

*Anwendung von Computern bei der Strukturaufklärung organischer Verbindungen mit Hilfe spektroskopischer Methoden**

Von F. ERNI und J. T. CLERC

Laboratorium für Organische Chemie, Eidgenössische Technische Hochschule Zürich

Summary

A system for computer-aided identification of organic compounds by means of spectroscopic methods is described. The unknown spectra are compared with a library of reference spectra. A highly flexible software automatically selects the appropriate search strategy and directs the updating of the library, thereby ensuring continuous adaption to varying needs.

Einleitung

Mit der zunehmenden Bedeutung der spektroskopischen Methoden, im besondern der Massen-, Infrarot-,

Kernresonanz- und UV-Spektroskopie für die Strukturaufklärung organischer Verbindungen¹, ist auch der Bedarf nach leicht zugänglichen Referenzdaten angestiegen. Bei der heute allgemein üblichen weitgehend empirischen Interpretation der Spektren wird meist so vorgegangen, daß aufgrund spektroskopischer und chemischer Evidenz (und nicht zuletzt auch aufgrund der chemischen Intuition) eine vorläufige Struktur postuliert

* Eingegangen am 17. September 1970.

¹ R. H. SILVERSTEIN und G. C. BASSLER, *Spectrometric Identification of Organic Compounds*, John Wiley & Sons, Inc., New York/London 1963.

wird. Dieser Strukturvorschlag wird nun daraufhin überprüft, ob er mit den experimentellen Befunden vereinbar ist. Bestehen keine Widersprüche, so wird die vorgeschlagene Struktur der Verbindung definitiv zugeordnet. Für die Überprüfung der Strukturvorschläge sind grundsätzlich zwei verschiedene Kategorien von Referenzdaten notwendig:

- a) Spektren von Verbindungen mit ähnlicher Struktur
- b) Struktur von Verbindungen mit ähnlichen spektralen Daten

Die heute zugänglichen Spektrensammlungen^{2,3,4} sind so organisiert, daß Referenzspektren von Verbindungen mit ähnlicher Struktur relativ leicht über Formelregister, Namenregister und eventuell auch über Stoffklassenregister zugänglich sind. Allerdings sind Kombinationen von zwei und mehr Suchkriterien ohne die Hilfsmittel der elektronischen Datenverarbeitung nur mit prohibitiv hohem Zeitaufwand zu verarbeiten.

Das Auffinden von Strukturen von Verbindungen, deren spektrale Daten denen der unbekannteren Verbindungen ähnlich sind, ist wesentlich schwieriger. Dafür sind mehrere Gründe maßgebend. Einerseits ist es heute noch kaum möglich, spektrale Daten so übersichtlich und logisch zu kategorisieren und in einem Katalog zu ordnen, wie dies z. B. für Bruttoformeln oder Stoffklassen möglich ist. Im weiteren ist der Informationsgehalt eines einfachen spektroskopischen Merkmals meist so gering, daß nur Kombinationen aus mehreren einfachen Merkmalen brauchbare Suchkriterien bilden. Schließlich kann die Signifikanz eines spektralen Merkmals nur im Zusammenhang mit der postulierten Struktur abge-

schätzt werden, so daß die Auswahl der Merkmalkombinationen notwendigerweise voreingenommen ist.

Dementsprechend wird im allgemeinen vom Suchen von Referenzverbindungen mit ähnlichen spektralen Daten abgesehen, obwohl damit auf wertvolle Informationen verzichtet wird. Eine Lösung des gestellten Problems ist zwar meist trotzdem möglich, jedoch mit wesentlich größerem Aufwand verbunden.

Die heute zur Verfügung stehenden elektronischen Datenverarbeitungsanlagen erlauben es, spektrale Daten von Referenzverbindungen so zu speichern, daß das weitgehend automatische Finden von Referenzen mit gegebenen spektralen Eigenschaften mit vertretbarem Aufwand möglich wird. Voraussetzung dazu ist allerdings, daß die Verschlüsselung der spektralen Daten, die Organisation ihrer Speicherung sowie die Suchstrategie für diesen Anwendungsbereich optimalisiert wird. Eine Speicherung der gesamten Information, die im Kurvenzug eines analog registrierten Spektrums enthalten ist, kann aus wirtschaftlichen Erwägungen nicht in Betracht gezogen werden, wie dies die folgende Überschlagerrechnung zeigt.

Der Informationsgehalt eines Infrarotspektrums ergibt sich bei einer Auflösung der Wellenlängenskala von etwa 1 : 4000 und einem dynamischen Bereich der Intensitätswerte von 1 : 250 ($1 : 2^8$) zu rund 32000 Bit. Für die übrigen Spektren liegt der Speicherbedarf in der gleichen Größenordnung. Damit ergäbe sich für eine voll dokumentierte Referenzverbindung (4 Spektren, Struktur, Name, Quelle usw.) ein Speicherbedarf von mindestens 150000 Bit. Das neue Rechenzentrum der ETH Zürich verfügt mit einem Doppelsystem CDC 6400/6500 und einem Zusatzkernspeicher (Extended Core Storage) über etwa 630K Worte Kernspeicher zu 60 Bit. Darin ließen sich theoretisch also rund 250 Referenzverbindungen unterbringen, eine für die Praxis völlig ungenügende Zahl. Auch die Speicherung auf einem Plattenspeicher führt zu unsinnigen Größenordnungen. Am ehesten realisierbar wäre die Speicherung auf Magnetbändern. Dabei wird aber die Zugriffszeit und damit die Verarbeitungszeit untragbar groß. Dementsprechend muß aus praktischen Gründen zum vorstehenden auf einen beträchtlichen Teil der spektroskopischen Information verzichtet werden. Naturgemäß wird man vor allem jene Merkmale berücksichtigen, die chemisch und spektroskopisch die höchste Aussagekraft aufweisen. Welche Merkmale das sind, ist nicht in jedem Fall sofort offensichtlich. Ihre Auswahl soll vorteilhaft durch Personen mit großer Erfahrung in der Anwendung spektroskopischer Methoden in der Strukturaufklärung organischer Verbindungen erfolgen, da die Brauchbarkeit einer Referenzsammlung, wie sie hier vorgeschlagen wird, in hohem Maße von der richtigen Auswahl dieser Merkmale abhängt. Sie müssen so ausgewählt werden, daß zwar möglichst alle gewünschten Verbindungen, aber möglichst nur diese gefunden werden. Im weiteren sind auch Suchmethodik sowie die Art

² *Catalog of Mass Spectral Data, Manufacturing Chemists Association, Research Project, Carnegie Institute of Technology, Pittsburg (Pa.);*

Selected Mass Spectral Data, American Petroleum Institute, Research Project 44, Carnegie Institute of Technology, Pittsburg (PA); ASTM Index of Mass Spectral Data, Heyden & Son Limited, London;

A. CORNU und R. MASSOT, *Compilation of Mass Spectral Data*, Heyden & Son Limited, London 1966;

E. STENHAGEN und S. ABRAHAMSON, *Atlas of Mass Spectral Data*, John Wiley, New York 1969.

³ *NMR Spectra Catalog*, herausgegeben von Varian Associates, Palo Alto, Band 1 1962, Band 2 1963;

F. BOVEY, *NMR Data Tables of Organic Compounds*, Band 1, Interscience/Wiley & Sons, New York/London/Sydney 1968;

W. BRÜGEL, *Kernresonanzspektren und chemische Konstitution*, Band 1: *Die spektralen Kernresonanzparameter von Verbindungen mit analysiertem Spektrum*, Verlag Dr. Dietrich Steinkopff, Darmstadt 1968;

M. G. HOWELL, A. S. KENDE und J. S. WEBB, *Formula Index to NMR Literature Data*, Plenum Press, New York, Band 1 1965, Band 2 1966;

Stadler NMR Reference Spectra, Heyden & Son Ltd., Spectrum House, Alderton Crescent, London;

H. A. SZYMANSKY und R. F. YELIN, *NMR Band Handbook*, IFTI/Plenum Press, New York 1968.

⁴ *Dokumentation der Molekülspektroskopie (DMS-Kartei)*, Institut für Spektrochemie und angewandte Spektroskopie, Dortmund, Verlag Chemie GmbH, Weinheim;

American Society for Testing Materials, Wyandotte Punched Card Index to Infrared Absorption Spectra, ASTM, Philadelphia; *Sadtler Standard Spectra*, Sadtler Research Laboratories, Philadelphia.

der Codierung von grundsätzlicher Bedeutung, da der zeitliche Aufwand und damit auch die Wirtschaftlichkeit weitgehend durch sie bestimmt wird.

Ein sinnvolles System zum automatischen Auffinden von Referenzen für die Strukturaufklärung organischer Verbindungen mit Hilfe von spektroskopischen Methoden muß den im folgenden zusammengestellten und kurz diskutierten Anforderungen genügen.

Limitierte Zielsetzung

Die Maschine soll und kann den Menschen nicht ersetzen; sie kann ihn jedoch ergänzen, indem sie ihm jene Arbeiten abnimmt, die sie aufgrund ihres logischen Aufbaues besser und schneller erledigen kann. Die Maschine verfügt über eine große Speicherkapazität, sie ist zuverlässig, rasch und unvoreingenommen; andererseits verfügt sie nur über geringe Fähigkeiten zur Zeichenerkennung (Pattern Recognition) und über keinerlei Phantasie und Intuition. Dementsprechend wird keine eindeutige Identifikation erwartet, sondern lediglich Typenvorschläge und Modellverbindungen. Die letzten, komplexen Entscheidungen müssen dem Menschen vorbehalten bleiben.

Benützerfreundlichkeit

Datenein- und -ausgabe sollen in einer Form erfolgen, die dem Chemiker und seiner Denkweise angepaßt ist. Das Format der Input-Daten soll eine einfache Kontrolle auf Fehler erlauben, also möglichst einfach, klar und übersichtlich sein. Selbstverständlich werden auch die Resultate in übersichtlicher und gefälliger Form dargestellt.

Anpassungsfähigkeit

Das System soll ohne grundlegende Änderung den Bedürfnissen verschiedenster Benutzerkreise angepaßt werden können. Die zu einer solchen Anpassung notwendigen Informationen liefert es selbst, indem es beispielsweise angibt, welche Verbindungen bzw. Verbindungstypen am häufigsten verlangt wurden und welche Merkmale am häufigsten zur Entscheidung wesentlich beitragen.

Erweiterungsfähigkeit

Es soll ohne wesentlichen Aufwand möglich sein, neue Merkmale hinzuzufügen oder einzelne Merkmale bei Bedarf zusammenzulegen oder weiter aufzuspalten. Ebenso muß die Möglichkeit bestehen, Referenzverbindungen ohne weiteres hinzuzufügen oder wegzulassen.

Wirtschaftlichkeit

Um die Computerbelegung möglichst niedrig zu halten, muß die Speicherorganisation dem verwendeten Computertyp und die Suchstrategie dem jeweiligen Problem

angepaßt sein. Von den in der Datenbank vorhandenen relevanten Verbindungen sollen jeweils möglichst alle gefunden werden (Recall hoch). Gleichzeitig soll der Anteil der nicht relevanten Referenzen klein bleiben (Präzision hoch). Auf besonders aufwendige periphere Geräte (Display-Systeme usw.) soll verzichtet werden können.

Beschreibung des Systems

Ein System, das den zusammengestellten Anforderungen weitgehend entspricht, befindet sich am Laboratorium für Organische Chemie der Eidgenössischen Technischen Hochschule in Zürich im Aufbau. Sein prinzipieller Aufbau soll im folgenden eingehend beschrieben werden.

Aus der Menge aller spektroskopischen Merkmale werden jene ausgewählt, von denen man annimmt, daß sie für das automatische Auffinden von Referenzverbindungen besonders geeignet sind. Diese Auswahl ist notwendigerweise subjektiv und wird dementsprechend je nach den erhaltenen Resultaten modifiziert werden müssen. Jedem der ausgewählten Merkmale wird nun eine Variable zugeordnet, deren Wert nur dann von Null verschieden ist, wenn die betrachtete Verbindung das entsprechende Merkmal aufweist. Der Wert aller dieser Variablen für eine gegebene Verbindung nennt man die Signatur dieser Verbindungen; die Signaturen aller Referenzverbindungen sind zusammen mit weiteren nützlichen Zusatzinformationen (Struktur, Name, Bruttoformel, Herkunft, Reinheit usw.) in einer Datenbank gespeichert. Soll nun zu einer unbekanntem Verbindung mit gegebenen spektralen Eigenschaften eine Referenzverbindung gefunden werden, so werden die Signaturelemente der Unbekannten mit jenen in der Datenbank verglichen. Jene Verbindungen, deren Signaturen am besten mit der Signatur der unbekanntem Verbindung übereinstimmen, sind die gesuchten Referenzen.

Kritisch für das Funktionieren des Systems ist die Bewertung des Übereinstimmungsgrades zweier Signaturelemente. Die hierzu verwendeten Kriterien sind im folgenden beschrieben.

Der Vergleich zweier Signaturelemente kann zu vier verschiedenen Resultaten führen, deren Bedeutung für den Grad der Übereinstimmung unterschiedlich ist:

1. Unbekannte und Referenz weisen beide das entsprechende Merkmal auf: «positive» Übereinstimmung.
2. Bei beiden Verbindungen fehlt das entsprechende Merkmal: «negative» Übereinstimmung.
3. Die Referenz weist ein bei der Unbekannten nicht vorhandenes Merkmal auf.
4. Das bei der Unbekannten vorhandene Merkmal fehlt in der Referenz.

Diesen vier möglichen Resultaten eines elementaren Vergleichs werden nun vier verschiedene Werte zuge-

ordnet, die den Grad der Übereinstimmung quantitativ beschreiben. Dem Fall 1 («positive» Übereinstimmung) kommt dabei der höchste, den Fällen 2 und 3 geringere, dem Fall 4 schließlich der kleinste Wert zu. Im Prinzip könnte nun der Übereinstimmungsgrad zweier Signaturen durch die Summe der den Resultaten der elementaren Vergleiche zugeordneten Werte beschrieben werden. Dem höchsten Wert dieser Summe entspricht die beste Übereinstimmung.

Da aber die einzelnen Elemente einer Signatur spektralen Merkmalen von unterschiedlicher Bedeutung entsprechen, ist es naheliegend, dieses einfache Übereinstimmungsmaß durch Berücksichtigung dieser Unterschiede weiter zu verbessern.

Im hier beschriebenen System geschieht dies durch zweimalige Gewichtung. Ein erster Gewichtungsfaktor soll die spektrale Signifikanz des entsprechenden Merkmals berücksichtigen. Seine Werte werden aufgrund der Erfahrung bei der Interpretation von Spektren willkürlich festgelegt.

Ein zweiter Satz von Wägefaktoren berücksichtigt den statistischen Informationsgehalt der Datenbank für das betreffende Merkmal, indem das Resultat eines Elementarvergleichs um so höher bewertet wird, je seltener das entsprechende Merkmal in der Datenbank auftritt. Die Werte für diese Gewichtungsfaktoren werden vom System aufgrund der jeweiligen Zusammensetzung der Datenbank selbst errechnet und sind dem Benutzer nicht zugänglich. In der Praxis ist allerdings die saubere Trennung von spektraler und statistischer Signifikanz einstweilen kaum möglich. Die Auftrennung der Bewertungsgrößen in je einen vom Benutzer und vom System unterhaltenen Anteil erscheint uns jedoch aus praktischen Gründen erwünscht.

Würden in dieser Weise alle einander entsprechenden Merkmale verglichen, so ergäbe dies schon bei einem bescheidenen Umfang der Datenbank unverantwortbar hohe Rechenzeiten. Dementsprechend ist auch die Suchstrategie so ausgelegt, daß mit minimalem Aufwand ein brauchbares Resultat erreicht wird.

Aus der Signatur der unbekannteren Verbindung und den beiden Gewichtungssätzen kann das System selbständig jene Reihenfolge der einzelnen Elemente der Signatur erstellen, die im Übereinstimmungsfall den raschesten Anstieg der Übereinstimmungssumme ergibt. Der Vergleich der einzelnen Merkmale erfolgt nun in dieser Reihenfolge. Nach jedem Elementarvergleich wird der aktuelle Stand der Übereinstimmungssumme mit dem zu diesem Zeitpunkt im Maximum erreichbaren Wert verglichen. Bleibt der tatsächlich erreichte Wert zu weit hinter dem Maximum zurück, so wird der Vergleich der beiden Signaturen als erfolglos abgebrochen. Damit kann erreicht werden, daß eklatante Unterschiede schon sehr früh erkannt werden und daß keine Rechenzeit an hoffnungslose Fälle vergeudet wird.

Die für das Durchsuchen der Datenbank notwendige Rechenzeit wird in hohem Maße auch durch die Größe

der Datenbank bestimmt; diese ist im wesentlichen durch die Anzahl der Signaturen und durch deren Länge gegeben. Dies zwingt zu einer Limitierung der Anzahl der in der Sammlung aufzunehmenden Verbindungen. Will man dennoch brauchbare Resultate erzielen, so muß der Auswahl dieser Verbindungen besondere Beachtung geschenkt werden.

Im Idealfall soll die Zusammensetzung der Datenbank bezüglich der darin vertretenen Verbindungstypen den Interessenbereich der Benutzer möglichst getreu widerspiegeln. Jene Stoffklassen, die von den Benutzern intensiv bearbeitet werden, sollen mit einer großen Anzahl auch spezieller Fälle in der Sammlung vertreten sein. Von den seltener vorkommenden Stoffklassen genügen einige typische Beispiele.

Das System kann nun selbständig die notwendigen Informationen darüber liefern, ob die getroffene Auswahl den Anforderungen aus dem Benutzerkreis entspricht. Dazu wird eine Statistik darüber erstellt, welche Verbindungen besonders häufig und welche selten oder nie als Referenzen ausgewählt wurden. Solche selten oder nie ausgewählte Verbindungen sind entweder falsch codiert oder gehören zu seltenen Stoffklassen. Sie können bis auf einige typische Vertreter weggelassen werden, ohne daß die Leistungsfähigkeit des Systems leidet. Sie sollen durch Vertreter jener Stoffklassen ersetzt werden, die überdurchschnittlich oft als Referenzen ausgewählt wurden.

Durch diese periodische Korrektur der Zusammensetzung der Datenbank ist eine laufende Anpassung auch an wechselnde Bedürfnisse der Benutzer gewährleistet, ohne daß die Größe der Datenbank und damit der Rechenaufwand ohne Grenzen ansteigt.

Die optimale Größe der Datenbank hängt von der Breite des Interessenbereichs der Benutzer ab. Aufgrund der Erfahrungen mit anderen Datensammlungen⁴ glauben wir, daß auch bei einem universellen Interessenbereich nicht mehr als 20000 Verbindungen notwendig sein werden. Bei sehr schmalen Interessenbereichen kann auch die Erstellung mehrerer, sich teilweise überlappenden Datenbanken in Betracht gezogen werden.

Ob aus der Fülle aller denkbaren spektralen Merkmale eine gute Auswahl getroffen wurde, läßt sich dadurch überprüfen, daß man durch das System eine Statistik darüber führen läßt, wie oft jedes Merkmal am Entscheidungsprozeß beteiligt war. Von Zeit zu Zeit wird diese Statistik abgefragt. Je nach dem Resultat wird man dann einzelne Merkmale feiner aufspalten bzw. mehrere Merkmale in ein einziges zusammenfassen. Gegebenenfalls können auch die Werte der die spektrale Bedeutung berücksichtigenden Wägefaktoren geändert werden.

Die mit Hilfe der beiden erwähnten Statistiken erhaltenen Bewertungen für die Auswahl der in die Sammlung aufgenommenen Verbindungen einerseits und der zur Beschreibung ihrer spektralen Eigenschaften verwendeten Merkmale andererseits sind aber nicht unabhängig voneinander, sondern beeinflussen sich gegen-

seitig in nicht leicht durchschaubarer Weise. Wir glauben jedoch, daß auch hier die Auftrennung in diese beiden Anteile praktische Vorteile bietet.

Resultate

Erste praktische Versuche mit dem am organisch-chemischen Laboratorium der ETH Zürich im Aufbau befindlichen System haben ermutigende Resultate gezeigt.

In der gegenwärtigen ersten Versuchsphase wurde vor allem auf Übersichtlichkeit und Transparenz der Programmierung Wert gelegt. Daher sind die Programme größtenteils in Fortran IV geschrieben worden, was naturgemäß zu einer geringen Effizienz und damit zu relativ hohen Rechenzeiten führt. In einer späteren Phase ist es vorgesehen, die ausgetesteten Programme zu optimieren, indem wesentliche Teile in eine niederere Programmiersprache übertragen werden. Dementsprechend wird die Effizienz noch weiter ansteigen. Alle im folgenden gemachten Angaben über Rechenzeiten beziehen sich auf die Ausführung der Programme auf einer Rechenanlage CDC 6500.

Zum jetzigen Zeitpunkt enthält die Sammlung erst etwa 900 Verbindungen*, von denen nur die Protonenresonanzspektren codiert sind. Als zusätzliche Information sind noch Quelle, Summenformel sowie die nach WISWESSER⁵ codierte Struktur vorhanden. Die Codierung der Massenspektren sowie der IR-Spektren ist im Gange. Der Aufbau der Datenbank wurde mit den Protonenresonanzspektren begonnen, da hier der Feinstruktur der Signale wegen mit den größten Codierungsschwierigkeiten zu rechnen war.

Die Vorversuche haben jedoch gezeigt, daß es vollauf genügt, zwischen scharfen Singletts einerseits und mehrfach aufgespaltenen oder breiten Signalen andererseits zu unterscheiden.

Eine feinere Unterteilung der Signaltypen (Singlett, Dublett, Triplet, Quartett, höheres Multiplett, breites Signal) führt zu keiner wesentlichen Änderung in Präzision und Recall. Die Rechenzeit für das Vergleichen der Signaturen stieg hingegen von etwa 21,5 sec auf etwa 50,5 sec pro 10000 Spektren an. Zur Codierung der Signallage ist der Bereich angegeben worden, in den der Schwerpunkt des Signals fällt. Die Grenzen dieser Bereiche wurden empirisch so festgelegt, daß Recall und Präzision möglichst hoch lagen. Die Bereichsgrenzen sowie die Belegung der einzelnen Bereiche sind in Tabelle 1 zusammengestellt. Die verwendeten Gewichte zur Bewertung der spektralen Signifikanz sind in Tabelle 2 wiedergegeben. Es handelt sich dabei um vorläufige Werte, die noch weiter verbessert werden sollen.

* Die Spektren entstammen dem *NMR Spectra Catalog*³ von BHACCA *et al.* sowie der Sammlung von Übungsbeispielen zur Vorlesung über Instrumentalanalyse am organisch-chemischen Laboratorium der ETH Zürich.

⁵ E. G. SMITH, *The Wiswesser Line-formula Chemical Notation*, McGraw-Hill, 1968.

Tabelle 1. Codierung von NMR-Spektren. Bereichsgrenzen und Belegung der Bereiche

Such 24 Version Juni 1970

Nr.	Statistik Bereich ppm	Versuch 7/4 Anzahl Merkmale		
		S	M	Signal
1	— 1,00 — 1,05	35	123	149
2	1,00 — 1,55	88	240	303
3	1,50 — 2,05	80	210	272
4	2,00 — 2,55	172	199	343
5	2,50 — 3,05	72	162	223
6	3,00 — 3,55	66	129	192
7	3,50 — 4,05	139	163	290
8	4,00 — 5,05	68	214	273
9	5,00 — 6,05	59	179	230
10	6,00 — 6,95	63	239	289
11	6,90 — 7,65	77	354	423
12	7,60 — 8,05	21	163	183
13	8,00 — 9,05	27	101	123
14	9,00 — 10,05	19	28	47
15	10,00 — 12,05	24	19	43
16	12,00 — 16,00	11	11	22

S: Singlett

M: Multiplett

Signal: Singlett und/oder Multiplett

Durch vorzeitigen Abbruch des Vergleichsprozesses bei schlechter Übereinstimmung zweier Signaturen kann die Effizienz des Systems stark verbessert werden. Der Abbruch erfolgt, wie weiter oben beschrieben, wenn die Übereinstimmungssumme unter einem gewissen Prozentsatz des bei völliger Übereinstimmung erreichbaren Wertes sinkt; die ersten paar Vergleiche (im hier beschriebenen System z. B. die ersten 5) werden aber in jedem Fall durchgeführt. Eine Erhöhung des Schwellenwertes bewirkt eine deutliche Verminderung der benötigten Rechenzeit. Die Zusammensetzung und die Rangfolge der ersten 20 gefundenen Referenzen änderte sich dabei nicht, so lange die Schwelle weniger als 60% des jeweils maximal erreichbaren Wertes betrug. Den Zusammenhang zwischen Rechenzeit, Anzahl wieder aufgefundener Referenzverbindungen, Anteil vorzeitig abgebrochener Vergleiche einerseits und der Schwellenhöhe andererseits zeigt Abb. 1.

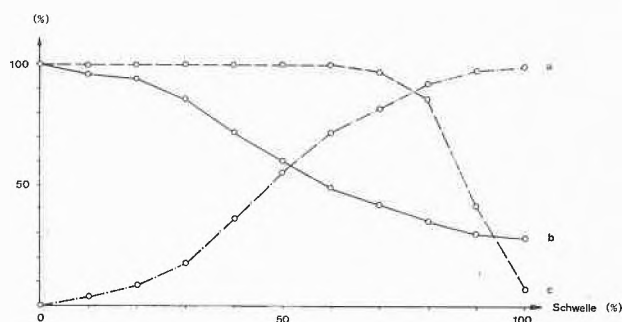


Abb. 1. Einfluß der Schwellenhöhe für vorzeitigen Abbruch des Vergleichsvorgangs. a) Anteil vorzeitig abgebrochener Vergleiche, b) Rechenzeit in Prozent des Maximalwertes, c) Wiederaufgefundene Referenzverbindungen in Prozent des Maximalwertes

Tabelle 2. Codierung von NMR-Spektren. Gewichtung der spektralen Signifikanz

Such 24 Version Juni 1970

Versuch 7/4

Nr.	Bereich ppm	Positives Übereinstimmen			Negatives Übereinstimmen			Merkmal nur bei Referenz		
		S	M	Signal	S	M	Signal	S	M	Signal
1	— 1,00 — 1,05	32	19	140	27	20	90	12	8	40
2	1,00 — 1,55	32	19	90	33	28	120	12	8	40
3	1,50 — 2,05	32	19	140	27	20	90	12	8	40
4	2,00 — 2,55	51	38	225	36	29	115	21	17	65
5	2,50 — 3,05	51	38	225	36	29	115	21	17	65
6	3,00 — 3,55	51	38	225	36	29	115	21	17	65
7	3,50 — 4,05	51	38	225	36	29	115	21	17	65
8	4,00 — 5,05	51	38	225	36	29	115	21	17	65
9	5,00 — 6,05	51	38	225	36	29	115	21	17	65
10	6,00 — 6,95	32	19	140	33	28	120	12	8	40
11	6,90 — 7,65	32	19	90	27	20	90	12	8	40
12	7,60 — 8,05	32	19	140	27	20	90	12	8	40
13	8,00 — 9,05	32	19	140	27	20	90	12	8	40
14	9,00 — 10,05	32	19	140	27	20	90	12	8	40
15	10,00 — 12,05	51	38	265	36	29	115	21	17	65
16	12,00 — 16,00	51	38	265	36	29	115	21	17	65

S: Singlett

M: Multipllett

Signal: Singlett und/oder Multipllett

Für den Fall «Merkmal fehlt bei Referenz» sind alle Gewichte = 0

Ein Beispiel für den Output des Systems ist in Tabelle 3 wiedergegeben. Sie zeigt die vom System aufgrund der NMR-Spektren vorgeschlagenen Referenzverbindungen zu 3-Brompropen.

Über die Resultate, die mit einer wesentlich erweiterten Datenbasis mit teilweise unvollständigen Signaturen erzielt werden können, soll zu einem späteren Zeitpunkt berichtet werden.

Tabelle 3. Beispiel für Output: Referenzverbindungen zu 3-Brompropen

Versuch 7/4

Gesuchte Verbindung Var24

Maximales S 2696

Name	Rel.S	Summenformel	Wiswesser-Notationen
Var	24	100,00	C 3 H 5 BR1 E2U1 -V
S	55	100,00	C 5 H 4 O 3 T5VOVY EHJ DU1 -V
Var	64	98,16	C 4 H 6 O 2 1U1V01 -V
Var	110	97,84	C 5 H 8 O 1 1U201U1 -V -V
Var	34	96,00	C 3 H 6 O 1 Q2U1 -V
Var	105	96,00	C 5 H 7 N 1 O 2 T5MV EHJ D01
Var	113	96,00	C 5 H 8 O 2 1UYV01 -V
Var	540	95,93	C 9 H13 N 3 O 5 T6NVNJ DZ A- BT50TJ OQ DQ E1Q -A+CD -B+BE
Var	250	95,68	C10 H 9 N 1 S 2 T56 BN DSJ CS2U1 -V
Var	537	95,68	C 9 H12 N 2 O 6 T6NVMVJ A- BT50TJ CQ DQ E1Q -A+C -B+BDE
S	36	95,68	C11 H14 O 2 1U2R C01 D01 -V
Var	2	94,47	C 2 H 3 CL 3 GYG1G
Var	134	94,47	C 6 H10 O 1 1U202U1 -V -V
Var	260	93,84	C10 H12 O 2 1U2R DQ C01 -V
Var	328	93,84	C17 H19 N 1 O 3 T56 +6 D0 CHJ G1U2U1V- AT6NTJ -T -T
Var	678	93,84	C19 H22 O 4 1U1YR BC1+ R B01 D01 E01
Var	458	93,73	C 6 H 8 N 2 ZR CZ
Var	170	93,63	C 7 H 9 N 1 O 1 T5NJ A BVH E
Var	531	93,63	C 9 H10 O 2 L45 EV BU FUTJ B D01
Var	595	93,53	C12 H12 O 5 T66 BVO GV JOT+J D EQ H1U2 -T

Name: VAR: Varian Katalog^a

S: Sammlung von Übungsbeispielen zur Instrumentalanalyse

Rel.S: Übereinstimmungssumme, in Prozent des maximal möglichen Wertes