

Von Wolfgang Bremser\*

Professor Helmut Dörfel zum 60. Geburtstag gewidmet

„Der Unterschied zwischen den älteren und den neueren Computersystemen läuft letzten Endes auf den Unterschied zwischen einem Insekt und einem Menschen hinaus. Das Insekt kommt „vollständig programmiert“ auf die Welt – programmiert durch Instinkte, denen es gedankenlos folgt. Der Mensch dagegen muß das richtige Verhalten erst erlernen – doch dieses Lernen hat verselbständigende Wirkung: Der Mensch kann nämlich aus eigenem Entschluß und eigener Erkenntnis seine bisherigen Handlungsprogramme verändern. Die Vorbereitung eines Computersystems der x-ten Generation wird bereits weit mehr der Erziehung eines Kindes als dem klassischen Programmieren einer Rechenmaschine ähneln.“

Stanisław Lem<sup>[1]</sup>

## 1. Strukturaufklärung in der Organischen Chemie

Die wachsende Bedeutung der Strukturaufklärung in der Organischen Chemie ist offenkundig. Neben der klassischen Analyse der Konstitution ist in den vergangenen Jahren immer stärker die Untersuchung von Konfiguration und Konformation in den Vordergrund getreten, gipfelnd in der genauen Abbildung ganzer Proteinstrukturen mit Hilfe der Röntgenstrukturanalyse oder der NMR-Spektroskopie. Auch natürliche und bei Synthesen anfallende Gemische werden routinemäßig charakterisiert, sei es nach vorgeschalteter Trennung, durch zeitaufgelöste Beobachtung chemischer Reaktionen oder durch Sequenzanalyse bei Hochpolymeren. Schließlich kann die Messung der Isotopenverhältnisse helfen, die Entstehungsgeschichte einer Substanz im Reaktionsgefäß oder im biologischen System relativ genau zurückzuverfolgen.

Die Leistungsfähigkeit der modernen Analysengeräte hat durch Elektronik und Rechnersteuerung einen derart hohen Stand erreicht, daß weitgehend vollautomatisch eine Flut von genauen und hochspezifischen Daten erzeugt wird, die auf eine detaillierte Auswertung warten. So werden in unserem Spektroskopielabor täglich ca. 40 MByte Daten an die Auswertung und Dokumentation übergeben und harren dort der Bearbeitung, eine gewaltige Aufgabe, die nur mit Hilfe der elektronischen Datenverarbeitung gelöst werden kann.

Aber auch der Bedarf an Detailinformationen über unsere Substanzen steigt von Tag zu Tag. Die Auflagen durch den Gesetzgeber, der Umweltschutz und der angestrebte hohe Qualitätsstandard, der im harten industriellen Wettbewerb die einzige Überlebenschance bietet, fordern täglich die Analytik von neuem. Außerdem ist eine Verbesserung der Stoffeigenschaften nur bei genauer Stoffkenntnis möglich, die gezielte Synthese eines Wirkstoffs nur bei Kenntnis des Wirkmechanismus. Unnötiger Aufwand bei der Synthese und dem Test der falschen, weil unwirksamen Verbindungen können oft durch bessere Analysendaten vermieden werden.

Nur durch den konsequenten Einsatz moderner Computertechnik bei der Aufnahme und Interpretation spektroskopischer Daten können diese stetig wachsenden Anforderungen erfüllt werden. Dieser Beitrag kann und will nicht die Gesamtheit des Problems behandeln. Vielmehr sollen neben einem kurzen Exkurs in die Leistungen und Probleme der Automation eines modernen Spektroskopielabors vor allem die Werkzeuge vorgestellt werden, die eine „intelligente“ Auswertung der Rohdaten ermöglichen. Dabei sollte man im Auge behalten, daß der Auftraggeber für eine spektroskopische Analyse meist nicht am Spektrum an sich, sondern an der daraus ableitbaren Strukturinformation interessiert ist. Gelingt es, die spektroskopischen Parameter direkt in Topologien und die zugehörigen Strukturbilder (Topographien) umzuwandeln, ist ein weiterer entscheidender Schritt zur Verbesserung der Leistungsfähigkeit der analytischen Methoden getan.

## 2. Automatisierte Spektrenaufnahme

Komponenten zur vollautomatischen Steuerung der Aufnahme von NMR- und IR- sowie Massenspektren gehören heute zur Grundausstattung der meisten käuflichen Spektrometer. Mit ihnen werden die Experimente definiert, die Meßparameter optimiert, die gewünschte Zahl von Läufen durchgeführt, die Daten aufsummiert, aufbereitet und in standardisierter Form auf dem Bildschirm abgebildet oder über den Drucker ausgegeben. Der ganze Ablauf wird in flexibler Dialogführung vom Benutzer vorbereitet und vom Computer abgearbeitet. Auch Serien von Experimenten, wie 2D-NMR-Aufnahmen, Messungen zur Bestimmung von Reaktionskinetiken<sup>[2]</sup>, Tochterscans in der Massenspektroskopie etc., können auf diese Weise unbeaufsichtigt erledigt werden (Abb. 1).

Als Beispiel für eine neuere Entwicklung sei die automatische Umschaltung zwischen chemischer Ionisation (CI) und Elektronenstoßanregung (EI) bei der Aufnahme von Massenspektren erwähnt. Das CI-Experiment ist notwendig, um die Molekülmasse und damit die Bruttoformel genau bestimmen zu können. Vor allem für eine vollautomatische Interpretation ist diese Information eine wichtige Voraussetzung (siehe Abschnitt 10). EI-Spektren dagegen liefern die charakteristische Fragmentierung des Moleküls

[\*] Dr. W. Bremser  
Molekülspektroskopie, ZHV/S  
Hauptlaboratorium der BASF Aktiengesellschaft  
D-6700 Ludwigshafen



Abb. 1. Blick in ein modernes Spektroskopielabor mit einem Hochfeld-NMR-Spektrometer einschließlich Probenwechselsystem im Vordergrund und einem Robotersystem zur automatischen Vorbereitung der Substanzproben im Hintergrund. Für die Zwischenkontrolle wird bewußt der Mensch eingeschaltet, der die Lösung inspiziert und in den Wechsler einsortiert.

und somit den Fingerabdruck der Struktur, der für die Bibliothekssuche benötigt wird. Mit der ACE-Technik (Alternate Chemical Ionisation/Electron Impact) kann zwischen beiden Ionisierungsarten während der Messung umgeschaltet werden. Während beim Standardverfahren eine mittelgroße Blende als Kompromiß zwischen dem kleinen CI- und dem großen EI-Spalt eingesetzt wird und daher Mischspektren erhalten werden, die für Identitätsvergleiche nicht geeignet sind, wird bei der in Zusammenarbeit mit einem Gerätehersteller entwickelten Modifikation<sup>[3]</sup> pneumatisch zwischen den optimalen Blenden umgeschaltet. Abbildung 2 zeigt, daß auf diese Art EI-Spektren ohne CI-Anteile zugänglich sind und daß auch während einer GC/MS-Kopplung oder einer fraktionierenden Verdampfung „gleichzeitig“ CI- und EI-Spektren gemessen werden können.

## 2.1. Probenwechselsautomaten

Für eine vollautomatische Messung unterschiedlicher Verbindungen ist ein automatischer Wechsel der vorbereiteten Substanzproben notwendig. Wechselsautomaten für Probenröhrchen (Abb. 1) in der NMR- und für KBr-Preßlinge in der IR-Spektroskopie gehören heute bei den meisten kommerziellen Systemen zum Lieferprogramm, obwohl die Aufgabe wegen der zahlreichen notwendigen Sicherheitskontrollen eigentlich recht kompliziert ist. Werden der IR-Spektroskopie oder der Massenspektrometrie Trenntechniken vorgeschaltet, bietet sich ein Probengeber (Autosampler) an, der mit seinem Spül-Trocken-Cyclus auch zum Füllen von IR-Küvetten verwendet werden kann<sup>[3]</sup>. In der Massenspektrometrie ist die Entwicklung automatischer Festkörperschleusen (vgl. AUDEVAP<sup>[4]</sup>) in vollem Gang. Um flüssige Proben auf KBr-Fenster aufzutragen und diese dann in das Spektrometer einzusetzen, verwendet man Roboter (siehe Abb. 3). Da diese Wechselsautomaten Tag und Nacht und am Wochenende eingesetzt werden können, ermöglichen sie einen großen Durchsatz an Proben und eine effiziente Nutzung der teuren Spektrometer. So können pro Gerät und pro Tag NMR-spektroskopisch über 200 Proben und massenspektrome-

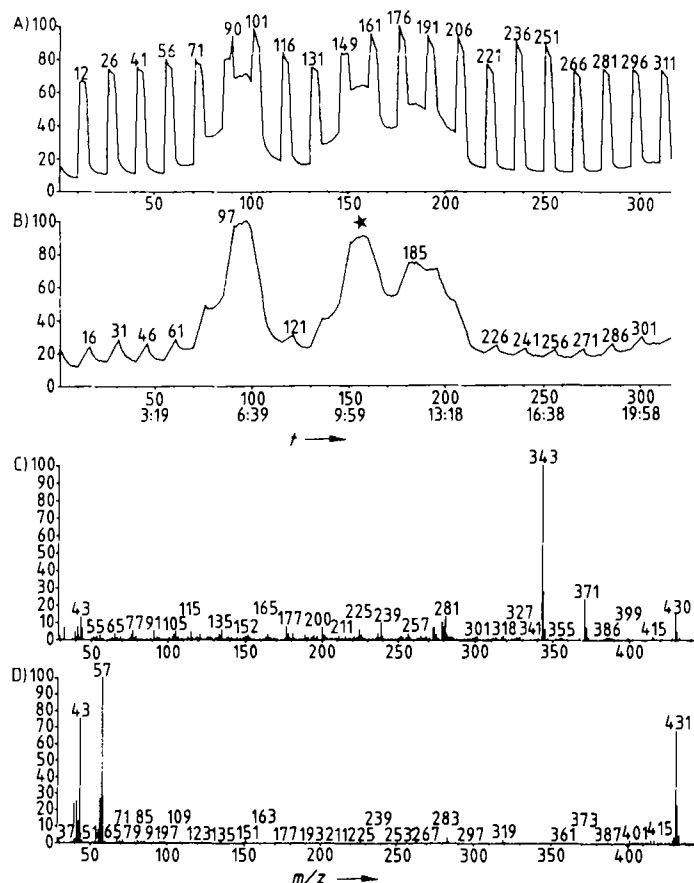


Abb. 2. Direktverdampfung eines festen Probengemisches unter Totalionenstrom-Kontrolle mit automatischer Umschaltung zwischen EI und CI. A, B) Rekonstruierter Totalionenstrom als Funktion der Zeit. Die Maxima in konstanten Abständen resultieren aus den Ionen des Reaktantengases während der CI-Einstellung. Den Verdampfungsverlauf gibt B) am besten wieder, das aus dem Ionenstrom unter EI-Bedingungen abgeleitet ist. C) EI-, D) CI-Spektren der Komponente, die zum zweiten Maximum des Verdampfungsprofils gehört (\*). Die Molekülmasse 430 aus dem EI-Spektrum wird durch das Signal für  $(M+H)^+$  im CI-Spektrum bestätigt. Das EI-Spektrum enthält keine wesentlichen Verunreinigungen durch das vorangegangene CI-Experiment.

trisch über 24 GC/MS-Kopplungen mit durchschnittlich zehn Komponenten bearbeitet werden.

Der Informationsausstoß dieser Systeme kann nur bewältigt werden, wenn auch zur Dokumentation und Interpretation der Meßergebnisse Rechner herangezogen werden.

## 2.2. Automatische Probenvorbereitung

Durch die in Abschnitt 2.1 geschilderten Möglichkeiten zur automatischen Abarbeitung der Sequenzen von teils komplizierten Experimenten wurde das Laborpersonal entlastet, der Probendurchsatz erhöht und die Standardqualität verbessert. Als wichtige manuelle Arbeit bleibt die Vorbereitung der Substanzproben vor allem für die IR- und NMR-Spektroskopie. Hier bietet sich ein Robotersystem an, das dank der Flexibilität seiner Bewegungen und Operationen schnell auf die bei substanzspezifischen Problemen notwendigen Entscheidungen reagieren kann.

Abbildung 1 zeigt im Hintergrund ein Robotersystem zur Präparation von Lösungen für die NMR-Spektroskopie (für Details siehe [3]). Das System führt für feste wie

flüssige Substanzen alle Schritte vom Öffnen des Probenfläschchens bis zum Verschließen des Meßröhrchens aus. Bis zu hundert Substanzen können in einem Lauf meßfertig gemacht werden, was dem typischen Tagespensum eines Mitarbeiters entspricht.

In der IR-Spektroskopie wird neben einem Automaten zur Präparation von KBr-Preßlingen<sup>[3, 5]</sup> auch ein Roboter zum Auftragen von Filmen auf KBr-Fenster eingesetzt (Abb. 3), der folgendermaßen vorgeht:

- Eine KBr-Scheibe wird aus dem Spender geholt und bereitgelegt.
- Das Probengefäß wird aufgeschraubt, und mit der Spritzenhand wird die Probe auf die KBr-Scheibe pipettiert. Eine zweite KBr-Scheibe wird aus dem Spender geholt und auf die erste gedrückt, anschließend werden beide KBr-Scheiben mit dem Probenfilm dazwischen in eine Halterung gesetzt und fest zusammengepreßt.
- Die Halterung wird in den Meßstrahl des Spektrometers gefahren und das Startsignal für die Messung ans Spektrometer geschickt. In der Zwischenzeit wird das Probengefäß wieder verschraubt und an seinen Platz zurückgestellt.
- Nach Beendigung der Messung wird die Halterung nach oben gefahren und die gemessene Probe abgelegt. - Nun kann von neuem begonnen werden.

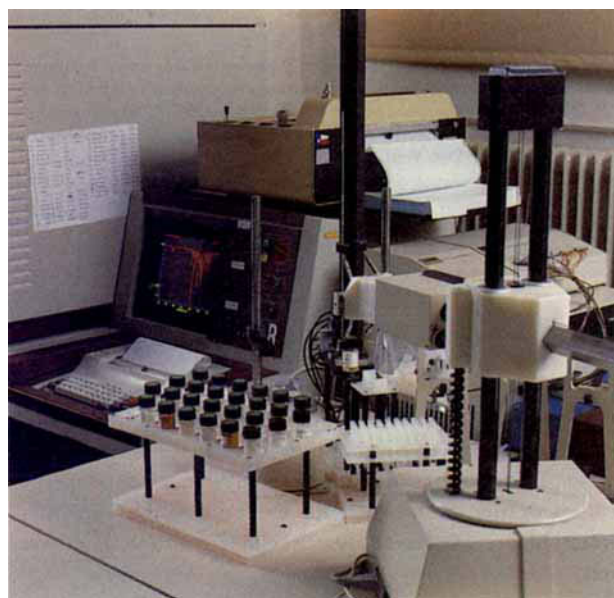


Abb. 3. Laborroboter (Zymark) zur Präparation von Filmen auf KBr-Fenstern. Ein Probenfläschchen wird gerade zur Präparation bereitgestellt. Links der Rechner des Spektrometers, rechts die optische Bank, in die der Roboter die vorbereiteten KBr-Fenster einsetzt.

Auf diese Art können in der IR-Spektroskopie 90% aller flüssigen Proben automatisch präpariert und vermessen werden; der Zeitaufwand für die Vorbereitung des Roboters beträgt etwa eine halbe Stunde. Mit Hilfe des Roboters ist es also gelungen, das zeitaufwendige Beschicken, Reinigen und Trocknen der Standardküvetten zu umgehen und auch für die routinemäßige Aufnahme von IR-Spektren flüssiger Substanzen die notwendigen technischen Voraussetzungen zu schaffen. Der Einsatz von Autosamplern mit Spül-Trocken-Cyclus<sup>[3]</sup> hat damit eine wichtige Ergänzung gefunden.

### 3. Informationsfluß im Labor

Abbildung 4 zeigt den Informationsfluß in einem IR-Spektroskopielabor. Die Substanzprobe wird vom Chemiker unter Angabe der gewünschten Experimente an das Labor geschickt. Bei festem Aggregatzustand wird die Verbindung mit KBr verrieben, in der automatischen Presse

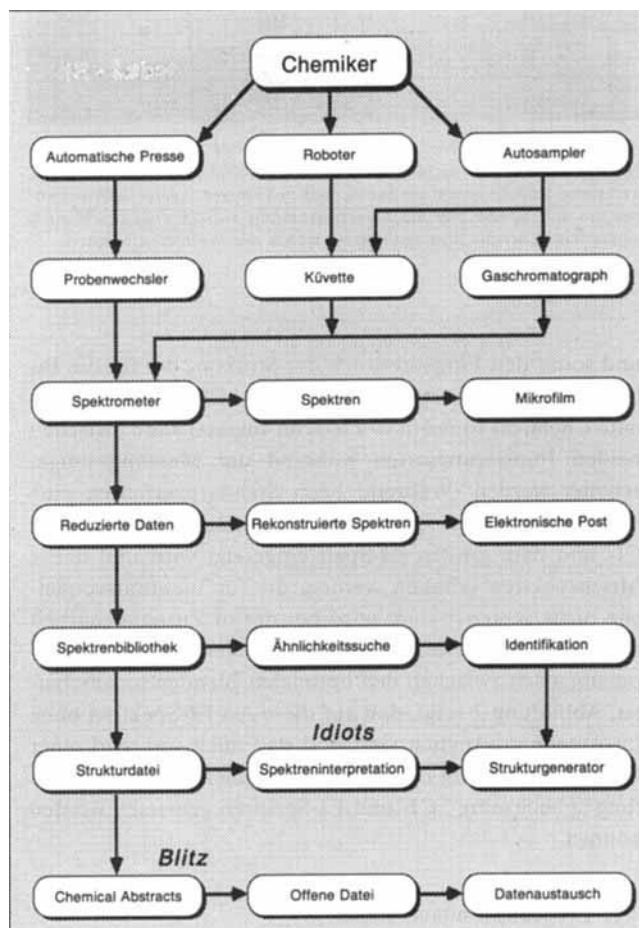


Abb. 4. Informationsfluß im IR-Spektroskopielabor. Ein Vergleich der in der Strukturdatei vorhandenen Strukturen mit denen von Chemical Abstracts über das Blitz-System zeigt, ob die Verbindungen bereits publiziert sind und daher die Datensätze für den Austausch mit Partnern bei der Entwicklung von wissensbasierten Expertensystemen freigegeben werden können.

zum Preßling geformt und in das Probenwechselrad des Spektrometers abgelegt. Flüssige Proben werden entweder vom Roboter auf KBr-Fenster aufgetragen (siehe Abschnitt 2.2), von einem Autosampler in eine Küvette gefüllt oder über eine kurze GC-Säule in die Gasküvette (light pipe) gebracht. Das automatische FTIR-Spektrometer nimmt die Meßdaten auf und zeichnet ein Spektrum, das nach Sicherung auf Mikrofilm im Original an den Auftraggeber gesandt wird. Im Spektrometerrechner werden die Daten auf Bandenlage, -intensität und -breite reduziert und in einer Spektrenbibliothek abgelegt. Aus dieser kondensierten Information kann jederzeit ein Spektrum rekonstruiert werden, das sich praktisch nicht vom Original unterscheidet. In Zukunft werden die Spektren direkt über LAN (= Local Area Network) an die Chemiker versandt werden.



Zur Auswertung vergleicht man die gemessenen IR-Spektren mit allen bis dahin gespeicherten Referenzspektren; dies ermöglicht eine schnelle und sichere Identifikation bekannter Produkte oder eine Aufklärung unbekannter Substanzen. Um diese zu charakterisieren, bietet sich eine rechnerunterstützte Spektreninterpretation an<sup>[6]</sup>, deren Varianten im folgenden erläutert werden sollen. Die Strukturdatei ermöglicht eine Transformation der spektralen Parameter in Strukturparameter, die der Strukturgenerator nutzt, um neuartige topologische Strukturen vorzuschlagen. Auch diese werden dokumentiert und dem Auftraggeber übersandt, so daß sich der Kreis des Analysenauftrags schließt.

#### 4. Abschätzung von chemischen Verschiebungen

Wie bereits gesagt, sollte die Spektralinformation  $\{\nu\}$  direkt in die vom Auftraggeber gewünschte Strukturinformation  $\{S\}$  umgesetzt werden können. Allerdings ist die Funktion  $\{S\}=f\{\nu\}$  ebenso unbekannt und komplex wie ihre Umkehrung  $\{\nu\}=f'\{S\}$ , wobei  $\{\nu\}$  die Gesamtheit der spektralen Parameter  $\delta$  (chemische Verschiebung),  $J$  (Kopplung) und  $\tau$  (Relaxationszeit) sowie deren Wechselwirkungen beschreibt.

Am ehesten ist diese Korrelation noch in der NMR-Spektroskopie anzugeben, in der die Beziehung  $\delta=f(S)$  seit den Inkrementenregeln von Shoolery et al.<sup>[7]</sup> und Grant et al.<sup>[8]</sup> auf ein Strukturelement  $S$  einschließlich der sphärisch angeordneten  $\alpha$ -,  $\beta$ -,  $\gamma$ - und  $\delta$ -Substituenten angewendet wird. Diese Denkweise liegt dem HOSE-Code<sup>[9]</sup> mit seiner hierarchisch geordneten sphärischen Darstellung der Umgebung (*Hierarchically Ordered Spherical Description of Environment*) zugrunde. Bei der Suche nach allen Referenzdaten für einen charakteristischen Knoten in der topologischen Strukturbeschreibung (Methylgruppe an Doppelbindung in Abb. 5) werden nach dem Aufbau der gewünschten Struktur die Nummer des Knotens eingege-

ben, der Teilstrukturcode berechnet und die entsprechenden Einträge im Register ausgegeben. Abbildung 5 zeigt die Einzelwerte mit den Kurzcodes für das jeweilige Lösungsmittel, den daraus resultierenden Mittelwert samt Standardabweichung und die Zahl der Referenzverbindungen, auf denen diese Abschätzung der  $^{13}\text{C}$ -chemischen Verschiebung beruht. Die Standardabweichung ist relativ groß, und die Einzelwerte können um zwei Maxima bei  $\delta = 12.5$  und  $17.8$  gruppiert werden. Die Inspektion der zugehörigen Strukturen (Abb. 6) liefert als Erklärung dafür die wohlbekannten sterischen Wechselwirkungen: Cisoide Methylgruppen absorbieren bei  $\delta \approx 12$ , transoide bei  $\delta \approx 18$ .

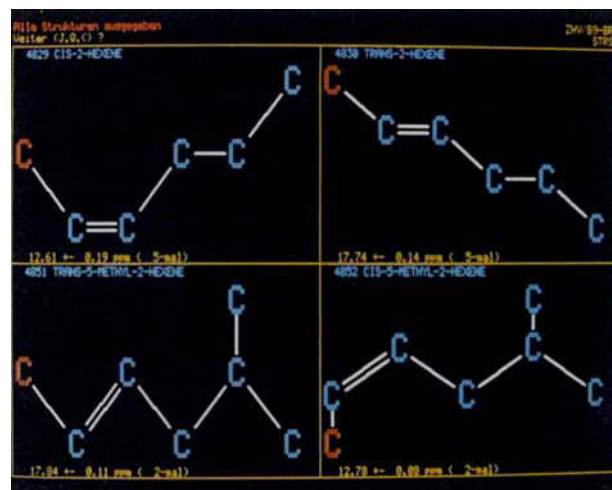


Abb. 6. Darstellung der ersten vier Referenzverbindungen mit Markierung der Atome, deren chemische Verschiebung gefragt ist, und Ausdruck der zugeordneten chemischen Verschiebungen. Sind mehrere Spektren gespeichert, ergeben sich Mittelwerte.

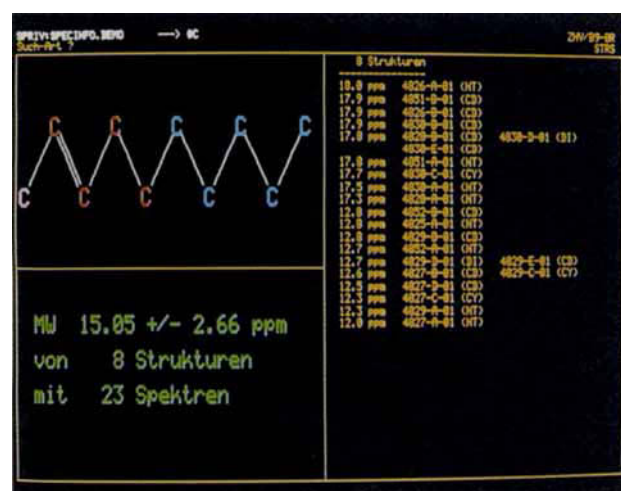


Abb. 5. Ergebnis der Suche nach Teilstrukturen mit einer Methylgruppe (lila) an einer isolierten Doppelbindung. Die Reichweite des HOSE-Codes mit vier Sphären ausgehend vom interessierenden Knoten (lila) der topologischen Strukturbeschreibung ist rot wiedergegeben. Eine Auswertung der chemischen Verschiebungen aus 23 Spektren von acht Verbindungen ergibt einen Mittelwert für  $\delta$  von 15,05 mit einer Standardabweichung von 2,66 ppm. Die rechts aufgelisteten Referenzverbindungen sind teilweise in Abbildung 6 dargestellt.

Aus diesem einfachen Beispiel lassen sich drei Erkenntnisse ableiten:

Stereochemische Wechselwirkungen haben einen großen Einfluß auf den spektralen Habitus; die topologische Strukturcodierung sollte deshalb auch die Stereochemie (zumindest die Konfiguration an Doppelbindungen) einbeziehen. Ganz allgemein muß gesagt werden, daß eine Strukturverschlüsselung, wenn sie nicht nur der Dokumentation und Wiederfindbarkeit, sondern auch der Erklärung und Modellierung physikalischer Eigenschaften dienen soll, möglichst genau die Wechselwirkungen im Molekül festhalten muß.

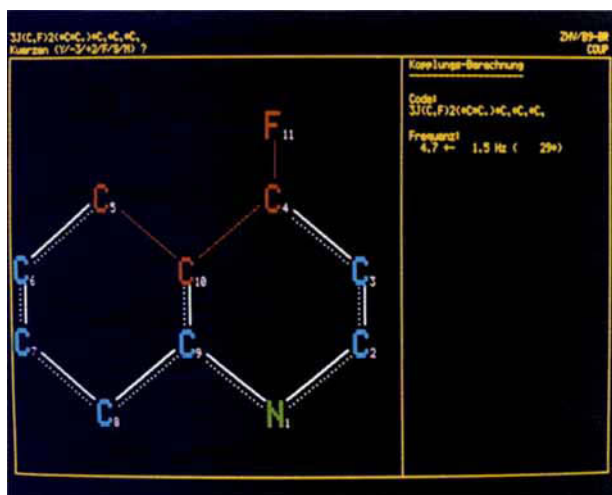
Ein bildliches Erfassen der Wechselwirkungen mit markierten Strukturbildern entspricht der Denkweise des Chemikers. Ein Expertensystem sollte nicht etwa einen Experten ersetzen, sondern dem Experten als kompetenter „Gesprächspartner“ dienen. Entscheidungshilfen des Systems für den Benutzer sollten benutzerfreundlich aufbereitet werden, und ein Blättern („Browsing“) durch die gespeicherten chemischen Strukturen, korreliert mit spektroskopischer Information, sollte zum kreativen Nachdenken anregen.

Wenn eine Korrelation Teilstruktur  $\leftrightarrow$  chemische Verschiebung möglich ist, kann auch das Gesamtspektrum als Summe der Teilinformationen vorausgesagt werden.

Es sollte jedoch immer die Grundgesamtheit, d. h. alle benutzten Informationen, auf Mehrfachmaxima getestet werden<sup>[10]</sup>, und Doppeleinträge sollten angegeben werden. Der erfahrene Spektroskopiker kann meist sofort entscheiden, welcher Effekt bei der Codierung der Daten unberücksichtigt blieb, ansonsten empfiehlt sich das Blättern in den Detailinformationen.

## 5. Erfassung von Spin-Spin-Wechselwirkungen

Um die Bausteine des Puzzles zweifelsfrei zusammensetzen zu können, ist der Spektroskopiker nicht nur an den einzelnen Molekülteilen, sondern auch an der Wechselwirkung zwischen ihnen interessiert. Besondere Bedeutung hat hier die Spin-Spin-Kopplung. Abbildung 7 zeigt die Anfrage nach der Größe der Kopplung zwischen einem <sup>19</sup>F-Atom an C4 von Chinolin und dem Atom C5.



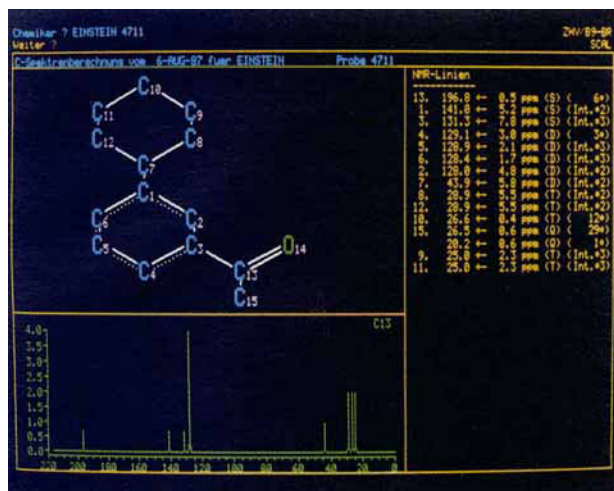


Abb. 10. Abschätzung des  $^{13}\text{C}$ -NMR-Spektrums von 3-Cyclohexylacetophenon. Außer dem rekonstruierten Spektrum werden die Einzeldaten aufgelistet, aus denen man einen Eindruck von der Zuverlässigkeit der Berechnung gewinnen kann. In der Standardausgabe, die den Auftraggebern zusammen mit den experimentellen Ergebnissen übersandt wird, sind alle wesentlichen Informationen für eine spätere Dokumentation enthalten.

Die Einzeldaten werden durch die Standardabweichung und die Zahl der Einträge oder eine Angabe, auf wieviele Sphären bei der Interpolation gekürzt wurde (INT\*2), charakterisiert und können jederzeit zu den Originaldaten zurückverfolgt werden. Doppelinträge (z. B. bei C15) resultieren aus Doppelmaxima der Grundgesamtheit, die sterische oder elektronische Wechselwirkungen, aber auch Dateifehler widerspiegeln. Dieses Verfahren kann auf alle Kerne angewendet werden – wie in Abbildung 11 für  $^{15}\text{N}$  gezeigt –, sofern die notwendige strukturorientierte Spektrendatei<sup>[11]</sup> vorhanden ist. In unserem Labor liegen die Datensätze für  $^{13}\text{C}$ ,  $^{31}\text{P}$ ,  $^{15}\text{N}$  und  $^{19}\text{F}$  vor, für  $^1\text{H}$  fehlt noch eine Zuordnung der Resonanzfrequenzen, und eine Spektrenberechnung ist hier wegen des starken Einflusses von Solvation und Konfiguration sowie wegen der Signalaufspaltung nur schwer möglich.

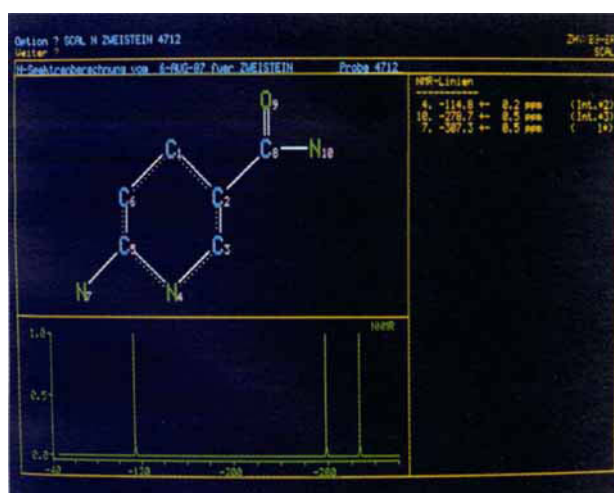


Abb. 11. Abschätzung des  $^{15}\text{N}$ -NMR-Spektrums von 6-Aminonicotinamid aufgrund einer Referenzdatei mit 1000  $^{15}\text{N}$ -NMR-Spektren [11].

Anders als die NMR-Spektroskopie ist die IR-Spektroskopie keine Atomspektroskopie, sondern eine echte Molekülspektroskopie. Auch hier ist eine Spektrensimulation

möglich, durch Vergleich mit den Spektren ähnlicher Verbindungen und eine geeignete Modifikation dieser Spektren. Ein Ergebnis wie das für Anthranilsäurehydrazid (Abb. 12) ist recht vielversprechend, vor allem, wenn man das Spektrum mit den experimentellen Spektren in Spektrenkatalogen<sup>[12]</sup> vergleicht.

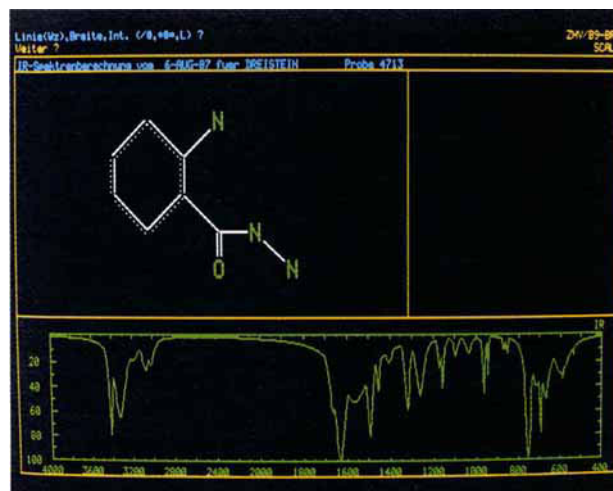


Abb. 12. Abschätzung des IR-Spektrums von Anthranilsäurehydrazid als KBr-Preßling.

Bei der Simulation von Massenspektren werden zunächst aufgrund von Ionisationspotentialen Aussagen über die wahrscheinlichen Orte der Ionisierung gemacht. Für jedes so erhaltene Radikalkation werden dann Fragmentierungs- und Umlagerungssequenzen erstellt, wobei jeder Schritt und jede Zwischenstufe entsprechend den physikalischen Parametern bewertet wird<sup>[13]</sup>.

Abbildung 13 zeigt vorläufige Ergebnisse für 3-Hydroxytetrahydrofuran. Als Ionisierungsorte kommen die

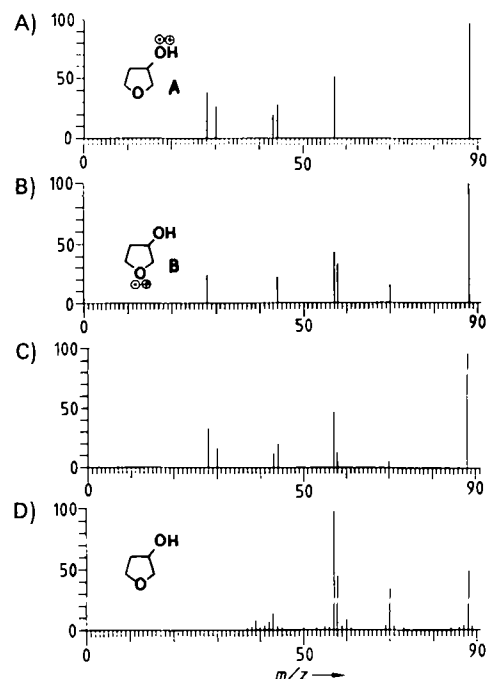


Abb. 13. Simulation des Fragmentierungsverhaltens von 3-Hydroxytetrahydrofuran. Die beiden möglichen Radikalkationen A und B liefern die Teilspektren A bzw. B, deren einfache Kombination das berechnete Massenspektrum C ergibt. Bis auf die im Experiment (Spektrum D) nicht erfaßten Massen bei  $m/z$  28 und 30 wird der spektrale Habitus gut wiedergegeben.



freien Elektronenpaare an den beiden Sauerstoffatomen in Frage, wodurch die Radikalkationen **A** bzw. **B** erhalten werden, die nach unterschiedlichen Mechanismen fragmentieren. Die Überlagerung der beiden simulierten Massenspektren ergibt Spektrum C, das recht gut mit dem experimentellen Spektrum D übereinstimmt.

## 7. Spektrensuche

Die Suche nach ähnlichen Verbindungen kann im Rechner auf verschiedene Arten erfolgen<sup>[14]</sup>. Das einfachste und oft sehr wirkungsvolle Verfahren einer Anfrage nach spektralen Mustern ist die Liniensuche über Schnittmengen. Wird nach den Merkmalen Quartett bei  $\delta = 14.0$ , Triplet bei 60.5 und Singulett bei 173.1 eines  $^{13}\text{C}$ -NMR-Spektrums bei einer Fensterbreite von 0.5 ppm gesucht, resultieren die in Abbildung 14 dargestellten Referenzspektren als die drei ähnlichsten.

Aufschlußreicher sind wiederum die zugehörigen Strukturen (Abb. 15), die erkennen lassen, daß es sich bei der

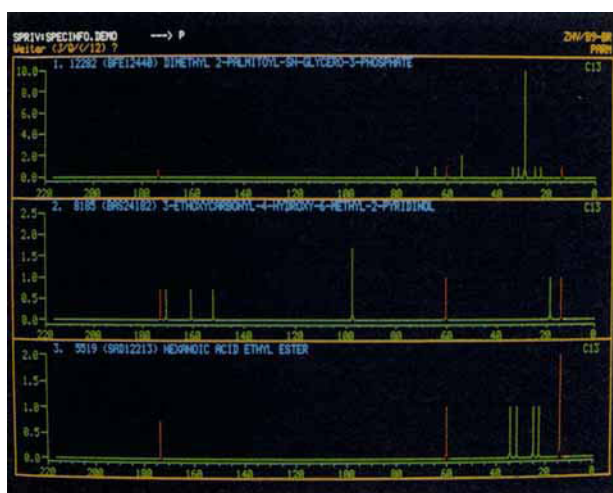


Abb. 14. Die drei ähnlichsten Referenzspektren geordnet nach steigender Summe der Abweichungen in den chemischen Verschiebungen. Die gesuchte Signalkombination ist rot eingezeichnet, die zugehörigen Verbindungen sind in Abbildung 15 dargestellt.

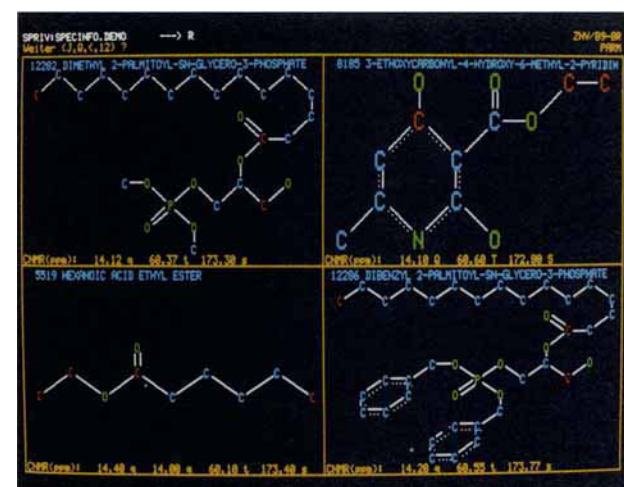


Abb. 15. Die Verbindungen zu den vier ähnlichsten Referenzspektren (vgl. Abb. 14) mit Auflistung der übereinstimmenden chemischen Verschiebungen und Markierung der entsprechenden Kohlenstoffatome (rot).

Analysensubstanz um einen Ethylester handeln könnte. Der Überblick wird dadurch erleichtert, daß durch die Zuordnung der chemischen Verschiebungen zu den Knoten der topologischen Matrix die Atome mit übereinstimmenden Resonanzfrequenzen kenntlich gemacht werden können. Trotzdem ist das Bild verwirrend, weil Strukturen mit rein zufälligen Ähnlichkeiten die Zusammenhänge verschleiern. Eine wirkungsvolle Unterdrückung dieses strukturellen Rauschens gelingt mit der Teilstrukturanalyse, die in Abschnitt 8 gezeigt werden wird.

Noch schwieriger ist die Situation in der IR-Spektroskopie und der Massenspektrometrie, da es dort kaum möglich ist, spektrale Eigenschaften mit definierten Strukturteilen zu verknüpfen. Die Suche nach IR-Spektren mit Banden bei 3305, 1255, 747 und 564  $\text{cm}^{-1}$  (Toleranz  $\pm 5 \text{ cm}^{-1}$ ) liefert die in Abbildung 16 gezeigten drei ähnlichsten Referenzspektren. Ein Vergleich der zugehörigen Verbindungen ohne Markierung der gemeinsamen Strukturteile ist verwirrend (Abb. 17). Hier kann nur eine Computeranalyse der Moleküle hinsichtlich übereinstimmender Teilstrukturen weiterhelfen.

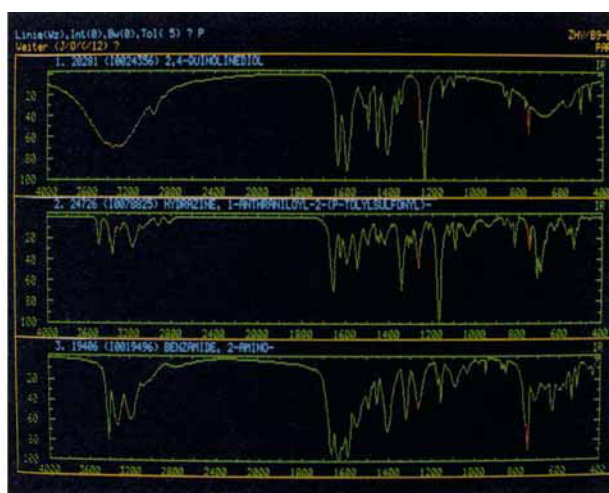


Abb. 16. Die drei ähnlichsten Referenz-IR-Spektren, geordnet nach steigender Abweichung der Bandenlagen. Die gesuchte Kombination von vier Wellenzahlen ist rot wiedergegeben, die zugehörigen Verbindungen sind in Abbildung 17 dargestellt.

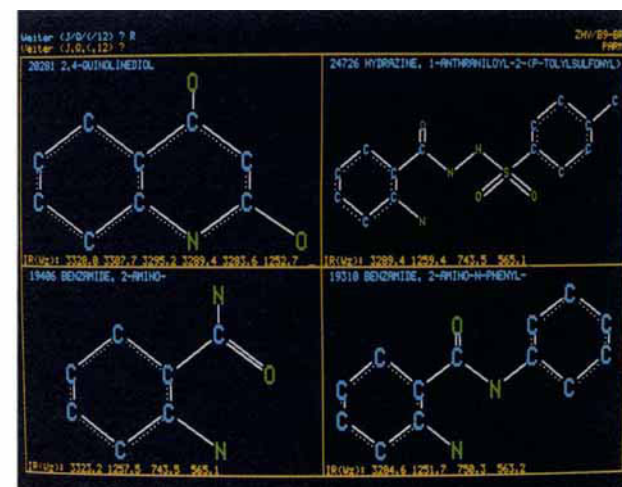


Abb. 17. Die Strukturen zu den vier ähnlichsten Referenzspektren (vgl. Abb. 16) mit Auflistung der übereinstimmenden Bandenlagen.

## 8. Teilstrukturanalyse

Bei der Teilstrukturanalyse werden die gefundenen Strukturen in ihre kleinsten Bausteine zerlegt und mit Hilfe des HOSE-Codes bezüglich ihrer Umgebung charakterisiert (wahlweise ein bis vier Sphären). Eine topologische Matrix mit 15 Knoten würde also in 15 HOSE-Codes und eine bestimmte Anzahl von HORD-Codes (*Hierarchically Ordered Ring Description*) für die Ringsysteme verwandelt. Diese Daten werden für die gewünschte Anzahl Verbindungen (typischerweise 20) erzeugt, zusammengefaßt und in Form eines Histogramms dargestellt (Abb. 18). Dabei beschreibt die grün wiedergegebene Information mit dem zugehörigen Wert in der dritten Zahlenkolonne die statistische Wahrscheinlichkeit, die jeweilige Teilstruktur bei insgesamt 20 Verbindungen in der Datei zu finden. Wird die Teilstruktur häufiger beobachtet (zweite Zahlenkolonne), so überschreitet die entsprechende rote Codierung die triviale Standardhäufigkeit (grün). Dies definiert die Wahrscheinlichkeit, mit der aufgrund der spektralen Ähnlichkeiten eine bestimmte Teilstruktur in der unbekannten Verbindung vorliegen sollte.

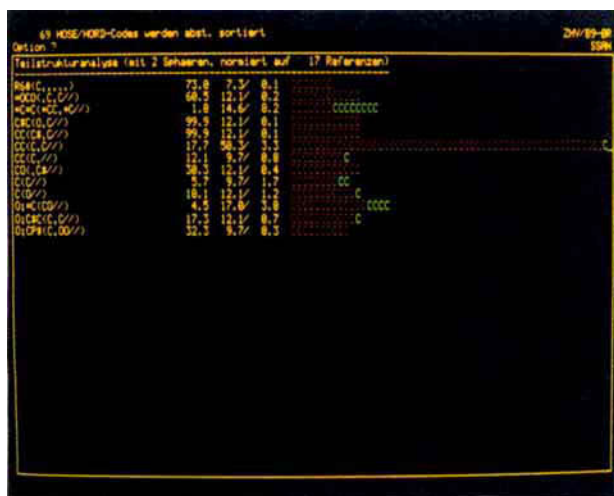


Abb. 18. Teilstrukturanalyse der Verbindungen, die bei einer Suche nach  $^{13}\text{C}$ -NMR-Referenzspektren mit gleichem spektralen Muster (siehe Abb. 14) erhalten wurden. Das Vorkommen der links aufgeführten Teilstrukturen wird als Histogramm dargestellt, wobei die grün wiedergegebene Information die Standardhäufigkeit in der  $^{13}\text{C}$ -NMR-Datei angibt und die roten Symbole die darüber hinausgehenden Knoten im Datensatz charakterisieren. Die erste Zeilenkolonne ist der Quotient aus Vorkommen und Standardhäufigkeit, die als zweite bzw. dritte Kolonne aufgelistet sind.

Es ist nicht notwendig, die Codierung der Teilstrukturen zu verstehen, um dieses Verfahren zu nutzen, obwohl das nicht schwierig wäre. Der Rechner kann die zu den spektralen Mustern gehörenden Teilstrukturen zeichnen<sup>[15]</sup>, und er kann die Informationen nutzen, um in den vorgeschlagenen Strukturen (Abb. 15, 17) die übereinstimmenden Teilstrukturen farblich hervorheben (Abb. 19 bzw. 20). Die rot wiedergegebenen Zentren treten dabei in 70%, die lila gezeichneten nur in 35% der 20 besten Referenzstrukturen auf. Die Darstellung (Abb. 19) hat sich gewandelt, die strukturellen Zusammenhänge sind deutlicher zu erkennen. Vor allem bei den IR-Vergleichsverbindungen (Abb. 20) erkennt man nun sehr schnell die Gemeinsamkeiten und damit die wahrscheinlichen Strukturbausteine. Es braucht sicherlich nicht hinzugefügt zu werden, daß der

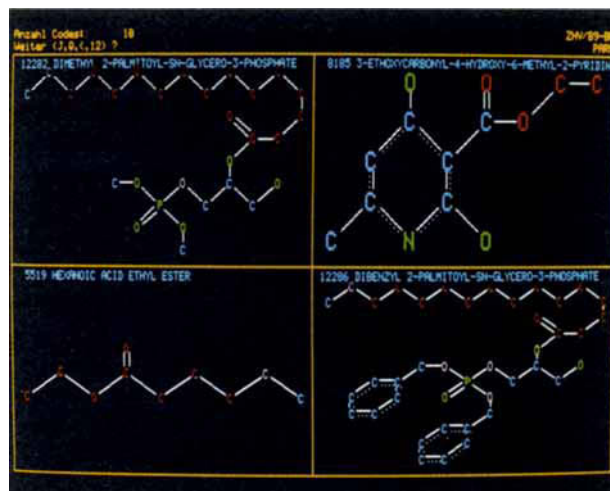


Abb. 19. Kennzeichnung der wahrscheinlichsten Teilstrukturen aufgrund der Teilstrukturanalyse (Abb. 18, vgl. Abb. 15). Die häufigsten Knoten sind rot gezeichnet, ebenso die Bindungen der dominierenden Ringsysteme (in Abb. 20). Weniger häufige, jedoch markante Strukturteile werden lila dargestellt. Die Schwellenhöhe (Standard: 70 bzw. 30%) für diese Darstellung kann im interaktiven Dialog leicht optimiert werden, um die gemeinsamen Strukturelemente sichtbar zu machen.

Rechner die Information, die er darstellen kann, auch „versteht“ und in weiterführenden Algorithmen verwenden kann.

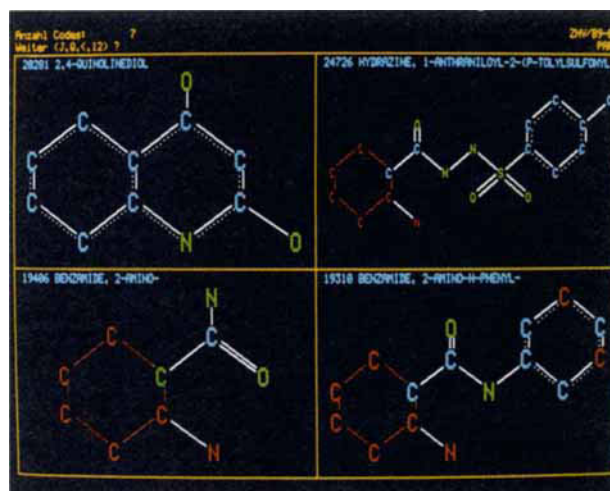


Abb. 20. Markierung der wahrscheinlichsten Teilstrukturen zur IR-Banden-suche von Abbildung 16.

Durch die Teilstrukturanalyse gelingt es also, spektrale Information, z.B. chemische Verschiebungen, Bandenlagen, Massen, Abspaltungen von Neutralteilen<sup>[15]</sup>, Koppungen, Aufspaltungsmuster etc., in Strukturinformation zu verwandeln. Die erhaltene Strukturinformation ist einheitlich, normiert und computerorientiert. Dies hat vor allem den Vorteil, daß man die Ergebnisse verschiedener Experimente und Methoden miteinander kombinieren, d.h. „Äpfel mit Birnen vergleichen“ kann. In der „Multidimensionalen Spektroskopie“<sup>[15]</sup> trägt jede Methode nur mit den statistisch relevanten, d.h. aussagekräftigen Befunden zum Gesamtergebnis bei. Das strukturelle Rauschen wird wirkungsvoll unterdrückt. Im endgültigen Strukturvorschlag läßt sich kennzeichnen, welche Technik welche Teilstrukturinformation bestätigt hat (Methodenmarkierung)<sup>[15]</sup>.



Es zeigt sich dabei, daß  $^{13}\text{C}$ -NMR-Spektroskopie und Massenspektrometrie für die Konstitutionsaufklärung am wichtigsten sind, weil sie vor allem durch das Molekülskelett bestimmt sind, das der Chemiker üblicherweise als das Hauptcharakteristikum der Struktur betrachtet. IR- und  $^1\text{H}$ -NMR-Spektroskopie mit ihren Aussagen über Substituenten und Konfiguration tragen zur Verfeinerung des Bildes bei.

Mit dem Konzept der überlappenden Teilstrukturen<sup>[16]</sup> lassen sich die Teilstrukturcodes leicht zu topologischen Strukturen zusammensetzen. Das Einfüttern zusätzlicher Information wird dadurch ermöglicht und die kombinatorische Explosion bei der Strukturgenerierung (siehe Abschnitt 9) eingedämmt.

## 9. Intelligente Spektrenauswertung

Der Begriff „Künstliche Intelligenz“ wird von verschiedenen Autoren in unterschiedlicher Weise definiert. „Intelligenz“ leitet sich ab vom lateinischen *intelligentia* = Vorstellung, Einsicht, Verstehen. „Im allgemeinen Verständnis ist Intelligenz die übergeordnete Fähigkeit (bzw. eine Gruppe von Fähigkeiten), die sich in der Erfassung und Herstellung anschaulicher und abstrakter Beziehungen äußert, dadurch die Bewältigung neuartiger Situationen durch problemlösendes Verhalten ermöglicht“<sup>[17]</sup>. In diesem Sinn sind die Werkzeuge, die wir bisher betrachtet haben, durchaus „intelligent“ und müssen nur noch zum Lösen neuartiger Probleme zusammengefaßt werden.

Tabelle 1. Strukturaufklärungsstrategien unterschiedlicher Intelligenz und Abstraktion mit zugehörigem Regelwerk und zugehöriger Wissensbasis.

Verfahren	Regeln	Wissen
Bibliothekssuche	Spektrum $\leftrightarrow$ Struktur	Spektralbibliothek
Ähnlichkeits-suche	Spektrum $\leftrightarrow$ Struktur	Spektralbibliothek
Interpretation	Teilspektrum $\leftrightarrow$ Teilstruktur	Toleranz
		Interpretationsregeln
Spektrenberechnung	Verschiebung $\leftrightarrow$ chem. Umgebung	$\delta = f(S)$
Strukturaufklärung	Verschiebung $\leftrightarrow$ chem. Umgebung	$S = f(\delta)$
	chemische Strukturen	
	Spektroskopie	

„Intelligente“ Systeme bestehen normalerweise aus einem Regelwerk (inference engine) und einer Wissensbasis<sup>[18]</sup>. Betrachten wir einmal die bisher besprochenen Methoden unter diesen Gesichtspunkten (Tabelle 1). Die Bibliothekssuche benötigt eine Sammlung von Referenzspektren und arbeitet mit der Grundregel, daß eine Beziehung zwischen Spektrum und Struktur besteht. Ähnliche Spektren sollten ähnliche Strukturen charakterisieren, eine vereinfachende Annahme, die nicht immer gültig ist.

Die Ähnlichkeitssuche ist mehr für die Suche nach ähnlichen als für die nach identischen Verbindungen optimiert. Notwendigerweise müssen daher größere Abweichungen toleriert werden. Deshalb werden im sequentiellen Durchlauf *alle* Eigenschaften des Spektrums (Lage, Intensität, Form und Wahrscheinlichkeit des Linienvorkommens) verglichen. Dies gelingt ausgezeichnet mit Bewertungsmatrizen (nach Clerc et al.<sup>[19]</sup>), denen jedoch größere Vergleichsalgorithmen (Vor- und Zwischenrecherche<sup>[14]</sup>)

vorgeschaltet sein müssen, um bei großen Dateien mit über 50000 Spektren eine Suche in akzeptabler Zeit durchführen zu können. Dieser Ansatz arbeitet also mit spektroskopischem Fachwissen, das aufgrund von Erfahrungen definiert, ob und welche Abweichungen toleriert werden können. So weiß z. B. der Spektroskopiker, daß sich im  $^1\text{H}$ -NMR-Spektrum beim Wechsel des Lösungsmittels Triplets in Dubletts von Dubletts verwandeln können und die Signale acider Protonen besonders stark verschoben werden. Auch der Rechner sollte bei Ähnlichkeitsbewertungen dieses Wissen anwenden.

Bei der Interpretation werden Teilspektren mit Teilstrukturen korreliert. Als Wissensbasis dient eine Bibliothek mit Interpretationsregeln<sup>[20]</sup>, die z. B. definieren, daß die Signale einer Ethoxygruppe (in Ethylestern) im  $^{13}\text{C}$ -NMR-Spektrum bei  $\delta = 60.4 \pm 0.5$  (Triplet) und  $14.2 \pm 0.5$  (Quartett) auftreten. Wann immer dieses spektrale Muster beobachtet wird, ist es sehr wahrscheinlich, daß eine Ethoxygruppe in der unbekannten Verbindung vorhanden ist, vor allem wenn weitere Bedingungen, wie  $^1\text{H}$ -NMR-Signale bei  $\delta = 4.1 \pm 0.2$  (Quartett, Intensität 2) und  $1.2 \pm 0.2$  (Triplet, Intensität 3), erfüllt sind. Abbildung 21 zeigt als Beispiel die Suche nach 4,6,6-Trichlor-7,7,7-trifluor-3,3-dimethyl-heptansäureethylester als fiktiver Unbekannter. Die Liste der vorgeschlagenen Strukturelemente enthält, nach absteigender Wahrscheinlichkeit geordnet, wesentliche Fragmente der gesuchten Struktur. In einem zweiten Lauf wird die Liste nochmals sequentiell abgearbeitet und auf Konsistenz geprüft, d. h. jede  $^{13}\text{C}$ -NMR-Resonanz darf

SPEKTRENINTERPRETATION						
MIT 1500 TRANSKRIPTEN						
STANDARDBEWERTUNG ? (J/N)						
=J						
ERGEBNIS: 17 PASSENDE TRANSKRIPTS (1423 VR-TREFFER)						
	GUETE	CNMR	HNMR	IR	T-NR.	NAME
1	1.933	1.997	0.968		103	ETHYL ESTER (-)
2	0.946	1.183	0.800		1112	CHLORO-METHYL (=)
3	0.946	1.183	0.800		1112	CHLORO-METHYL (=)
4	0.916	1.308	0.701		104	ETHYL ESTER (*)
5	0.449	0.915	0.491		9	METHYL (*), TWO ORTHO-SUBS1
6	0.339	0.782	0.434		1393	3-HYDROXY-BUTANOIC ACID ESTER
7	0.278	0.775	0.359		1042	2-HYDROXY-ACETIC ACID ESTER
8	0.268	1.049	0.256		1306	AMINOMETHYL (*)
9	0.191	0.806	0.237		1398	1,2-ETHYLENEDIAMINE-N1,N2-DIYL (*,*)
10	0.103	1.039	0.100		1197	2-HYDROXY-ETHYL-AMIDE
PASSENDE TRANSKRIPT-KOMBINATION						
	GUETE	T-NR.	NAME			
	1.997	103	ETHYL ESTER (-)			
	1.183	1112	CHLORO-METHYL (=)			
	1.183	1112	CHLORO-METHYL (=)			
RESTLICHE LINIEN						
	171.2	S	1			
	122.4	S	1			
	84.3	S	1			
	63.9	D	1			
	39.1	S	1			
	24.4	Q	1			
	23.3	Q	1			
TRANSKRIPT-STRUKTUREN ?						
=J						
103. TRANSKRIPT						
	C					
	C	O				
1112. TRANSKRIPT						
	C					
	CL					

Abb. 21. Beispiel zur kombinierten Interpretation von  $^1\text{H}$ - und  $^{13}\text{C}$ -NMR-Spektren. Die Strukturbausteine, die nach dem ersten Durchlauf vorgeschlagen wurden, lassen sich in diesem Fall nicht zu einem konsistenten Datensatz vereinigen, weil außer der Ethoxygruppe noch keine der im Molekül vorhandenen Teilstrukturen als Regel codiert wurde.

be der Korrelation nur einmal belegt werden. Im Gegensatz zu den Schulbeispielen<sup>[20]</sup> ist es bei dieser ungewöhnlichen Struktur recht schwierig, aus den Teilergebnissen die unbekannte Verbindung zusammenzusetzen. Das Verfahren hat drei wesentliche Nachteile:

- Die Teilstrukturen sind in der Datei vordefiniert, und nur diese Teilstrukturen können vom Rechner gefunden werden. Bei der Vielfalt der denkbaren Strukturelemente ist die Zahl der augenblicklich benutzten 1500 Regeln sicherlich zu klein, sie wird jedoch nie groß genug für die Praxis werden können. Das in Abschnitt 8 geschilderte Verfahren der Teilstrukturanalyse ist daher bestimmt der flexiblere und erfolgversprechendere Weg.
- Die Regeln wurden unter Zurateziehen der Referenzspektren weitgehend handcodiert. Dieses mühsame Verfahren muß periodisch wiederholt werden, sollen die Regeln den neuesten Stand des Wissens (= der Spektrenbibliothek) widerspiegeln. Nur vollautomatisch aus einer strukturorientierten Spektrenbibliothek abgeleitete Regeln<sup>[21]</sup> werden in der Zukunft bestehen. Dies gilt vor allem für die IR-Spektroskopie und die Massenspektrometrie, in denen die Beziehungen zwischen Teilstrukturen und Teilspektren sehr komplex sind und erst durch statistische Aufarbeitung des Spektrenmaterials sichtbar gemacht werden können (vgl. Programmpaket *Idiot*<sup>[21]</sup>).
- Die wahrscheinlichen Strukturfragmente müssen per Hand zusammengefügt werden. Diese Aufgabe ist in der Regel zeitaufwendig und sollte deshalb einem Rechner übertragen werden (siehe Abschnitt 10). Andererseits ist der Begriff „Teilstruktur“ nicht sauber definiert, und die Übersetzung der teils großen, teils kleinen, teils unscharfen Topologiefragmente in die Sprache des Strukturgenerators bereitet Schwierigkeiten.

Während die Computer-Interpretation im wesentlichen Denkanstöße im Sinne des „Brain Storming“ liefert, ist die Spektrenberechnung in der NMR-Spektroskopie eine relativ exakte Methode. Wir befinden uns nun auf dem höchsten Niveau der Abstraktion und korrelieren chemische Verschiebungen mit chemischen Umgebungen. Die Wissensbasis, ein Indexfile, in dem die mittleren Resonanzfrequenzen für alle in den Strukturen der Datei vorhandenen chemischen Umgebungen gespeichert sind, wird vollautomatisch aus der Spektrendatei erzeugt und spiegelt immer den neuesten Stand des Wissens wider. Die Anwendung wurde bereits geschildert (Abb. 10 und 11).

Das umgekehrte Verfahren, die Strukturaufklärung, ist der Traum des Chemikers und erst in Teilbereichen verwirklicht. Als Wissensbasis wird hier die invertierte Information benötigt, d.h. für jede chemische Verschiebung müssen alle zugehörigen chemischen Umgebungen gespeichert sein. Schließlich sollten auch die Grundregeln der Strukturchemie und der Spektroskopie enthalten sein, vor allem, um eine effektive Kombination der multidimensionalen Informationen zu ermöglichen.

## 10. Vollautomatische Strukturaufklärung

Die wesentlichen Bausteine und der Informationsfluß eines Systems zur automatischen Strukturaufklärung sind

am Beispiel des Programmpakets ACCESS<sup>[16]</sup> (=Automatic Computer Correlation and Evaluation of Spectra based on Substructures) dargestellt (Abb. 22). Die spektralen Pa-

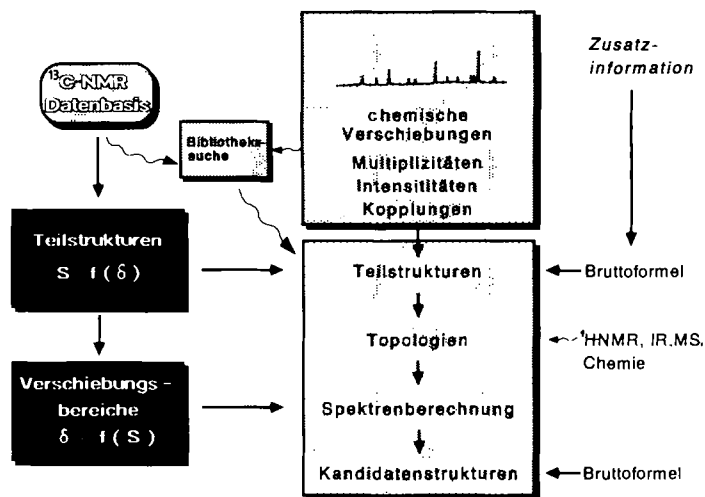


Abb. 22. Informationsfluß bei der automatischen Strukturaufklärung mit ACCESS.

rameter (chemische Verschiebung, Multiplizität, Intensität, Kopplung) einer unbekannten Verbindung werden analysiert und mit der Teilstrukturdatei  $S = f(\delta)$  verglichen. Die Bruttoformel, die leicht (vor allem, wenn die Zahl der C-Atome bekannt ist) aus hochaufgelösten Massenspektren gewonnen werden kann, dient zur Eliminierung von Teilstrukturcodes, die aufgrund der Elementzusammensetzung nicht in Frage kommen. So brauchen z.B. keine  $\text{SiCH}_3$ -Gruppen berücksichtigt zu werden, wenn laut Massenspektrum oder Elementaranalyse kein Silicium vorhanden ist. Oder aromatische Systeme können ausgeschlossen werden, wenn aufgrund der Zahl der Wasserstoffatome nur drei Doppelbindungen/Ringe zugelassen sind.

Aus den gesammelten und nach ihrer Wahrscheinlichkeit geordneten Teilstrukturen werden nun alle nur denkbaren Topologien aufgebaut. Bei wachsender Molekülgröße und/oder zahlreichen Heteroatomen und Doppelbindungen kommt es zu einer kombinatorischen Explosion, so daß der hierarchische Baum frühzeitig beschnitten werden muß, um Konvergenz zu erzwingen. Dies geschieht am besten durch Einbeziehung zusätzlicher Meßergebnisse, etwa durch Teilstrukturanalyse der Verbindungen, die bei einer Bibliothekssuche aufgrund der IR- oder Massenspektren<sup>[15]</sup> erhalten wurden. Auch Wissen über die chemische Reaktion und die Reaktionspartner kann hier einfließen. Der Chemiker kann ähnlich wie bei der Syntheseplanung mit dem Computer den Fortschritt des Programms beobachten und helfend eingreifen, wenn „unsinnige“ Vorschläge auftauchen. Da es für den Computer keine unsinnigen Vorschläge gibt, ist hier große Vorsicht geboten, damit nicht Wunschenken die Aussage des unvoreingenommenen „Gesprächspartners“ verfälscht. Besteht jedoch ein Informationsdefizit, so sollte der Rechner das Wissen des Menschen berücksichtigen können.

Auch zeigt sich, daß durch zusätzliche vom Computer vorgeschlagene Experimente die Wahrscheinlichkeit ein-

# STRUKTURAUFKLAERUNG

- 1 BRUTTOFORMEL  
=C11 H16 CL3 F3 O2  
ES GIBT 1 C-DOPPELBINDUNGEN ODER RINGE  
1-SPHAEREN-CODES SUCHEN.....  
2-SPHAEREN-CODES SUCHEN.....
- 2 BERUECKSICHTIGUNG DER ERGEBNISSE ANDERER SUCHEN ?  
=S
- 3 WIE HOCH SOLL JEDER DIESER CODES BEWERTET WERDEN ? (RETURN=1.00)  
=0.8
- 4 NUMMER DER STARTLINIE ? (RETURN=AUTOMATISCH)  
=

## SAHO - SPEKTREN - SUCHE

MIT 77486 SPEKTREN

ERGEBNIS: 1 SPEKTNUM

ES WERDEN 11 HOSE-CODES MIT LINIENLAGE GESPEICHERT

WIE HOCH SOLL JEDER DIESER CODES BEWERTET WERDEN ? (RETURN=1.00)  
=0.8

DATEI MIT 239235 ERWARTUNGSBEREICHEN VON 84335 SPEKTREN

NUMMER DER STARTLINIE ? (RETURN=AUTOMATISCH)  
=

DIE VERARBEITUNG BEGINNT MIT DER 2. LINIE (122.4 PPM)

ES WERDEN NUR DIE 8 BESTEN CODES JEDER LINIE VERARBEITET

ZU DIESEN EINGABEDATEN WURDE KEINE STRUKTUR GEFUNDEN. VIELLEICHT SOLLTEN SIE DIE ZAHL DER VERARBEITETEN CODES BZW. DIE ZUSATZREIEN VERGROESSERN ODER DIE MINDESTBEWERTUNG VERKLEINERN. ES IST ABER AUCH MOEGLICH, DASS ES SICH UM EINE UNGEWUENHLICHE STRUKTUR HANDELT UND KEINE PASSENDEN TEILSTRUKTUREN IN DER DATEI VORHANDEN SIND.

BESTE NAEHERUNG

F CL  
F 2 3.6  
F CL

DIE ANZAHL ZU VERARBEITENDE CODES PRO LINIE WURDE AUF 16, DIE MINDESTBEWERTUNG AUF 1.00 GESETZT

DIE VERARBEITUNG BEGINNT MIT DER 2. LINIE (122.4 PPM)

ES WERDEN NUR DIE 16 BESTEN CODES JEDER LINIE VERARBEITET

1. LOESUNGSVORSCHLAG: BRUTTOFORMEL C11 H16 CL3 F3 O2

CL CL D  
F 3 1.0.5.11  
2 6 7  
F F 4.8  
CL 10 9



## VERGLEICH MIT GESCHAETZTEM CNMR-SPEKTNUM

	EXP.	ABW.
C 1 (S) 171.0 PPM	0.1 PPM ( 4 MAL)	171.2 PPM 0.2 PPM
C 2 (S) 122.4 PPM	0. PPM ( 1 MAL)	122.4 PPM 0. PPM
C 3 (S) 84.3 PPM	0. PPM ( 1 MAL)	84.3 PPM 0. PPM
C 4 (D) 64.0 PPM	0. PPM ( 1 MAL)	63.9 PPM 0.1 PPM
C 5 (T) 60.4 PPM	0.3 PPM ( 386 MAL)	60.4 PPM 0. PPM
C 6 (T) 43.9 PPM	0.8 PPM ( 3 MAL)	44.9 PPM 1.0 PPM
C 7 (T) 43.5 PPM	0. PPM ( 1 MAL)	43.5 PPM 0. PPM
C 8 (S) 38.9 PPM	0.2 PPM ( 3 MAL)	39.1 PPM 0.2 PPM
C 9 (Q) 23.9 PPM	0.5 PPM ( 6 MAL)	24.4 PPM 0.5 PPM
C10 IST SYMMETRISCH ZU C 9		23.3 PPM 0.6 PPM
C11 (Q) 14.2 PPM	0.1 PPM ( 2000 MAL)	14.3 PPM 0.1 PPM

SYMMETRIE-ABWEICHUNG: 1.1 PPM (WURDE ZUR STANDARDABWEICHUNG ADDIERT)

DIESER VORSCHLAG HAT EINE STANDARDABWEICHUNG VON 1.59 PPM

2. LOESUNGSVORSCHLAG: BRUTTOFORMEL C11 H16 CL3 F3 O2

F F D  
F F 1  
2 CL 0 11  
F 3 5  
CL 6 7  
4.8  
CL 10 9

## VERGLEICH MIT GESCHAETZTEM CNMR-SPEKTNUM

	EXP.	ABW.
C 1 (S) 170.6 PPM	0.5 PPM ( 136 MAL)	171.2 PPM 0.6 PPM
C 2 (S) 122.4 PPM	0. PPM ( 1 MAL)	122.4 PPM 0. PPM
C 3 (S) 84.3 PPM	0. PPM ( 1 MAL)	84.3 PPM 0. PPM
C 4 (D) 69.5 PPM	1.6 PPM ( INTERP.)	63.9 PPM 5.6 PPM
C 5 (T) 63.6 PPM	1.0 PPM ( INTERP.)	60.4 PPM 3.2 PPM
C 6 (T) 43.5 PPM	0. PPM ( 1 MAL)	44.9 PPM 1.4 PPM
C 7 (T) 40.1 PPM	5.1 PPM ( INTERP.)	43.5 PPM 3.4 PPM
C 8 (S) 36.5 PPM	1.4 PPM ( INTERP.)	39.1 PPM 2.6 PPM
C 9 (Q) 28.5 PPM	2.4 PPM ( INTERP.)	24.4 PPM 4.1 PPM
C10 IST SYMMETRISCH ZU C 9		23.3 PPM 5.2 PPM
C11 (Q) 20.7 PPM	0.2 PPM ( 822 MAL)	14.3 PPM 6.4 PPM

SYMMETRIE-ABWEICHUNG: 1.1 PPM (WURDE ZUR STANDARDABWEICHUNG ADDIERT)

DIESER VORSCHLAG HAT EINE STANDARDABWEICHUNG VON 4.69 PPM  
DIESER VORSCHLAG WURDE VERWORFEN

--> 1 KZYKLEN DURCHGEFUEHRT (14.59 S CPU)

UND WEITER ?

=

--> 2 KZYKLEN DURCHGEFUEHRT (25.74 S CPU)

UND WEITER ?

=

SE	STR.#	IDENT.	BRUTTOFORMEL
1.59	1		C11 H16 CL3 F3 O2
4.69	2		C11 H16 CL3 F3 O2

(VERWORFEN)

Abb. 23. Beispiel einer automatischen Strukturaufklärung mit ACCESS. Die augenblickliche Programmversion kann noch nicht erkennen, daß C9/C10 aufgrund der Wechselwirkung mit dem benachbarten Chloratom unterschiedliche chemische Verschiebungen aufweisen sollten. Man beachte die schnelle Konvergenz des Programms in nur 25 s CPU. Die Abbildung beginnt in der linken Spalte oben!

zelter Teilstrukturen erhöht werden kann, so daß sie über das strukturelle Rauschen herausgehoben werden. Dadurch wird die Konvergenz beschleunigt, das Ergebnis klarer. Deshalb wird gerne mit den gleichen Daten eine Bibliothekssuche mit anschließender Teilstrukturanalyse vorgeschaltet. Die Bibliothekssuche sieht die spektralen Merkmale im Zusammenhang und kann deshalb spezifischere Strukturinformationen liefern.

Die Spektren aller vom Rechner vorgeschlagenen Strukturen werden mit Hilfe der Verschiebungsbereichsdatei berechnet und mit den experimentellen Spektren verglichen. Es entsteht eine nach der Wahrscheinlichkeit geordnete Liste von Kandidatenstrukturen, die nochmals auf Übereinstimmung mit der Bruttoformel überprüft werden. Strukturen mit großer Abweichung zwischen Berechnung und Experiment werden verworfen. Aus einer Detailspektion der wahrscheinlichsten Strukturen folgt häufig, daß weitere Meßdaten für eine endgültige Klärung notwendig sind.

Abbildung 23 veranschaulicht den Ablauf einer automatischen Strukturaufklärung (mit ACCESS) am Beispiel des 4,6,6-Trichlor-7,7,7-trifluor-3,3-dimethylheptansäureethylesters. Nach Eingabe der Bruttoformel ① ermittelt der Rechner die Zahl der Doppelbindungen oder Ringe. Auf Wunsch können die Ergebnisse anderer Suchverfahren berücksichtigt werden ②; in diesem Fall wurde eine Bibliothekssuche über Hash-Code durchgeführt (SAHO=Spectral Appearance in Hierarchical Order<sup>[14, 22]</sup>), und die gefundenen Teilstrukturen wurden mit einer Wahrscheinlichkeit von 0.8 bewertet ③.

Nun beginnt die eigentliche Suche in einem Indexfile der <sup>13</sup>C-NMR-Verschiebungsbereiche für 240 000 Teilstrukturen, generiert aus 84 000 Datensätzen von <sup>13</sup>C-NMR-Spektren mit topologischen Strukturen. Die Startlinienposition für den Zusammenbau der ermittelten Strukturteile durch den Strukturgenerator wird vom Rechner selbst gewählt ④. Die Linie bei δ=122.4 mit der großen Kopplung zu F ist offensichtlich am charakteristischsten. In einem ersten schnellen Durchlauf mit nur acht Teilstrukturcodes pro Linie wird nur ein Teil des Moleküls aufgebaut, für eine Konvergenz fehlen noch wichtige Strukturbausteine. Dieses grobe Prüfverfahren ist mit der „Blitzschacheinstellung“ bei kommerziellen Schachcomputern zu vergleichen, bei der die oberflächliche Bewertung eine schnelle Übersicht über die Komplexität des Problems ermöglicht.



Der Rechner schaltet dann automatisch um auf eine detaillierte Bearbeitung des Problems mit je 16 Teilstrukturen und einer Mindestbewertung der Wahrscheinlichkeit von 1.0. Nun resultieren zwei Strukturvorschläge, von denen der zweite aufgrund der starken Abweichung (4.7 ppm) zwischen Berechnung und Experiment verworfen wurde. Nach 2 K Cycles, d.h. 2000 erfolglosen Versuchen, eine neue Struktur aufzubauen, hat der Rechner alle Möglichkeiten durchgespielt und bricht seine Auswertung mit einer Zusammenstellung des Ergebnisses ab ⑥. Die Struktur wurde korrekt vorhergesagt, die Alternative kann leicht aufgrund der Lage von C11 ausgeschlossen werden.

## 11. Strukturelle Ähnlichkeit

Die Frage der Ähnlichkeit von chemischen Strukturen kann nicht allgemein beantwortet werden, da sie von den Nebenbedingungen der Fragestellung abhängt. So klassifizieren z.B. der Wirkstoffchemiker, der Pharmazeut und der Spektroskopiker ganz unterschiedliche Moleküle als ähnlich. Auch die häufig auf eine Methode spezialisierten Spektroskopiker sehen Strukturunterschiede unter dem Aspekt der Unterscheidungsfähigkeit „ihrer“ Technik und werden deshalb Ähnlichkeiten unterschiedlich wichten. Ein Maß für strukturelle Ähnlichkeit/Unähnlichkeit als Funktion der Methode ist jedoch wichtig, um bei Bedarf an zusätzlicher Information das richtige Experiment zu wählen und zwischen Alternativstrukturen zu entscheiden.

Wichtigste und in Synthesen meist unveränderte Molekülbausteine sind die Ringsysteme. Ringsuchen<sup>[22]</sup> und die Einbeziehung des Ringcodes HORD in die Teilstrukturanalyse<sup>[16]</sup> sind daher notwendige Strategien. Für kondensierte Ringsysteme bietet sich der CAS-Ring-Index<sup>[23]</sup> an, der in abgewandelter Form als Suchkriterium verwendet wird<sup>[22]</sup>. Erfolgsversprechend ist eine Analyse der gegebenen Struktur hinsichtlich ihrer charakteristischen Bausteine.

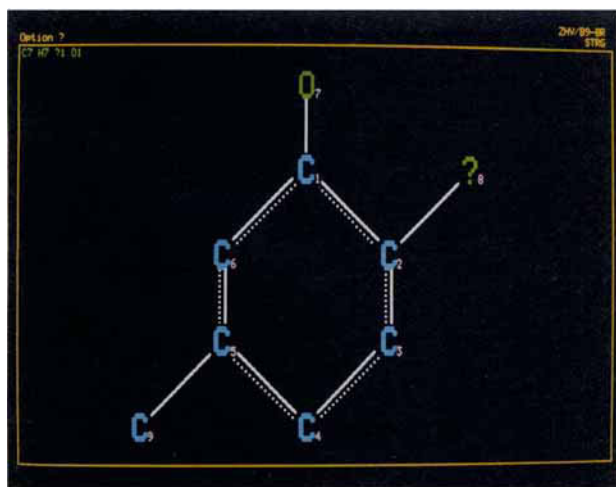


Abb. 24. Anfrage für eine Suche nach ähnlichen Strukturen mit einem nicht festgelegten Substituenten.

ne. Diese sind in einer Bibliothek als Kurzcodes vordefiniert<sup>[24]</sup> und können methodenspezifisch bewertet werden.

Abbildung 24 und 25 veranschaulichen das Verfahren. Die Zielstruktur 3-Methylphenol mit unbekanntem Substituenten in Position 5 wird eingegeben und einer Fragmentanalyse unterworfen (derzeit werden 240 Fragmente verwendet). Abbildung 25 zeigt die für die Zielstruktur geforderten Molekülbausteine und ihr Vorkommen in der Strukturdatenbank.

Fragment	Occurrence
134 1= AROM(1,2,3,4)	264
189 1= O=C(1)	474
97 3= >C=C(1)	7474
71 1= >C=C(1)	1754
130 1= O=C(1)	1324
130 1= O=C(1)	7711
239 1= C(1)	7604
189 1= O=C(1)	18174

Abb. 25. Kurzcodes [24] als Suchkriterium für die Anfrage in Abbildung 24. Diese Kontroll-Liste kann vom geübten Benutzer vor der Suche ediert, d.h. ergänzt oder verkürzt werden. Rote Einträge sind in vorausgehenden, spezielleren Kurzcodes bereits enthalten und werden deshalb bei der Suche nicht berücksichtigt.

Die roten Einträge sind redundant, weil im übrigen Datensatz schon enthalten, und daher wird nach ihnen nicht gesucht. Die resultierenden ähnlichen Strukturen zeigt Abbildung 26. Da das Verfahren im Augenblick für die IR-Spektroskopie optimiert wurde, weisen alle diese Verbindungen trotz der drastischen Unterschiede in Zusammensetzung und Konstitution ähnliche IR-Spektren auf.

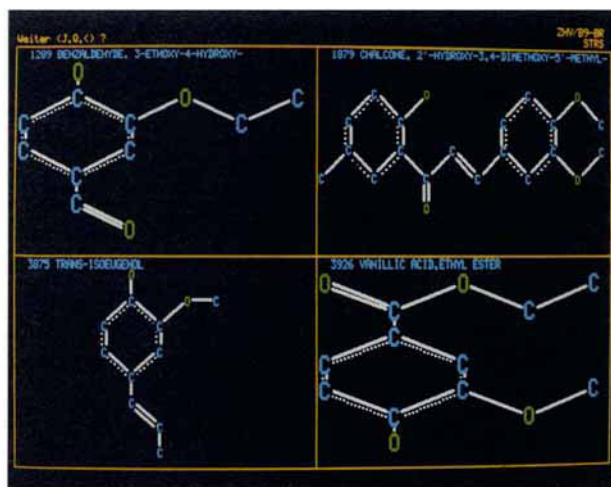


Abb. 26. Vier Strukturen, die die Ähnlichkeitsanfrage in Abbildung 24 und die automatisch erzeugten Kurzcodes in Abbildung 25 erfüllen.

## 12. Ausblick

Trotz der Schwierigkeiten, die intelligente Systeme bei der Interpretation von Spektren in der Vergangenheit hatten und heute noch haben, besteht die begründete Hoffnung, daß diese Systeme nach der Spektrenaufnahme bald auch die Spektrenauswertung revolutionieren werden. Deshalb sei zum Abschluß die Entwicklung skizziert, die wohl in den nächsten fünf Jahren ablaufen dürfte.

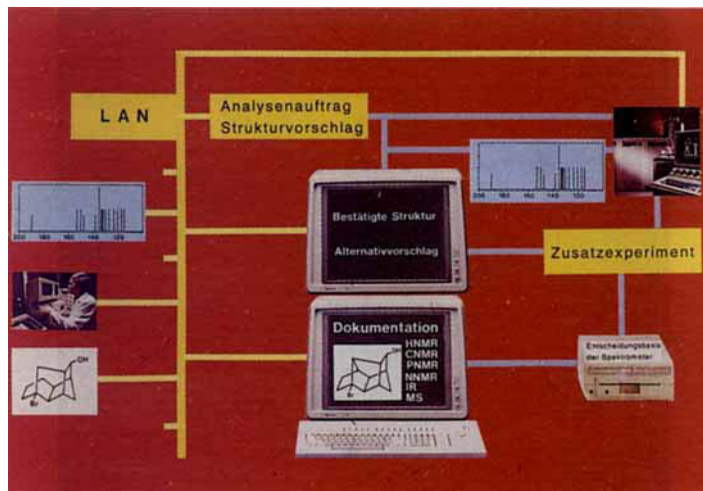


Abb. 27. Informationsfluß in einem Spektroskopielabor der neunziger Jahre. Die Einzelkomponenten sind im wesentlichen bereits vorhanden und brauchen nur noch zu einem arbeitsfähigen System zusammengestellt zu werden.

Abbildung 27 zeigt den Informationsfluß in einem Spektroskopielabor der Zukunft. Die Computer an den Arbeitsplätzen der Chemiker sind über LAN mit dem Spektroskopielabor verbunden; die Chemiker senden ihren Analysenauftrag einschließlich Strukturvorschlag direkt an das Informations-Managementsystem des Labors. Entsprechend Aufgabenstellung und Kapazität wird das geeignete Gerät gewählt und die Substanzprobe dorthin angefordert. Nach der Messung überprüft der Spektrometerrechner, ob die Meßergebnisse in Einklang mit der vorgeschlagenen Struktur sind. Wenn ja, erhält der Auftraggeber die Bestätigung und auf Wunsch auch das Spektrum, das er in sein elektronisches Laborjournal hineinkopieren kann. Bei Erfüllung der Qualitätsnormen wird der Datensatz vervollständigt und in die Datenbank als Wissensbasis aufgenommen. Werden Unstimmigkeiten zwischen Experiment und Berechnung beobachtet, schlägt das zentrale Informationssystem SPECINFO<sup>[25]</sup> Alternativstrukturen vor, die dem Chemiker zur Prüfung übersandt und eventuell durch automatisch durchgeführte Messungen bestätigt werden. Für diese Experimente empfiehlt es sich, das gleiche Spektrometer zu verwenden, da dann die Substanzprobe nicht transportiert werden muß. Erst wenn die Suche immer noch nicht konvergiert, muß der menschliche Experte eingreifen und partnerschaftlich mit dem wenig „kreativen“, jedoch ungeheuer „fleißigen“ und „belesenen“ Partner Computer das Problem angehen. Wie in Abschnitt 9 erwähnt, sollte ein Teil des großen Wissens und der Erfahrungen des Menschen in ein Spezialprogramm übertragen

werden, das im Spektrometer vorhanden ist und direkt intelligente Entscheidungen treffen kann.

Entscheidend für die Leistungsfähigkeit des gesamten Systems ist die Wissensbasis. Sie sollte zugleich breit angelegt und auf die speziellen Probleme des Arbeitskreises optimiert sein, d. h. käufliche Datenbanken und eigene Meßdaten vereinigen.

Eine kurze Übersicht über die kommerziell erhältlichen Referenzdaten findet man in<sup>[20]</sup>. Wichtig für kleine Laboreinheiten ist die Möglichkeit, auf <sup>13</sup>C-NMR-, IR-<sup>[26]</sup> und MS-Datenbanken<sup>[27]</sup> und die zugehörigen Expertensysteme zugreifen zu können, auch wenn das Hinzuladen hausinterner Informationen, z. B. spezieller Meßergebnisse aus der Chemie der eigenen Arbeitsgruppe nicht möglich ist. Ein guter Kompromiß sind häufig die für Kleinrechner entwickelten Systeme, die teils von Geräteherstellern direkt als Bestandteil der Analysengeräte angeboten werden, teils auch auf Auswerterechnern vom Typ IBM-PC<sup>[28]</sup> eingesetzt werden können. Die Leistungsfähigkeit der Programme und Dateien ist jedoch recht unterschiedlich<sup>[29]</sup>.

In der Zukunft muß steigender Wert auf die Qualität der Referenzdaten gelegt werden, denn fehlerhafte Dateieinträge führen zu falschen Voraussagen. Wie Lem in dem als Motto dieses Beitrags gewählten Text erwähnt, führt „Lernen“ zur Verselbständigung. Wir werden daher sicherlich noch viele, hoffentlich freudige Überraschungen bei der Arbeit mit Computern erleben (Abb. 28).

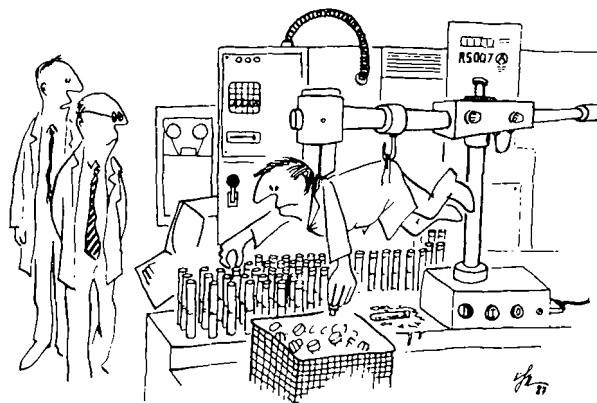


Abb. 28. „Einige Anlaufschwierigkeiten konnten wir auf diese Weise beheben.“

*Ein so ehrgeiziges Projekt wie die automatisierte Strukturauflösung kann nur in enger Zusammenarbeit mit engagierten Kollegen durchgeführt werden, die alle entscheidend zum augenblicklichen Leistungsstand des Gesamtsystems beigetragen haben: M. Beckmann, W. Fachinger, Prof. Dr. J. Gasteiger, A. Hardt, C. Heistermann, C. Köhler, Dr. R. Neudert, Dr. M. Passlack, R. Rupprecht, C. Schäfer, H. Wagner und F. Wolf. Der Forschungsleitung der BASF, Prof. Dr. H. Dörfel, Dr. U. Schenck und Dr. H. W. Schmidt, danken wir für die kontinuierliche Unterstützung und dem Bundesministerium für Forschung und Technologie sowie dem Fachinformationszentrum Energie, Physik, Mathematik GmbH für die langjährige finanzielle Förderung.*

Eingegangen am 27. August 1987 [A 657]

- [1] S. Lem: *Also sprach Golem*, Insel-Verlag, Frankfurt 1984.
- [2] R. Neudert, E. Ströfer, W. Bremser, *Magn. Reson. Chem.* 24 (1986) 1089.
- [3] W. Bremser, R. Neudert, *Eur. Spectrosc. News* 75 (1987) 10.
- [4] Finnigan MAT, Bremen, AUDEVAP Application Note (1982).
- [5] Maschinenfabrik Herzog GmbH, Osnabrück; M. Passlack, *Bruker Rep. J* (1986) 11.
- [6] M. E. Munk, M. Forkas, A. H. Lipkis, B. D. Christie, *Mikrochim. Acta* 1986 II, 199; K. Funatsu, C. A. Del Carpio, S. Sasaki, *Comput. Enhanced Spectrosc.* 3 (1986) 133.
- [7] J. N. Shoolery, L. F. Johnson, W. A. Anderson, *J. Mol. Spectrosc.* 5 (1960) 110.
- [8] D. M. Grant, E. G. Paul, *J. Am. Chem. Soc.* 86 (1964) 2948.
- [9] W. Bremser, *Anal. Chim. Acta* 103 (1978) 355.
- [10] W. Bremser, B. Franke, H. Wagner: *Chemical Shift Ranges in Carbon-13 NMR Spectroscopy*. Verlag Chemie, Weinheim 1982.
- [11] R. Neudert, G. Goizet, W. Bremser, M. Martin, unveröffentlicht.
- [12] Sadtler Grating Spectra Nr. 11671, Sadtler Research Laboratories, Philadelphia, PA, USA (1968).
- [13] J. Gasteiger, *Nachr. Chem. Tech. Lab.* 34 (1986) 228; J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gaun, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki, *Top. Curr. Chem.* 137 (1987) 19.
- [14] W. Bremser, H. Wagner, B. Franke, *Org. Magn. Reson.* 15 (1981) 178; W. Bremser, *Fresenius Z. Anal. Chem.* 286 (1977) 1.
- [15] R. Neudert, W. Bremser, H. Wagner, *Org. Mass Spectrom.* 22 (1987) 321.
- [16] W. Bremser, W. Fachinger, *Magn. Reson. Chem.* 23 (1985) 1056.
- [17] *Meyers Neues Lexikon, Band 4*, Dudenverlag, Mannheim 1979, S. 164.
- [18] J. W. Frazer, *Mikrochim. Acta* 1986 II, 163.
- [19] J. T. Clerc, F. Erni, *Top. Curr. Chem.* 39 (1973) 91.
- [20] W. Bremser, *Nachr. Chem. Tech. Lab.* 31 (1983) 456.
- [21] M. Passlack, W. Bremser in J. Zupan (Hrsg.): *Computer-Supported Spectroscopic Databases*, Ellis Horwood, Chichester 1986.
- [22] W. Bremser, L. Ernst, W. Fachinger, R. Gerhards, A. Hardt, P. M. E. Lewis: *Carbon-13 NMR Spectral Data*, 4. Auflage, VCH Verlagsgesellschaft, Weinheim 1987.
- [23] *Ring Systems Handbook*, Chemical Abstracts Service, Columbus, OH, USA 1984.
- [24] M. Passlack, W. Bremser, M. Beckmann, H. Wagner: *Infrared Spectral Data*, VCH Verlagsgesellschaft, Weinheim 1986.
- [25] W. Bremser, W. Fachinger, H. Wagner: *SPECINFO-Multidimensionales Expertensystem zur Strukturaufklärung*, unveröffentlicht.
- [26] Datenbank INKADATA, Fachinformationszentrum Energie, Physik, Mathematik GmbH, D-7514 Eggenstein-Leopoldshafen 2.
- [27] Datenbank MASSLIB, Fachinformationszentrum Chemie, D-1000 Berlin 12.
- [28] B. Wittig, M. Zeitler, W. Schmidt, W. Bremser, W. Fachinger: *Datendokumentation II*, Wissenschaftliche Software, VCH Verlagsgesellschaft, Weinheim 1987.
- [29] J. T. Clerc, E. Pretsch, M. Zürcher, *Mikrochim. Acta* 1986, 217.