SpecTool: A Hypermedia Toolkit for Structure Elucidation

Marc Cadisch, Margit Farkas, Jean-Thomas Clerc, and Ernö Pretsch*

Department of Organic Chemistry, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland, and Department of Pharmacy, University of Bern, CH-3012 Bern, Switzerland

Received February 10, 1992

A collection of reference data and spectra and computational tools needed for the interpretation of spectra has been compiled and organized within the frame of HyperCard.

INTRODUCTION

The topic of this paper is computer-aided spectrum interpretation. Often a full spectrum estimation is required, i.e., a structure or, in most cases, a constitution (connectivity) is entered and the output is a predicted spectrum.

Besides such a full spectrum estimation in his daily work, the chemist often needs a less comprehensive estimation of just parts of the spectrum. Here the goal is not to obtain estimated spectra for whole structures but to investigate one feature after another of the experimental spectrum in order to derive structural elements or to check consistency with corresponding parts of a proposed structure. This is usually done by examining data for structurally similar compounds and using the knowledge about possible biases, exceptions, etc.[1,2]

The various spectroscopic techniques provide partly supplementary and partly overlapping pieces of information. Therefore a multimethodical view allows the most powerful way of spectrum interpretation.[1-3]

Trivial consistency checks, logical combinations, and other straightforward calculations are also routinely needed during spectrum interpretation, in addition to spectrum prediction. Examples for such tasks are the computation of all possible molecular formulas for a given molecular mass and element range and other constraints,[4] the calculation of isotope distributions[5] and the determination of masses of all possible fragments obtained by direct bond cleavages[6] in mass spectrometry, the calculation of the expected number of signals in the ^{13}C NMR spectrum for a compound,[7] the calculation of a 1D[8] or 2D NMR spectrum[9,10] given the spectral parameters, or the systematic assembly of molecular fragments to entire molecules.[11] All these types of computations have in common that no heuristic rules are coded into the corresponding programs. Thus, they do not make decisions about accepting or discarding structures. Making decisions remains the task of spectroscopists based on the facts derived by such programs and on their expert knowledge.

In this contribution a new medium is presented which supports chemists in making their estimations and decisions based on reference data and on facts obtained by using the included computational tools. The system supports a multimethod view by integrating mass spectrometry, ^1H and ^{13}C NMR spectroscopy, and infrared and UV/visible spectroscopy. A set of carefully selected reference data and spectra with versatile user-friendly features to browse through them is provided together with computational tools in a single hypermedia environment.

HYPERMEDIA

Hypermedia[12-14] is an information management technology that links text, graphics, and sound in an associative fashion. It allows users to navigate through the pieces of information in a nonlinear way. This is contrary to a book where the pages are ordered due to the nature of the medium in just one (sequential) way. A hypermedia system containing only text is called hypertext.

In hypermedia, the information is cut into chunks of variable size. Each chunk is connected to at least another related piece of information by a link. The user can access all nodes following these links. Most nodes have many more than one link attached (only one link per node would be a linear sequence of the data), so the user has the choice of which way to follow. He can browse through the system in a way appropriate for his needs.

SpecTool[15] uses the hypermedia software HyperCard, which is provided free by Apple Computer with new Macintosh computers. HyperCard shows windows that look like note cards as metaphors for a node. The cards are collected in files, called stacks in HyperCard. A card in HyperCard contains, besides graphics, a variable number of objects:

buttons: the mouse clicked within the area of a button triggers an action

fields: places where texts are displayed

Programs written in the built-in programming language HyperTalk,[16] which has some object-oriented features, may be attached to each object.

Hypermedia offers a vehicle for fast, intuitive, nonlinear access to large quantities of data. Recently, numerous hypermedia applications have been developed due to the availability of suitable hardware and software, and the technology gained a great deal of attention.

NODES

Both the contents of the nodes and the types of available links are essential to characterize a hypermedia application. Although they intimately belong together, a clearer understanding is possible by presenting them in two consecutive sections.

Contrary to a sheet of paper, a node is dynamic: a part of the objects can be seen by the user while another part might be hidden, depending on the situation. Hidden information might be displayed upon a user action. The information stored on a card can fill the screen many times.

The nodes which build SpecTool can be assigned to one of four groups according to their contents. These will be briefly described in the next sections:

Data Nodes. Data nodes mainly contain assigned spectroscopic parameters for selected reference compounds, generalized parameters for a group of compounds (e.g., ranges of parameters for a compound class), or other heuristic rules. The contents of data nodes are thus very much like those of

0095-2338/92/1632-0286$03.00/0

printed volumes containing selected reference data.[3] There are, however, two important differences. First, one can cover a much larger amount of data and still have a handy "volume" in a hypermedia application. Contrary to books, the access time to a particular page increases only very slightly with the number of pages. Secondly, in contrast to a printed volume, a hypermedia application allows various kinds of presentations of the same data according to the context or to the need of the user. The user may, for example, choose between graphical or numerical representation where this makes sense. He may change the display of tables. For example, the table entries can be reordered according to sorted values of a user selectable column. Furthermore, various tools have been developed to locate information in large tables. Parts of available information (like references or measuring conditions) are only displayed over the other data if the user needs them. Thus, a concentrated presentation of the relevant information is possible without a loss of further data that are only relevant in special cases. A more detailed description of the data presentation tools will be given elsewhere.

Spectra Nodes. Spectral data provide a concentrated presentation of interpreted features of spectra. During spectrum interpretation, however, often a look at the original spectral curves (with or without assignments) is of great help since the visual comparison of curves can be much more informative than the inspection of numerical data. For infrared spectroscopy a selected set of spectra are, for example, available as a part of an excellent textbook.[17] A number of printed spectra catalogs have been published for the various spectroscopic techniques (for references see ref 3). Finally computerized databases are also available for all relevant spectroscopic methods (see, for example ref 18). In spite of the availability of a huge amount of spectra, the access to typical spectra of a selected compound class is a time-consuming (and often expensive) task today.

In SpecTool a set of selected spectra, characteristic for the various compound classes, is being implemented. Browsing through spectra of structurally related compounds is directly supported. Furthermore, it is always possible to switch between data and spectra for the selected compound type.

Tool Nodes. Programs for calculations which are typically necessary during spectrum interpretation are integral parts of SpecTool. For nondemanding calculations, like the estimation of ^{1}H NMR chemical shifts by using additivity rules[3] or the calculation of coupling constants using a modified Karplus equation,[19] the programs are written in HyperTalk. They are attached to the objects of a node. The node itself serves thus as a user interface. More demanding calculations use so-called external commands (XCMDs). These are routines written in common programming languages such as FORTRAN, Basic, Pascal, or C. They are called in the same way as a HyperTalk command. Again, the user interface is a Hypercard node. An example is an XCMD for the calculation of all possible elemental compositions of a fragment of a given mass and element range.[20] Finally, control can be passed to an external program by calling it from a SpecTool node. After performing the calculations, the user gets back to the same node. Virtually any program can thus be attached to SpecTool. Programs implemented so far include MOLFORM,[4] C13SHIFT,[21] an isotope pattern calculation program (according to ref 5), and a program determining all possible fragments of a molecule obtained by one or two bond cleavages with or without H rearrangements (ref 22, designed after ref 6). More programs are currently being developed.

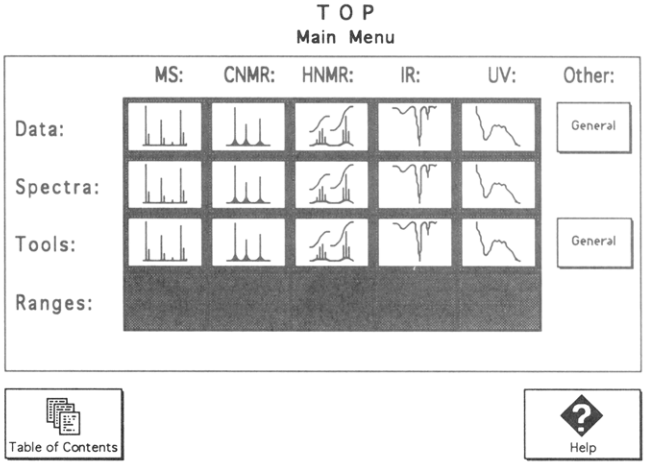

Figure 1. TOP-card is the central navigational card and corresponds to the main menu of SpecTool.

Navigation Nodes. A considerable part of the over 1000 nodes implemented so far is made purely for navigation purposes. They will be described in the next section.

User Nodes. Besides the nodes designed by the authors, the user can make new ones. They will be logically attached to the node at which they were created. The user node can be of any of the four types discussed. Creating user nodes requires some knowledge of HyperCard. Only authoring knowledge (i.e., the ability to make new card objects and to enter information into them) is needed for making data or spectra nodes. At least some HyperTalk experience is required for making tool and navigation nodes.

LINKS

The real power of a hypermedia application originates from the logical structures. The nodes can be looked at as pages of a book (even if some of these "pages" are computational nodes), whereas the navigation structures define the various orders in which the pages can be accessed, i.e., the kinds of possible browsing. Several navigation structures designed over the same set of nodes result in something like a set of books, each one containing the same pages but in various orders.

For the present purpose of spectrum interpretation, the following orders are typically needed and have therefore been implemented into the system:

- The possibility to browse within a particular spectroscopic method through all available compound classes
- The possibility to browse through all related spectroscopic data (MS, ^{13}C NMR, IR, etc.) of a compound or compound class
- The possibility to browse between spectra and data for a given spectroscopic method and compound class.

Further parts of the navigation system correspond to various tables of contents and various structured index systems. Thus, the links also define the possible entries into the system. A printed volume has one linearly ordered table of contents and usually one index with alphabetically ordered entries. Hierarchical sets of tables of contents are implemented in SpecTool which allow a very efficient and fast access to the data and tools. Each node is a table of content of the nodes on the next lower hierarchical level. Every node typically contains 20–30 items so that some 10 000 nodes can be accessed within only three steps. At each level, only a manageable number of items has to be investigated during this procedure. SpecTool contains *several sets* of such hierarchical tables of contents for the same data and program nodes.

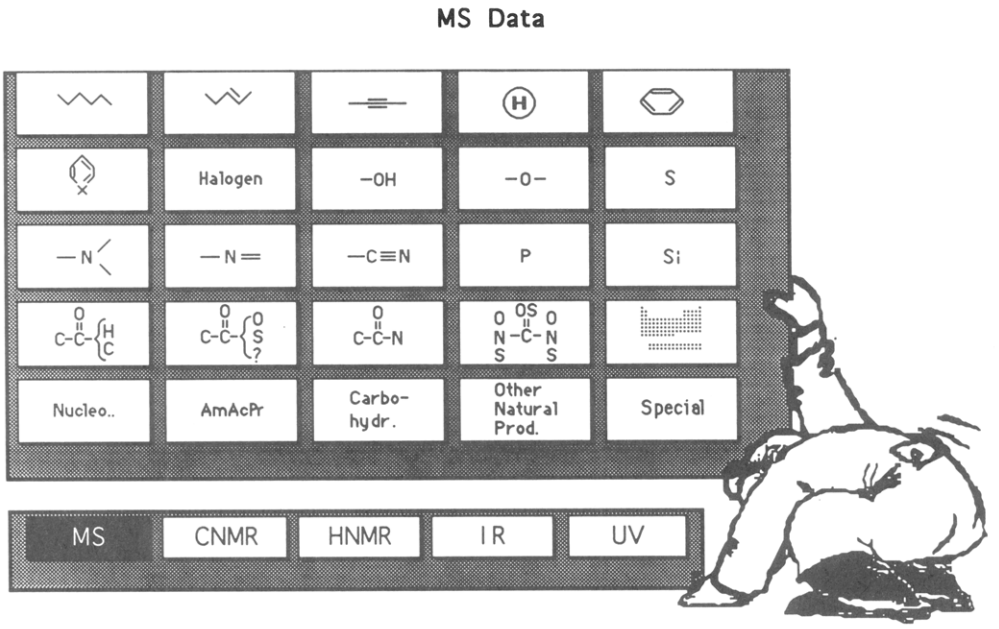

Figure 2. Navigational card offering 24 classes of compounds and a node for other data section entries.

Further entry points are spectral parameters (e.g., an m/z value in MS or a chemical shift in NMR). A list of structural fragments or types is displayed after the selection of a value. Clicking with the mouse at a line of the list performs a jump to the corresponding reference data.

Finally, the user can place bookmarks anywhere in the system and assign a name to them. Bookmarks can be used by a pull-down menu for direct jumps. Any number of bookmark collections can be saved.

These sets of navigation networks allow accessing the nodes in a way which is highly efficient and highly redundant at the same time. The high efficiency means that any node can be accessed within no more than four steps from any other node. The high redundancy means that any node can be accessed by many different ways, depending on the problem to be solved.

All the links discussed have been implemented in various ways, including objects of cards, items of pull-down or pop-up menus, and palettes. Several of these possibilities have been realized for important links. Thus, not only the way of accessing a node is redundant but also the tool used for changing nodes.

FILE STRUCTURE

The file structure of the system has been designed to meet two demands. First, it has to be transparent for development and updating purposes, and secondly, it has to allow straightforward implementation of various logical structures. At present the system consists of 32 stacks and four external applications with their files. These can be grouped into six categories:

1. Control and organization stacks
2. Navigation stacks
3. Stacks containing spectroscopic information
4. Stacks containing information created by the user
5. Help system
6. Applications

At present SpecTool consists of 1050 nodes which use a disk space of 6.2 MB. The data nodes contain an enlarged set of data available as a printed volume.[3]

EXAMPLES

The real flavor of a hypermedia application cannot be presented in a printed medium. Some examples were collected

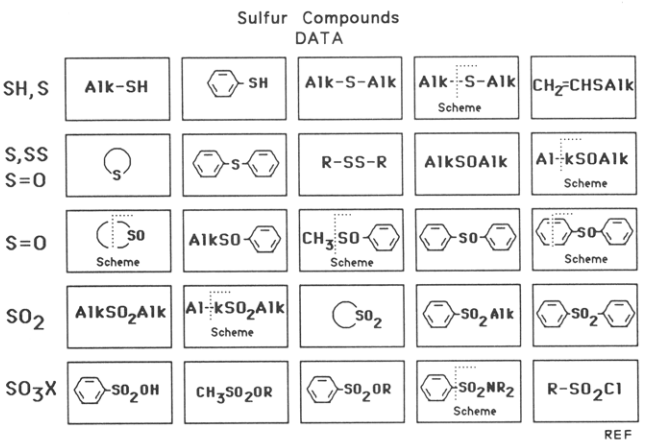

Figure 3. Navigational card pointing to available data cards for functional groups containing sulfur.

in order to highlight a typical usage of the system to give at least some ideas.

In Figures 1–4 four nodes are shown which a user would typically browse through if mass spectrometric reference data for aliphatic thioethers are required. The first node (Figure 1) is the so-called TOP-card, which is the center of the system. This node is displayed after starting SpecTool. The TOP-card can be reached from anywhere in the system with a single jump. The 20 buttons shown correspond to jumps to other navigation nodes.

By clicking at the "MS-Data" button, the system selected the card shown in Figure 2. This card could also be reached in one jump from anywhere in the system. The 25 buttons on the upper part of this card correspond to 24 compound classes and one more node for entries into the data section by other points of view than compound types. For all five spectroscopic methods the ordering at this level is the same for both data and spectra, i.e., the same types of 25 nodes can be addressed from these nodes. Thus, the "Sulfur"-node for MS-data (Figure 3) is one of 10 sulfur nodes (spectra and data nodes for the five methods). Direct jumps are possible within the five data nodes and the spectra nodes as well as between the data and spectra nodes for a selected spectroscopic method.

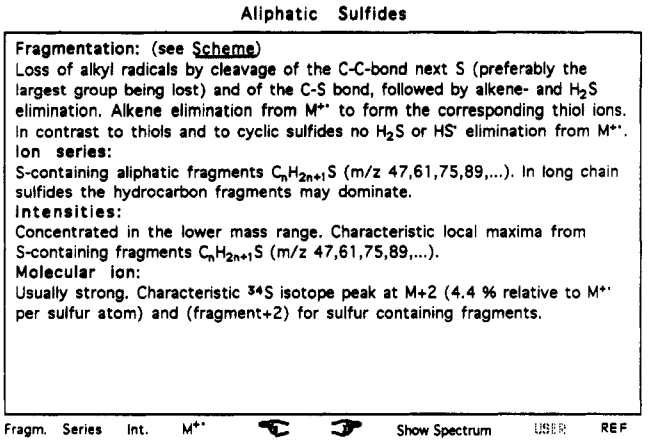

Figure 4. Data card showing heuristic rules for aliphatic sulfides in MS.

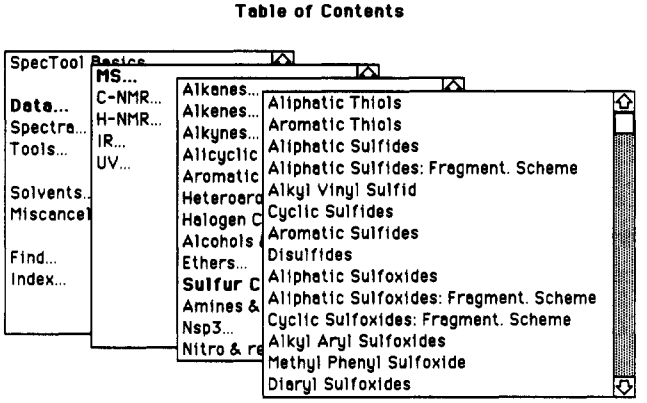

Figure 5. Navigational card offering access by tables of contents.

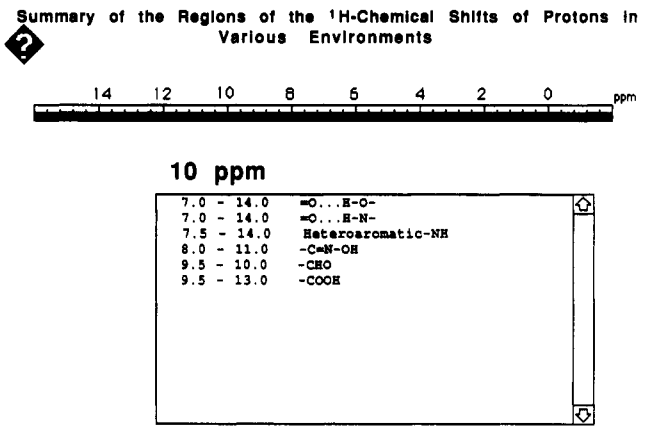

Figure 6. Card containing ^{1}H NMR shift ranges. The field in the lower part of the card contains groups corresponding to the chemical shift, selected on the top part. The lines of this field can be used as pointers to the corresponding data cards.

Figure 3 shows which MS-data nodes are available for sulfur compounds. Selection of aliphatic thioethers leads to Figure 4. In this figure, fragmentation rules are collected together with a typical spectrum. Upon clicking at "Show spectrum", the spectrum will be displayed on the top part above the rules field, whose size is then reduced to one-third of the card.

Direct jumps are possible to another MS-data card showing fragmentation schemes for thioethers. All other MS compound classes (as selectable from the node on Figure 2) are also directly accessible. Altogether, direct jumps are possible to over 50 different places (not counting possible bookmarks). This number is quite typical for any node in the system and indicates how dense the navigation network is.

An alternative way to reach the MS-data card of aliphatic thioethers is by selecting the node's name in the hierarchical structured index system (Figure 5).

An example for entries into the system based on spectroscopic data instead of partial structures is shown in Figure 6 for ^{1}H NMR spectroscopy. Here the user enters an observed chemical shift by clicking at the appropriate position on the scale. Partial structures corresponding to the shift value are listed in the table at the lower part of the card. By clicking at any line, a direct jump is invoked to a card which lists available references for the corresponding compound class.

DISCUSSION

The hypermedia application SpecTool is a new kind of medium in analytical chemistry. It is neither a database, an expert system, nor a collection of programs, but it is somehow related to all three categories.

In contrast to expert systems, SpecTool does not make any decision based on the data. There is no rule-generation software included. Its navigation structure exhibits, however, an expert knowledge in the sense that it proposes various browsing routes typically used by experts.

In a database the order of the items is not relevant so that usually nobody would browse through it. Browsing through records found by a search is usually limited to a linear order in case of databases. In contrast to databases, SpecTool exhibits a dense network of connections which allows browsing according to various points of views.

SpecTool comprises several programs, besides data, graphics, and sounds (as browsing aids), as an integral part, in contrast to many hypermedia applications. Another distinct property of SpecTool is the possibility to extend it. The user can add more data, spectra, and programs that are automatically integrated into the existing navigation structure.

The collection of spectra into a database opened up new dimensions. Possibilities, like, for example, the search for compounds with similar spectra, were not even imaginable before databases existed. The same holds for the collection of interpreted spectroscopic data, interpreted spectra, and computational tools into a hypermedia environment. It adds new possibilities, neither available in printed media nor in spectroscopic databases. Spectroscopic databases are almost replacing printed spectra collections. Similarly, hypermedia applications like SpecTool could successively replace printed collections of interpreted spectral data.

ACKNOWLEDGMENT

This work was partly supported by the Swiss National Science Foundation.

REFERENCES AND NOTES

(1) Silverstein, R. M.; Bassler, G. C.; Morrill, T. C. *Spectrometric Identification of Organic Compounds*, 5th ed.; Wiley: New York, 1991.
(2) Clerc, J. T.; Pretsch, E.; Seibl, J. *Structural Analysis of Organic Compounds by Combined Application of Spectroscopic Methods*; Elsevier: Amsterdam, 1981.
(3) Pretsch, E.; Clerc, J. T.; Seibl, J.; Simon, W. *Tables of Spectral Data for Structure Elucidation of Organic Compounds*, 2nd ed.; Springer-Verlag: Berlin, 1989.
(4) Fürst, A.; Clerc, J. T.; Pretsch, E. A. Computer Program for the Computation of the Molecular Formula. *Chem. Intell. Lab. Sys.* **1989**, *5*, 329–334.
(5) Kubinyi, H. Calculation of isotope distributions in mass spectrometry. A trivial solution for a non-trivial problem. *Anal. Chim. Acta* **1991**, *247*, 107–119.
(6) Schubert, V.; Bremser, W.; Neudert, N.; Kubinyi, H.; Gasteiger, J.; Varmuza, K. Das Expertensystem Specinfo. *Nachr. Chem. Tech. Lab.* **1989**, *37*, 721–728.

(7) Shelley, C. A.; Munk, M. E. Signal Number Prediction in Carbon-13 Nuclear Magnetic Resonance Spectrometry. *Anal. Chem.* **1978**, *50*, 1522–1527.
(8) Diehl, P.; Kellerhals, H.; Lustig, E. Computer Assistance in the Analysis of High-Resolution NMR Spectra. In *NMR Basic Principles and Progress*; Diehl, P., Fluck, E., Kosfeld, R., Eds., Springer-Verlag: Berlin, 1972; Vol. 6.
(9) Widmer, H.; Wütrich, K. Simulation of Two-Dimensional NMR Experiments Using Numerical Density Matrix Calculations. *J. Magn. Reson.* **1986**, *70*, 270–279.
(10) Studer, W. SMART, a General Purpose Pulse Experiment Simulation Program Using Numerical Density Matrix Calculations. *J. Magn. Reson.* **1988**, *77*, 424–438.
(11) Gray, N. A. B. *Computer-Assisted Structure Elucidation*; Wiley: New York, 1986.
(12) Parsaye, K.; Chingnell, M.; Khoshafian, S.; Wong, H. *Intelligent Databases: Object Oriented, Deductive Hypermedia Technologies*; John Wiley & Sons: New York, 1989.
(13) Woodhead, N. *Hypertext & Hypermedia: Theory and Applications*; Addison Wesley: Reading, MA, 1991.
(14) Bielawski, L.; Lewand, R. *Intelligent Systems Design: Expert Systems, Hypermedia and Database Technologies*: John Wiley & Sons: New York, 1991.
(15) Farkas, M.; Cadisch, M.; Pretsch E. Hypermedia Tools for Structure Elucidation Based on Spectroscopic Methods. In *Scientific Computing and Automation (Europe) 1990*; Karjalainnen, E. J., Ed.; Elsevier: Amsterdam, 1990.
(16) Winkler, D.; Kamins, S. *Hypertalk 2.0: The book*; Bantam Books: New York, 1990.
(17) Colthup, N. B.; Daly, L. H.; Wiberly, S. E. *Introduction to Infrared and Raman Spectroscopy*, 2nd ed.; Academic Press: New York, 1991.
(18) Zupan, J., Ed.; *Computer-supported spectroscopic databases*. Ellis Horwood: Chichester, 1986.
(19) Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. The relationship between proton–proton NMR coupling constants and substituent electronegativities. *Tetrahedron* **1980**, *36*, 2783–2792.
(20) Gloor, A. Diploma Thesis, ETH-Zurich, 1992.
(21) Fürst, A.; Pretsch, E. A Computer Program for the Prediction of ^{13}C-NMR Chemical Shifts of Organic Compounds. *Anal. Chim. Acta* **1990**, *229*, 17–25.
(22) Brodmeier, T. Diploma Thesis, ETH-Zurich, 1991.