

# Neue Möglichkeiten zur Recherche von organisch-chemischen Reaktionen: Ein Vergleich der «in-house»-Datenbanksysteme REACCS, SYNLIB und ORAC

Engelbert Zass\* und Stefan Müller

In der Chemie-Dokumentation wurden in den letzten Jahren wesentliche Fortschritte durch die Verwendung von Computern möglich. Während frühere maschinelle Informationssysteme nur Spezialisten vorbehalten waren, sind die modernen Online-Datenbanken – wie bisher ja auch die gedruckte Sekundär- und Tertiärliteratur – im Prinzip jedem Chemiker zugänglich. Voraussetzung für die breite Anwendung ist aber eine einfache Bedienung, das heisst vor allem die Möglichkeit, Fragen in der Sprache des Chemikers, der Strukturformel, ohne komplizierte Terminologie, Nomenklatur oder Computerbefehle zu stellen. Für die Suche nach Reaktionen war im Gegensatz zu (Sub)struktur-Recherchen das Angebot in diesem Sinne bisher unbefriedigend. Dies hat sich mit den hier vergleichend betrachteten neueren Reaktions-Datenbanksystemen REACCS, SYNLIB und ORAC grundlegend geändert.

## 1. Einleitung

Für die Suche nach Informationen über organisch-chemische Reaktionen steht den Chemikern schon seit langem die gedruckte Sekundärliteratur mit Referateorganen wie *Chemical Abstracts* (CA)<sup>[1]</sup>, den «klassischen» Handbüchern *Theilheimer's Synthetic Methods of Organic Chemistry*, *Houben-Weyl: Methoden der organischen Chemie*, *Beilsteins Handbuch der Organischen Chemie*, und Serien wie *Organic Syntheses*, *Organic Reactions*, *Reagents for Organic Synthesis* (Fieser/Fieser et al.) zur Verfügung. Dieses Informationsangebot wird durch mehrbändige Werke wie z. B.

*The Chemistry of Functional Groups* (Patai et al.), *Methodicum Chemicum* (Korte et al.), *Formation of C–C Bonds* (Mathieu/Weill-Raynal), durch zahllose Monographien über einzelne Reagentien und Reaktionen, und durch spezielle Dienste für laufende Information wie *Journal of Synthetic Methods*, *Current Chemical Reactions* oder die Abstracts in *Synthesis* ergänzt. Der Zugriff nach Reaktionen erfolgt hierbei in allen Fällen mittels gedruckter Register über Namen und/oder Summenformel der beteiligten chemischen Verbindungen oder allgemeiner über Schlagwörter für Verbindungsklassen, Reaktionstypen und Namenreaktionen.

Das Problem für den Benutzer besteht nicht nur darin, aus diesem vielfältigen Angebot die für die Beantwortung seiner Frage erfolgversprechendsten Quellen auszuwählen, sondern nicht zuletzt auch

darin, die primäre Fragestellung – fast immer in Form chemischer Strukturen und deren Umwandlungen – in die «Sprache(n)» (= Indexierung) der jeweiligen Quelle zu «übersetzen». Beides verlangt erhebliche Kenntnisse, viel Erfahrung, und meistens viel Zeit. Während das Vorgehen bei Reaktionen definierter chemischer Verbindungen – z. B. Oxidation von Lanosterol – über den (systematischen!) Namen oder über die Summenformel noch relativ einfach ist, sind die häufigen allgemeineren Fragen – z. B. nach Methoden zur Reduktion eines Ketons in Gegenwart eines Lactons – schon schwieriger in die erforderliche Terminologie umzusetzen und damit zu suchen. Bei Fragen nach Ausbeuten, Reagentien, Reaktionsbedingungen (Temperatur, Druck, pH-Wert), Lösungsmitteln, Katalysatoren lassen einen die meisten Quellen im Stich. Solche Informationen sind entweder mangels entsprechender Register nicht direkt suchbar, obwohl im Prinzip vorhanden (z. B. im «Theilhei-



Engelbert Zass: Geboren 1950 in Porz-Urbach (Köln). Studium der Chemie an der Universität Köln 1968–71, Diplomarbeit 1972 bei Prof. Emanuel Vogel (Synthese überbrückter [14]Annulene); Promotion 1977 an der Eidgenössischen Technischen Hochschule Zürich (ETHZ) bei Prof. Albert Eschenmoser (Komplexierung von Chlorophyll-Derivaten). 1978 Assistent, seit 1984 Oberassistent am Laboratorium für Organische Chemie der ETHZ mit Lehrauftrag für Benützung der chemischen Literatur. Arbeitsgebiet: Chemie-Information und -Dokumentation.



Stefan Müller: Geboren 1959 in Berlin. Studium der Chemie an der Universität Köln, 1979–82, nach dem Vordiplom an der Eidgenössischen Technischen Hochschule Zürich bis zum Diplom 1983; seit 1984 als Doktorand in der Gruppe von Prof. Dieter Seebach am Laboratorium für Organische Chemie.

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

mer»), oder sie fehlen in der Sekundärliteratur ganz, wie mit bestimmten Ausnahmen in CA<sup>[2]</sup>. Hinzu kommt, dass gedruckte Register *eindimensional* sind, d. h. die Suche nur jeweils nach einem (alphabetisierten) Sachverhalt erlauben; die meisten Fragen enthalten aber mehrere Konzepte (Dimensionen) wie z. B. "Baeyer-Villiger-Oxidation mit H<sub>3</sub>PO<sub>5</sub> unter milden Bedingungen".

## 2. Computerunterstützte Reaktions-Recherchen

Eine wesentliche Verbesserung brachte auch für Reaktions-Recherchen<sup>[3]</sup> der Einsatz von Computern<sup>[4]</sup>. Heute stehen die zwei wichtigsten Sekundärpublikationen für Reaktionen in maschinenlesbarer Form als öffentlich *online* (d. h. interaktiv) abfragbare Datenbanken zur Verfügung, nämlich CA als CA SEARCH und CRDS (*Chemical Reaction Documentation Service* von DERWENT = "Theilheimer" + "J. Synth. Meth."). Eine "Beilstein"-Online-Datenbank ist in Entwicklung<sup>[5]</sup>. In solchen (Literatur-)Datenbanken, deren Inhalt praktisch mit der gedruckten Version übereinstimmt, ist der Zugriff zur Information wesentlich erweitert und, zumindest was den Suchvorgang an sich betrifft, einfacher. So sind in CA SEARCH *alle* Begriffe aus dem Titel und den Register-eintragungen (und nicht wie im gedruckten Register nur die "Index Headings") zu Verbindungen, Verbindungsklassen, Namenreaktionen etc. direkt suchbar. Wichtiger noch ist die Möglichkeit zur *mehrdimensionalen* Recherche durch Verknüpfung mehrerer Schlagwörter (Konzepte) und deren Synonyme z. B. mit logischen Operatoren (AND, OR, NOT), mit denen sich Begriffe im inhaltlichen Zusammenhang suchen lassen<sup>[6]</sup>. Die in vielen Fällen für Reaktions-Recherchen unzureichende Indexierung, z. B. in den verbindungsorientierten CA<sup>[7]</sup>, wird aber natürlich durch Online-Suche nicht geändert. Auch die «Übersetzung» der Fragestellung ist wie bei der Suche in gedruckten Quellen nach wie vor notwendig, wenn sie auch online durch Hilfsmittel und vor allem durch das interaktive Suchverfahren (mit direkter Modifizierung der Fragestellung aufgrund der gefundenen Resultate und relativ mühevoller Wiederholung der Suche) erleichtert wird. Für Reaktions-Recherchen in CRDS z. B. kann man ausser den im "Theilheimer" zu verwendenden Schlagwörtern für Verbindungsklassen Strukturelemente und Bindungsänderungen bei der Reaktion präziser durch einen aus dem "Ringcode"<sup>[8]</sup> entwickelten Code eingeben. Diese «Übersetzung» einer *Struktur* ist jedoch aufwendig und fehleranfällig; zudem betragen die Kosten für die Benutzung dieser Datenbank mit ca. 65 000 Reaktionen US \$100 pro Anschaltstunde plus Gebühren für Telekommunikation!

Die bisher unbefriedigende Situation wird noch deutlicher, wenn man sie mit

den modernen Methoden der Struktur- und vor allem der Substruktur-Recherche vergleicht. Denn in den entsprechenden Online-Systemen CAS ONLINE/STN oder DARC werden an einem grafikfähigen Terminal direkt (Sub)strukturen eingegeben; Verschlüsselung und Suche in den z. B. mehr als 7 Millionen Strukturen des *CAS Registry Systems* übernimmt der Computer, die Strukturen der gefundenen Verbindungen erscheinen sofort auf dem Bildschirm, anschliessend kann noch eine Online-Literatur-Recherche durchgeführt werden<sup>[9]</sup>. Aufgrund dieser Erfahrungen und der Informationsbedürfnisse des Chemikers (Organikers) müsste eine moderne *Reaktions-Datenbank* unter anderem folgende Voraussetzungen erfüllen:

*Inhalt:*

- kritisch und repräsentativ ausgewählt;
- Strukturen für Edukte/Produkte/Reagentien einschliesslich Stereochemie;
- Reaktionsbedingungen (z. B. Lösungsmittel, Katalysator, Temperatur, Druck, pH-Wert), Ausbeuten;
- allgemeine Beschreibung des Reaktionstyps und der beteiligten Verbindungsklassen durch Schlagwörter, Kommentar zu Anwendbarkeit, Selektivität, gegebenenfalls Mechanismus;

*Zugriff:*

- interaktiv, einfach und benutzerfreundlich, d. h. vor allem direkte graphische Eingabe von (Sub)strukturen;
- gesamte Information suchbar, Kombination von Struktur-, Schlagwort- und Daten-Recherche;
- schnelle Suche auch in grösseren Datenbeständen, sofortige Ausgabe der Resultate in einer dem Chemiker vertrauten Darstellung;
- mit verbreteter und nicht zu teurer Hardware (Rechner, Terminal) betreibbar.

Die Datenbanksysteme REACCS, SYNLIB und ORAC erfüllen grundsätzlich diese Bedingungen. Allen dreien ist die einfache *Struktureingabe* gemeinsam, die an einem graphischen (Tektronix-PLOT10-kompatiblen) Terminal oder an einem Mikrocomputer mit hochauflösendem Graphik-Bildschirm und entsprechender Emulations-Software mit einem Lichtgriffel oder durch eine mittels Tastatur, graphischem Tablett oder «Maus» bewegte Zeichenmarke (Cursor) erfolgt und so weitgehend das Formel-Zeichnen mit Papier und Bleistift simuliert. Ähnlich einfach und benutzerfreundlich geschieht auch die *Steuerung* des Systems: Statt wie bei den vorhererwähnten Literatur-Datenbanken Befehle über die Tastatur einzugeben (die man jeweils wissen, d. h. erlernen muss), werden dem Benutzer hier auf dem Bildschirm «Menus» mit Befehlen angeboten, aus denen er die gewünschte Operation (Suche, Ergebnisausdruck etc.) durch «Aktivieren» der entsprechenden Stelle des Menus mit Lichtgriffel oder Cursor wählt.

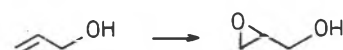
Während die zuvor genannten Literatur- und Struktur-Datenbanken bei Bezah-

lung von Benutzungsgebühren auf einem zentralen Rechner öffentlich online zugänglich sind, werden Programme und Reaktions-Datenbanken dieser drei Systeme lokal («in-house») auf einen Minicomputer – alle Systeme laufen auf DEC VAX-Rechnern mit dem Betriebssystem VMS – geladen und online abgefragt, so dass praktisch keine direkten benutzungsabhängigen Kosten entstehen. Der Unterschied beruht auf der Tatsache, dass diese Programme zur Speicherung (firmen)interner Informationen entwickelt worden sind. Ausser Software für Aufbau und Abfrage interner, selbst erstellter Reaktions-Dateien bieten alle drei Systeme zusätzlich fertige Datenbanken an, was sie in Verbindung mit günstigen Konditionen für akademische Benutzer auch für Hochschulen attraktiv macht. Im folgenden wird zuerst jedes der Systeme anhand eines Beispiels in seinen wesentlichen Funktionen kurz vorgestellt, anschliessend wollen wir ihre Eigenschaften vergleichend diskutieren.

## 3. REACCS (Reaction Access System)

Das Programm für REACCS wurde von Molecular Design Ltd. unter der Leitung von *W. T. Wipke* entwickelt und erstmals 1982 bei Eastman Kodak in USA zur Speicherung firmenrelevanter Reaktionen verwendet<sup>[10]</sup>. REACCS ist neben dem älteren MACCS (Molecular ACCESS System<sup>[11]</sup> für Struktur-Datenbanken) das Kernstück eines umfangreichen integrierten Programm- und Datenbank-Angebots zur «in-house»-Speicherung und -Verarbeitung von Strukturen, Reaktionen, Text und Daten<sup>[12]</sup>. Als Datenbanken für REACCS werden THEILHEIMER (zur Zeit Vol. 15–35, 35 000 Reaktionen mit 70 000 Strukturen), ORGSYN («Org. Synth.», Vol. 1–60, 4667 Reaktionen mit 4997 Strukturen), ein CURRENT LITERATURE FILE<sup>[13]</sup> mit zur Zeit 16 000 Reaktionen aus der Literatur von 1983–1985, und FCDRX ("Fine Chemicals Directory" mit 86 000 Katalog-Eintragungen für käufliche Chemikalien als Edukte) angeboten<sup>[14]</sup>.

*Beispiel A*



Eine Suche nach Epoxidierungen von (substituierten) Allylalkoholen (*Beispiel A*) in der THEILHEIMER-Datenbank läuft folgenderweise ab: Bei Verbindungsaufnahme mit dem Rechner («Login») wählt man die gewünschte Datenbank, und es erscheint das [MAIN]-Menu<sup>[15]</sup> auf dem Bildschirm, das unter anderem Befehle (Optionen) für Wahl oder Wechsel der Datenbank und den (hier nicht interessierenden) Aufbau eigener Reaktions-Dateien enthält (Fig. 1). Die Kopfzeile aller REACCS-Menus bietet die Optionen MAIN, BUILD, SEARCH, VIEWL,

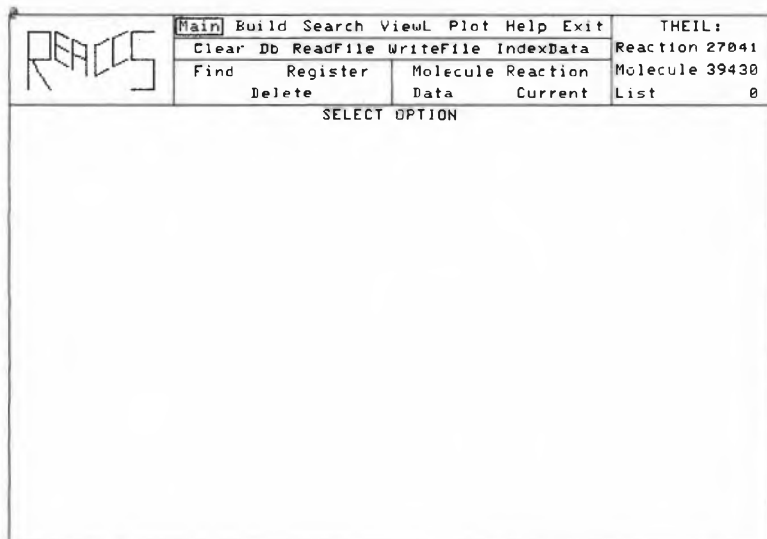


Fig. 1. REACCS: [MAIN]-Menu mit leerem Zeichenfeld<sup>[29]</sup>.

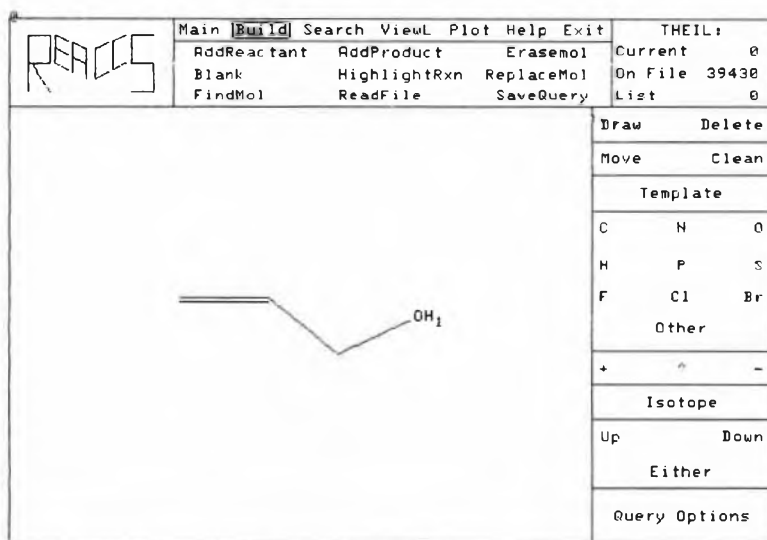


Fig. 2. REACCS (Beispiel A): [BUILD]-Menu (mit DRAW OPTIONS-Teilmenu) mit kompletter Substruktur vor Definition als Edukt mit ADDREACTANT<sup>[29]</sup>.

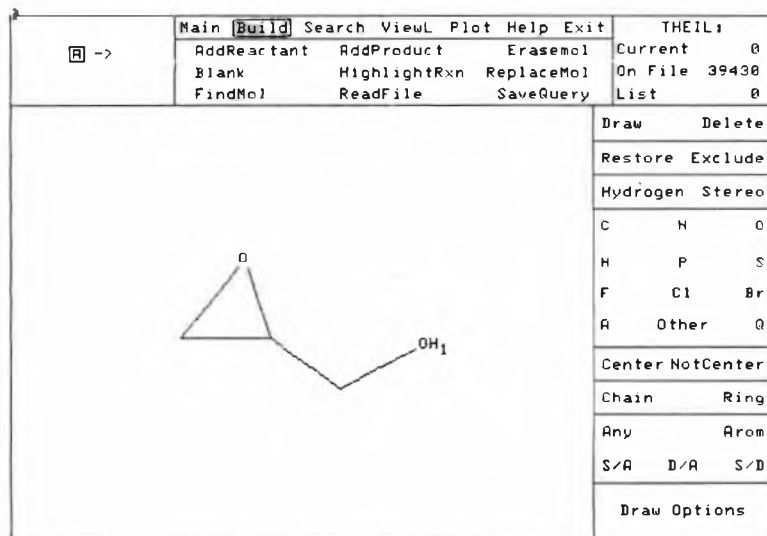


Fig. 3. REACCS (Beispiel A): [BUILD]-Menu (mit QUERY OPTIONS-Teilmenu) mit kompletter zweiter Substruktur vor Definition als Produkt mit ADDPRODUCT<sup>[29]</sup>.

PLOT zum Aufruf der entsprechenden Menus, HELP für Erklärungen und EXIT zum Verlassen des Systems. Zur Eingabe der Fragestellung wird – je nach verwendetem Terminal mit Lichtgriffel oder Cursor – BUILD aktiviert, worauf das in Fig. 2 gezeigte [BUILD]-Menu erscheint. Nach Aktivieren von DRAW beginnt man auf den leeren Schirm zu zeichnen, indem man nacheinander Positionen von Atomen markiert, die dabei laufend automatisch durch Bindungen verknüpft werden und sofort auf dem Schirm erscheinen. Das so eingegebene Skelett besteht aus C-Atomen und Einfachbindungen. Zum Einzeichnen der Doppelbindung im Edukt von *Beispiel A* wird einfach der Zeichenvorgang über der Bindung wiederholt. Dann aktiviert man auf dem rechten Teilmenu (Fig. 2) das "O" und überschreibt die entsprechende Position in der Struktur durch Markieren. Zur Eingabe des H-Atoms aktiviert man QUERY OPTIONS, worauf das rechte Teilmenu ausgetauscht wird. Dieser neue Teil (Fig. 3) enthält spezielle Optionen zur Definition von Bindungen und Atomen, unter anderem HYDROGEN. Letztere wird nun aktiviert, auf die Meldung "ENTER NO. OF H's:" mit der Tastatur "1" eingegeben und dann das O-Atom in der Struktur markiert. Mit DRAW OPTIONS wechselt man zurück ins ursprüngliche Menu und sieht das Formelbild in Fig. 2. Zum Aufbau der Reaktions-Fragestellung wird jetzt diese Struktur mit ADDREACTANT als Edukt definiert. Anschliessend wird analog die Produkt-Substruktur entworfen. Dazu kann man entweder die noch auf dem Schirm verbliebene Edukt-Substruktur modifizieren (hier Doppelbindung mit DELETE in Einfachbindung umwandeln usw.) oder die Struktur neu zeichnen, oder mit TEMPLATE einen Speicher (File) mit «vorfabrizierten» Strukturelementen (Ringe, funktionelle Gruppen) aufrufen und daraus einen Cyclopropanring als «Ausgangsmaterial» wählen. Fig. 3 zeigt die Substruktur nach Definition des H-Atoms. Nach Eingabe dieser kompletten Produkt-Substruktur mit ADDPRODUCT aktiviert man nun SEARCH und erhält dadurch das in Fig. 4 gezeigte [SEARCH]-Menu mit der Fragestellung. Durch RSS (Reaction Substructure Search) wird der eigentlichen Suchprozess eingeleitet, der innert ca. 0.5 Minuten ein Ergebnis lieferte, das mit "[R] 5 HITS RSS / [R] 5 HITS LIST 1" gemeldet wurde; die Antworten waren automatisch in «Liste 1» gespeichert. Zum Betrachten der Resultate wechselt man ins [VIEWL]-Menu (VIEWList), um dort mit FIRST, NEXT oder PREVIOUS die gefundenen Reaktionen «durchzublätern», welche alle die Fragestellung als Substruktur enthalten und relevant waren. Fig. 5 zeigt eine der Antworten; da acyclische Strukturfragmente als *acyclisch* oder *cyclisch* angesehen werden, ist diese auch richtig (mit CHAIN oder RING in [BUILD]/QUERY OPTIONS hätte man das vorher näher spezifizieren bzw. ausschliessen können).

Exakt nur die eigentlich skizzierte Reaktion, d. h. die Epoxidierung von Allylalkohol selbst, wäre dann gefunden worden, wenn man statt der Reaktions-Substruktur-Recherche mit FIND CURRENT in [MAIN] eine exakte Reaktions-Recherche («match») durchgeführt hätte (vgl. Tabelle 1).

Die Fragestellung ist nach der Suche noch gespeichert und kann z. B. nach «Aufruf» auf den Schirm mit QUERY in [VIEWL] für eine analoge Recherche in ORGSYN oder CURRENT LITERATURE FILE verwendet werden. Man kann auch in [BUILD] die Frage gegebenenfalls modifizieren, z. B. indem man im QUERY OPTIONS-Teilmenü mit CENTER bzw. NOTCENTER Bindungen definiert, die an der Reaktion beteiligt sein müssen bzw. nicht beteiligt sein dürfen. Diese Einschränkung ist zur Erhöhung der Relevanz vor allem bei kleinen, häufigen Strukturfragmenten nützlich, aber auch nicht ungefährlich, da die Definitionen der reagierenden Bindungen nicht immer ganz eindeutig sind, und durch «Markierungs»-Fehler in der Datenbank relevante Zitate verlorengehen, wie wir in einigen Fällen feststellten (Faustregel: Immer minimale Beschreibung wählen!).

In REACCS werden also Reaktionen durch (Sub)strukturen für Edukt(e) und/oder Produkt(e) in einer dem Chemiker geläufigen und unmittelbar verständlichen Weise eingegeben und gesucht, wobei man durch Definition der reagierenden Bindungen nötigenfalls noch einschränken kann. Die derzeitige Version 6.0 des Programms kann ausser auf DEC VAX-Computern (Betriebssystem VMS) noch auf PRIME- und IBM-Rechnern (unter MVS/TSO bzw. VM/CMS) benutzt werden; als Terminals sind unter anderem DEC VT100/VT640, HP 2623A und die Tektronix 41XX-Serie geeignet<sup>[12]</sup>. Für Hochschulen sind REACCS-Software und CURRENT LITERATURE FILE gratis, wenn sie sich in bestimmtem Umfang an dessen Aufbau beteiligen. Ohne diese Verpflichtung zahlen sie ca. 8% des kommerziellen Preises für die Software<sup>[16]</sup>. In nächster Zeit wird die THEILHEIMER-Datenbank vollständig mit Vol. 1–35 zur Verfügung stehen (insgesamt 44000 Reaktionen); unter anderem ist noch die volle Integration von REACCS (für Reaktionen) und MACCS (für Strukturen) in einem System geplant<sup>[14]</sup>. Eine Mikrocomputer-Version von REACCS wird bald verfügbar sein<sup>[12,17]</sup>.

#### 4. SYNLIB

##### (SYNthesis LIBrary)

Dieses System ist von W. C. Still (Columbia University, New York) zusammen mit Smith Kline & French Laboratories (Philadelphia) seit 1981 als eine möglichst einfach interaktiv abzufragende «in-house»-Reaktions-Datenbank entwickelt worden<sup>[18]</sup>. Unser Beispiel A kann in SYNLIB auf folgende Weise gesucht werden:

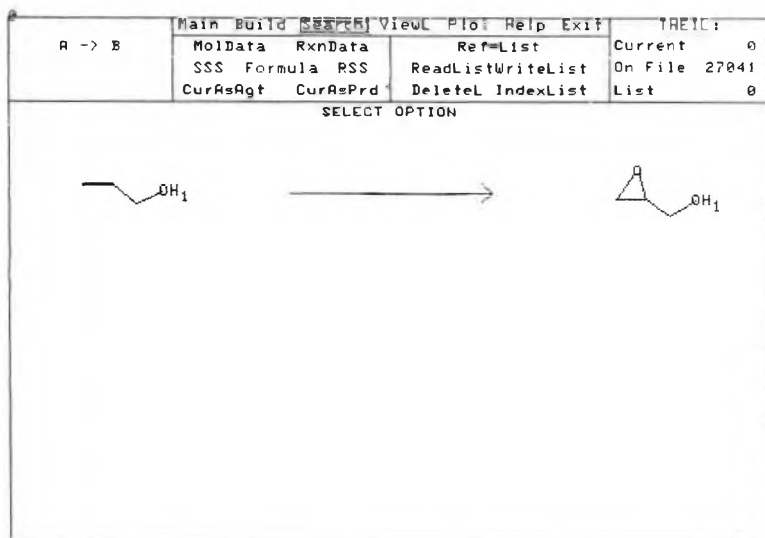


Fig. 4. REACCS (Beispiel A): [SEARCH]-Menu mit vollendeter Fragestellung vor Starten der Reaktions-Recherche mit RSS (reaction substructure search)<sup>[29]</sup>.

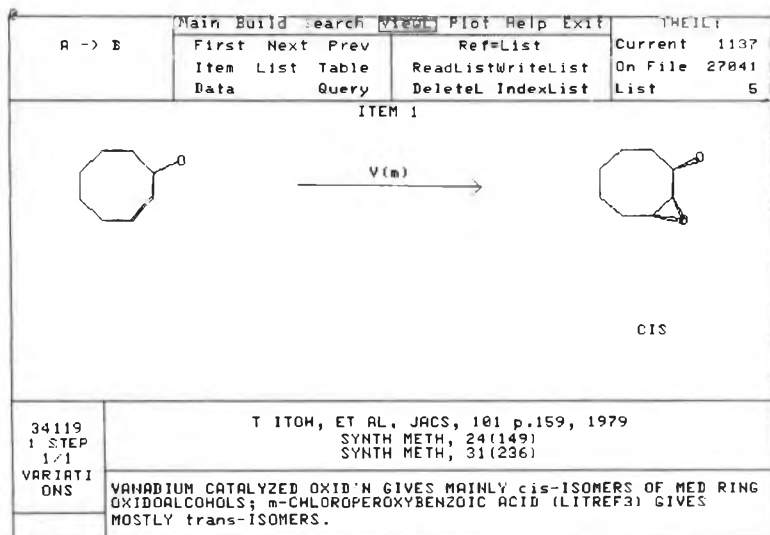


Fig. 5. REACCS (Beispiel A): [VIEWL]-Menu mit einer der gefundenen Reaktionen («Theilheimer», Vol. 34, 119: Literaturzitate, Kommentar)<sup>[29]</sup>.

Beim «Login» wählt man das mit dem Programm gelieferte, zur Zeit 25600 Reaktionen aus der organisch-synthetischen Literatur enthaltende «master file» (oder eine selbst erstellte Reaktions-Datenbank) und gelangt ins [INPUT MODE]-Menu. Man aktiviert DRAW und zeichnet wie schon in Abschnitt 3 erläutert die Produkt-Substruktur, wobei auch hier eine Cyclopropan-Schablone (Template) zur Verfügung steht. Das H-Atom muss wie bei ORAC (vgl. Abschnitt 5) als «Heteroatom» explizit eingezeichnet werden; eine Definition als «Attribut» wie vorher bei REACCS gibt es nicht. Durch Aktivieren von 0 und SUBS nacheinander deklariert man dann alle Atome als an der Reaktion direkt oder indirekt beteiligt (Markierung mit \*, Fig. 6). Da SYNLIB zur Zeit nur Substrukturen von Produkten suchen kann, eine solche Recherche allein aber vermut-

lich auch zahlreiche nicht interessierende Reaktionen geliefert hätte, wählen wir mit RXN das [REACTION MODE]-Menu, auf dem ein Reaktionspfeil mit der Produkt-Substruktur auf beiden Seiten erscheint. Die «Edukt-Seite» wird mit DRAW und DELETE entsprechend modifiziert (Fig. 7), und dann kehrt man mit REQR (required) und RTRN (return) zum [INPUT MODE] zurück. Aus dieser graphischen Information ermittelt das Programm nun die in der Reaktion geknüpften/gebrochenen Bindungen bzw. die addierten/eliminierten Atome als «Transformations-Statistik» zur Beschreibung der Reaktion und Einschränkung der Suche. Wenn man dann aus dem [INPUT MODE] mit SRCH in das [SELECT SEARCH]-Menu wechselt (Fig. 8), bieten sich zwei Suchverfahren an: BROAD SEARCH findet alle Reaktionen, in deren

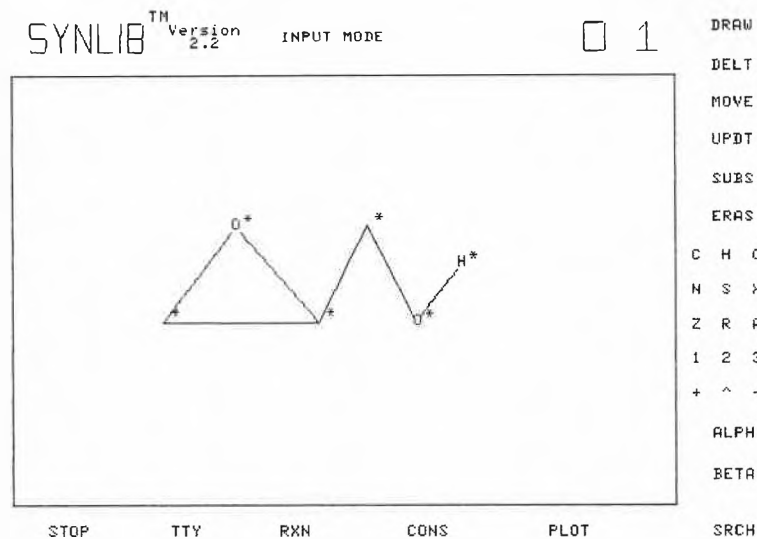


Fig. 6. SYNLIB (Beispiel A): [INPUT MODE]-Menu mit eingezeichneter Produkt-Substruktur (alle Atome durch 0 SUBS mit \* als Suchbedingung markiert) [29].

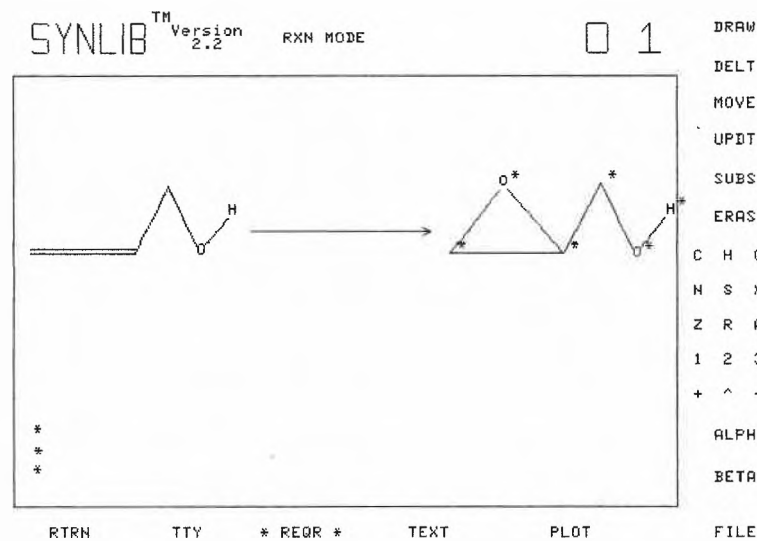


Fig. 7. SYNLIB (Beispiel A): [RXN MODE]-Menu mit kompletter Reaktion (mit REQR = required als Suchbedingung definiert) [29].

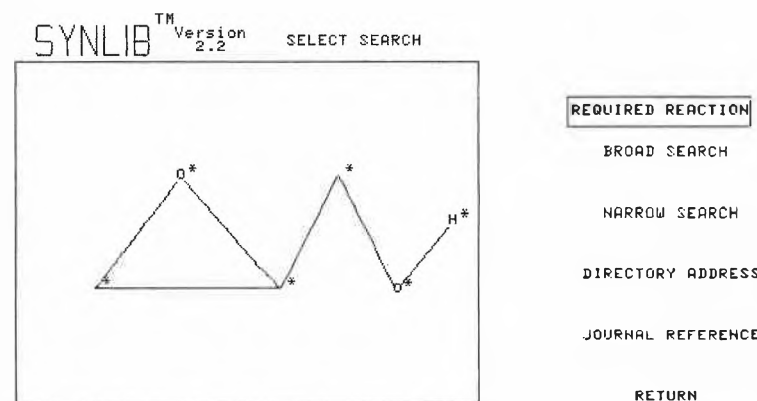


Fig. 8. SYNLIB (Beispiel A): [SELECT SEARCH]-Menu mit (markierter) Produkt-Substruktur und Anzeige einer als Suchbedingung definierten Reaktion (= «Transformation», vgl. Fig. 7) [29].

Produkten die in der eingegebenen Substruktur mit \* markierten, relevanten Atome (vgl. Fig. 6 und Fig. 8) unter anderen ebenfalls markiert sind (d. h. bei Datei-Erstellung als für die Reaktion wesentlich befunden wurden). Bei NARROW SEARCH wird *zusätzlich* verlangt, dass *umgekehrt* alle markierten Atome des Reaktionsprodukts in der Substruktur der Fragestellung vorkommen. In unserem Beispiel war wegen der Bedingung «required reaction» (vgl. Fig. 8) ausser der beschriebenen Produkt-Substruktur-Recherche ja *zusätzlich* verlangt, dass die gefundenen Reaktionen (unter anderem) die ermittelte Transformations-Statistik befolgen. Nach Aktivieren von NARROW SEARCH erschien die erste gefundene Reaktion fast sofort auf dem Schirm. Mit dem Befehl COUNT wurde die Suche dann fortgesetzt und die Zahl der Antworten mit fünf angegeben, die alle relevant waren (Fig. 9).

SYNLIB bietet auf dieser Stufe noch eine originelle Möglichkeit, das primäre Suchergebnis weiter einzuschränken, was vor allem bei BROAD-Recherchen mit relativ vielen Antworten nützlich ist: Mit SRCH ruft man die nächste Reaktion auf und inspiziert so sukzessive die Resultate. Dabei können unerwünschte Reaktionstypen mit OMIT eliminiert werden, weil das Programm dann die Transformations-Statistik der gezeigten Reaktion ermittelt und temporär alle gefundenen Antworten mit der identischen Statistik, d. h. vom gleichen Typ, aus dem Antwort-File eliminiert. Umgekehrt führt FOCUS zur Einschränkung auf die gewünschten Reaktionen. Diese Einschränkungen lassen sich durch RECOVER rückgängig machen, es gehen also keine Ergebnisse verloren. Mit RTRN kehrt man schliesslich zum [INPUT MODE] zurück, wo die noch vorhandene ursprüngliche (Produkt-)Substruktur für eine erneute Suche modifiziert oder mit ERASE gelöscht werden kann. Aktivierte man statt NARROW bei diesem Beispiel BROAD SEARCH, so wurden *zusätzlich* 21 weitere relevante Reaktionen gefunden, die ausser den verlangten noch andere funktionelle Gruppen enthalten (4 der insgesamt 30 Reaktionen waren nicht relevant).

SYNLIB operiert also mit einer Kombination aus Substruktur-Suche (zur Zeit nur für Produkte) und Suche nach Strukturänderungen durch die (formale) Transformations-Statistik, entweder mit REQUIRED REACTION bei der eigentlichen Suche oder zur nachträglichen Verfeinerung der Ergebnisse mit OMIT/FOCUS. Neben diesen zwei impliziten, d. h. vom Programm aus Substrukturen abgeleiteten Strukturänderungen kann man im [CONSTRAINT MODE]-Menu, das mit CONS aus dem [INPUT MODE] zugänglich ist, *explizit* die wichtigsten gebrochenen/geknüpften Bindungen bzw. addierten/eliminierten Atome definieren: für Beispiel A B(roken) C=C, F(ormed) C-O (Fig. 10). Darüber hinaus sind weitere Einschränkungen wie z. B. verbotene funktio-

nelle Gruppen, minimale Ausbeute, allgemeine Reaktionsbedingungen wählbar.

SYNLIB versucht, zur einfachen Benutzung mit einem Minimum an Optionen auszukommen (besonders im Vergleich mit REACCS), was natürlich etwas auf Kosten der Flexibilität geht. Der Suchvorgang ist effizient, aber nicht so leicht durchschaubar. Die aktuelle, seit Frühjahr 1984 erhältliche Version 2.2 von SYNLIB benötigt ca.  $10 \cdot 10^6$  32-Bit-Worte Speicherplatz, wovon etwa 3/4 durch die Reaktions-Datenbank belegt werden<sup>[20]</sup>. Hochschulen erhalten Programm, Datenbank und Ergänzungen (ca. 4000 Reaktionen pro Jahr) *gratis*, wenn sie beim weiteren Ausbau mitarbeiten, indem sie in bestimmtem Umfang wichtige Reaktionen aus der Literatur über organische Synthesen auswählen und auf SYNLIB-Formularen zur Eingabe in die Datenbank vorbereiten<sup>[19]</sup>. Von diesem Angebot haben vor allem in USA schon viele Institute Gebrauch gemacht. Als Terminal können unter anderem Tektronix 4010 oder 4014 und deren Emulationen benutzt werden; gut geeignet sind z. B. DEC VT100/VT640 oder Apple Macintosh mit Emulations-Software Tekalike<sup>[20]</sup>. Die nächste Version von SYNLIB (3.0) soll durch höhere Suchgeschwindigkeit und Erweiterungen der Transformations-Statistik die Selektivität verbessern. Mit der Version 3.0 wird man auch Edukt-Substrukturen suchen sowie Resultate speichern und kombinieren können. Neben der Anpassung an weitere Terminals sind Mikrocomputer-Versionen von SYNLIB geplant (die erste für IBM PC AT soll spätestens im Sommer 1986 erhältlich sein)<sup>[17, 20]</sup>.

## 5. ORAC

### (Organic Reaction Access by Computer)

Das dritte System wird seit Oktober 1983 an der Wolfson Unit for Computer-Aided Design of Organic Synthesis der University of Leeds unter der Leitung von A. P. Johnson entwickelt, eine erste Version wurde im Frühjahr 1984 vorgestellt<sup>[21, 22]</sup>. Es gibt einen Zusammenhang zwischen der Entwicklung von ORAC und Syntheseplanungs-Programmen wie LHASA<sup>[22, 23]</sup>, die bekanntlich auch interne Reaktions-Datenbanken brauchen. Derzeit enthält ORAC 10000 Reaktionen aus der Literatur, das Ziel für 1986 ist 25000, mit einem weiteren jährlichen Zuwachs von 20000 Reaktionen für die nächsten Jahre. Hochschulen, die sich am Aufbau der Datenbank beteiligen, erhalten diese und das Programm *gratis*<sup>[24]</sup>. Als Terminal eignen sich unter anderem DEC VT100/640 oder 650, HP 2647A und Mikrocomputer wie IBM PC mit Emutek-Software<sup>[24]</sup>.

Nach «Login» in ORAC erscheint das [INTRODUCTION]-Menu mit den Optionen SEARCH FOR REACTIONS, HELP und EXIT sowie ADD NEW REACTION für den Aufbau eigener Da-

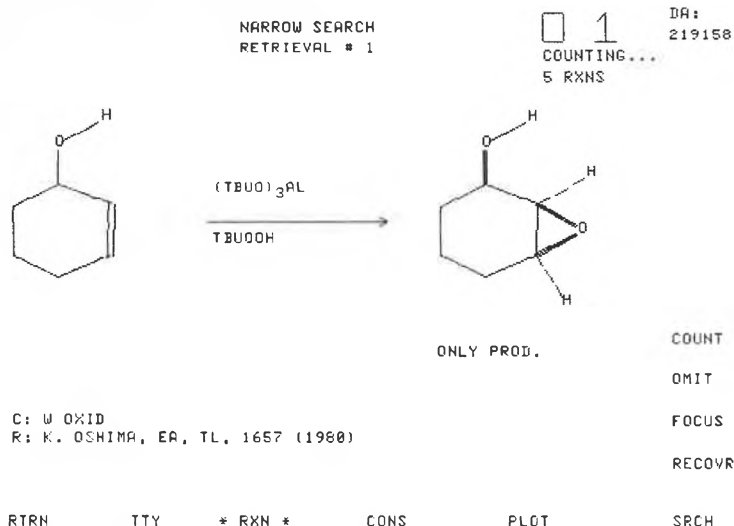
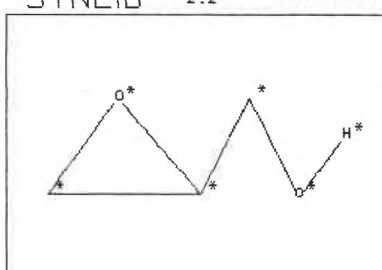


Fig. 9. SYNLIB (Beispiel A): [NARROW SEARCH]-Menu mit erster gefundener Reaktion und Anzahl der insgesamt gefundenen Antworten (nach Aktivierung von COUNT)<sup>[29]</sup>.

SYNLIB<sup>TM</sup> Version 2.2

CONSTRAINT	MODE	UPDT	RESET	RTRN
AROMATICITY?	*YES			NO
FORBIDDEN_FB?	*YES			NO
FULL_STRUCT?	YES			* NO
INTER/INTRA?	*YES			NO
C=C STEREO?	*YES			NO
STRATEGIC BOND				
TEXT				
SET MAX NO OF STEPS				
	1	2	3	4 5 6 7
MIN YLD?	1%			40% 70%
DATE	BEFORE		AFTER	
ATOM	*ADD		DELT	
	O N X		SZ C	



RX	CNDX	*REQU	FORB
	U	* M	S
ACID	BASE	OMET	OXID REDU NUCL
ELEC	****	FRAD	PHOT HEAT
BOND	*FORM	*BREAK	
C-C	BC=C	FC-O	C=O C-N
C=N	C*SZ	Y*Y	TRI

Fig. 10. SYNLIB: [CONSTRAINT MODE]-Menu mit (markierter) Produkt-Substruktur und explizit definierten Bindungsänderungen (B(roken) C=C, F(ormed) C-O; alternative Formulierung der Fragestellung zu Beispiel A)<sup>[29]</sup>.

teien. Durch Aktivieren der ersten Option gelangt man in das zentrale [QUERY MENU] (Fig. 11). Zur Suche in unserem Beispiel A wird SUBSTRUCTURE aktiviert, worauf das entsprechende Menu erscheint. Wie bei den anderen Systemen gibt man die Produkt-Substruktur ein (DRAW ist automatisch aktiv), aktiviert dann PRODUCT auf der rechten Menu-Seite zur Definition der «Rolle» der Substruktur und RXBOND (reaction bond) zur Beschreibung der Bindungsänderungen in der Reaktion. Dabei wird die «Bindungsordnung» definiert, welche eine Bindung vor (wenn wie hier Substruktur als «Product» angegeben) oder nach der Reaktion (bei Eingabe einer Edukt-Substruktur) hat: Auf die Frage «CHANGED BOND

TO (Z, S, D, T, A/YES/NO/ANY):» wurde zuerst «z(ero = Bindung vor der Reaktion nicht vorhanden)» eingetippt, danach die beiden «neuen» Bindungen des Oxiran-Rings markiert, RXBOND erneut aktiviert, dann auf die wiederholte Frage mit «d(ouble bond im Edukt)» geantwortet und die dritte Ringbindung markiert (Fig. 12). Alle Definitionen werden durch ein «\$» an der Bindung angezeigt; diese Festlegung kann jeweils mit QUERY überprüft werden. Die so beschriebene «Reaktion» wurde durch ENTER (Rückkehr zum [QUERY]-Menu) und danach SEARCH gesucht. Im Mittelteil von [QUERY MENU] erscheint in diesem Fall die Meldung «SET 1 8 (hits)». Die Resultate können mit DISPLAY inspiziert wer-

QUERY OPTIONS	CARD SETS	HIT COUNT	COMMANDS
AUTHORS JOURNAL NAME YEAR OF PUBLICATION ORAC CARD NUMBER REACTANT PRODUCT REACTION SUBSTRUCTURE REACTION KEY REAGENT KEY SOLVENT KEY YIELD TEMPERATURE ACTUAL REAGENTS NAME OF REACTION			SEARCH DISPLAY CLEAR SETS WIPE SETS RESTORE SETS SAVE SETS DEFINE SETS RESTART QUERY HELP STOP
			LOGIC
			AND OR AND NOT

Fig. 11. ORAC: [QUERY]-Menu mit Such-Optionen (links), Antwortfeld (Mitte), Steuerbefehlen sowie logischen Operatoren (rechts) und Textfeld (unten) [29].

SHIFT	ROTATE	SCALE	SOLID	DASH	WEDGE	HELP	ENTER	EXIT
UPDATE	REDRAW	ERASE	DRAW	DELETE	MOVE	STORE	FETCH	TEMPLATE
C H O N F Cl Br I S P B Si : o - + [Diagrams of rings and bonds]								
-NOO -COO -SO <sub>2</sub> C A1 A2 A3 A4 A Q Symbol Rings Groups								
Changed bond to (Z,S,D,T,A/Yes/No/Any):Z Changed bond to (Z,S,D,T,A/Yes/No/Any):D								
								REACTANT > PRODUCT R_OR_P R_AND_P R_NOT_P P_NOT_R QUERY RING AROMATIC ALTBOND FUSION RXBOND RXATOM

Fig. 12. ORAC (Beispiel A): [SUBSTRUCTURE]-Menu (> PRODUCT) mit kompletter Produkt-Substruktur und durch RXBOND definierten Bindungsänderungen zur Beschreibung der Reaktion («prompts» nach Aktivierung von RXBOND und zugehörige Tastatur-Eingabe "z(ero)" oder "d(ouble)" im Textfeld; entsprechende Bindungen mit \$ markiert) [29].

Card:4760 TANAKA S, YAMAMOTO H, NOZAKI H, SHARPLESS K B  
Box :1  
Skip:1 J AM CHEM SOC., 5254, 11974)  
No:6 of:8 SHARPLESS

Other examples given, yields not quoted but probably high.  
>99 % Erythro isomer produced in this case. Lower temperatures favour greater stereoselectivity using vanadium catalysis. Use of Mo(CO)<sub>6</sub> as catalyst requires reflux temperatures.

Compiled by: HAYWARD J Experimental: Not given

NEXT BACK SKIP KEEP STOP KEYS PLOT SITE VIEW HELP

Fig. 13. ORAC (Beispiel A): [DATA CARD] einer der gefundenen sieben relevanten Reaktionen (vgl. Fig. 19) [29].

den; Fig. 13 zeigt eine der sieben relevanten Antworten. Die nicht relevante Reaktion war, wie auch eine in SYNLIB, die Epoxidierung der analogen Carbonsäure – wir hatten ja bei der Substruktur-Recherche in allen Systemen Substitution am  $\alpha$ -C-Atom nicht ausgeschlossen! In ORAC kann man Reaktionen auch anders suchen: Bei Fragen mit *exakten* Strukturen für Produkt und/oder Edukt benutzt man REACTANT bzw. PRODUCT (Menu ähnlich SUBSTRUCTURE). Bei REACTION gibt man nur die an der Reaktion beteiligten Atome und Bindungen von Edukt und Produkt ein. Dieses Verfahren ist wegen der Definition von «beteiligt» oft problematisch (vgl. die Markierung reagierender Bindungen in REACCS) und empfiehlt sich vor allem für allgemeine Fragestellungen wie Umwandlungen funktioneller Gruppen oder Umlagerungen; die Suche ist extrem schnell. Die von uns am häufigsten angewendete Methode war SUBSTRUCTURE, denn hier wird eine Substruktur in ihrer «Rolle» bei einer Reaktion gesucht: «Reactant» bzw. «Product» bzw. logische Kombinationen davon benutzt man zur Definition von Substrukturen, die in der Reaktion umgewandelt bzw. gebildet bzw. nicht verändert werden (siehe Optionen auf der rechten Menüseite in Fig. 12). Ausserdem kann durch explizite, spezifische Definition der Bindungsänderungen bzw. der an der Reaktion (nicht)beteiligten Atome oder Bindungen die gesuchte Reaktion weiter definiert werden.

ORAC sucht also offensichtlich Reaktionen mit einer Kombination aus Substruktur-Recherche und vor allem (in REACTION sogar ausschliesslich) Suche nach Bindungsänderungen. Diese Formulierung der Fragestellung unterscheidet sich wesentlich von der Eingabe bei REACCS, die direkter, aber zur Definition einer Strukturumwandlung weniger genau ist; im Vergleich mit der grundsätzlich ähnlichen Suche via die überwiegend implizite Transformations-Statistik in SYNLIB besticht ORAC mit RXBOND, RXATOM durch eine präzise, explizite Beschreibung der Strukturänderungen. Die dafür erforderliche, in der Reaktionsdatenbank gespeicherte «Korrespondenz» der reagierenden Atome – Fig. 14 zeigt eine weitere Antwort aus unserer Recherche in entsprechender Darstellung – zeichnet ORAC ebenso aus wie Indexierung und einfache Text-Suche (vgl. Abschnitt 6). Die neueste Version 6.0 (Februar 1986) bringt weitere Datenfelder sowie Verbesserungen der Programm-Struktur. Damit werden für spätere, bereits in Entwicklung befindliche Versionen (6.x) Ergänzungen möglich: (Sub)struktursuche aller Reagentien (bisher nur über Summenformel mit ACTUAL REAGENT suchbar, vgl. Abschnitt 6), Zwischenprodukte und Synthons; automatische Erkennung von tautomeren Strukturen; Suche nach allen Stereoisomeren exakter Strukturen; hierarchische Ordnung von Schlag-

wörtern; einfachere Benutzung der Templates<sup>[21]</sup>.

## 6. Vergleich der Systeme REACCS, SYNLIB und ORAC

Aus der kurzen Beschreibung der drei Datenbanksysteme und ihrer Handhabung am gleichen Reaktionsbeispiel (*Beispiel A*) in den Abschnitten 3–5 wird bereits deutlich, dass trotz vieler Gemeinsamkeiten (z. B. der Bedienung) jedes System die Recherchen auf unterschiedliche Weise durchführt. Im Aufbau unterscheiden sich die Systeme beträchtlich: Während bei REACCS jedes der fünf Menus jederzeit aufgerufen werden kann (Fig. 15), wird bei SYNLIB und ORAC der Benutzer durch eine hierarchische Anordnung (Fig. 16 und 17) mehr angeleitet, was für Anfänger einfacher und bequemer sein mag. Die einzelnen Menus sind ungeachtet der Unterschiede alle sehr benutzerfreundlich gestaltet, die Optionen in den Menus erklären sich praktisch von selbst. Als Gebrauchsanweisung dienen gute Handbücher in englischer Sprache mit zahlreichen Beispielen. Für REACCS und ORAC gibt es auch online «help messages», die pauschal oder spezifisch in jedem Menu mit HELP aufgerufen werden können; ein «Steckenbleiben» ist damit (fast) ausgeschlossen. Bei SYNLIB fehlt dies, was uns angesichts der einfachen Systemstruktur und des besonders ausführlichen Handbuchs zur Zeit nicht als wesentlicher Mangel erscheint. Texteingaben über die Tastatur werden fast nur erforderlich, wenn man nach Schlagwörtern und Daten sucht, oder zum Speichern/Aufrufen/Löschen von Strukturen und Reaktionen. In diesen Fällen stellt das System eine Frage («prompt»). Bei ORAC und REACCS fallen diagnostische Fehlermeldungen positiv auf, beispielsweise wenn man vorschriftswidrige oder unvollständige Operationen unternimmt. Spätestens beim Versuch, eine Struktur mit fünfbindigem Kohlenstoffatom suchen zu lassen, erscheint auch in SYNLIB eine Fehlermeldung wegen der Valenzüberschreitung. Fehlbedienungen, die Anfänger oft befürchten, sind so weitgehend ausgeschlossen. Da die Aktivierung einer Menu-Option akustisch («beep») oder optisch quittiert wird – gleichsam als «Treffermeldung» – und die Option dann markiert bleibt, geht man im Programm nicht «verloren». Soll eine gewählte Option nicht gelten, so wird einfach ein anderer Befehl aktiviert, mit der «carriage return»-Taste umgeht man die Beantwortung einer Frage. «Notbremsen» gibt es auch, denn mit der Taste CONTROL C (Rückkehr zum Betriebssystem des Rechners) lassen sich Suchvorgänge und überhaupt alle Operationen unterbrechen. In REACCS und SYNLIB führt dies aber zum Verlust vorhergehender Resultate, während sie in ORAC danach noch verfügbar sind (man kann hier mit CONTROL Z nur eine Suche unterbrechen und kehrt dann zum [QUERY MENU] zurück). Die Suchzei-

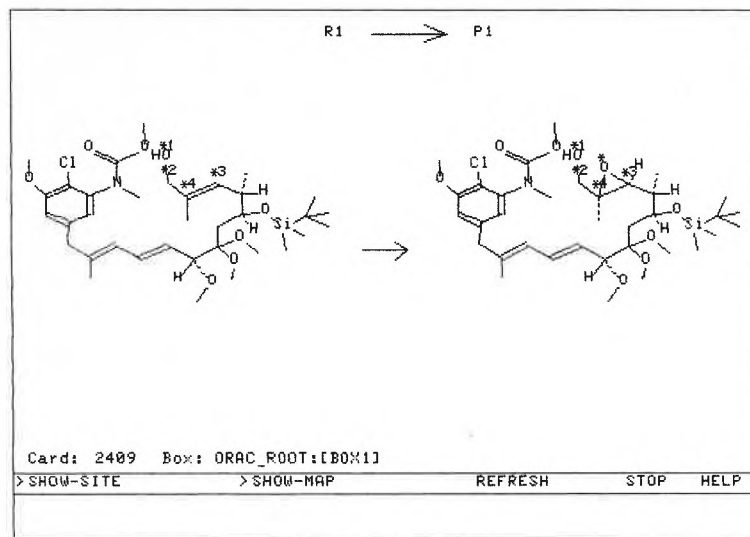


Fig. 14. ORAC: [VIEW]-Menu einer Reaktion zu Beispiel A mit durch SHOW-SITE und SHOW-MAP hervorgerufener Markierung reagierender (\*) und korrespondierender (Nr.) Atome<sup>[29]</sup>.

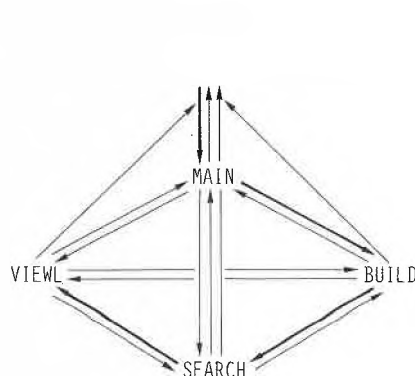


Fig. 15. Menus in REACCS (ohne [PLOT] und [HELP]; verstärkte Linien: Ablauf der Recherche zu Beispiel A).

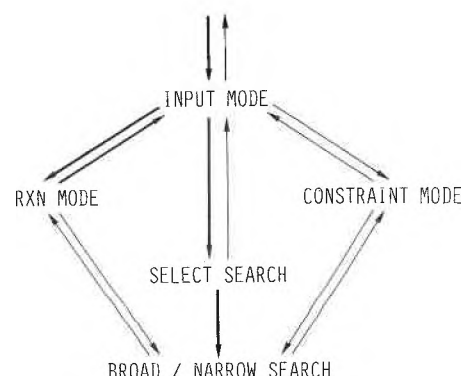


Fig. 16. Hierarchie der Menus in SYNLIB (ohne [PLOT] und TTY; verstärkte Linien: Ablauf der Recherche zu Beispiel A).

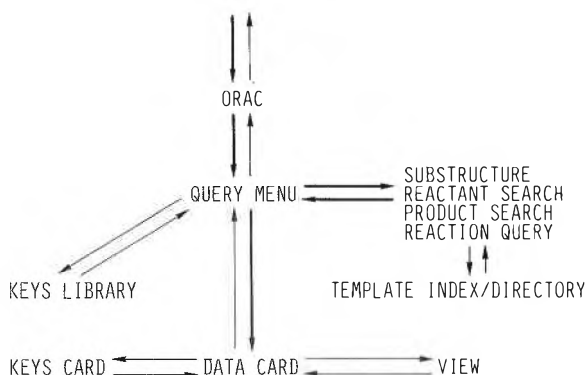


Fig. 17. Hierarchie der Menus in ORAC (ohne [HELP]; verstärkte Linien: Ablauf der Recherche zu Beispiel A).

ten haben in allen Systemen meistens die Grössenordnung 0.5–1.5 Minuten; exakte Struktursuchen in REACCS/ORAC, REACTION in ORAC sind sogar noch schneller (innert Sekunden), und nur bei allgemeinen Fragestellungen (z. B. Substruktur C=C=C-C) oder vielen Benutzern am Rechner dauert es länger. Das

auch terminal-abhängige «Füllen» des Bildschirms ist übrigens bei der Darstellung von Resultaten in REACCS unter vergleichbaren Bedingungen deutlich langsamer<sup>[25]</sup>.

Die Eingabe der Fragestellung erfolgt in allen Systemen durch Zeichnen von Strukturen auf einem graphischen Terminal,

wobei auf dem Bildschirm mit Lichtgriffel oder Cursor Atome markiert und dabei fortschreitend mit Bindungen verknüpft werden ( $n + 1$  Markierungen für  $n$  Bindungen); das Unterbrechen an Verzweigungen («lift pen») wird durch zweimaliges Markieren des selben Atoms (CONTINUOUS DRAW in REACCS, DRAW in ORAC) oder in SYNLIB durch erneutes Aktivieren von DRAW bewirkt. In allen Systemen sind Atome a priori Kohlenstoffatome («default»), alle Bindungen Einfachbindungen. Die wichtigsten Heteroatome sind in den Menüs vorhanden, seltenere werden über die Tastatur eingegeben. Mehrfachbindungen erzeugt man durch mehrmaliges «Überschreiben». Solche Änderungen oder Löschen von Bindungen sowie Atomen mit DELETE werden bei REACCS unmittelbar auf dem Schirm sichtbar, bei ORAC erscheint auf manchen Terminals, bei SYNLIB immer nur eine Markierung («D» für Doppelbindung «xx» für gelöschte Bindungen/Atome). Erst mit UPDATE vollzieht sich das «Umzeichnen» der Struktur. REACCS hat die differenziertesten Funktionen für das Zeichnen und vor allem für die Manipulation von Strukturen, Fragmenten, funktionellen Gruppen und einzelnen Atomen. Diese Vorzüge – auf Kosten der Einfachheit – fallen allerdings wesentlich nur bei der Eingabe von Strukturen für das Erstellen von Reaktions-Datenbanken, kaum bei der Recherche ins Gewicht. ORAC verfügt hingegen nur über die wichtigsten Funktionen zum «Schönen» von Strukturen: automatische Skalierung, Ausgleich von Bindungslängen und -winkeln mit REDRAW (REACCS: CLEAN), SHIFT/ROTATE/MOVE für Translation/Rotation von Strukturen und Translation einzelner Atome. In SYNLIB gibt es nur MOVE; das Zentrieren und die Skalierung werden automatisch mit UPDATE erreicht, nicht aber Bindungsausgleich. Grössere Strukturen wie Steroide lassen sich durch Benutzung von Templates (vgl. Abschnitt 3) einfacher und schneller konstruieren als durch direktes Zeichnen. Während in ORAC und REACCS die Verknüpfung von Templates für Ringe und/oder funktionelle Gruppen sozusagen nach dem Baukastenprinzip durch sukzessiven Aufruf möglich ist, kann in SYNLIB nur ein einziges Template aufgerufen und dann verändert werden. ORAC hat die grösste Auswahl an Templates; ausser dem in Fig. 12 links gezeigten Teil mit Heteroatomen und den wichtigsten Ringen stehen noch fünf direkt zugängliche Teilmenüs für Heterocyclen sowie eines für funktionelle Gruppen zur Verfügung, dazu kommen mehrere mit TEMPLATE aufrufbare Files mit weiteren, grösseren Strukturen (z.B. Steroide, Aminosäuren). Alle Systeme gestatten es, zusätzliche Files mit eigenen Templates zu speichern und zu benutzen.

Folgende Optionen für Erweiterungen oder Einschränkungen von Substrukturen sind für Recherchen noch wichtig: Da bei

Substruktur-Recherchen im Rahmen der freien Valenzen beliebige Substitution erlaubt ist, muss dies gegebenenfalls mit H-Atomen ausgeschlossen werden (vgl. *Beispiel A*). Dazu werden in REACCS die Anzahl H-Atome mit der Option HYDROGEN eingetippt und die betreffenden Atome markiert, hingegen müssen in SYNLIB und ORAC die H-Atome (umständlicher) explizit eingezeichnet werden. Gruppen von Atomen als Alternativen für bestimmte Positionen in Substrukturen können frei definiert werden; in ORAC (AROMATIC) und SYNLIB (R, A) sind «Atome» als beliebige aliphatische oder aromatische Reste definierbar. ORAC und REACCS haben die «allgemeinen Atome» A (= alle ausser H) und Q (= alle ausser C oder H), SYNLIB hat X für die Halogene; allein<sup>[30]</sup> REACCS ermöglicht den Abschluss bestimmter Atome an einer Position und die Spezifikation «Isotop». Als Variablen für Bindungen in Substrukturen lässt ORAC mit ALTBOND den allgemeinen Typ «any» sowie beliebige Kombinationen bestimmter Bindungstypen zu, REACCS neben «any» die Paare «single/aromatic», «double/aromatic», «single/double», SYNLIB nur die Kombinationen der Variablen Z «single/double», «double/triple» (die Typen «any» und «aromatic» fehlen als explizite Definition). Man kann einschränken, dass in einem Strukturfragment bestimmte Bindungen (REACCS) oder Bindungen sowie Atome (ORAC) zu einem Ring gehören oder nicht gehören müssen; ohne diese Spezifikation können acyclische Strukturfragmente ja auch Teil eines Rings sein (vgl. *Beispiel A*). In ORAC lässt sich mit FUSION weiter spezifizieren, ob eine Bindung zwei Ringen gemeinsam sein soll. So fanden wir in *Beispiel B* durch Markierung der Bindungen als «not fused» in SUBSTRUCTURE (R NOT P) genau eine Reaktion, bei der das isolierte Ringsystem entsteht, während ohne diese Einschränkung drei weitere Reaktionen zur Bildung komplexerer Ringsysteme gefunden wurden<sup>[26]</sup>.

#### Beispiel B



Die Stereochemie an Doppelbindungen wird von allen drei Datenbanksystemen bei entsprechender Eingabe erkannt und meistens selektiv gesucht. Für Chiralitätszentren werden die gleichen Konventionen zur Darstellung, aber mit unterschiedlichen Befehlen zur Markierung der Bindungen in REACCS/SYNLIB/ORAC verwendet: UP/BETA/WEDGE für Bindungen aus der Zeichenebene heraus, DOWN/ALPHA/DASH für solche unter die Ebene. REACCS erkennt in Struktursuchen die korrekte absolute, in Substruktur-Recherchen jedoch nur die relative Konfiguration. ORAC berücksichtigt in Substruktur-Recherchen die absolute (ein

Asymmetriezentrum) bzw. relative Konfiguration (Strukturen mit mehreren Asymmetriezentren). *Beispiel C* zur *trans*-Hydroxylierung disubstituierter *cis*-Olefine brachte im THEILHEIMER (REACCS) fünf Reaktionen, davon vier mit korrekter Stereochemie und eine ohne Spezifikation in der Produktformel (aber:  $\text{KMnO}_4/\text{NaOH} = \text{cis}$ ). In SYNLIB fanden wir vier *cis*- und nur eine *trans*-Hydroxylierung, in ORAC eine *trans*-Hydroxylierung (neben vier *cis*-Hydroxylierungen in der Datenbank). Aufgrund einer (beschränkten) Zahl von Beispielen stellten wir fest, dass im THEILHEIMER-File nicht selten die Stereochemie-Information fehlt, in ORAC ist die vollständige Beschreibung der asymmetrischen Zentren einschliesslich H-Atome gemäss den vom System verwendeten Konventionen kritisch für korrekte Resultate, und bei SYNLIB konnten wir überhaupt keine Differenzierung nach Konfiguration an tetraedrischen Zentren entdecken.

#### Beispiel C



Wie bereits erwähnt, erlaubt SYNLIB nur eine beschränkte explizite Definition der Bindungsänderungen bei der Reaktion im [CONSTRAINT MODE], die ausserdem nicht mit bestimmten Bindungen der Substruktur korreliert werden können. Auch die Transformations-Statistik ist ja an sich keine präzise Beschreibung. In REACCS können Bindungen in Edukt- oder Produkt-(Sub)struktur mit CENTER als an der Reaktion beteiligt und mit NOT-CENTER als nicht beteiligt deklariert werden. In der Option HIGHLIGHTRXN in [BUILD] erkennt das System mit «Perceive» Reaktionszentren sogar automatisch, die Brauchbarkeit ist aber noch beschränkt. Ausser den erwähnten Definitionsproblemen kann sich auf eine Suche auch negativ auswirken, dass die Korrespondenz der Reaktionszentren nicht gespeichert ist: Bei der Suche nach Wittig-Reaktionen (*Beispiel D*) fanden wir in REACCS mit markierten C=C- und C=O-Bindungen 472 Reaktionen, darunter sehr viele Aldol-Kondensationen; durch Einschränkung mit der Substruktur P-C-C als zweitem Edukt verminderte sich die Zahl auf 11 Wittig-Reaktionen. In SYNLIB führte die ursprüngliche Fragestellung zu 14 Reaktionen (BROAD SEARCH, 6 relevant<sup>[27]</sup>), in ORAC mit REACTION (Fig. 18, implizite Definition aller eingegebenen Atome/Bindungen als an der Reaktion beteiligt) zu 7 Reaktionen (alle relevant). ORAC hat ja als einziges System die Korrespondenz der Reaktionszentren in der Datenbank gespeichert (vgl. Fig. 14) und ermöglicht daher nicht nur qualitative (beteiligt/nicht beteiligt), sondern exakte

quantitative Beschreibungen von reagierenden Bindungen (RXBOND in [SUBSTRUCTURE]). Bei allen drei Systemen kann in den gefundenen Reaktionen die bei der Datei-Erstellung gewählte Definition der Reaktionszentren in den Antworten zum Vergleich «eingebildet» werden.

Beispiel D



Alle drei Systeme bieten die Möglichkeit, zusätzlich zu den mitgelieferten auch eigene Reaktions-Datenbanken zu erstellen und abzufragen. Die Datenstruktur der Programme, d. h. Anzahl und Format der Datenfelder, die Art und Umfang der abfragbaren Information bestimmen, sind bei SYNLIB und ORAC festgelegt; bei REACCS können diese Spezifikationen vom Benutzer geändert werden. Gerade für Schlagwort- und Datensuche muss hier ein Kompromiss zwischen einfacher Benutzung (also Suchmöglichkeiten im Menu wie bei ORAC, vgl. Fig. 11) und Flexibilität getroffen werden, welcher bei ORAC und vor allem bei SYNLIB den ersten, bei REACCS hingegen den zweiten Aspekt betont. Die erwähnten umfangreichen Möglichkeiten zur Eingabe von Strukturen und Daten in REACCS kommen hier voll zum Tragen. Um die Menus nicht zu überladen, sind bei REACCS manche Funktionen nur über Tastatur-Befehle zu erreichen. In REACCS und SYNLIB müssen die Suchen in jeder Datenbank mit der gleichen Fragestellung einzeln nacheinander durchgeführt werden, während ORAC normalerweise alle Datenbanken einschliesslich der «private files» automatisch sequentiell absucht.<sup>[30]</sup>

Die unterschiedlichen Suchverfahren für Reaktionen in den drei Systemen, so wie sie sich dem Benutzer darbieten, haben wir bereits in den Abschnitten 3–5 diskutiert. Zusammenfassend ist festzustellen, dass REACCS Reaktionen struktur-orientiert sucht, ORAC und besonders SYNLIB hingegen sich auf die Struktur-, d. h. Bindungs-Änderungen («Transformation») in der Reaktion konzentrieren; Tabelle 1 stellt die Möglichkeiten der Suche von Reaktionen, Substrukturen (freie Valenzen beliebig substituiert) und Strukturen («match», freie Valenzen automatisch mit H aufgefüllt) gegenüber.

REACCS offeriert als einziges System die Möglichkeit, wie in einer Struktur-Datenbank exakte Struktur- (FIND CURRENT in [MAIN]) bzw. Substruktur-Recherchen (SSS = SubStructure Search) an allen in der Reaktions-Datenbank als Edukt, Produkt, Reagentien etc. vorkommenden Verbindungen durchzuführen, und zwar entweder über eine eingegebene Struktur oder über die vollständige bzw. partielle Summenformel (FORMULA)<sup>[28]</sup>. Zur Beantwortung der Frage nach «Bedin-

The screenshot shows a menu with the following options: SHIFT ROTATE SCALE SOLID DASH WEDGE HELP ENTER EXIT; UPDATE REDRAW ERASE DRAW DELETE MOVE STORE FETCH TEMPLATE. Below these are lists of atoms (C, H, O, N, F, Cl, Br, I, S, P, B, Si), symbols (\*, o, -, +), and rings (diamond, pentagon, hexagon, heptagon, octagon, nonagon, decagon). There are also options for functional groups (-NOO, -COO, -SO2C) and a section for 'R # \* Symbol' and 'Rings Groups'. A chemical structure of a carbonyl group is shown. At the bottom, a reaction scheme shows the conversion of a carbonyl group to an alkene.

Fig. 18. ORAC (Beispiel D): [REACTION]-Menu nach Eingabe der Fragestellung und vor Verlassen des Menus mit ENTER zur Durchführung der Suche (Eingabe: Edukt-Substruktur zeichnen → REACTANT aktivieren → Substruktur zum Produkt modifizieren → PRODUCT aktivieren)<sup>[29]</sup>.

Tabelle 1. Vergleich der Möglichkeiten zur Suche nach Reaktionen in REACCS, SYNLIB und ORAC.

	REACCS	SYNLIB	ORAC
via Eingabe von Reaktionen (Edukt → Produkt)			
identische Reaktion	FIND CURRENT	–	REACTANT AND PRODUCT <sup>[a]</sup>
Reaktion vom gleichen Typ	RSS	SUBS/RXN (SUB/CONS)	SUBSTRUCTURE (REACTION) <sup>[b]</sup>
via Eingabe von Substrukturen			
Edukt	RSS <sup>[c]</sup>	–	SUBSTRUCTURE
Produkt	RSS <sup>[c]</sup>	SUBS	SUBSTRUCTURE
Reagens/Lösungsmittel etc.	<sup>[c]</sup>	–	–
via Eingabe von (exakten) Strukturen			
Edukt	CURASAGT <sup>[d]</sup>	–	REACTANT
Produkt	CURASPRD <sup>[d]</sup>	–	PRODUCT
Reagens/Lösungsmittel etc.	CURASAGT <sup>[d]</sup>	–	–

<sup>[a]</sup> Kombination von zwei Recherchen, vgl. Text.

<sup>[b]</sup> SUBSTRUCTURE: Substruktur (Produkt oder Edukt) + Bindungsänderung bei der Reaktion, beteiligte/unbeteiligte Atome bzw. Bindungen; REACTION: Eingabe beschränkt auf an der Reaktion beteiligte Bindungen und Atome.

<sup>[c]</sup> RSS = Menu-Option (Substruktur mit ADDREACTANT oder ADDPRODUCT in ihrer «Rolle» definiert); zusätzliche Tastatur-Befehle für Reaktions-Recherchen: «SSS as catalyst», «SSS as solvent».

<sup>[d]</sup> Suche nach allen Reaktionen, in denen die eingegebene Struktur (vom Programm mit H zur exakten Struktur «aufgefüllt») in einer bestimmten «Rolle» vorkommt: CURASAGT = «current structure as agent» (Edukt, Reagens, Lösungsmittel, Katalysator); CURASPRD = «current structure as product» im [SEARCH]-Menu; zusätzliche Differenzierung über die Tastatur-Suchbefehle «current as reactant», «... as solvent», «... as catalyst».

gungen zur Oxidation der 3-OH-Gruppe in Lanosterol» setzten wir in REACCS die FORMULA-Suche ein und fanden für C<sub>30</sub>H<sub>50</sub>O kein Lanosterol, aber Cycloartenol. Die auf dem Schirm als Resultat gezeigte Struktur wurde nun in [BUILD] mit ADDREACTANT als Edukt-Substruktur eingegeben (Cycloartenol ist in der Datenbank wie alle Strukturen ohne H-Atome gespeichert, die vom Programm nur bei Bedarf hinzugefügt werden) und die so definierte «Reaktion» mit RSS gesucht. Ergebnis: ein Literaturhinweis mit der Oxidation von Cycloartenol. In SYNLIB und ORAC kann man das Zeichnen des Moleküls nicht so elegant umgehen. Wir beschränkten uns hier aber auf das Wesentlichste, d. h. das Ring-A-Fragment (Benutzung von Steroid-Templates wäre eine Al-

ternative gewesen), und zwar in SYNLIB (SUBS, RXN) als entsprechende Produkt-Substruktur und Reaktion, in ORAC wurde in [SUBSTRUCTURE] (Reactant) die Edukt-Substruktur mit Änderung der C–O-Bindung (RXBOND: «d(ouble)») und Definition der einen C–C-Bindung mit FUSION als «fused» eingegeben. Im ersten Anlauf war das Ergebnis in beiden Fällen Null, ohne die im zweiten Versuch gelöschten geminalen Methylgruppen an C-4 fanden wir in ORAC die Oxidation von 5β-Cholestan-3β,19-diol am C-3-OH, in SYNLIB (BROAD SEARCH) nur drei irrelevante Zitate (formal korrekte Transformations-Statistik, aber keine einfache Oxidation vom Alkohol zum Keton). Substrukturen können auch aus mehreren Fragmenten bestehen, wie im Beispiel E

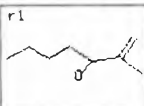
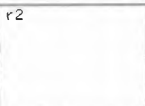
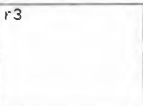
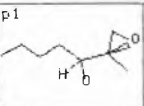
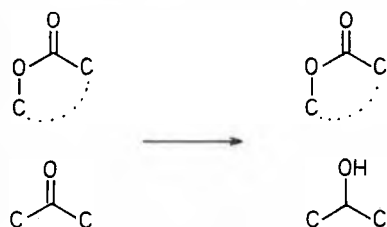
							
alt1	alt2	int1	int2	synthon			
Card: 4760 Journal: J AM CHEM SOC Year: 1974 Page: 6254 Yield: 90 Temp: 0 Title: SHARPLESS		Authors: TANAKA S YAMAMOTO H NOZAKI H SHARPLESS K B MICHAELSON R C Reaction Keys: EPOXIDATION ALLYLIC CATALYSIS OXYGENATION STERESELECTIVE Solvent Keys:	Reagent Keys: OXIDANT HYDROPEROXIDE CATALYST VANADIUM Reagents: t-BuOOH 1 VO(acac)2 1 TOLUENE				
NEXT	BACK	SKIP	KEEP	STOP	DATA	PLOT	HELP

Fig. 19. ORAC: [KEYS CARD] zu der in Fig. 13 (Reaktion aus Beispiel A) gezeigten [DATA CARD] mit Strukturen, Indexierung und Daten<sup>[29]</sup>.

zur selektiven Reduktion eines Ketons in Gegenwart eines Lactons. In REACCS und ORAC wurden die Bindungen/Atome wie vorher beschrieben als reagierend oder nicht-reagierend sowie im Lacton-Fragment als «Ring» definiert. In ORAC genügte die «Produktseite» (Substruktur mit Bindungsänderung) zur Beschreibung, in REACCS erfolgte die Eingabe der ganzen Reaktion wie im Formelbild dargestellt. In SYNLIB wurde ebenfalls (SUBS/RXN) die ganze Reaktion eingegeben, hier aber ohne die Möglichkeit, explizit nicht-reagierende oder Ring-Bindungen zu definieren. Ergebnisse: keine Antwort in REACCS, ein irrelevantes Zitat in SYNLIB, vier relevante Reaktionen in ORAC. Die Eingabe von Fragestellungen mit mehreren Edukten oder Produkten gelingt auf unterschiedliche Weise. In REACCS werden die (Sub)strukturen einzeln gezeichnet und mit ADD... nacheinander zur Reaktion aufgebaut, in SYNLIB gibt man mehrere «Fragmente» gleichzeitig ein, und in ORAC müssen zwei separate (Sub)strukturen definiert und eine Suche mit logischer Kombination ausgeführt werden: z. B. SUBSTRUCTURE → erstes Edukt zeichnen → ENTER → SUBSTRUCTURE → zweites Edukt zeichnen → ENTER → AND (im [QUERY MENU]) → SEARCH.

#### Beispiel E



Zur Einschränkung oder Erweiterung kann man also Suchen oder deren Resultate miteinander kombinieren. Dies geschieht einmal durch «Nachsuche» in einer Menge von bereits mit einer ersten Recherche gefundenen Antworten, bei SYNLIB mit OMIT/FOCUS (vgl. Abschnitt 4), in REACCS durch Definition einer Antwort-«Liste» mit REF = LIST als «Sub-Datei» für die nächsten Recherchen; in ORAC kann man dazu unmittelbar Such-Optionen (auch (Sub)struktur und Text) mit AND kombinieren und dann erst SEARCH aktivieren (vgl. oben). Eine Alternative dazu nur bei REACCS und ORAC ist die logische Verknüpfung von Resultaten (unabhängiger) Suchen mit AND, OR, NOT.

Text-Recherchen nach Schlagwörtern für Reaktionstyp und allgemeine Reaktionsbedingungen, Ausbeute, Temperatur, Druck, pH-Wert, nach bibliographischen Daten (Autor, Jahr, Zeitschrift) von Zitaten etc. (gegebenenfalls auch nach Summenformeln oder Namen von Reagentien, Lösungsmitteln) sind zwar grundsätzlich weniger wichtig als die Suche nach (Sub)strukturen, dennoch werden sie entweder für sehr allgemeine Fragestellungen oder vor allem zur Einschränkung von Ergebnissen benötigt. ORAC bietet dafür sowohl die einfachsten Suchmöglichkeiten als auch den notwendigen Inhalt der Datenbank (Indexierung). Ein Datenbank-Eintrag in ORAC besteht nicht nur aus der in Fig. 13 gezeigten [DATA CARD], sondern man kann mit KEYS die zugehörige [KEYS CARD] (Fig. 19) aufrufen, die alle zur Reaktion suchbaren Schlagwörter in den entsprechenden Kategorien zeigt und die Indexierung illustriert. Der direkte Zugriff auf Text/Daten ist im [QUERY MENU] (vgl. Fig. 11) ersichtlich. Eine Re-

cherche nach «Baeyer-Villiger-Oxidationen» lief folgendermaßen ab: [QUERY MENU] → REACTION KEY aktiviert → «prompt» «REACTION KEY» → «Baeyer-Villiger» eingetippt → SEARCH → drei Reaktionen. Eine ergänzende Suche mit NAME OF REACTION ergab noch weitere sechs relevante Reaktionen (die Indexierung ist also nicht immer unproblematisch). Für die Datentypen JOURNAL, REACTION KEY (Reaktionstyp), REAGENT KEY (Reagensklasse, im Unterschied zu ACTUAL REAGENTS, vgl. unten), SOLVENT KEY wird ein kontrolliertes Vokabular verwendet, das mit «?» nach dem «prompt» in einem [KEYS LIBRARY]-Menu aufgerufen werden kann. Mit YEAR (Publikationsjahr), TEMPERATURE und YIELD sind sowohl einzelne Daten als auch Bereiche suchbar. REACCS bietet zwar vielfältige und flexible, d. h. vom Benutzer für eigene Datenbanken veränderbare Datentypen und -felder zur Speicherung und Abfrage von Text, der Zugriff ist jedoch etwas umständlicher, und vor allem enthalten THEILHEIMER und ORGSYN relativ wenige Schlagwörter bzw. Daten. Zur Suche wurde RXNDATA in [SEARCH] aktiviert, und dann auf einen «prompt» Datentyp und der zu suchende Text bzw. Wert eingegeben: «RNAME = BAEYER@» (@ = «Maske» für beliebige Zeichen zur Erfassung verschiedener Schreibweisen), die 11 Reaktionen wurden mit «YIELD > 90» auf 3 reduziert. Datentypen und das (kritische) Eingabeformat kann man sich mit INDEXDATA bzw. INDEXFULL zeigen lassen. SYNLIB verfügt über wenige Möglichkeiten zur Daten-Recherche im [CONSTRAINT MODE]-Menu (vgl. Fig. 10) wie MIN.YLD., MAX.NO.OF STEPS, DATE/BEFORE/AFTER, sonst aber nur mit TEXT über einen langsamen, unspezifischen «stringsearch» zum Suchen von «Zeichenketten» im eher mageren Textteil eines SYNLIB-Zitats (vgl. Fig. 9): z. B. [CONSTRAINT MODE] → TEXT → Eingabe «BAEYER» fand 15 Reaktionen, 3 nicht relevante, darunter eine «Baeyer»-Reaktion, aber auch 2 relevante «Baeyer-Villiger»-Reaktionen.

Reagentien und Lösungsmittel sind in REACCS über die Struktur suchbar (vgl. Tabelle 1), während man in SYNLIB und (zur Zeit noch) in ORAC den Weg über eine Textsuche gehen muss. Dies ist bei Lösungsmitteln, z. B. HMPA/HMPT (in ORAC unter SOLVENT KEY beide Abkürzungen durch Querverweise verknüpft, in SYNLIB «stringsearch» mit «HMP»), kaum problematisch. Vinylithium als Reagens wurde in SYNLIB und ORAC via Formel (korrekte Schreibweise?) nicht gefunden, in REACCS via Struktur (vgl. Tabelle 1) aber in fünf Reaktionen. Bei einer Recherche zur Verseifung von Nitrilen zu primären Amiden mit Ameisensäure wurde die entsprechende Substruktur-Recherche C=C=N → C-CO-NH, (in REACCS z. B. 27 solche Umsetzungen,

ohne Markierung der reagierenden Bindungen sogar 43) durch Kombination mit RXNDATA... "REGNO = 11" (REACCS, interne Nummer von Ameisensäure, gefunden durch vorherige exakte Struktursuche via Struktur- oder Summenformel) auf ein relevantes Zitat eingeschränkt. In ORAC ging dies entsprechend durch Kombination mit ACTUAL REAGENTS: "HCOOH", in SYNLIB durch «stringsearch» von "HCOOH" mit TEXT im [CONSTRAINT MODE]. Für allgemeine Fragen nach Reagentklassen wie Säure, Base, Peroxid sind SYNLIB ([CONSTRAINT MODE]) und vor allem ORAC (REAGENT KEY) besser gerüstet als REACCS.

Die in einer Recherche gefundenen Reaktionen werden anschliessend auf dem Schirm inspiert und gegebenenfalls entweder über einen an das Terminal direkt angeschlossenen grafikfähigen Matrixdrucker/Bildschirm-Kopierer oder, mit PLOT-Befehlen formatiert, auf einem Plotter/Laserdrucker ausgedruckt. REACCS (maximal 99 «Lists») und ORAC (maximal 99 «Sets») speichern Resultate jeweils für die Sitzung (d. h. in REACCS bis zum Datenbank-Wechsel); selektive Löschen sowie permanente Speicherung sind möglich. Will man ein anderes Resultat als das letzte betrachten, so muss man die entsprechende «List» bzw. «Set» aktivieren. SYNLIB hat leider zur Zeit nur die Antworten der letzten Recherche verfügbar, was unter anderem nachträgliche Kombinationen ausschliesst. Die erste Antwort wird bei SYNLIB automatisch schon während der Suche im [SEARCH MODE] und bei ORAC sofort nach Ende der Suche ausgegeben, wenn man diese mit DISPLAY statt mit SEARCH gestartet hatte (sonst kehrt man am Suchende ins [QUERY MENU] zurück und aktiviert hier DISPLAY<sup>[30]</sup>). In REACCS wechselt man in das [VIEWL]-Menu und aktiviert FIRST<sup>[25]</sup>. «Durchblättern» der Resultate in REACCS/SYNLIB/ORAC geschieht vorwärts mit NEXT/SRCH/NEXT oder rückwärts mit PREVIOUS/1 SRCH/BACK. Mit FIRST in REACCS sowie 2 SRCH in SYNLIB springt man zum Anfang zurück. Vor- und Rückwärts-Springen zum Betrachten jeder zweiten, dritten, fünften... Antwort kann man in ORAC durch definieren eines SKIP-Wertes; in REACCS ist eine bestimmte Antwort mit ITEM über ihre Nummer (Nr.) erhältlich.

## 7. Fazit

REACCS, SYNLIB und ORAC, die in Systemaufbau, Datenbankinhalt und -zugriff deutliche Unterschiede aufweisen, erfüllen die an ein modernes Reaktions-Datenbanksystem zu stellenden Anforderungen. Sie bringen gegenüber den bisher gebräuchlichen «klassischen» Quellen einen echten Durchbruch in der Informationsbeschaffung und haben im Unterschied zu den meisten bis heute bekannten

Online-Datenbanken das Potential, zum wirklich alltäglichen Hilfsmittel des Chemikers zu werden, besonders wegen ihrer einfachen, benutzerfreundlichen Bedienung, den nicht von der Anschlusszeit dominierten Kosten – und vor allem, weil sie leicht und schnell Informationen liefern, die organisch-synthetisch tätige Chemiker sehr häufig benötigen. Hoffentlich nehmen auch im deutschsprachigen Raum Hochschulen das für sie günstige Angebot wahr und nützen zudem die Chance, einen Beitrag zum Aufbau ausgewogener, kritisch bewerteter Reaktions-Datenbanken zu leisten. Kriterien für ein Reaktions-Datenbanksystem sind sowohl die Suchmöglichkeiten des Programms als auch der Inhalt der Datenbank, d. h. Qualität von Auswahl und Beschreibung der Reaktionen, die sich ergänzen müssen. Wir haben uns in diesem Fortschrittsbericht auf ersteres beschränkt, da sich alle Datenbanken, besonders die von ORAC, noch im Aufbau befinden und noch nicht einigermaßen repräsentative «Basis-Sammlungen» darstellen, die zu vergleichen sinnvoll wäre. REACCS hat hier den Vorzug, neben seinem CURRENT LITERATURE FILE<sup>[13]</sup> auch das Repertoire von "Theilheimer" und "Org. Synth." verfügbar zu machen. Die jeweils zu den Reaktionen gespeicherten Informationen, bei ORAC etwa die Schlagwörter, sind aus den Abbildungen dieses Beitrags ersichtlich. In REACCS und ORAC wurden die Strukturen von Edukt und Produkt vollständig gespeichert, während in SYNLIB unbeteiligte Strukturbezirke leider oft durch Reste R abgekürzt sind. Dies führt mitunter zu Informationsverlusten; z. B. fanden wir die selektive Reduktion eines Methoxymethyl-Esters in Gegenwart eines Methyl-Esters in der Datenbank deshalb nicht über die übliche Fragestellung, weil hier R = MEM nur im Text vorkam.

Im Hinblick auf die ständige Weiterentwicklung der Systeme kann dieser Vergleich nur eine «Momentaufnahme» sein. Nach einer auf zwei Monate beschränkten, aber intensiven Evaluation kommen wir zum Schluss, dass gegenwärtig keines der Systeme als Ganzes sich so wesentlich in der Leistungsfähigkeit unterscheidet, dass eine allgemeine Empfehlung gerechtfertigt wäre. Aufgrund unterschiedlicher Anforderungen und Gewichtung der Eigenschaften, dabei unvermeidlich auch durch subjektive gefärbte Gewohnheiten und Vorstellungen beeinflusst, benutzen wir in unseren beiden (Hochschul-)Arbeitsgruppen zur Zeit ORAC bzw. SYNLIB.

*Technische Einzelheiten:* REACCS (Programmversion 6.0 mit den Datenbanken THEIL («Theilheimer»), Vol. 15–35, 27041 Reaktionen mit 39430 Strukturen) und ORGSYN («Org. Synth.», Vol. 1–60, 4667 Reaktionen mit 4997 Strukturen) und SYNLIB (Version 2.2, 25600 Reaktionen) waren auf dem Rechner DEC VAX 8600 des Rechenzentrums der ETH Zürich mit dem Betriebssystem VMS 4.2 imple-

mentiert; unsere Terminals waren über das ETH-Lokalnetz KOMETH (asynchron, 9600 Baud) mit dem Rechner verbunden. Der Zugriff auf ORAC (Version 5.3, 5000 Reaktionen) erfolgte auf DEC VAX 11/750 unter VMS 4.1 der Wolfson CADOS Unit an der University of Leeds via KOMETH/TELEPAC-Gateway und das öffentliche Datennetz. Als Terminals verwendeten wir DEC VT100/VT640 Retrographics (alle Systeme, Lichtgriffel oder Cursor-Steuerung über Tastatur für ORAC), Hazeltine IMLAC II (REACCS, Lichtgriffel; Leihgabe der Molecular Design MDL AG, Basel), für ORAC und SYNLIB die Terminals DEC VT240 (mit Drucker DEC LA50) und HP 2647A (mit Drucker HP 2631G), beide im Tektronix-PLOT10 Emulations-Modus mit Cursor-Steuerung über die Tastatur, sowie Apple MacIntosh mit Tekalike PLOT10 Emulation (Cursor mit «Maus» gesteuert).

*Unser Dank gilt Dr. A. Kos und der Fa. Molecular Design MDL AG (Basel) für die Leihgabe von REACCS und des IMLAC-Terminals, Dr. A. P. Johnson (Wolfson CADOS Unit, University of Leeds) für den Zugriff auf ORAC sowie Dr. D. Chodosch (Smith Kline & French Laboratories, Philadelphia) für das im Rahmen des «academic program» der Gruppe von Prof. Dr. D. Seebach (Laboratorium für Organische Chemie ETHZ) überlassene SYNLIB. Alle Genannten unterstützten uns bereitwillig auch durch Weitergabe von Informationen und Diskussionsbeiträge. Dr. W. B. Schweizer aus unserem Laboratorium danken wir für die Implementierung von REACCS und SYNLIB am Rechenzentrum der ETHZ, das uns Rechenzeit auf der VAX 8600 und den TELEPAC-Gateway zur Verfügung stellte. S. M. dankt dem Verband der Chemischen Industrie (Frankfurt am Main) für ein Promotionsstipendium.*

Eingegangen am 20. Januar 1986 [FR 22]

- [1] A. J. Beach, H. F. Dabek jr., N. L. Hosansky, *J. Chem. Inf. Comput. Sci.* 19 (1979) 149.
- [2] Zur Indexierungspolitik von *Chemical Abstracts* (CA) für Verbindungen in Reaktionen vgl. *Appendix II* in den "Index Guides".
- [3] Zur Problematik der Speicherung, Klassierung und Suche von Reaktionen vgl. J. Ash, P. Chubb, S. Ward, S. Welford, P. Willett: *Communication, Storage, and Retrieval of Chemical Information*, Horwood/Wiley, Chichester (1985), S. 203–220; J. H. Winter, in: *Tagungsbericht 1. Vortragsstagung Fachgruppe Chemie-Information (Frankfurt/Main, 27./28.10.1983)*, Gesellschaft Deutscher Chemiker, Frankfurt am Main (1983), S. 102; J. Valls, in W. T. Wipke, S. R. Heller, R. J. Feldman, E. Hyde (Ed.): *Computer Representation and Manipulation of Chemical Information*, Wiley, New York (1974), S. 83; H. J. Ziegler, *J. Chem. Doc.* 6 (1966) 81. – Zur Verwendung von Computern für die «Erzeugung» statt Speicherung von Reaktionen siehe J. Bauer, R. Herges, E. Fontain, I. Ugi, *Chimia* 39 (1985) 43.
- [4] Zur Recherche von Reaktionen im seit ca. 1958 entwickelten System GREMAS, das leider nur den Mitgliedsfirmen der IDC (Internationale Dokumentationsgesellschaft für Chemie, Frankfurt

- am Main) zugänglich ist, siehe R. Fugmann, G. Kusemann, J. H. Winter, *Inf. Process. Manage.* 15 (1979) 303.
- [5] C. Jochum, in: *Tagungsbericht 2. Vortragstagung Fachgruppe Chemie-Information (Aachen, 25.3.1985)*, Gesellschaft Deutscher Chemiker, Frankfurt am Main (1986), S. 51.
- [6] Beispiele für Online-Reaktions-Recherchen in CA oder CRDS findet man z. B. in D. Rehm, F.-P. Montforts, M. Ockenfeld, G. Wess: *Online-Recherchen in Datenbanken des Chemical Abstracts Service*, Verlag Chemie, Weinheim (1982), S. 98–103; E. Zass, in E. Ziegler (Ed.): *Computer in der Chemie* (2. Aufl.), Springer-Verlag, Berlin (1985), S. 24, 44–46; E. Zass, *Nachr. Chem. Tech. Lab.* 32 (1984) 424; H. Mlodzik, in: *1. Deutsches Online-Informationstreffen (Köln 1980)*, Learned Information Ltd., Oxford (1981), S. 85; A. Finch, *J. Chem. Inf. Comput. Sci.* 26 (1986) 17.
- [7] CA entwickelt zur Zeit eine spezielle, mit (Sub)struktur-Recherchen abfragbare Reaktions-Datenbank unter Einbeziehung von Reagentien, Lösungsmitteln etc. Vgl. dazu D. Weisgerber, in: *Tagungsbericht 2. Vortragstagung Fachgruppe Chemie-Information (Aachen, 25.3.1985)*, Gesellschaft Deutscher Chemiker, Frankfurt am Main (1986), S. 119; P. E. Blower, R. C. Dana: «Creation of a Chemical Reaction Database from the Primary Literature», Vortrag bei der Chemical Structure Association Conference *Modern Approaches to Chemical Reaction Searching*, York, 10.7.1985, im Druck (wir danken Dr. D. W. Weisgerber, Chemical Abstracts Service, Columbus, für eine Kopie des Manuskripts).
- [8] O. Schier, W. Nübling, W. Steidle, J. Valls, *Angew. Chem.* 82 (1970) 622; *Angew. Chem. Int. Ed. Engl.* 9 (1970) 599; K.-H. Bork, *Chem. Ztg.* 96 (1972) 330.
- [9] P. G. Dittmar, N. A. Farmer, W. Fisanick, R. C. Haines, J. Mockus, *J. Chem. Inf. Comput. Sci.* 23 (1983) 93; R. Attias, *ibid.* 23 (1983) 102; J.-E. Dubois, Y. Sobel, *ibid.* 25 (1985) 326; E. Zass, *Nachr. Chem. Tech. Lab.* 32 (1984) 502.
- [10] *Chem. Eng. News* 60 (15) (1982) 26.
- [11] S. Anderson, *J. Mol. Graph.* 2 (1984) 83; G. W. Adamson, J. M. Bird, G. Palmer, W. A. Warr, *J. Chem. Inf. Comput. Sci.* 25 (1985) 90.
- [12] Über Einzelheiten informiert auf Anfrage: H. R. Kottmann, Molecular Design MDL AG, Wallstrasse 8, CH-4002 Basel, Tel. (061) 23 29 29.
- [13] P. S. Zurer, *Chem. Eng. News* 63 (33) (1985) 21, 45.
- [14] A. Kos (Basel), persönliche Mitteilung.
- [15] Zur Beschreibung des Ablaufs wird folgende Notation verwendet: [MENU-NAME], MENU-BEFEHL/OPTION, "MELDUNGEN DES SYSTEMS", "Eingabe über Tastatur".
- [16] Für eine Benutzung von THEILHEIMER oder ORGSYN sind in jedem Fall entsprechende jährliche Lizenzgebühren zu entrichten. Zuständig für das «academic program» ist Marjorie Faltens, Molecular Design Ltd., 2132 Farallon Drive, San Leandro, CA 94577 (USA).
- [17] *Chem. Eng. News* 63 (34) (1985) 13.
- [18] D. F. Chodosh, W. L. Mendelson, *Drug Inf. J.* 17 (1983) 231; *Pharm. Technol.* 7 (1983) 90.
- [19] Informationen und Kontakt: Dr. D. Chodosh, Distributed Chemical Graphics, 10788 Lockart Road, Philadelphia, PA 19116 (USA).
- [20] D. Chodosh (Philadelphia), persönliche Mitteilung.
- [21] A. P. Johnson (Leeds), persönliche Mitteilung.
- [22] A. P. Johnson, *Chem. Br.* 21 (1985) 59.
- [23] E. J. Corey, A. K. Long, S. D. Rubenstein, *Science* 228 (1985) 408; A. K. Long, S. D. Rubenstein, L. J. Joncas, *Chem. Eng. News* 61 (19) (1983) 22.
- [24] Über Einzelheiten informiert auf Anfrage: Dr. A. P. Johnson, Wolfson Unit for Computer-Aided Design of Organic Synthesis, Department of Organic Chemistry, The University of Leeds, Leeds LS2 9JT (U.K.), Tel. (0532) 43 17 51 Ext. 6071 oder 6058.
- [25] Inzwischen wurde das «Füllen» des Bildschirms beschleunigt; bei einer Suche erscheint automatisch das erste Ergebnis auf dem Schirm, während die Suche fortgesetzt wird.
- [26] Durch den Modus "P(roduct) NOT R(eactant)" ist in ORAC die Bildung des Ringsystems (Ringabschluss) impliziert. In REACCS fanden wir bei Eingabe der Produkt-Substruktur in einer Reaktions-Recherche (Edukt nicht definiert) 10 Literaturzitate, die dann mit RXNDATA ... KEYWORDS = RCL ("ring closure" in THEILHEIMER, vgl. die Diskussion über Text-Recherchen) auf zwei eingeschränkt wurden, von denen eines die Bildung des isolierten Ringsystems beschreibt.
- [27] Das Resultat der Recherche wurde im [CONSTRAINT MODE] mit TEXT "PH(enyl)3P=C" auf fünf relevante Reaktionen beschränkt (vgl. Abschnitt 4).
- [28] Ergebnis-«Listen» aus solchen (Sub)struktur-Recherchen können mit denen aus Reaktions-Recherchen nicht logisch kombiniert werden.
- [29] Der Ausdruck für diese Abbildung wurde mit dem Terminal HP2647A und dem Matrix-Drucker HP2631G hergestellt.
- [30] Änderungen in ORAC Version 6.0: Ausschluss von Atomen an einer bestimmten Position möglich; Wahl der Reaktions-Files (zur Zeit zwei «Boxes» mit je 5000 Reaktionen) vor der Suche (alle oder nur bestimmte); Suche wird nur mit SEARCH, nicht mehr mit DISPLAY ausgelöst, das anschliessend zum Betrachten der Resultate aktiviert werden muss.