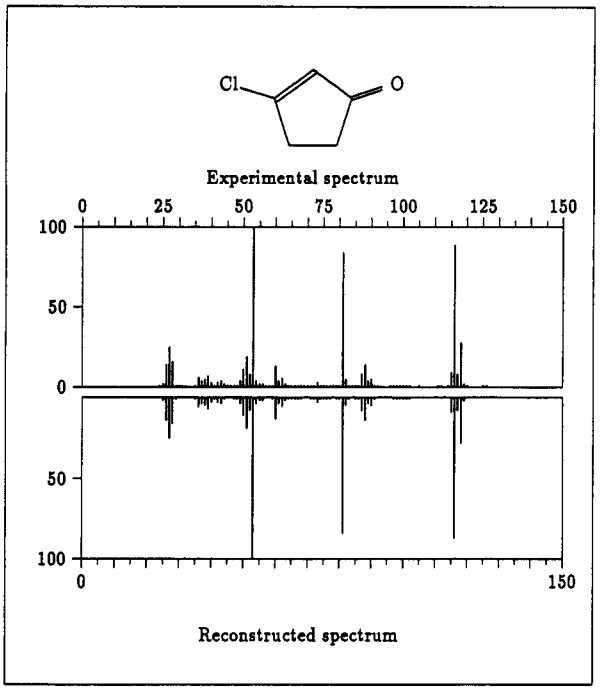


Figure 19. Experimental and reconstructed mass spectrum of 3-chlorocyclopent-3-enone. The explanation factor is 98.0%.

electrons interconverting such structures.

CONCLUSIONS

The door has been opened to the automatic extraction of knowledge about the events occurring in the mass spectrometer directly from the spectra. The fragmentation and reaction schemes derived from the mass spectra provide deep insight into the details of reactions in the mass spectrometer.

Collecting the instances of the different elementary reaction types from a series of mass spectra in separate databases gives information that can be processed to knowledge by statistical analyses. The knowledge acquired in this process can be used for the simulation of mass spectra and, thus, help in enhancing the power of mass spectroscopy as a tool for structure elucidation.

ACKNOWLEDGMENT

We gratefully acknowledge support of this work by Bundesminister für Forschung und Technologie. The work benefitted from helpful discussions with Dr. W. Bremser (ROW), Prof. Dr. H. Kubinyi and Dr. R. Neudert (both BASF), Prof. Dr. K. Varmuza (Technische Universität Wien), and Dr. M. Weller (Chemical Concepts). Stimulating discussions and continuing support by Dr. V. Schubert (GMD-PTF) is particularly acknowledged.

REFERENCES AND NOTES

(1) Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. *Applications of Artificial Intelligence for Organic Chemistry—The Dendral Project*; McGraw-Hill: New York, 1980.

(2) McLafferty, F. W.; Stauffer, D. B. Retrieval and Interpretative Computer Programs for Mass Spectrometry. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 245–252.

(3) Domokos, L.; Henneberg, D.; Weimann, B. Optimization of Search Algorithms for a Mass Spectra Library. *Anal. Chim. Acta* **1983**, *150*, 37.

(4) Neudert, R.; Bremser, W.; Wagner, H. Multidimensional Computer Evaluation of Mass Spectra. *Org. Mass Spectrom.* **1987**, *22*, 321–329.

(5) Funatsu, K.; Miyabayashi, N.; Sasaki, S. Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18–28.

(6) Hanebeck, W.; Rafeiner, K.; Schulz, K.-P.; Röse, P.; Gasteiger, J. *Software-Development in Chemistry 4*; Springer-Verlag: Heidelberg, 1990; pp 187–195.

(7) Gasteiger, J.; Hutchings, M. G.; Christoph, B.; Gann, L.; Hiller, C.; Löw, P.; Marsili, M.; Saller, H.; Yuki, K. A New Treatment of Chemical Reactivity: Development of EROS, an Expert System for Reaction Prediction and Synthesis Design. *Top. Curr. Chem.* **1987**, *137*, 19–73.

(8) Röse, P.; Gasteiger, J. Automated Derivation of Reaction Rules for the EROS 6.0 System for Reaction Prediction. *Anal. Chim. Acta* **1990**, *235*, 163–168.

(9) (a) Gasteiger, J. Automatic Estimation of Heats of Atomization and Heats of Reaction. *Tetrahedron* **1979**, *35*, 1419–1426. (b) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228. (c) Gasteiger, J.; Hutchings, M. G. Quantification of Effective Polarisability. Applications to Studies of X-ray Photoelectron Spectroscopy and Alkylamine Protonation. *J. Chem. Soc. Perkins 2* **1984**, 559–564.

(10) (a) Hutchings, M. G.; Gasteiger, J. Residual Electronegativity—An Empirical Quantification of Polar Influences and Its Application to the Proton Affinity of Amines. *Tetrahedron Lett.* **1983**, *24*, 2541–2544. (b) Gasteiger, J.; Hutchings, M. G. Quantitative Models of Gas-Phase Proton Transfer Reactions Involving Alcohols, Ethers, and Their Thio Analogs. Correlation Analyses Based on Residual Electronegativity and Effective Polarizability. *J. Am. Chem. Soc.* **1984**, *106*, 6489–6495. (c) Hutchings, M. G.; Gasteiger, J. A Quantitative Description of Fundamental Polar Reaction Types. Proton and Hydride Transfer Reactions Connecting Alcohols and Carbonyl Compounds in the Gas Phase. *J. Chem. Soc. Perkins 2* **1986**, 447–454.

(11) Simon, V. VERGIL (Versatile Eros Reactivity functions Generation and Interpretation Language), Language Description, Version 1.0, TU München, Dec 1991.

(12) Hanebeck, W.; Gasteiger, J. Rapid Calculation of Ionization Potentials of Organic Molecules. *J. Comput. Chem.*, submitted.