

# Online-Datenbank-Recherchen an der Hochschule

## Erfahrungen in der Abteilung für Chemie der ETH Zürich

Engelbert Zass\*

Die Nutzung von Online-Literatur- und -Struktur-Recherchen ist für Chemiker an den Hochschulen im Gegensatz zur chemischen Industrie immer noch zu wenig verbreitet. Dies beruht unter anderem auf einer offenbar falschen Einschätzung der Leistungsfähigkeit und des Kosten/Nutzen-Verhältnisses dieses für Forschung und Lehre wichtigen modernen Hilfsmittels. An der Abteilung für Chemie der ETH Zürich hat man seit Februar 1979 – als eines der ersten Hochschul-Institute in Europa – Zugriff auf eine grosse Zahl öffentlich verfügbarer Datenbanken, und inzwischen sind Online-Recherchen dort allgemein als unverzichtbares Hilfsmittel anerkannt. Diese Entwicklung hat auch zu Lehrveranstaltungen in Chemie-Information unter Einschluss der «klassischen», trotz Datenbanken immer noch notwendigen gedruckten Quellen geführt. Die Erfahrungen und Massnahmen an der ETH Zürich sollten zum grössten Teil auch auf andere Hochschulen übertragbar und anwendbar sein.

### 1. Einleitung

Die Informationsbedürfnisse der Forschungschemiker in Industrie und Hochschule sind sehr ähnlich, die Methoden zu ihrer Befriedigung unterscheiden sich aber deutlich: In der *Industrie* gibt es eigene Dokumentationsabteilungen, in denen sich Spezialisten um die Informationsversorgung kümmern, während an der *Hochschule* die meisten Fragen vom Fragesteller alleine, allenfalls mit Hilfe eines Bibliothekars, beantwortet werden müssen.

Das in der Industrie offenbar besser entwickelte Informations- und Kostenbewusstsein führte schon früh zur Entwicklung von internen mechanisierten Dokumentationssystemen und aufgrund der so gesammelten Erfahrungen auch zur schnellen Übernahme der Benutzung von kommerziellen Online-Suchmöglichkeiten, als diese ab Mitte der siebziger Jahre verfügbar wurden. Als Konsequenz dieser Entwicklung ist festzustellen, dass nun schon viele Chemiker über die in den La-

boratorien vorhandenen Terminals/Mikrocomputer ihre Fragen *online* zum Teil selbst, d. h. nicht mehr über die Dokumentationsabteilung beantworten<sup>[1]</sup>.

An den Hochschulen<sup>[2]</sup> sucht man Informationen mangels Zugriff auf Online-Datenbanken – dieser Mangel ist durch finanzielle Beschränkungen, aber auch durch Unkenntnis über die Leistungsfähigkeit von Online-Recherchen oder sogar eine ablehnende Haltung gegenüber dieser neuen Methode bedingt – immer noch von Hand in *gedruckter* Sekundärliteratur oder Handbüchern, oder informiert sich durch informelle Kommunikation mit Kollegen. Die Erfahrung zeigt, dass man sich der Grenzen dieses Verfahrens meist nicht bewusst ist oder nicht realisiert, dass sich mit modernen Methoden viele Fragen leichter, schneller (und damit schlussendlich billiger, denn auch an der Hochschule ist Zeit Geld, gerade wenn Sach- und Personalmittel so knapp sind wie heute) oder gar überhaupt erst so beantworten lassen.

### 2. Beispiele für Online-Recherchen

Zwei Beispiele aus unserer Praxis sollen noch einmal<sup>[3]</sup> die Möglichkeiten von Online-Recherchen zeigen, die für uns inzwischen vor allem in der Forschung zu einem unverzichtbaren Werkzeug geworden sind.



Engelbert Zass, Dr. sc. nat., ist Oberassistent am Laboratorium für Organische Chemie der ETH Zürich und befasst sich dort seit 1979 mit der Durchführung und Betreuung von Online-Recherchen; seit 1984 hat er einen Lehrauftrag für Benützung der chemischen Literatur (zu Person und Arbeitsgebiet vgl. auch *Chimia* 40 (1986) 38). Vor dem Hintergrund des in letzter Zeit stark gestiegenen Interesses an Online-Recherchen in der Chemie an Hochschulen – dies manifestiert sich unter anderem durch einschlägige Veranstaltungen im Rahmen des «3e cycle en chimie» an den Universitäten Bern, Genève und Lausanne im Herbst 1986 – berichtet er von seinen persönlichen Erfahrungen auf diesem Gebiet.

Für eine Literatur-Recherche zur Frage «Reduktion von Enaminoketonen» müsste man im gedruckten *Chemical Abstracts General Subject Index* der Zeitspanne 1967–1984 in drei Sammel- und sechs Band-Registern unter so unterschiedlichen Eintragungen wie «Reduction, of enamino ...», «Reduction, electrochemical, of enamino ...», «Ketones, reactions, enamino ...», «Enamines, oxo ...», «Amines, reactions, enamino ...», «Amines, reactions, oxo ...» Tausende von Register-eintragungen durchsuchen, da *Chemical Abstracts (CA)* der Verbindungsklasse «Enaminoketone» keinen eigenen Deskriptor («Index Heading») zuordnet. Auf diese Weise, die natürlich eingehende Kenntnisse über die von *CA* verwendete Terminologie und Indexierungspolitik voraussetzt, hätte man in der angegebenen Zeitspanne 7 Zitate gefunden. Eine entsprechende Online-Recherche z. B. im *CA File* des Systems *STN/CAS ONLINE* ergab aber mit dem Suchprofil «(redn OR reduc?)(L)(enamin!keton? OR enamino? OR enamini!(L)keton?)» 12 Zitate, die alle relevant waren. Bei der maschinellen Suche findet man mit einem solchen oft unmittelbar die Begriffe der Fragestellung benutzenden Suchprofil unter Berücksichtigung verschiedener Synonyme (statt des oft unzureichenden Vokabulars der *CA Index Headings*), Schreibweisen und *CA*-Abkürzungen (Eingabe mit «Masken» wie «?» = «mehrere beliebige Zeichen» für Singular/Plural, Substantiv/Adjektiv; «!» = «ein beliebiges Zeichen» für enamino/enamine) wesentlich mehr Zitate, weil hier *alle* Begriffe im Zusammenhang einer Register-eintragung (= Verknüpfung mit *(L)-Operator*) unabhängig von ihrer lexikalischen Anordnung, die ja die Suche in gedruckten Registern begrenzt, gesucht

\* Korrespondenz: Dr. E. Zass  
Laboratorium für Organische Chemie  
Eidgenössische Technische Hochschule Zürich  
ETH-Zentrum, Universitätstrasse 16  
CH-8092 Zürich

werden. Ein weiterer wichtiger Vorzug von Online-Recherchen ist, dass man mit Hilfe von Operatoren viele Begriffe verknüpfen und damit im logischen oder textlichen Zusammenhang suchen kann. Oft sind ja in noch grösserem Mass als bei obigem Beispiel zur Beschreibung einer Frage mehrere Konzepte miteinander zu verknüpfen, von denen jedes durch eine Reihe von Synonymen und verschiedenen Schreibweisen ausgedrückt werden muss.

Maschinelle Methoden bieten also bei *Literatur-Recherchen* mit Schlagwörtern viele Vorteile und mehr Komfort gegenüber Suchen in gedruckter Sekundärliteratur<sup>[4]</sup>. Eigentlicher Kernpunkt der Chemie-Information und vieler Fragestellungen in der Chemie sind aber nicht Begriffe, sondern die im täglichen Gebrauch des Chemikers durch universell verständliche Strukturformeln dargestellten chemischen *Verbindungen*. Einzelne, vollständig definierte Verbindungen waren schon immer in den bekanntesten gedruckten Registern von z. B. *CA* und «*Beilstein*» im Prinzip suchbar, wenn auch wegen der Komplexität und der Änderungen der verwendeten Nomenklatur(en) oft nur mühsam und über den Umweg der (Summen)formelregister. *Substruktur-Recherchen*, d. h. die Suche nach einer Gruppe von Verbindungen, die über eine gemeinsame, definierte Teilstruktur hinaus noch beliebige andere, nicht festgelegte Strukturelemente enthalten, sind hingegen erst durch unmittelbar auf der Struktur (Topologie) statt der Nomenklatur basierende (Sub)struktur-Suchsysteme<sup>[5]</sup> wie *DARC*<sup>[6]</sup> oder *CAS ONLINE*<sup>[7]</sup> möglich geworden. Voraussetzungen dafür waren ausser den entsprechenden Suchprogrammen umfangreiche, maschinenlesbare Strukturspeicher wie das *CAS Registry System*<sup>[8]</sup>, der *ISI Index Chemicus*<sup>[9]</sup>, oder *SANSS* (structure and nomenclature search system) im Datenbanksystem *CIS*<sup>[10]</sup>. Durch direkte Eingabe einer Strukturformel über ein (bevorzugt, aber nicht notwendigerweise grafikfähiges) Terminal lassen sich z. B. Fragen nach Literatur über die Biosynthese von allen Naturstoffen mit der in Fig. 1 angegebenen Substruktur oder nach spektroskopischen Daten solcher Verbindungen relativ leicht und schnell beantworten. Eine durchschnittliche Substruktur-Recherche im *CA Registry File* von *STN/CAS ONLINE* dauert ca. 15–20 Minuten, davon entfallen auf den eigentlichen Suchprozess in über acht Millionen Verbindungen nur etwa 5 Minuten; die Kosten dafür (ohne Kommunikationsgebühren) betragen etwa 250 DM (für Hochschulen im Rahmen des «Academic Program» nur 10% dieses Preises!<sup>[11]</sup>), wobei die feste (zeitunabhängige) Gebühr pro Suche fast 200 DM beträgt.

Der dargebotene Strauss von Suchmöglichkeiten enthält jedoch leider einige «Dornen». Jedes der in Fig. 1 angegebenen vier Datenbank-Systeme hat nämlich seine eigene, in Einzelheiten deutlich unterschiedliche «Retrievalsprache» für Ein-

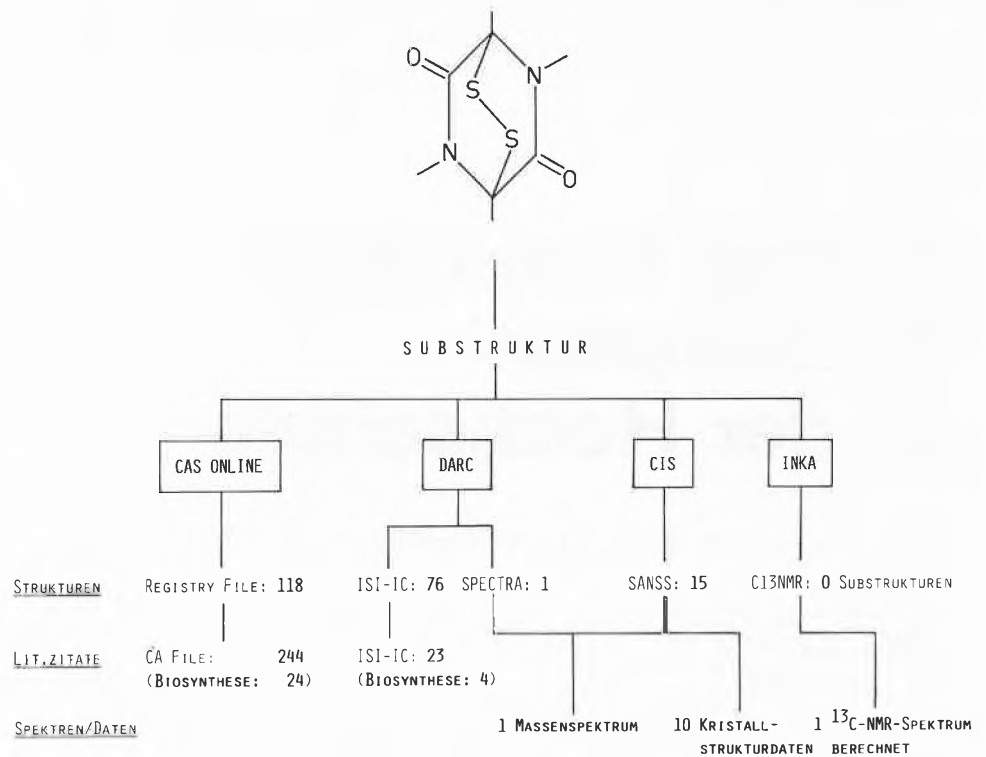


Fig. 1. Substruktur-Recherchen nach Verbindungen, Literatur (gesamte und solche zur Biosynthese) und Daten in den Systemen *STN/CAS ONLINE* (*CA*), *DARC/Télésystèmes Questel*<sup>[5]</sup> (*ISI Index Chemicus*<sup>[9]</sup>), *Massenspektren des US National Bureau of Standards*, *CIS*<sup>[10]</sup> (*Massenspektren des US National Bureau of Standards*, *Cambridge Crystallographic Database*<sup>[12]</sup>) und *INKA* (*C13NMR*<sup>[13]</sup>).

gabe und Suche. Ausser diesen Systemkenntnissen muss der Benutzer z. B. noch wissen und bei der Struktureingabe berücksichtigen, dass von *CA* die Bindungen der Lactam-Gruppe in der Teilstruktur je nach Substitution am Stickstoff-Atom (formal) unterschiedlich definiert werden: unsubstituiert = tautomeres System mit delokalisierten («normalized») Bindungen, substituiert = lokalisierte («exact») Einfach- und Doppelbindungen.

Wenn auch manche Fragen nur durch Online-Recherchen überhaupt oder mit vertretbarem Aufwand, vor allem an Zeit, beantwortet werden können, so sind doch die «klassischen», gedruckten Referateorgane und Handbücher keinesfalls einfach durch ein Terminal mit Zugriff auf entsprechende Datenbanken zu ersetzen und dadurch überflüssig geworden. Gerade bei orientierenden Suchen, z. B. nach Reaktionen, sind Handbücher wie «*Houben-Weyl*» oder gute Monographien unersetzlich. Zur laufenden Literaturüberwachung etwa auf dem Gebiet der präbiotischen Chemie werden neben der Lektüre («browsing») einschlägiger Spezialzeitschriften wie *Journal of Molecular Evolution*, *Origins of Life*, *BioSystems* auch die Schlagwortregister der neuesten gedruckten *CA*-Hefte unter «evolution» sowie «prebiotic» durchgeschaut und sofort die «Spreu vom Weizen» getrennt – ein automatisches Online-Suchprofil (SDI = selective dissemination of information) würde hier zuviel Ballast liefern. Ein solches Profil dient jedoch dazu, gezielt die Literatur über uns in

diesem Zusammenhang interessierende Verbindungen und auch Publikationen von anderen auf dem Gebiet tätigen Wissenschaftlern zweiwöchentlich automatisch auszudrucken. Um bisher noch unbekannte Verbindungen zu erfassen, werden periodisch im *CA Registry File* entsprechende Substruktur-Recherchen durchgeführt. Bei vielen Informationsproblemen lautet die Frage nach dem Vorgehen nicht «online oder gedruckt suchen?», sondern man muss oft alle verfügbaren Methoden und Quellen benutzen, um ans Ziel zu kommen.

Für die laufende Sammlung von Information über Methanbakterien ist ein automatisches Suchprofil die Methode der Wahl, weil hier wegen der Breite des Gebiets von der Biotechnologie über Molekularbiologie, Mikrobiologie bis hin zur präparativen organischen Chemie eine Verfolgung der neuesten Entwicklungen über die Primärliteratur praktisch nicht mehr möglich ist, da sich die Beiträge auf zu viele Zeitschriften verteilen. Aus dem gleichen Grund ist aber auch die Abfrage von *CA* allein nicht hinreichend, sondern durch ein zweites automatisiertes Suchprofil in *BIOSIS* (*Biological Abstracts* online) zu ergänzen<sup>[14]</sup>.

### 3. Organisation von Online-Recherchen

Unser Datenbankanschluss wurde in der Bibliothek der Abteilung für Chemie eingerichtet. Dies hat nicht nur den Vorteil

einer Betreuung, Verwaltung und Kontrolle durch den Bibliothekar, sondern vereint räumlich maschinelle Methoden mit gedruckter Sekundärliteratur und deren auch zur Vorbereitung von Online-Recherchen benötigten Suchhilfen (*CA Index Guide*, *BIOSIS Search Guide*, etc.). Da das Ergebnis der meisten Online-Recherchen nicht aus der erwünschten Information selbst, sondern aus einem Hinweis auf die Primärliteratur besteht, ist auch deswegen die Nähe zur Bibliothek mit ihren Beständen an etwa 400 laufenden Periodika von Vorteil. Der im Frühjahr 1986 bezogene Erweiterungsbau der Chemie-Bibliothek hat spezielle Räume für die Aufstellung der Terminals und Vorbesprechungen zu den Recherchen. Dies steht im Zusammenhang mit einer Neudefinition der Rolle unserer Bibliothek als Informationszentrum, die über die traditionelle Funktion als Archiv für gedruckte Literatur hinausreicht<sup>[15]</sup>.

Wir begannen Anfang 1979 mit dem System *DIALOG*, das auch heute noch neben den für Chemiker wichtigen *CA* die grösste Zahl von Online-Datenbanken eines einzelnen Hosts (Datenbank-Anbieter) verfügbar hat. Im folgenden Jahr kam *SDC ORBIT* hinzu, um den dort exklusiv verfügbaren *Chemical Reaction Documentation Service (CRDS)*<sup>[16]</sup> von *Derwent*, die Online-Version des «*Theilheimer*» und des *J.Synth.Methods*, abfragen zu können, und für spektroskopische sowie toxikologische Daten das System *CIS*<sup>[10]</sup>. Zugriff auf *CAS ONLINE* und damit zum Strukturspeicher (*Registry File*) von *CA* erhielten wir 1981, zu *DARC/Télesystèmes Questel* (über die ETH-Hauptbibliothek, die schon 1978 eine Informationsvermittlungsstelle für Online-Recherchen eingerichtet hatte) 1982. Der Host *DATA-STAR* folgte im gleichen Jahr, so dass wir für Literatur-Recherchen in *CA* nicht mehr auf amerikanische Hosts mit den damals noch relativ teuren Kommunikationsgebühren (vgl. Abschnitt 5) angewiesen waren. Seit 1984 haben wir, ebenfalls via ETH-Hauptbibliothek, auch Zugriff auf die *C13NMR-Datenbank*<sup>[13]</sup> von *INKA* in Karlsruhe.

Als Terminals stehen uns für Recherchen in der Chemie-Bibliothek zur Zeit ein (Druckerterminal) *Digital DecWriterIII* und das grafikfähige *HP 2647A* (mit Matrixdrucker und Graphik-Tablett) zur Verfügung. Das für Substruktur-Recherchen 1981 beschaffte *HP-Terminal* wird ausser dafür ungeachtet seiner vielfältigen Möglichkeiten – z. B. Aufzeichnen der Recherche auf Magnetband und nachträgliches Redigieren – nur von sehr wenigen verwendet, da die meisten Benutzer für Literatur-Recherchen das einfachere zu bedienende *DEC-Terminal* vorziehen. Dies zeigt, dass leistungsfähigere, aber anspruchsvollere Systeme durchaus nicht selbstverständlich den Bedürfnissen besser entsprechen müssen. Beim unmittelbar bevorstehenden Umsteigen auf *IBM PC AT 03* als intelligente, grafikfähige Terminals werden wir diese Erfahrung beherzigen, vor allem bei

der Auswahl der notwendigen Software, die von eher einfachen (nicht grafikfähigen) Kommunikationsprogrammen<sup>[17a, 18a]</sup> über Graphik-Terminal-Emulationen wie *EMU-Tek*, *PC-Plot* bis hin zu speziell für Online-Recherchen entwickelter Software<sup>[18b, 19a, 20a]</sup> reicht.

Mit dieser Umrüstung folgen auch wir dem allgemeinen Trend, für Online-Recherchen statt Terminals die leistungsfähigeren oder relativ zu Graphik-Terminals auch billigeren Mikrocomputer einzusetzen<sup>[17]</sup>. Damit bietet sich die Möglichkeit zu «uploading» (lokale Vorbereitung und Eingabe von Fragestellungen, die dann nach Anschluss an den Host mit hoher Geschwindigkeit gesendet werden können) und zu «downloading»<sup>[20]</sup> (Zitate oder Strukturformeln werden nicht unmittelbar ausgedruckt, sondern elektronisch gespeichert und dann editiert und/oder in eine lokale Datenbank geladen).

«Uploading» spart Kosten, da nicht mehr so viel teure Anschlusszeit durch langsame Eingabe verlorengeht, problematisch ist aber, dass bei zuviel Vorbereitung der interaktive Charakter, das «Frage-und-Antwort-Spiel» als wesentliches Element der Online-Recherche, zu stark eingeschränkt werden kann.

«Downloading» erfordert fast immer die Zahlung zusätzlicher Gebühren, will man nicht die Vertragsbedingungen mit dem Host bezüglich Urheberrecht verletzen. Den Vorteilen des Aufbaus einer persönlichen Datenbank ohne manuelle Eingabe stehen Probleme gegenüber, die nach meiner Meinung oft unterschätzt werden: Solche lokalen Literaturdateien erfordern ständige Wartung und Aktualisierung, damit sie brauchbar bleiben. Eine derartige Datenbank bietet, da heute meines Wissens für Graphik (Formeln!) noch kaum wirklich zufriedenstellende Software existiert, weder die Flexibilität einer guten Handkartei noch (wahrscheinlich) ihre ständige Verfügbarkeit, und andererseits derzeit nicht die Geschwindigkeit und Suchmöglichkeiten der Online-Systeme, vor allem bei grösseren Dateien. Die Nützlichkeit einer sorgfältig kontrollierten und selektionierten Computer-Datei als «authority file» z. B. für das Publizieren mit einem Textverarbeitungs-System soll hier natürlich nicht bestritten werden. Eine Diskette mit schnell geladenen Zitaten löst aber noch keine Informationsprobleme, denn sie speichert Bits und nicht a priori «Information»; die bei einer Handkartei implizierte Informationsverarbeitung und Selektion übernimmt der Computer nicht!

#### 4. Zugang zu Online-Recherchen

Im Prinzip kann jeder Dozent, Assistent oder Doktorand in der Abteilung für Chemie, der diese Suchmethode erlernen will oder sich über entsprechende Kenntnisse ausweisen kann, in der Bibliothek selbst in Datenbanken recherchieren, wobei der Bibliothekar für die Abwicklung der Suche –

Festlegung eines Termins, Herstellung der Verbindung zum Datenbanksystem – verantwortlich ist. Er übernimmt auch in den meisten Fällen die eventuell notwendige Hilfestellung am Terminal oder Anleitung zur Vorbereitung einer Recherche.

In der Literatur wird seit langem diskutiert, ob Online-Recherchen eher von Informations-Spezialisten oder vom Fragesteller selbst («end-user») durchgeführt werden sollen, wobei sich mit der Zeit das Gewicht zu letzterem hin verschoben hat<sup>[21]</sup>. Die Eingabe/Ausgabe von Strukturformeln bleibt natürlich den Chemikern vorbehalten. Auch bei Schlagwort-Recherchen wäre es wegen der komplexen Fachterminologie, die ja für die Erstellung des Suchprofils bekannt sein muss, und der Beurteilung der Relevanz der Ergebnisse – und einem sich daraus ergebenden allfälligen Entschluss zur Änderung der Fragestellung – am besten, wenn der Chemiker jeweils selbst die Suche übernimmt<sup>[22]</sup>. Die Voraussetzungen dafür, dass er dazu noch nicht einmal die Bibliothek aufsuchen müsste, sind mit der zunehmenden Computerdichte von der rein technischen Seite her gegeben, da alle diese Computer mit entsprechender Emulations-Software auch als oft sogar graphische und «intelligente» Terminals fungieren können; die Verbindung zum Host erfolgt dabei entweder über Modems/Telephone oder noch einfacher über lokale Computernetze wie *KOMETH* an der ETH und deren «gateways» zum öffentlichen Datennetz (*Telepac*, bzw. *Datex* in der BRD). In der Industrie ist diese Entwicklung schon wesentlich weiter fortgeschritten; es ist aber sehr wichtig anzumerken, dass hier von den Forschungschemikern überwiegend interne Datenbanken abgefragt werden, der Zugriff auf die (teuren) öffentlichen Datenbanken jedoch nur beschränkt zugelassen ist. Da die Hochschulen fast ausschliesslich (vgl. aber Abschnitt 7) auf letztere angewiesen sind, steht einer breiten, unkontrollierten Benutzung erst einmal die Gefahr der Kostenexplosion entgegen, wenn hier überall «geprübelt» würde. Der Datenbankzugriff ist durch Lösungsworte abgesichert, pro Lösungswort kann nur ein Benutzer gleichzeitig, dann im Prinzip aber von überall her und jederzeit suchen, wobei der Anschlussinhaber (d. h. die Hochschule) für die Bezahlung der Rechnungen und die Sicherheit des Passwortes haftet! Wie man sich leicht vorstellen kann, führt dies bei der Personalrotation an Hochschulen zu beträchtlichen Problemen der Sicherheit und der Administration, wenn man nicht für die nötige Kontrolle sorgt.

Das gewichtigere Problem ist aber, dass die derzeit öffentlich online verfügbaren Datenbank-Systeme als überwiegend «schnelle mechanische Suchhilfen» von der Abfragesprache (Suchmöglichkeiten, Zugriff zur Information) und zum Teil auch von der Indexierung her<sup>[23]</sup> zu wenig benutzerfreundlich, d. h. zu wenig fehlertolerant und «intelligent» sind, obwohl sie leider gerade bei Anfängern den gegenteili-

gen Eindruck zu erwecken scheinen. Daher braucht man für effiziente, zuverlässige Online-Recherchen nicht nur eine entsprechende (Grund)Ausbildung, sondern wegen dieser selten offensichtlichen Schwächen der Systeme und deren laufender Erweiterung und Verbesserung eine gewisse Erfahrung und ständige Übung, die man nur durch häufiges Suchen erhält. Der einzelne Doktorand hat dazu aber nach bisherigen Erfahrungen zu selten Gelegenheit (und auch Zeit).

Zur Illustration zwei Beispiele: Eine naheliegende, «simple» Formulierung der Frage nach der Reduktion von Enaminoketonen, nämlich «reduction AND enamino-ketone» (d. h. ohne Berücksichtigung von Singular/Plural, verschiedenen Schreibweisen etc.), gab nur *eines* statt 12 Zitate (vgl. Abschnitt 2). Sucht man Literatur zur Synthese von Reserpin, so geschieht dies online mit der *CAS Registry Number* (als «Synonym» des systematischen *CA*-Namens in der Online-Version von *CA*) und dem Suffix «P» für «preparation», und zwar entweder «50-55-5P» (*STN/CAS ONLINE, DIALOG*) oder «50-55-5-P» (*DATA-STAR*). Erfolgt die Eingabe nicht exakt so wie angegeben, dann ist das Resultat auf jeden Fall *Null*, wobei man nicht durch eine entsprechende Meldung auf Formatfehler (hier Bindestriche, sonst oft Leerzeichen) aufmerksam gemacht wird. In einem weniger offensichtlichen Fall als diesem lässt sich ein gelegentlicher Benutzer zur (vielleicht erwünschten?) Überzeugung verleiten, es gäbe keine publizierten Synthesen dieser Verbindung. Ein unerfahrener Benutzer verpasst wahrscheinlich, selbst wenn er keine «trivialen» Eingabefehler macht, neben den 25 Publikationen zu obiger Frage zwei weitere Totalsynthesen des *racemischen* Reserpins, das eine andere *Registry Number* (75331-15-6) als das «natürliche» Diastereomer hat.

Frustrationen und Systemfeindlichkeit durch Verallgemeinerung unbefriedigender Einzelergebnisse infolge mangelnder Anleitung und Betreuung oder, noch schlimmer, Fehlinformationen als vermeintlich sichere Grundlage für Forschungsplanung könnten die Folge unbetreuter und somit oft unkritischer Recherchen sein.

Bisher hat nicht nur die Zahl der für Chemiker relevanten Datenbanken, sondern auch die Zahl der Hosts und damit der zu erlernenden «Sprachen» für die Abfrage zugenommen. Diese an sich erfreuliche Zunahme des Informationsangebots führt zu einer Komplizierung der Situation und der Entscheidungsprozesse in der Informationssuche, vor allem dann, wenn Fragen ausserhalb eines Routinebereichs auftreten oder die zunehmenden interdisziplinären Forschungsprojekte in der Chemie neben den unentbehrlichen *CA* die Nutzung zusätzlicher spezieller, bisher unbekannter, da z. B. nicht gedruckt in der eigenen Bibliothek verfügbarer Quellen erfordern. Ob gewisse neuere Entwicklungen

wie z. B. die Vereinigung der Hosts *SDC ORBIT* und *Pergamon InfoLine*, deren verschiedene Datenbanken in Zukunft mit nur einer «Sprache» statt zweien abgefragt werden können, oder die Übernahme der *DATA-STAR*-«Retrievalsprache» durch *FIZ Technik* in Frankfurt für seine Datenbanken eine Trendwende ankündigen, lässt sich noch nicht sagen<sup>[24]</sup>.

Bisher hat auch die Komplexität einzelner Systeme durch ständige Verbesserungen noch zugenommen, wobei z. B. bei *CAS ONLINE* der Umfang der Handbücher für (Sub)struktur-Recherchen – 1981: 1 Band, 122 Seiten; 1985: 4 Bände, total 608 Seiten – noch schneller gewachsen zu sein scheint als die Leistungsfähigkeit. Eine kurzfristige Lösung der geschilderten Benutzerprobleme sehe ich nur in intensiver Ausbildung und Anleitung/Hilfestellung durch einige erfahrene «Spezialisten», ausser für einfache Recherchen oder solche vom immer gleichen Typ. Die Entwicklung «intelligenter» sogenannter «Front-Ends»<sup>[25]</sup> befindet sich noch in den Anfängen, und bisherige Ansätze konzentrieren sich auf die Lösung eher taktischer Fragen; menu-gesteuerte Programme wie *SciMate*<sup>[19a]</sup>, die im wesentlichen Grundfunktionen wie Suchen, Ausdrucken etc. in die jeweilige «Retrievalsprache» übersetzen, lösen weder das Problem der Datenbankauswahl noch das der Frageformulierung (vgl. aber die «Scripts» in *Searchmaster*<sup>[19b]</sup>). Die Entwicklung leistungsfähiger Suchhilfen wird nicht nur durch technische, sondern vor allem auch durch die mit dem Online-Informationsmarkt verbundenen ökonomischen und politischen Faktoren behindert.

Online-Recherchen müssen in der Hochschule eine Doppelfunktion erfüllen: Einerseits dienen sie nur als eines von vielen Hilfsmitteln zur effizienten Beschaffung der für die *Forschung* benötigten Informationen, andererseits müssen Chemiker im Rahmen einer modernen *Ausbildung* eigene praktische Erfahrungen mit dieser in der Industrie an Bedeutung noch zunehmenden Methode an sich sammeln. Die Informationsversorgung für die *Forschung* wäre am besten durch eine kleine Gruppe von «Online-Experten», etwa vergleichbar den Serviceleistungen für Spektroskopie an vielen Hochschulen, gewährleistet. Die zweite Aufgabe, die Lehre, verlangt hingegen, dass möglichst alle Doktoranden einmal selbst recherchieren, was aber aus den erwählten Gründen beträchtliche, grundsätzliche Probleme bezüglich der Effizienz und auch der Kosten bringt. Wir versuchen, dieses Dilemma durch einen *Kompromiss* zu lösen, der uns auch auf andere Hochschulen übertragbar scheint: Für jede Forschungsgruppe führen ein bis zwei Assistenten oder Doktoranden die Recherchen für die gesamte Gruppe aus. Diese «Rechercheure» vereinen damit eine eingehende Kenntnis der Forschungsgebiete und Fragestellungen mit einer meistens hinreichenden Anzahl von Suchen.

In diesem Zusammenhang wurde nicht

nur an der ETH Zürich beobachtet, dass normalerweise das zuerst erlernte System mit seiner Befehlssprache als «Online-Muttersprache» möglichst beibehalten wird – der Wechsel von *DIALOG* zum billigeren *DATA-STAR* benötigte Zeit und auch einen gewissen Druck, um die «Rechercheure» zum Umsteigen zu bewegen. Andere, speziellere Hosts und Datenbanken wie z. B. *CAS ONLINE, CRDS* bei *SDC ORBIT* oder *CIS* werden nur von wenigen, besonders Interessierten benutzt. Viele Chemiker sind anscheinend auf *CA* als nicht nur wichtigste, sondern einzige Quelle ausgerichtet. Zwar lassen sich die meisten (Literatur-)Fragen durch «Rechercheure» im derzeitigen Standardsystem *DATA-STAR* in *CA* beantworten, zur Betreuung und Durchführung von Suchen in speziellen, seltener benötigten Datenbanken/Systemen müssen noch zusätzlich einige «Spezialisten» bereitstehen. Damit nicht durch die ständige Rotation an der Hochschule immer wieder wichtige Erfahrungen verlorengehen, sollten dies Hochschullehrer oder Assistenten sein, die dann auch die Informations-Ausbildung für Chemiker übernehmen können.

## 5. Kosten und Nutzung

In unserer Abteilung wie auch an anderen Stellen im Schulratsbereich (ETHZ, EPFL, Annexanstalten) werden Online-Recherchen von der *Eidgenössischen Drucksachen- und Materialzentrale* finanziert. An chemischen Instituten einiger Schweizer Universitäten, die unserem Beispiel gefolgt sind, müssen Online-Recherchen hingegen aus allgemeinen Etatmitteln des Instituts oder der einzelnen Forschungsgruppen bezahlt werden (z. B. Université de Genève, Université de Fribourg<sup>[26]</sup>). Diese zusätzliche Belastung der ohnehin knappen allgemeinen Etats ist einer notwendigen Verbreitung von Online-Recherchen abträglich; man muss dafür Auswege durch spezielle Kredite, über vermehrte Bibliotheksmittel oder eine Berücksichtigung von Informations-Kosten in der Projektförderung finden. So führen z. B. Informationsvermittlungsstellen an allen österreichischen Universitäten für Hochschul-Angehörige *gratis* Online-Recherchen durch. Infolge der geschilderten Besonderheiten der Chemie-Information lassen sich Fragen aus diesem Gebiet besser innerhalb chemischer Institute als durch zentrale Stellen beantworten; wir arbeiten in diesem Sinne mit der Datenbankstelle der ETH-Hauptbibliothek zusammen.

Wir versuchen, durch Anleitung und Kontrollen einen sinnvollen und *kostenbewussten* Einsatz von Datenbanken zu erreichen, ohne durch zu starke Restriktionen oder bürokratische Hindernisse den Zugang zur Online-Information in nachteiliger Weise einzuschränken. Der Bibliothekar sorgt jedoch dafür, dass nicht aufwendige Online-Recherchen durchgeführt wer-

den, wenn in der Bibliothek vorhandene gedruckte Quellen (vgl. Abschnitt 3) eine bessere Problemlösung ermöglichen. *Essentiell ist die sorgfältige Vorbereitung von Online-Recherchen mit entsprechenden Suchhilfen* (z. B. *CA Index Guides*). Vermeidbare Kosten entstehen etwa durch das Ausdrucken von Hunderten nur zum Teil relevanten Zitaten statt einer Optimierung und sinnvollen Einschränkung des Suchprofils oder der Fragestellung, da ja in *CA* und den meisten anderen Datenbanken für jedes ausgedruckte Zitat (bzw. Verbindung) Gebühren berechnet werden. Auch das relativ kostspielige Ausdrucken des Abstract-Texts im *CA File* von *STN/CAS ONLINE* ist meistens nicht gerechtfertigt, da man ja für die Strukturformeln sowieso auf die wohl überall vorhandene gedruckte Version zurückgreifen muss.

Gerade an der Hochschule muss darüber hinaus noch vermehrt darüber aufgeklärt werden, dass Information nicht gratis ist und bei guter Qualität auch nicht sein kann, denn nicht nur Online-Recherchen mit ihren direkt quantifizierbaren und ersichtlichen Kosten, sondern auch Bibliotheken und andere Informationsquellen kosten viel Geld. Die Zeit, während der Doktoranden in der Bibliothek in gedruckten Quellen suchen – und möglicherweise aus grundsätzlichen Schwierigkeiten (vgl. das in Abschnitt 2 angeführte, repräsentative Beispiel) oder wegen mangelnder Kenntnis (d. h. Ausbildung) nicht das Gewünschte finden – ist ebenfalls Geld; von Frustrationen nach stundenlangem Wälzen von Registern gar nicht zu reden. Der relative Komfort, die Schnelligkeit und Leistungsfähigkeit der Online-Recherche haben nun einmal ihren Preis. Diese Aspekte der Informationsbeschaffung werden aber an der Hochschule ganz im Gegensatz zur kosten- und leistungsbevorzugten Einstellung in der Industrie oft übersehen.

Bis 1982 stieg bei uns die Online-Anschlusszeit auf 180% (bezogen auf 1979 = 100%), während sich die laufenden Kosten seit 1979 sogar fast verdreifachten. Im Jahr 1983 trat dann eine vorläufige «Sättigung» und sogar Rückgang um ca. 10% ein; 1984 wurde wieder der Wert von 1982 erreicht, wobei die Kosten aber auf niedrigerem Niveau stagnierten. Dieser Effekt beruht einerseits auf der bevorzugten Abfrage des relativ zu *DIALOG* oder *SDC* günstigeren Systems *DATA-STAR* für Literatur-Recherchen und ab Mitte 1984 auch auf dem (ersten) «Academic Program» von *CAS ONLINE*, da wir für Substruktur-Recherchen statt Monatsrechnungen von bis zu über \$1000 nur noch einen Pauschalpreis von monatlich \$500 zahlen mussten. Bei den *Kommunikationskosten*, die bis 1983 etwa ein Drittel der Gesamtkosten betrugten, bewirkten andererseits zwei Änderungen eine deutliche Reduktion auf einen Anteil von ca. 10% im Jahre 1984: Erstens wurde der *DATA-STAR*-Computer in Bern<sup>[41]</sup> direkt über eine normale Telephonleitung ausgewählt,

zweitens stand für Verbindungen nach den USA (*DIALOG*, *SDC*) statt der Vermittlung über *Radio Schweiz* mit Durchschnittskosten von 1.75 sFr./min die wesentlich günstigere Verbindung via das schweizerische Paketvermittlungsnetz *Telepac* zur Verfügung. Auch die Möglichkeit, *CAS ONLINE* statt in Columbus/USA über den *STN*-Knoten Karlsruhe/BRD anzuwählen, trug zur Kostenreduktion bei. Von 1984 nach 1985 stieg die Online-Anschlusszeit um ca. 3%, während sich die Kosten um ca. 5% (Dollarkurs!) erhöhten.

Dividiert man durch die Anzahl aller in Forschung und Lehre an der Abteilung für Chemie tätigen Doktoranden, Assistenten und Dozenten, so ergeben sich ca. 1.5 Online-Stunden und Kosten unter 500 sFr. pro Chemiker und Jahr. Solche Informationskosten sind in Relation zu den Lohn-, Material- und Infrastrukturkosten zu stellen. Das häufig gehörte Argument, dass Hochschulen sich Online-Recherchen nicht leisten können, ist vor diesem Hintergrund wohl nicht mehr haltbar. Man sollte eher fragen, ob die Hochschulen es sich leisten können, in Forschung und Lehre *nicht* die modernsten und effizientesten Mittel der Informationsversorgung zu nutzen.

In der Abfrage der verschiedenen *Hosts* haben sich seit 1979 einige Änderungen ergeben: 1981 z. B. war *DIALOG* noch das «Standardsystem» mit 77% der Anschlusszeit, schon 1984 übernahm aus den genannten Kostengründen *DATA-STAR* diese Rolle mit einem Anteil von 42% gegenüber noch 18% für *DIALOG*. In den Rest teilten sich 1984 *SDC ORBIT* mit 8% (1981: 12.5%) und *STN/CAS ONLINE* mit 32% (1981: 7%). *DATA-STAR* bietet ein grosses Datenbank-Sortiment für Chemie und verwandte Gebiete<sup>[41]</sup>, verbunden mit einem Hochschulrabatt und billigen Kommunikationskosten. Noch wesentlich günstiger ist infolge des erwähnten «Academic Program» (Literatur- und (Sub)struktur-Recherchen zu 10% des normalen Preises<sup>[11]</sup>) das Suchen in *CA* bei *STN/CAS ONLINE*, da *CAS* als Datenbank-Produzent natürlich einen konkurrenzlosen Preis offerieren kann. Da sich kürzlich das Angebot bei *STN* stark erweitert hat – ausser den *CA*-Literatur- und -Strukturdatenbanken stehen jetzt z. B. auch *BIO-SIS*, *DECHEMA*, *COMPENDEX* und Patentdatenbanken zur Verfügung – wird ein Umsteigen auf dieses System als «Standard» erwogen. Die Vielzahl der abgefragten *Hosts* ist nicht nur durch die historische Entwicklung bedingt, sondern trotz der damit verbundenen Schwierigkeiten unterschiedlicher «Retrievalssprachen» notwendig, um auf exklusiv angebotene Datenbanken zugreifen zu können oder spezielle Suchmöglichkeiten auszunutzen. Bei Aufschlüsselung der Anschlusszeiten nach *Datenbanken* dominieren erwartungsgemäss diejenigen von *CA* mit 81% (1984; 1981: 80.5%); davon entfielen 47.5% auf Literatur-, 4.5% auf Nomen-

klatur-, und 29% auf (Sub)struktur-Recherchen. Demgegenüber hatte die zweitwichtigste Datenbank, *BIO-SIS*, 1984 und 1981 nur einen Anteil von 4%, gefolgt von *SCISEARCH* (*Science Citation Index* online) mit 3.5 bzw. 2.5%, *CRDS* (1.5 bzw. 1%) und 77 weiteren Datenbanken mit einem Anteil von insgesamt 9% (1984).

## 6. Online-Recherchen und die Ausbildung in Chemie-Information

Voraussetzungen für die erfolgreiche Durchführung von Online-Recherchen sind *kritische Distanz zu den eigenen Resultaten* und

1. Kenntnis der Datenbank (z. B. *CA*), ihres Inhalts, ihrer Indexierungs- und Auswahlpolitik;
2. Sachkenntnis (Terminologie des Fachgebiets);
3. Kenntnis des Datenbanksystems (Zugriffsmöglichkeiten: «Retrievalssprache», Datenstruktur).

Die Probleme liegen, wie bereits erwähnt, meistens nicht in der Bedienung des Terminals (obwohl manchem Benutzer ein Schreibmaschinen-Kurs gut tun würde) oder der «Retrievalssprache», sondern in den gesuchten, mit der gedruckten Version der Sekundärliteratur fast immer identischen Inhalten von (Literatur)datenbanken [auf Probleme bei den für Chemiker grundsätzlich einfacheren (Sub)struktur-Recherchen wurde bereits hingewiesen]. Der Schwerpunkt einer Ausbildung muss daher bei den Inhalten und Informationsstrukturen liegen, und nicht bei den technischen Aspekten, die nur selten Schwierigkeiten bereiten.

Ein Blick in die Vorlesungsverzeichnisse von Hochschulen in der Schweiz, Deutschland<sup>[2]</sup> und Österreich zeigt, dass nicht nur Vorlesungen oder Praktika über die modernen maschinellen Methoden der Chemie-Information meist fehlen (mit bemerkenswerten Ausnahmen z. B. an der Universität Frankfurt/Main oder der Technischen Universität Wien), sondern selbst Lehrveranstaltungen über «konventionelle», d. h. gedruckte Quellen Mangelware sind<sup>[27]</sup>. Deshalb haben wir die Nutzung von Online-Diensten als Ausgangspunkt genommen, um den gesamten Problembereich der Chemie-Information in der Ausbildung aufzugreifen<sup>[28-30]</sup>.

Wegen des kompakten Studienplans an der ETH Zürich, der in acht Semestern zum Diplom führt, sind zur Zeit eigene Lehrveranstaltungen über Chemie-Information für *Studenten* nicht vorgesehen; Grundkenntnisse über *gedruckte Quellen* (*CA*, «*Beilstein*», «*Gmelin*») werden aber teilweise im Rahmen anderer Vorlesungen oder Praktika vermittelt. Die tägliche Erfahrung mit *Studenten* und *Doktoranden* in der Bibliothek zeigt jedoch, dass dies noch nicht in ausreichendem Masse geschieht. Gerade auch im Hinblick auf die später für *Doktoranden* erfolgende Aus-

bildung in Online-Recherchen sind solide, schon während des Studiums gelegte Grundlagen und praktische Erfahrungen wichtig, denn wer z. B. Aufbau und Indizierung der gedruckten CA nicht kennt, der wird später auch online kaum erfolgreich suchen. Eine Online-Ausbildung für Studenten, wie sie an einigen Universitäten in den USA erfolgt<sup>[29]</sup>, scheint mir aus organisatorischen und finanziellen Gründen bei uns noch nicht möglich. Diese Ausbildung findet erst in höheren Semestern, also während des *Nachdiplomstudiums* für *Doktoranden* statt. Eine seit Sommer-Semester 1984 als Lehrauftrag durchgeführte *Vorlesung* «Einführung in die Benützung der chemischen Literatur» (SS + WS, je 1 Wochenstd.) behandelt systematisch die wichtigen Gebiete der Chemie-Information, im ersten Teil (SS) eine allgemeine Einführung sowie Autoren- und vor allem Schlagwort-Recherchen, im zweiten Teil die Suche nach (Sub)strukturen, Reaktionen und Daten. Sie ist problemorientiert und nicht wie etwa die meisten Literaturführer quellenorientiert (Besprechung von CA, «Beilstein», etc.) aufgebaut; wir integrieren die Anleitung zur Auswahl und Benutzung gedruckter und maschinenlesbarer Quellen mit dem Schwerpunkt auf letzteren. Anhand von Beispielen aus der Praxis, die durch Projektionsfolien (vergrößerte Kopien des Originalausdrucks von Online-Recherchen) präsentiert werden, erarbeiten wir verallgemeinerbare *Suchstrategien* (vgl. Fig. 2).

Die Vermittlung von «Strategien» zur Lösung von Informationsproblemen ist bisher oft gegenüber technischen oder «taktischen» Aspekten wie Bedienung des Terminals, Befehlssprachen der Systeme, Aufbau von Datenbanken vernachlässigt worden. Gerade an der Hochschule muss sie ein wichtiger Teil der Informationsausbildung sein, denn diese Grundlagen werden in den systemorientierten Kursen der Datenbank-Anbieter (und oft auch in Kursen der Datenbank-Produzenten) zu wenig behandelt. Wegen des Mangels an geeigneten Lehrbüchern werden zur Vorlesung Unterlagen verteilt, auf denen die wichtigsten Informationen und Strategien zusammengefasst sind, z. B. durch Flussdiagramme wie das in Fig. 2 gezeigte. Die Theorie bleibt, nicht zuletzt aus Zeitgründen, auf das zum Verständnis und den praktischen Einsatz notwendige beschränkt; z. B. werden die Grundlagen von Recherchen in «inverted files», Paketvermittlungsnetze (*Telepac*, *Datex*) oder Verschlüsselungsmethoden für Strukturen nur kurz und mit Hinweis auf weiterführende Literatur behandelt. Bisher war es aus organisatorischen Gründen praktisch nicht möglich, die von den Hörern gewünschten Demonstrationen und Übungen «live» am Terminal durchzuführen. Nachdem inzwischen die technischen Möglichkeiten (über KOMETH mit Hosts verbundene Mikrocomputer als Terminals) und vor allem im Erweiterungsbau der Chemie-Bibliothek durch einen Seminarraum auch die räumli-

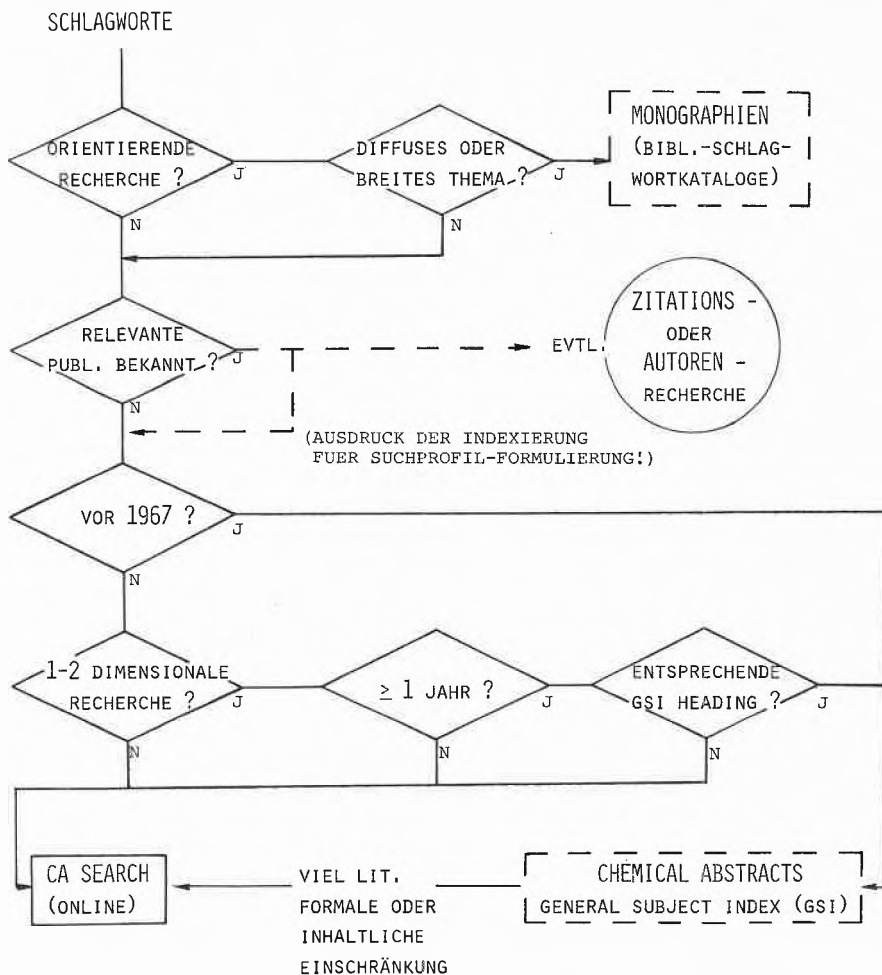


Fig. 2. Flussdiagramm zum Vorgehen («Strategie») bei Schlagwort-Recherchen in *Chemical Abstracts* (aus den Unterlagen zur Vorlesung «Einführung in die Benützung der chemischen Literatur»).

chen Voraussetzungen gegeben sind, planen wir ab SS 1987 solche Übungen zur praktischen Vertiefung des Lehrstoffs.

Eine Anleitung zur Abfrage einzelner Datenbanksysteme muss gewissermassen auf dem «taktischen» Niveau die in der Vorlesung gelegte Basis für diejenigen ergänzen, die tatsächlich selber Online-Recherchen durchführen wollen. Dies wurde bisher meistens durch eine Einzelausbildung am Terminal, vor allem für unser derzeitiges Standardsystem *DATA-STAR* durch den Bibliothekar *Josef Meienberger*, erreicht, und soll in Zukunft vermehrt durch gezielte (interne) Kurse ergänzt werden, die wir schon früher mehrmals mit Erfolg durchgeführt haben. Wir wollen die verbesserten Möglichkeiten der neuen Chemie-Bibliothek auch als Gastgeber für öffentliche Kurse der Produzenten (z. B. *CA*, *Derwent*) oder Hosts von Datenbanken (*DATA-STAR*, *DIALOG*, etc.) ausnützen; dieses Jahr werden *FIZ Chemie (Berlin)* und *STN Karlsruhe* bei uns Kurse abhalten. Solche Kurse sind vor allem der Weiterbildung der «Spezialisten» vorbehalten – wenn sie auswärts stattfinden aus Kostengründen, bei einer Veranstaltung

im Haus wegen des zumeist beschränkten Platzangebots. Sie können die genannten eigenen Ausbildungsaktivitäten für ein breiteres Publikum ergänzen, aber nicht ersetzen, vor allem auch deswegen, weil ein Hochschullehrer – zweckdienliche Ausbildung und vor allem Erfahrung vorausgesetzt – näher auf chemie- und sogar institutsspezifische Probleme eingehen und sich auf die oft sehr heterogenen Bedürfnisse und Vorbildung der Benutzer besser einstellen kann. Die meistens zu umfangreichen und daher für den Anfänger abschreckenden oder verwirrenden Handbücher sollten dabei durch an die Hörer verteilte, aus der Praxis unter besonderer Berücksichtigung der Chemie entstandene Kurz-Anleitungen ersetzt werden.

Wegen der ständigen Änderungen bei Datenbanken und Systemen ist die laufende Weiterbildung der «Spezialisten» und der Austausch der notwendigen Information mit allen online recherchierenden Chemikern sehr wichtig; die wenigsten lesen z. B. die wichtigen «Newsletter» der Anbieter. Dies kann informell (wie bisher bei uns) oder auch in Form periodischer Seminare geschehen.

## 7. Ausblick: Trends und Wünsche

Die Möglichkeiten zu *Literatur-* und *(Sub)struktur-*Recherchen sind auf einem befriedigend hohen Angebotsniveau, wenn man einmal von den geschilderten Problemen des Zugriffs und der Indexierung abseht. Das erwähnte «Academic Program» macht zwar *CA* als wichtigste Quelle der Chemie-Information für Hochschulen billig zugänglich, man bedauert aber, dass wichtige und zum Teil sehr teure Datenbanken wie *BIOSIS*, *SCISEARCH* (*Science Citation Index* online) und Patentdatenbanken bisher nicht diesem Beispiel gefolgt sind. Somit bleiben wesentliche Bereiche der *Forschungsinformation* für viele Hochschulen schwer erschwinglich, und es wird vielleicht eine zu einseitige Ausrichtung auf *CA* erzeugt. Noch unbefriedigender ist diesbezüglich die Lage in der *Informationsausbildung*, da selbst das «Academic Program» oder die z. B. bei *STN/CAS ONLINE* oder *DIALOG* vorhandenen «Übungsdatenbanken» mit einem Subset der vollständigen Literatur- bzw. Strukturdatenbanken noch zu teuer sind. In allen Fällen sind *zeitabhängige* Gebühren für Suche und Telekommunikation zu bezahlen, die bei einem für Trainingszwecke wünschenswerten intensiven Gebrauch und/oder einer grossen Zahl von Benutzern (auch noch durch Passwörter begrenzt) schnell unerschwinglich werden; dies macht auch eine Ausdehnung der Online-Ausbildung auf Studenten problematisch. Ein Ausweg aus diesem Dilemma wären *lokale* Implementierungen solcher Datenbank-Subsets z. B. auf im Hochschulbereich weitverbreiteten Rechnern, unter Umständen sogar auf leistungsfähigen Mikrocomputern, denn dann wird zahlreiches und langes Üben nicht mit steigenden Kosten «bestraft».

Für *Reaktionen* hat sich die Situation mit den (von uns kürzlich evaluierten<sup>[31a]</sup>) «in-house»-Reaktionsdatenbank-Systemen *REACCS*, *SYNLIB* und *ORAC*<sup>[31]</sup> signifikant verbessert. *CAS* wird demnächst mit einer eigenen Reaktionsdatenbank in *STN/CAS ONLINE* das Angebot noch erweitern<sup>[32]</sup>. Zur Nutzung von *REACCS*, *SYNLIB* und *ORAC* im Rahmen der jeweiligen «Academic Programs» sind zwar von den Hochschulen Gegenleistungen in Form der Erfassung von Reaktionen aus der laufenden Literatur zu erbringen<sup>[31a]</sup>, dafür ist dann aber die Abfrage dieser lokal auf dem eigenen Rechner geladenen Datenbanken beliebig oft ohne direkt zeitabhängige Kosten möglich. Da zugleich diese Systeme durch Menu-Steuerung an Stelle von Befehlseingabe sehr benutzerfreundlich sind, entfallen die zuvor gegen eine unkontrollierte Suche in den öffentlich online verfügbaren Datenbanken vorgebrachten Argumente (vgl. Abschnitt 4), so dass Doktoranden und sogar Studenten diese Systeme jederzeit und weitgehend selbständig abfragen können. Zur praktischen Demonstration der beschrittenen Zugriffspfade zur Reaktions-Information und wegen der Ergänzung der unter-

schiedlichen Reaktions-Dateien haben wir alle drei Systeme an der ETH Zürich implementiert. Diese für Ausbildung und die Bereitstellung von Information für die Forschung gleichermaßen hervorragend geeigneten Quellen mit schnellem und leichtem Zugriff auf insgesamt über 100 000 Reaktionen stehen uns im Prinzip seit Frühjahr 1986 zur Verfügung; ein intensiver Zugriff beginnt aber erst jetzt mit der kürzlichen Anschaffung von *Macintosh-Plus*-Computern als unter anderem «intelligente» graphische Terminals (mit Terminal-Emulations-Software *Versa-Term-PRO 2.0*) in den Arbeitsgruppen des Laboratoriums für Organische Chemie<sup>[33]</sup>, so dass heute noch zu wenig Erfahrungen über die Akzeptanz und breite Nutzung vorliegen.

Weitere Beispiele für kommerziell verfügbare «in-house»-Datenbanken sind die <sup>13</sup>C-NMR-Spektren von 20 000 Verbindungen von *Sadtler*, die *GenBank Nucleotide Sequence Database*<sup>[34a]</sup> mit mehr als 6000 Nucleotidsequenzen (beide für IBM PC) oder das *Chemical Hazard Response Information System* als Datenbank der *US Coast Guard* mit unter anderem toxikologischen Daten<sup>[34b]</sup> (öffentlich online im System *CIS* verfügbar). Das Angebot an solchen (Teil)datenbanken wird mit fortschreitender Hardware-Entwicklung zunehmen; welche Rolle Bildplatten in der Verbreitung dezentraler Informationssysteme spielen werden, ist noch nicht abzusehen<sup>[35]</sup>. Es ist aus technischen und auch finanziell-politischen Gründen klar, dass sich diese Lösungen zumindest zur Zeit nicht auf die um Grössenordnungen umfangreicheren Literatur- oder gar die mit relativ aufwendiger Hardware suchbaren Struktur-Datenbanken<sup>[7]</sup> etwa von *CA* übertragen lassen, so dass wir für Forschungsinformation in der näheren Zukunft nach wie vor auf die existierenden (zentralen) Online-Hosts angewiesen sein werden.

Für *Daten* und *Spektren* von Verbindungen besteht eine Lücke zwischen *CA* mit seiner umfassenden Abdeckung der gesamten Chemie, aber mangelhafter oder fehlender Indexierung für Daten (und als *Sekundär*-Quelle natürlich ohne die Daten selbst), und den spezialisierten, auf relativ wenige (Grössenordnung 50 000) Verbindungen beschränkten (spektroskopischen) Datenbanken (vgl. Fig. 1), die dafür aber unmittelbaren Zugriff auf Daten einschliesslich deren graphischer Darstellung und Manipulierung, zum Teil sogar deren Interpretation (*C13NMR*<sup>[13]</sup>) erlauben. In diese Lücke werden bald «*Beilstein*»- und «*Gmelin*»-Online-Datenbanken stossen<sup>[36]</sup>. Dies wird nicht nur unsere Möglichkeiten für gezielte Daten-Recherchen wesentlich erweitern, sondern auch schlussendlich die gesamte ältere Literatur (vor 1960<sup>[37]</sup>) online zugänglich machen.

Die bisherige Diskussion bezog sich stets auf der Sekundärliteratur entsprechende oder diesem Bereich zuzuordnende Datenbanken. Inzwischen ist zwar auch ein Teil

der *Primärliteratur*, nämlich die Zeitschriften der *American Chemical Society*<sup>[38]</sup> ab 1982 online als Volltext-Datenbank (aber ohne die wichtige graphische Information wie Formeln) bei *STN* suchbar. In Anbetracht der Verfügbarkeit der gedruckten Zeitschriften scheint mir eine Recherche in diesem File *CJACS* trotz eines «Academic Program» mit reduzierten Preisen nur in sehr wenigen Ausnahmefällen sinnvoll. Schon bei Literatur-Recherchen im *CA File* unter Einbeziehung des (such- und ausdrückbaren) Abstract-Textes findet man sehr häufig zu viele nicht relevante Zitate, in Volltext-Datenbanken dürfte sich dieser Ballastanteil noch deutlich erhöhen. Unser Hauptproblem ist der *gezielte* Zugriff auf die relevante Information aus der Masse der Publikationen, «Computer Power» und maschinell abfragbare Primärliteratur sind da kein Ersatz für besser indexierte Sekundärliteratur<sup>[39]</sup>.

An den Hochschulen müssen diese Entwicklungen der Chemie-Information besser als bisher verfolgt und schneller als bisher in die Informationsversorgung für die Forschung und in eine moderne Ausbildung integriert werden. In den Richtlinien der American Chemical Society für «Undergraduate Professional Education in Chemistry»<sup>[40]</sup> steht zu «Chemical Literature and Information Retrieval»:

“Students preparing for professional work in chemistry must learn how to retrieve specific information from the enormous and rapidly expanding chemical literature. The increasing volume and complexity of the literature means that students can no longer acquire skills in information retrieval without some formal instruction. . . .

It is highly desirable that students gain some experience with on-line interactive computer files. It is essential, however, that students understand the organization and use of printed information sources in order to use the computer readable files to best advantage.”

Diese Bedingungen sind an den meisten Hochschulen in der Schweiz, Deutschland und Österreich derzeit offensichtlich weder für Studenten noch für Doktoranden erfüllt; dass es auch in anderen Ländern und sogar in den USA mit der Informationsausbildung für Chemiker häufig nicht viel besser bestellt ist<sup>[29,30]</sup>, muss uns Ansporn und nicht etwa «Trost» sein.

*Wir danken der Eidgenössischen Druck-sachen- und Materialzentrale (EDMZ), Bern, für die grosszügige Finanzierung der Online-Recherchen an der Abteilung für Chemie der ETHZ, und PD Dr. Annetta Weber von der Datenbankstelle der ETH-Hauptbibliothek für gute Zusammenarbeit. Der Autor bedankt sich beim Bibliothekar der Chemie-Bibliothek, Josef Meienberger, für jahrelange erfolgreiche und erfreuliche Zusammenarbeit, seinen ständigen Einsatz*

und seine Hilfsbereitschaft, die wesentlich dazu beigetragen haben, dass Online-Recherchen den Stellenwert erreicht haben, den sie jetzt an unserer Abteilung besitzen.

Eingegangen am 20. März 1987 [FR 35]

- [1] M.K. Huber, in H. Grunewald (Ed.): *Chemistry for the Future* (Proc. 29th IUPAC Congr. 1983), Pergamon, Oxford (1983), p. 439; P. Rhyner, *Chimia* 37 (1983) 263.
- [2] Vgl. U. Jurgeleit, *Nachr. Dok.* 36 (1985) 271.
- [3] E. Zass, in E. Ziegler (Ed.): *Computer in der Chemie* (2. Aufl.), Springer-Verlag, Berlin (1985), p. 24; D. Rehm, F.-P. Montforts, M. Ockenfeld, G. Wess: *Online-Recherchen in Datenbanken des Chemical Abstracts Service*, Verlag Chemie, Weinheim (1982); W. Krietsch, *Chem. Labor Betr.* 37 (1986) 163; J. Harnisch, *Kontakte (Merck)* (1984) {3} 3; E. Zass, *Nachr. Chem. Tech. Lab.* 32 (1984) 424, 502, 578.
- [4] G. Naber, *Database* 8 (1985) {1} 20.
- [5] H. R. Pichler: *Online-Recherchen für Chemiker*, VCH Verlagsgesellschaft, Weinheim (1986); A. B. Wagner, *Online Rev.* 10 (1986) 173.
- [6] R. Attias, *J. Chem. Inf. Comput. Sci.* 23 (1983) 102; M. Pfau, *Actual. Chim.* (1985) {4} 45.
- [7] P. G. Dittmar, N. A. Farmer, W. Fisanick, R. C. Haines, J. Mockus, *J. Chem. Inf. Comput. Sci.* 23 (1983) 93.
- [8] P. G. Dittmar, R. E. Stobaugh, C. E. Watson, *J. Chem. Inf. Comput. Sci.* 16 (1976) 111; F. A. Tate, *Chemistry* 41 (1968) {7} 18.
- [9] Online z. Z. nicht mehr verfügbar; E. Garfield, M. Sim, *Pure Appl. Chem.* 49 (1977) 1803; E. Garfield, *Curr. Contents* 24 (1984) {26} 3; ISI = Institute for Scientific Information.
- [10] CJS = Chemical Information System; S. R. Heller, *J. Chem. Inf. Comput. Sci.* 25 (1985) 224; S. R. Heller, *Drexel Libr. Quart.* 18 (1982) {3-4} 39; S. R. Heller, G. W. A. Milne, *Anal. Chim. Acta* 122 (1980) 117.
- [11] Auskunft erteilt FIZ Chemie, Steinplatz 2, D-1000 Berlin 12.
- [12] F. H. Allen et al., *Acta Crystallogr. B* 35 (1979) 2331; F. H. Allen, O. Kennard, R. Taylor, *Acc. Chem. Res.* 16 (1983) 146; C. Kratky, H. Bernhard, *Fakten, Daten, Zitate* 4 (1984) {1} 6; die *Cambridge Crystallographic Database* ist in vielen Hochschulen lokal (auf dem eigenen Rechner) verfügbar.
- [13] W. Bremser, *Nachr. Chem. Tech. Lab.* 31 (1983) 456; W. Bremser, H. Wagner, B. Franke, *Org. Magn. Reson.* 15 (1981) 178.
- [14] Für eine vergleichende Recherche zu diesem Thema in *CA/BIOSIS* vgl. E. Zass, *Naturwissenschaften* 69 (1982) 276.
- [15] Um das auch im Erweiterungsbau der Chemie-Bibliothek nicht zuletzt infolge der Zusammenlegung mehrerer bisher im Bereich der Abteilung für Chemie bestehender (Instituts)Bibliotheken zu einer leistungsfähigen Bibliothek bald wieder akute Platzproblem langfristig zu lösen, stehen die Zeitschriften auch auf Mikrofilm zur Verfügung, so dass ältere oder selten gebrauchte Jahrgänge sukzessive ins Archiv verlagert werden können, ohne den Zugriff zu erschweren.
- [16] A. Finch, *J. Chem. Inf. Comput. Sci.* 26 (1986) 17.
- [17] Siehe z. B. a) S. J. Kolner, *Online* 9 (1985) {1} 37, {2} 39, {3} 44, {4} 27, {6} 42, 10 (1986) {4} 32; b) A. Mullen, E. Möller, M. Blunck, *Online Inf.* 8 (1984) 305; H. Jüngling, *Nachr. Dok.* 36 (1985) 144.
- [18] a) J. Bruman, *Proc. 4th Natl. Online Meet. New York* (1983) 75; S. Clancy, *Database* 9 (1986) {4} 52; b) INFOLOG (Information und Kommunikation) D. Rieth, Hartmann-von-Aue-Strasse 9, D-7801 Au bei Freiburg, GENESYS IuD-Software (GENESYS GmbH, Schulstrasse 5, Niederroth, D-8062 Markt Indersdorf).
- [19] a) E. Garfield, *Curr. Contents* 23 (1983) {12} 5, {14} 5; C. Stout, T. Marcinko, *Online* 7 (1983) {3} 112; b) K. R. Walton, *Online* 10 (1986) {5} 70.
- [20] a) O. Oberhauser, *ABI-Technik* 6 (1986) 177; b) E. Mortensen, *Online Inf.* 8 (1984) 331; C. J. Armstrong, J. A. Large, *ibid.* 9 (1985) 231.
- [21] M. Ockenfeld, *Online Inf.* 5 (1981) 307; S. G. Faibisoff, J. Hurych, *Spec. Libr.* 72 (1981) 347; K. R. Walton, P. L. Dedert, *Online* 7 (1983) {5} 42; J. S. Haines, *ibid.* 6 (1982) {6} 14; M. Eisenberg, in C. Keren, L. Perlmutter (Ed.): *The Application of Mini- and Microcomputers in Information, Documentation, and Libraries*, Elsevier/North-Holland, Amsterdam (1983), p. 611; J. A. Hunter, *Proc. 4th Natl. Online Meet. New York* (1983) 223; J. A. Hunter, *Online* 8 (1984) {3} 36; C. L. Borgman, D. O. Case, D. Ingebræten, *Online Rev.* 9 (1985) 307; R. E. Buntrock, A. K. Valicenti, *J. Chem. Inf. Comput. Sci.* 25 (1985) 415; M. A. S. Palma, C. Sullivan, *ibid.* 25 (1985) 422; M. Ojala, *J. Am. Soc. Inf. Sci.* 37 (1986) 197; W. Sewell, S. Teitelbaum, *ibid.* 37 (1986) 234.
- [22] R. Fugmann, *Angew. Chem.* 94 (1982) 581; *Angew. Chem. Int. Ed. Engl.* 21 (1982) 608.
- [23] In gewissem Sinne gilt dies sogar bei den für Chemiker «einfacheren» (Sub)struktur-Recherchen, bei denen unter anderem Nichtbeachten von CA-Bindungsdefinitionen eine häufige Ursache von unvollständiger Information ist; vgl. dazu die Diskussion zu Fig. 1.
- [24] Die existierenden oder geplanten «gateways» (automatischen Verbindungen) zwischen verschiedenen Hosts wie etwa *Pergamon InfoLine* und *ESAIRS* oder *BRS (Inf. Intell. Online Newslett.* 7 (1986) {6-7} 1) vereinfachen die Administration (nur noch eine Rechnung) und die Verbindungsaufnahme zum Host, aber bisher nicht die Abfrage.
- [25] C. Tenopir, *Libr. J.* 110 (1985) {12} 40; D. T. Hawkins, L. R. Levy, *Online* 9 (1985) {6} 30, 10 (1986) {1} 33, {3} 49.
- [26] Persönliche Mitteilungen von Prof. Dr. W. Oppolzer (Genève) bzw. Dr. R. Neier (Fribourg).
- [27] Eine kurze und deshalb kaum ausreichende Behandlung dieses Gebiets innerhalb anderer Lehrveranstaltungen (vgl. dazu die Diskussion in ref. [29]) lässt sich natürlich so nicht ermitteln. Wesentlich besser sieht es bei Lehrveranstaltungen über allgemeine Computernutzung (einschliesslich Programmierkurse) oder speziellen EDV-Einsatz in der Chemie aus.
- [28] S. H. Wilen, *J. Chem. Inf. Comput. Sci.* 24 (1984) 112; Y. Wolman, *ibid.* 24 (1984) 135; J. D. Graybeal, *ibid.* 24 (1984) 193.
- [29] A. N. Somerville, *J. Chem. Inf. Comput. Sci.* 25 (1985) 314.
- [30] H. Skolnik, *J. Chem. Inf. Comput. Sci.* 24 (1984) 95; G. McKinney, A. P. Mosby, *Online Rev.* 10 (1986) 107.
- [31] a) E. Zass, S. Müller, *Chimia* 40 (1986) 38; b) Beiträge von W. T. Wipke et al. (p. 92), D. F. Chodosh (p. 118), A. P. Johnson, A. P. Cook (p. 184), in P. Willett (Ed.): *Modern Approaches to Chemical Reaction Searching*, Gower, Aldershot (1986).
- [32] P. E. Blower, R. C. Dana, in ref. [31]b, p. 146.
- [33] Die Benutzung einer «Maus» zur Abfrage menu-gesteuerter Datenbank-Systeme ist nach unserer Erfahrung allen anderen Möglichkeiten wie Tastatur, Graphik-Tablett, Lichtgriffel oder «Touch Screen» vorzuziehen.
- [34] a) H. S. Bilofsky, C. Burks, J. W. Fickett, W. B. Goad, F. I. Lewitter, W. P. Rindone, S. D. Swindell, C.-S. Tung, *Nucleic Acids Res.* 14 (1986) 5; b) *Chem. Eng. News* 65 (1987) {2} 23.
- [35] N. K. Herther, *Online* 9 (1985) {6} 17; M. Q. Adams, *ibid.* 10 (1986) {4} 60; vgl. aber auch *ibid.* 10 (1986) {5} 11.
- [36] C. Jochum, *Mitt. GdCh-Fachgruppe Chemie-Inf.* 11 (1987) 12; C. Jochum, in: *Tagungsbericht 2. Vortragstagung Fachgruppe Chemie-Information (Aachen, 25.3.1985)*, Gesellschaft Deutscher Chemiker, Frankfurt am Main (1986), p. 51; L. Kiessling, B. Roth, *ibid.*, p. 58; vgl. auch *Chimia* 41 (1987) 41.
- [37] Vgl. dazu das «Pre-1965 Registration Project» von CAS, das zur Zeit alle in *Chemical Abstracts* ab 1962 (ab 7th Collective Index) vorkommenden Verbindungen erfasst hat und im *CA Registry File* als (Sub)struktur suchbar macht; eine Erweiterung zurück bis 1920 hängt von der Finanzierung ab, siehe *CAS Report* 21 (1986) 11.
- [38] S. W. Terrant, L. R. Garson, B. E. Meyers, S. M. Cohen, *J. Chem. Inf. Comput. Sci.* 24 (1984) 230; *CAS ONLINE News* 6 (1986) {5} 4.
- [39] R. Fugmann, G. Ploss, *Angew. Chem.* 85 (1973) 978; *Angew. Chem. Int. Ed. Engl.* 12 (1973) 882.
- [40] ACS Committee on Professional Training: *Undergraduate Professional Education in Chemistry: Guidelines and Evaluation Procedures*, American Chemical Society, Washington DC (1983), p. 13.
- [41] Vgl. H. Ochsner, *Chimia* 41 (1987) 132.