

Computergestützte Planung organisch-chemischer Synthesen: die zweite Programmgeneration**

Wolf-Dietrich Ihlenfeldt und Johann Gasteiger*

Professor Paul von Ragué Schleyer zum 65. Geburtstag gewidmet

Die Planung von Synthesen in der Organischen Chemie wurde in den letzten Jahrzehnten zunehmend auf solidere Fundamente gestellt, und es wurden weithin gültige Grundregeln formuliert. Fast gleichzeitig mit der Systematisierung des Gebietes wurde auch das Potential zur Automatisierung der Syntheseplanung durch den Einsatz von Computern erkannt. Trotz mittlerweile mehr als zwei Jahrzehnte dauernder Anstrengungen mehrerer Gruppen hat sich jedoch die klassische, dem Synthon- und Transformprinzip folgende computergestützte Syntheseplanung mit großen Bibliotheken nicht durchsetzen können – ganz im Unterschied zu Reaktionsdatenbanken, deren Nutzung erstaunlich schnell zur Routine geworden ist. Offen-

sichtlich befriedigen die alten Ansätze zur computergestützten Syntheseplanung nicht die Bedürfnisse des präparativ arbeitenden Chemikers. Dieser bisherige Mangel an Erfolg bedeutet jedoch nicht das Ende der computergestützten Syntheseplanung an sich. Die in den Programmen der ersten Generation verborgenen konzeptionellen Mängel sowie Probleme technischer wie psychologischer Art werden in diesem Beitrag kritisch analysiert. Inzwischen wird an Systemen der zweiten Generation gearbeitet, die auf neue Weise dem präparativ arbeitenden Chemiker Hilfestellung geben sollen. Dabei wird darauf geachtet, der typischen, keineswegs systematischen, sondern oftmals sprunghaften Planungsrichtung, (Sub)Ziel und Me-

thoden wechselnden Arbeitsweise des planenden Chemikers in größerem Maße entgegenzukommen. Neue Werkzeuge, die den Denkschemata und Gewohnheiten eines Chemikers besser entsprechen als die alten Methoden, geben dem Benutzer die Möglichkeit, innovative und hoch interaktive Synthesen organischer Moleküle zu entwerfen. Unsere Gruppe ist weiterhin aktiv auf diesem noch keineswegs abgeschlossenen Gebiet, und wir stellen in diesem Beitrag ein System mit einer Reihe von neuen Konzepten vor und demonstrieren diese an praktischen Beispielen.

Stichworte: Computerchemie · Syntheseplanung · WODCA

1. Die erste Generation

Einer der wichtigsten konzeptionellen Fortschritte in der präparativen organischen Chemie in den letzten drei Jahrzehnten war die Systematisierung und Kategorisierung der Synthese. Kein Chemiker durchläuft heute seine Universitätsausbildung, ohne unter Zuhilfenahme von Coreys Synthonkonzepten^[1–4] einfache Modellsynthesen nach dem Baukastenprinzip entworfen zu haben. Die klare logische Struktur des Synthonkonzeptes und alternativer Ansätze hat seit deren Anfängen Chemiker und Informatiker herausgefordert, diese Prinzipien als Algorithmen in Computerprogrammen zu implementieren und damit den

Computer zum Werkzeug für die Syntheseplanung zu machen^[5]. Parallel dazu hat die beneidenswerte Situation der Chemie, daß nahezu alle relevante Literatur in computerlesbarer Form abstrahiert wird, Forscher dazu angespornt, den riesigen Schatz an experimenteller Erfahrung mit Computerhilfe durchsuchbar zu machen und damit den traditionellen Gang in die Bibliothek auf der Suche nach Hinweisen zur praktischen Durchführbarkeit einer geplanten Reaktion oder Synthesestufe neu zu gestalten.

1.1. Computergestützte Syntheseplanung

In den letzten Jahrzehnten haben sich zwei Hauptarten von Programmen zur Unterstützung der Syntheseplanung in der Organischen Chemie herausgebildet: Die eine Gruppe sind Syntheseplanungsprogramme, die auf dem Synthon- oder Transformprinzip oder verwandten Konzepten oder auf der Erzeugung einer formalen Reaktion basieren. Die zweite Gruppe sind Re-

[*] Prof. Dr. J. Gasteiger, Dr. W.-D. Ihlenfeldt
Computer-Chemie-Centrum, Institut für Organische Chemie
Universität Erlangen-Nürnberg
Nägelsbachstraße 25, D-91052 Erlangen
Telefax: Int. + 9131/856566
E-mail: gasteiger@eros.ccc.uni-erlangen.de

[**] Wichtige in diesem Beitrag verwendete Akronyme sind am Ende des Textes zusammengestellt.

aktionsdatenbanken, die eine große Zahl von tatsächlich durchgeführten Reaktionen nach vielen Kriterien absuchbar machen und somit den präparativ arbeitenden Chemiker über Problemlösungen bei verwandten Aufgaben informieren. Eine dritte, weniger prominente, aber nichtsdestotrotz wichtige Klasse von Programmen sind Reaktionsvorhersageprogramme, die chemische Reaktionen modellieren und Informationen über die Produkte und Durchführungsbedingungen einer Reaktion durch Interpolation und Extrapolation über einen Trainingsdatensatz hinaus liefern können. Während quantenchemische Rechnungen für die Vorhersage und Optimierung von Stoffeigenschaften immer wichtiger werden, haben diese Methoden auf dem außerordentlich komplexen Gebiet der Reaktionsvorhersage für praktisch interessierende Systeme (von wenigen Ausnahmen abgesehen) noch nicht Fuß fassen können. In den nächsten beiden Abschnitten geben wir einen kurzen historischen Überblick über die klassischen Entwicklungen.

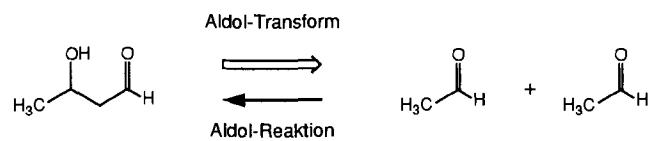
1.2. Syntheseplanungs- und Reaktionsvorhersageprogramme

Die Entwicklung von Syntheseplanungs- und Reaktionsvorhersageprogrammen ist eng gekoppelt sowohl an die Fortschritte in der Systematisierung von Reaktionen als auch an die stürmische Entwicklung von Computern und deren Peripherie. Der heutige Leser staunt, mit wie wenig Kernspeicher und mit welch primitiven Ein- und Ausgabegeräten die Pioniere auskamen und

welche Klimmzüge notwendig waren, um Strukturen unter diesen Bedingungen zu codieren. Die Gesamtzahl der Programme, deren Einsatzgebiet wenigstens peripher die Planung von Synthesen ist, soll mittlerweile mehr als 50 betragen^[5]. Die Zahl der Systeme, deren Ziel die Planung organisch-chemischer Synthesen in voller Breite ist, ist allerdings bedeutend kleiner; sie dürfte eher bei etwa 10 liegen.

1.2.1. Das Synthonprinzip

Die Entwicklung des Synthonprinzips, das heutzutage als selbstverständlich und offensichtlich angesehen wird, war ein fast ikonoklastischer Paradigmenwechsel. Synthesen wurden nicht mehr unter den Aspekten Namensreaktionen und einzelne Stufen gesehen, sondern unter denen globaler Gerüsttransformationen und Muster von Ladungen und Polaritäten, losgelöst von der Art des Substituenten, der das Ladungsmuster aufprägt. Ein zentraler Punkt dieser Analyse ist, daß die Synthesereaktionen in ihrer Umkehrrichtung (retrosynthetisch) betrachtet werden. Diese Umkehrreaktionen (Retroreaktionen) werden üblicherweise *Transforms* genannt. Schema 1 zeigt dies für die



Schema 1. Retrosynthetische Betrachtungsweise: die Aldolreaktion und das entsprechende retrosynthetische Transform.

Johann Gasteiger wurde 1941 in Dachau geboren. Nach dem Studium der Chemie an den Universitäten München und Zürich promovierte er 1971 bei Prof. Rolf Huisgen. Nach einem Jahr als Postdoc bei Prof. A. Streitwieser, Jr., an der University of California in Berkeley trat er eine Stelle an der Technischen Universität München an und habilitierte sich dort 1979. 1991 wurde ihm in Anerkennung seiner Verdienste um die Computerchemie die Gmelin-Beilstein-Denkünze der GDCh verliehen. 1994 wurde er als C3-Professor für Organische Chemie an die Universität Erlangen-Nürnberg berufen. Gemeinsam mit Prof. Dr. Paul von Ragué Schleyer und Priv.-Doz. Dr. T. Clark gründete er das Computer-Chemie-Centrum der Universität Erlangen-Nürnberg. Seine Forschungsschwerpunkte sind die Entwicklung von Computerprogrammen zur Syntheseplanung, Reaktions- und Reaktivitätsvorhersage, zur Analyse und Simulation von Massen- und IR-Spektren sowie zum Auffinden neuer Leitstrukturen mit biologischen Eigenschaften. Wissenserwerb aus Datenbanken der Chemie sowie die Anwendung von neuronalen Netzen und genetischen Algorithmen auf chemische Fragestellungen sind weitere Arbeitsgebiete.



J. Gasteiger

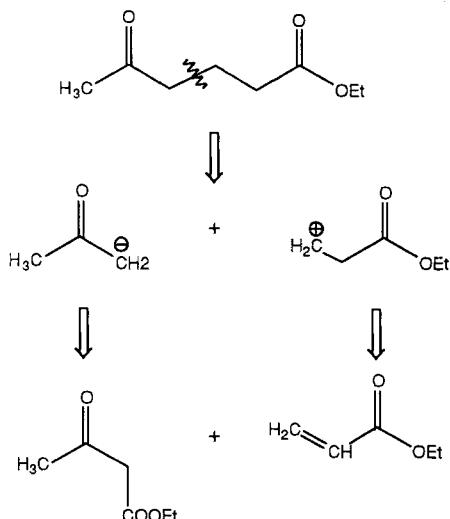


W.-D. Ihlenfeldt

Wolf-Dietrich Ihlenfeldt wurde 1963 in Hamburg geboren. Er studierte Chemie in Hamburg, Heidelberg und an der TU München, promovierte 1991 bei Prof. Gasteiger über computergestützte Syntheseplanung und befaßte sich dann bis Oktober 1994 an der Toyohashi University of Technology (Toyohashi, Japan) mit Arbeiten an neuartigen netzwerkgestützten Chemieinformationssystemen. Zur Zeit arbeitet er im Auftrag des FIZ Chemie, Berlin, an der Evaluierung von innovativen Informationsvermittlungs- und -erzeugungsmodellen in der Chemie. Sein derzeitiger Interessenschwerpunkt ist die Verknüpfung von globalen Computernetzen und dynamischen, offenen Datenrepräsentations- und Datenmanipulationstechniken mit chemiespezifischer Information und Informationsverarbeitung, eine Arbeit an der Nahtstelle zwischen Chemie und Informatik.

Aldolreaktion. Synthesen nach diesem Prinzip werden nach Lehrbuch in zwei Stufen geplant:

1. Die strategischen Bindungen werden definiert und die Transformationen formuliert, die diese Bindungen in der gewünschten Weise knüpfen oder lösen. Durch den Bruch einer strategischen Bindung kommt man zu (meist geladenen) Fragmenten, den *Synthonen*, aus denen dann durch Hinzufügen oder Entfernung von Atomen, Ionen oder Atomgruppen konkrete Moleküle, die Synthesevorstufen oder Ausgangsmaterialien, erhalten werden (Schema 2).



Schema 2. Zerlegung einer Zielverbindung an einer strategischen Bindung zu den (geladenen) Synthonen und deren Umwandlung in Reagentien.

2. Die tatsächlich ausgewählten Reaktionen sind im einfachsten Fall bekannte Reaktionen aus der Literatur oder Generalisierungen von bei Verbindungen mit ähnlich wirkendem Substitutionsmuster erfolgreich durchgeföhrten Reaktionen oder aber formale Bindungsumordnungsprozesse. Dieser Mehrstufenprozeß bietet einige Ansatzpunkte zur Algorithmisierung. Die einfachste Methode ist die systematische „Anwendung“ von Reaktionen oder deren abstrahierter Synthonbeschreibung. Bei der enormen Vielfalt an bekannten Reaktionen ist allerdings eine Steuerung und Vorauswahl unumgänglich. Die Wahl der strategischen Bindungen und die Bewertung der ausgewählten Reaktionen ist gegenüber der systematischen Transformation von Strukturen qualitativ weitaus schwieriger. Konsekutive und parallele Anwendung dieser Vorgehensweise führen zu einem fächerförmigen Schema an Schritten und Sequenzen von Retroreaktionen, dem *Synthesebaum* (Abb. 1).

Ein Synthesebaum kann leicht sehr komplex werden. Selbst wenn für jede Verbindung nur 10 alternative Retroreaktionen in Erwägung gezogen werden müssen, hat man nach drei Stufen bereits $10^3 = 1000$ und nach fünf Stufen $10^5 = 100\,000$ Reaktionen und Vorstufen, die zu betrachten, zu bewerten und gegeneinander abzuwegen sind, wobei eine Fünfstufensynthese wahrlich keine umfangreiche Synthese ist. Dies zeigt die ganze Problematik einer systematischen retrosynthetischen Vorgehensweise. Sie verlangt nach einer extrem sorgfältigen Bewertung und nach rigorosem frühzeitigem Verwerfen möglicher,

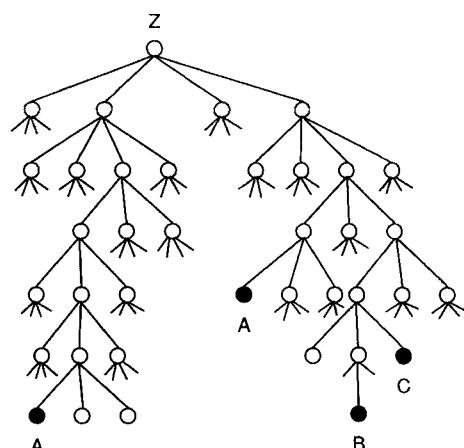


Abb. 1. Ein Synthesebaum als Sequenz von Retroreaktionen, die durch systematische retrosynthetische Analyse erhalten wurden. A, B und C sind käufliche Ausgangsmaterialien, Z ist das Zielmolekül.

jedoch unergiebiger Reaktionsschritte. Das wird sicherlich nicht selten mit dem Preis der Nichtentdeckung interessanter Synthesemöglichkeiten bezahlt.

1.2.2. Darstellung von Reaktionen im Computer

Die klassischen Programme zur Syntheseplanung und Reaktionsvorhersage^[5] verwenden relativ wenige und einfache Modelle zur internen Repräsentation ihrer Datenbasis. Im Mittelpunkt steht die Beschreibung des Reaktionszentrums und der daran durchgeföhrten Operationen. Die Beschreibung kann eine komplette Reaktion mit allen Bindungsverschiebungen umfassen, auf Halbreaktionen (d. h. einem angenommenen Dissoziationschritt und einem separaten Rekombinationschritt) basieren oder von den Mikroschritten der einzelnen Bindungsbrüche und Bindungsneubildungen ausgehen. Manche Programme speichern darüber hinaus noch Zusatzinformationen, z.B. über störende Substituenten, allgemeine Formulierungen zur Abschätzung der Durchführbarkeit und Nützlichkeit einer bestimmten Reaktion oder gar physikochemische Modelle der Reaktivität einzelner Bindungen des Reaktionszentrums.

Wichtig ist, daß hier immer nur ein isolierter Teil einer Reaktion, die bereits beschrieben sein kann, gespeichert wird. Die Wissensbasis der klassischen Syntheseplanungsprogramme enthält keine einzelnen, dokumentierten Reaktionen mit Hintergrundinformationen, sondern Destillate aus einer Vielzahl von Literaturquellen und praktischen Erfahrungen, die je nach Systemphilosophie in geschätzte oder statistisch validierte Parameter münden. Die Wissensbasis wird allerdings bei moderneren Systemen außerhalb des Kernsystems gehalten und ist erweiterbar und editierbar und insoweit mit einer Datenbank vergleichbar. Anders als bei Reaktionsdatenbanken handelt es sich jedoch um bearbeitete Information. Jedes Schema, jedes Reaktionsmodell entspricht der kondensierten Essenz einer (normalerweise) großen Zahl von in der Literatur erwähnten Reaktionen, während eine Reaktionsdatenbank Einzelumsetzungen enthält. Der Prozeß der Datenkomprimierung und -auswahl ist bis in die heutige Zeit nicht automatisiert.

1.2.3. Repräsentative Systeme

Als erstes System, das auf dem Synthonkonzept basiert, wird allgemein LHASA^[6] (der Nachfolger von OCSS^[7]) angesehen, das heute noch weiterentwickelt wird. Ein weiterer wichtiger Schritt wurde mit SECS^[8] und dessen Derivat CASP gemacht, das von einem Konsortium deutscher und Schweizer chemischer Firmen entwickelt wurde. Einige weitere repräsentative Vertreter sind SYNCHEM^[9], SYNGEN^[10], SCANCHEM^[11], COMPASS^[12], SYNSUP-MB^[13], AIPHOS^[14], LILITH^[15], TRESOR^[16] und SYNTHON^[17]. Diese Systeme nutzen entweder eine Datenbasis an komplett formulierten Reaktionen, zu denen klassische Namensreaktionen und durch kontinuierliche Literaturarbeit auch ein enormer Fundus an Exotika oder allgemeine Synthonschemata oder Halbreaktionen gehören. Die Bewertung einer Reaktion erfolgt durch Analyse der Umgebung des Reaktionszentrums, eventuell störender Funktionalität an unbeteiligten Positionen des Moleküls und durch allgemeine Nützlichkeitserwägungen. Jede Reaktion und jedes Synthon-schemata wurden per Hand codiert (AIPHOS verfügt über Mechanismen zum Lernen aus Reaktionsdatenbanken, und für SECS gibt es Zusatzprogramme für diesen Zweck^[18]), und für die einzelnen Reaktionstypen wurden Bewertungsfaktoren abgeschätzt. Weitere Module ermöglichen eine Steuerung des Programmablaufs, z.B. durch das Erkennen strategischer Bindungen nach allgemeinen, von Reaktionen losgelösten Heuristiken, die Anwendung von Reaktionen mit besonders großem Synthesepotential^[16a] oder die Verwendung allgemeiner strategischer Konzepte^[19].

Eine stärker formalisierte Vorgehensweise wurde bei der Entwicklung von EROS^[20] gewählt. Reaktionen werden dabei als Bindungsumordnungsschemata formuliert, so wie der Chemiker mit gebogenen Pfeilen die mechanistischen Elektronenverschiebungen im Zuge einer Reaktion darstellt. Mit einigen wenigen derartigen Elektronenverschiebungsschemata konnte ein Großteil der organisch-chemischen Reaktionen abgedeckt werden. Die zentrale Aufgabe bestand nun darin, Methoden zu entwickeln, die in geeigneter Weise festlegen konnten, welche Bindungsumordnungsprozesse einzusetzen sind und auf welche Bindungen in einem konkret vorgegebenen Molekül diese angewendet werden sollen.

Die Auswahl der in einer Reaktion zu brechenden und knüpfenden Bindungen basiert hauptsächlich auf quantitativen Abschätzungen elektronischer und energetischer Effekte an den einzelnen Atomen und Bindungen eines Moleküls^[20a]. Nach Vollzug einer Reaktion wird eine Gesamtbewertung durchgeführt, die Faktoren wie Reaktionsenthalpie^[21], Ladungsverteilung^[22], induktiven^[23], Resonanz-^[22] und Polarisierbarkeitsef-fekt^[24] umfaßt. Die Planungsrichtung – vorwärts bei Problemen der Reaktionsvorhersage, retrosynthetisch in der Syntheseplanung – entscheidet, wie die einzelnen Faktoren zu einer Gesamtbewertung zusammengefaßt werden.

Angesichts dieser starken Betonung der physikochemischen Bewertung von Reaktionen war es ganz natürlich, daß bei der Weiterentwicklung von EROS zunächst mehr Gewicht auf die Reaktionsvorhersage gelegt wurde. Die weitere Verfeinerung der Bewertungsmechanismen führte zu einem Punkt, an dem es nicht mehr praktikabel schien, beide Problemkreise – Reaktionsvorhersage und Syntheseplanung – mit einem einzigen Pro-

grammsystem zu behandeln. Die Aufgaben werden seither von getrennten Programmsystemen wahrgenommen: EROS wurde völlig neu entworfen und ist ab der Version 6.0 im wesentlichen nur noch zur Reaktionsvorhersage zu verwenden^[25a]. Die Mechanismen zur Bewertung der einzelnen chemischen Reaktionstypen werden außerhalb des Programms als Regeln in Dateien gehalten und sind dadurch leicht zu modifizieren und weiterzuentwickeln^[25b]. Die Reaktionsmodelle für diese Regeln werden durch umfangreiche statistische Analysen^[25b] gewonnen oder durch neuronale Netze repräsentiert^[26]. Die Erarbeitung und Validierung von Reaktionsmodellen für EROS 6 ist aufwendig, aber die erhaltenen Modelle sind oft sehr präzise^[27].

Zur Syntheseplanung wurde eine radikal andere Vorgehensweise gewählt: Das System WODCA (*Workbench for the Organization of Data for Chemical Applications*) vereinigt eine Vielfalt von Methoden und Werkzeugen, die für unterschiedliche Aufgabenstellungen während der Planung von Synthesen eingesetzt werden können. Dies wird in späteren Abschnitten näher erklärt.

Ein anderer Ansatz zur Reaktionsvorhersage liegt dem Programm CAMEO zugrunde^[28], bei dem durch Abschätzung Hunderter von Reaktivitätsparametern ein weiter Bereich von Reaktionen modelliert wird, ohne daß vordefinierte Bindungs-umordnungsschemata zum Einsatz kommen. Allerdings werden die allgemeinen Reaktionsbedingungen und akzeptablen Reaktionsklassen bei einer Anfrage vorgegeben. Die hier zum Einsatz kommenden Parameter sind wesentlich leichter zu erhalten (sie sind teilweise einfach abgeschätzt) als diejenigen im EROS-Sy-tem, sind aber häufig nicht streng statistisch validiert und damit weniger zuverlässig. Trotzdem liefert dieses System im allge-menen zutreffende Voraussagen über Haupt- und Nebenprodukte unter den angegebenen Bedingungen – nicht aber über konkrete Produktverhältnisse oder über die Kinetik – für ein sehr breites Spektrum von Reaktionen, die mittlerweile den Bereich „normaler“ Organischer Chemie fast vollständig abdecken, viel vollständiger als z.B. das EROS-System, dem zur Zeit nur Reaktionsmodelle für einige ausgewählte Klassen zur Verfügung stehen.

Eine streng formale Behandlung von Reaktionen liegt dem Dugundji-Ugi-Matrizenmodell der konstitutionellen Chemie zugrunde^[29]. Auf dieser Grundlage wurde ein Reaktionsklassifizierungsschema entwickelt, das in ein Programm zur formalen Reaktionsgenerierung, IGOR, implementiert wurde^[30]. Damit lassen sich für ein vorgegebenes Reaktionsschema alle formal möglichen Belegungen mit Atomen und Bindungen erzeugen. Das zu IGOR komplementäre Programm RAIN wiederum erzeugt aus einem vorgegebenen Ensemble von Molekülen dazu isomere Ensembles und kann somit zur Struktur- und Reaktionsgenerierung eingesetzt werden^[31]. In einem zweiten Ver-wendungsmodus formuliert RAIN aus zwei vorgegebenen En-sembles von Molekülen ein Netz von Reaktionen, das beide verbindet^[31].

Sowohl IGOR als auch RAIN stützen sich bei ihrer Wahl von Strukturen und Reaktionen ausschließlich auf formale Krite-riern wie erlaubte Valenzzustände und Zahl und Strukturmerkmale der Zwischenstufen. Dies bringt den Vorteil mit sich, daß alle denkbaren Struktur- und Reaktionsmöglichkeiten erhalten werden, sowohl bekannte als auch präzedenzlose. Die Auswahl der konkret zu realisierenden Reaktionen wird aber dem Benut-

zer überlassen, da auf eine detaillierte Bewertung z.B. durch physikochemische Abschätzungen an Strukturen und Reaktionen oder durch Vergleich mit bekannten Reaktionen oder Reaktionstypen verzichtet wird^[29].

Ein weiteres wichtiges Programm, das sich jedoch schwer in ein Schema pressen läßt, ist CHIRON^[32]. Dieses Programm ist im wesentlichen ein Such- und Analysesystem, mit dem Strukturen überlagert werden können, um festzustellen, welche Strukturteile, insbesondere unter Berücksichtigung der stereochemischen Verhältnisse, am besten aufeinanderpassen. CHIRON ist somit kein Syntheseplanungs- oder Reaktionsvorhersagesystem und verfügt auch nicht über viele traditionelle Datenbank-Suchverfahren; es ist jedoch für Planungsstudien insoweit von Nutzen, als Ausgangsmaterialien mit günstiger Konfiguration gefunden werden können. Die besondere Stärke von CHIRON liegt in dem großen, handgenerierten Katalog von etwa 2000 Verbindungen mit definierter Konfiguration, der nicht nur kommerziell vertriebene Verbindungen enthält, sondern auch viele aus der Literatur entnommene Verbindungen, die in wenigen Stufen aus einfach erhältlichen Grundsubstanzen synthetisiert werden können. Dieser einmalige Katalog steht auch unter unserem System WODCA, welches später erläutert werden wird, zur Verfügung.

1.3. Reaktionsdatenbanken

Seit Ende der achtziger Jahre ist die Zahl der verfügbaren Reaktionsdatenbanken stark gewachsen^[33–38]. Während es keine wirklich auf breiter kommerzieller Basis vermarkteten Syntheseplanungsprogramme außerhalb des rein didaktischen Sektors zu geben scheint, werden Reaktionssuchsysteme und Reaktionsdatenbanken seit längerem angeboten.

1.3.1. Prinzipien und Verfügbarkeit

Es wurde bereits betont, daß der Informationsgehalt von Reaktionsdatenbanken elementarerer Natur ist als der von Transformbibliotheken. Zugleich ist jedoch die Informationsmenge größer. Einträge einer Reaktionsdatenbank enthalten typischerweise komplette Literaturangaben mit genauen Reaktionsbedingungen, Ausbeuten und Verweisen auf die übrigen Reaktionen, die in einer Sequenz dem Eintrag vorangehen oder folgen. Wenn erst einmal ein Zusammenhang zwischen einer geplanten Reaktion und einem Präzedenzfall hergestellt wurde, läßt sich die Schwierigkeit der praktischen Durchführung einfacher abschätzen. Die Herstellung dieses Zusammenhangs ist allerdings nicht immer einfach. Der Fokus des Interesses ist typischerweise das Reaktionszentrum, also die Atome und Bindungen, die direkt am Bindungsumordnungsprozeß beteiligt sind. Während das Vorhandensein möglicherweise störender übriger Gruppen durch einfache Substruktursuche festgestellt werden kann, ist die Identifizierung des Reaktionszentrums – automatisch oder manuell – nicht immer eindeutig möglich, und nicht immer ist der Bindungsumordnungsprozeß mit den wenigsten Schritten der real ablaufende. Trotzdem gibt es gewaltige Fortschritte auf diesem Gebiet. Ein weiterer Schwachpunkt vieler solcher Systeme ist das Fehlen von Einträgen, die Negativbeispiele darstellen.

Die Suche in einer Reaktionsdatenbank wird heute in vielen Laboratorien routinemäßig durchgeführt. Reaktionsdatenbanken sind aber natürlich nur dann eine Hilfe, wenn der Auftraggeber einer Suche ungefähr weiß, nach welchen Reaktionstypen er zu suchen hat. Während Reaktionsvorhersage- und Syntheseplanungsprogramme der klassischen Auslegung zumindest theoretisch ihre Wissensbasis systematisch und erschöpfend abarbeiten können, muß einer Reaktionsdatenbank durch den Benutzer die Linie einer Synthese in groben Zügen vorgegeben werden.

1.3.2. Anwendung von Reaktionsdatenbanken

Der Umfang, in dem Reaktionen aus der Literatur systematisch erfaßt werden, ist stark gewachsen. Spätestens seit der Einführung von CASREACT^[34], der ChemInform-RX-Reaktionsdatenbank^[35] und von ChemReact wird hier ein breites Feld abgedeckt, welches ständig mit neuen Daten ergänzt wird. Auch klassische, molekülorientierte Datenbanken wie Beilstein verfügen über Suchverfahren, die über Umwege zu gleichartigen Ergebnissen wie die direkte Suche in reaktionsorientierten Datenbanken führen^[39]. Wichtige Grundstücke an Basisreaktionen sind ebenfalls erhältlich^[40]: die Kataloge des Theilheimer, der CRDS (Chemical Reactions Documentation Service), das JSM (Journal of Synthetic Methods) und die Organic-Synthesis-Buchreihe.

Für Firmen mit einem großen Schatz an eigenen Reaktionsdaten bieten die Reaktionsdatenbanken eine Reihe von speziellen Vorteilen gegenüber der Wartung von Transformsystemen. Der wichtigste ist, daß bei ersteren ausschließlich mit „harten“, uninterpretierten Daten gearbeitet wird und jede einzelne Reaktion, die technisch völlig unproblematisch eingegeben werden kann oder gar direkt aus computerisierten Laborjournalen stammt, für sich einen Wert hat, während für einen Transform- oder gar Synthoneintrag eine einzelne Beispielreaktion nicht ausreicht. Transforms, die aus zu wenigen Reaktionen oder aus einem zu engen Anwendungsräumen vorschnell codiert wurden, schaden nach allgemeiner Erfahrung mehr, als sie nutzen, weil zu viele überbewertete, aus einer einzigen Reaktion ungerechtfertigt extrapolierte Ergebnisse resultieren. Selbst wenn der Neueintrag an sich für Synthesen wertvoll ist, so ist es – vor allem wenn die Eingabe von einem Team mit unterschiedlicher Erfahrung und unterschiedlichem Hintergrund durchgeführt wird – sehr schwer, den neuen Eintrag relativ zu den alten richtig zu bewerten.

Die automatische Extraktion von Transformbibliotheken aus Reaktionsdaten steckt noch in den Kinderschuhen. Ein Hauptproblem dabei ist die Gruppierung von Beispielreaktionen zu Klassen, um so aussagekräftige Verallgemeinerungen über Reaktionen ableiten zu können^[18, 41]. Wenn diese Verfahren so weit ausgereift sind, daß sie vollautomatisch auf großen Reaktionsdatenbanken agieren können, ließe sich automatisch Wissen zur Reaktionsvorhersage und Syntheseplanung ableiten, was die Bedeutung der Reaktionsvorhersage, ja möglicherweise sogar des Synthon/Transform-Konzeptes, innerhalb von Syntheseplanungssystemen erhöhen würde. Hierbei trifft man dann allerdings immer wieder auf ein Problem, das schon bei der manuellen Ableitung von Transformschemata ein großes Hindernis war und bereits bei der einfachen Reaktionsdatenbank

erwähnt wurde: Weder in der Literatur noch in Reaktionsdatenbanken findet man in größerem Maße Angaben über Reaktionen, die *nicht* ablaufen (also Negativbeispiele), und häufig wird nichts über die Schwierigkeiten, eine Reaktion zum Laufen zu bringen oder vernünftige Ausbeuten zu erhalten, berichtet. Genaue solche Informationen sind jedoch für die Eingrenzung des Verwendungsbereichs und die Abschätzung der generellen Nützlichkeit einer Reaktion oder eines Transforms unverzichtbar. Das Problem tritt auch bei der einfachen Anfrage in einer Reaktionsdatenbank auf, ist dort aber nicht so schwerwiegend, weil Kontextinformation zur Verfügung steht.

In jedem Fall aber wird eine automatische Klassifizierung des Inhalts von Reaktionsdatenbanken oder der Serien an Antwortreaktionen, die durch eine Suche erhalten wurden, in der Zukunft unumgänglich werden. Denn mit der Zunahme der Zahl an Reaktionen in einer Datenbank – und es gibt schon solche mit mehr als einer Million Reaktionen (CASREACT)^[34] – erhält der Benutzer häufig eine so große Zahl an Antworten, daß er nicht mehr gewillt ist, alle der Reihe nach durchzuarbeiten. Hier muß eine hierarchische Reaktionsklassifizierung, also ein weiterer Schritt der computergestützten Nachbearbeitung, einsetzen, um dem Benutzer einen strukturierten Zugang zum umfangreichen Faktenmaterial ermöglichen. Ansätze hierfür wurden bereits entwickelt^[41].

1.4. Endpunkte der ersten Programmgeneration

Syntheseplanung mit dem Computer ist von Anfang an bei der breiten Masse der Chemiker auf Skepsis bis Ablehnung gestoßen. Jetzt, da auch nach 25 Jahren Entwicklungsarbeiten durch eine Reihe von Forschergruppen noch kein System breit in die Praxis eingeführt ist, scheint diese kritische Haltung bestätigt zu werden. Selbst bei denjenigen Wissenschaftlern, die anfänglich mit Enthusiasmus die Syntheseplanung mit dem Computer begrüßt haben, ist eine weithin spürbare Ernüchterung zu beobachten. Deutlich dokumentiert dies beispielsweise die Einstellung der Weiterentwicklung des Programms CASP, das unter den Transformsystemen die weitaus größte Sammlung an Reaktionstypen enthält. Das Fehlen von überzeugenden Resultaten, die mit dieser Wissensbasis gewonnen wurden, und der generelle Mangel an Akzeptanz durch den präparativ tätigen Chemiker markieren einen Endpunkt der Entwicklung bei Systemen dieser Art. Das bedeutet nicht, daß CASP nicht mehr genutzt werden würde, sondern daß eine Erweiterung der Transformdatenbank innerhalb des bestehenden Schemas als zu aufwendig und nicht sinnvoll angesehen wurde und damit das Konzept dieses Systems ausgereizt ist. Versuche, aus Reaktionsdatenbanken automatisch Transforms für dieses Programm zu generieren, sind ebenfalls gescheitert. Das letzte von einem Konsortium getragene Großprojekt, für das weiterhin erhebliche Mittel bereitstehen, ist das japanische System AIPHOS^[42]. Letzten Endes werden aber auch hier Synthons und Transforms auf strategische Bindungen angewendet, so daß mit den gleichen wohlbekannten Schwächen wie bei den anderen Systemen gerechnet werden muß, sofern nicht noch grundlegend neue Konzepte hinzukommen. Einige noch nicht bewertbare Entwicklungen sind zur Zeit in der Planungs- und Implementierungsphase^[42].

2. Neue Ansätze

Über den Erfolg eines Systems entscheiden die Benutzer, nicht die Entwickler. Wenn die Benutzer wirklich freie Wahl zwischen einer Vielzahl von Systemen haben, sieht die Realität aus wie in der Statistik des niederländischen CAOS/CAMM-Center^[43] (Tabelle 1). Die Dominanz der Reaktionsdatenban-

Tabelle 1. Benutzerstatistik des niederländischen CAOS/CAMM-Center für die Zeit von Januar bis Mai 1993.

| Programm | <i>N</i> [a] | Programm | <i>N</i> [a] |
|----------|--------------|----------|--------------|
| LHASA | ca. 25 | ORAC | 877 |
| CAMEO | 7 | REACCS | 103 |
| CHIRON | 12 | SYNLIB | 123 |
| SYNGEN | 8 | | |

[a] *N* = Zahl der Aktivierungen.

ken wird schon aus diesen Zahlen deutlich^[44]. Bei der Analyse der Daten in Tabelle 1 muß man zudem berücksichtigen, daß bei *allen* Einträgen eine unbekannte, aber nicht unerhebliche Zahl von Erkundungen durch Erstbenutzer anzunehmen ist, die sicherlich überwiegend auf Neugierde zurückzuführen sind. Nach Meinung des Autors der Statistik sind damit die registrierten Aufrufe der Syntheseplanungssysteme fast nur aus Neugierde getätigten worden.

Die Zugriffe auf ORAC umfassen auch Struktursuchen im Fine Chemicals Directory (FCD), einer umfassenden Liste kommerziell erhältlicher Chemikalien. Die Mehrheit der Aufrufe von CAMEO sollen pK_a -Abschätzungen und nicht Reaktionsvorhersagen zum Ziel gehabt haben. Diese Zahlen belegen zumindest in der Tendenz, daß normale Benutzer mit den zur Zeit verfügbaren Syntheseplanungssystemen wenig anzufangen wissen.

2.1. Konzeptionelle Probleme der klassischen Systeme

Aus den Zahlen in Tabelle 1 wird deutlich, daß die bisher vorhandenen Programmsysteme entweder noch nicht ausgereift genug sind oder an den Bedürfnissen des präparativ arbeitenden Chemikers vorbei entwickelt wurden. Die Methoden, die Chemiker bei der Syntheseplanung einsetzen, sind natürlich abhängig von persönlichen Stilen; es läßt sich aber doch eine allgemeine Tendenz formulieren:

Chemiker planen Synthesen nicht nach starren Schemata. Chemiker betrachten bei der Planung einer Synthese einzelne Reaktionen oder sogar Synthone oft zu allerletzt. Am Anfang der Entwicklung einer Strategie steht die Suche nach größeren Strukturfragmenten, aus denen das Molekül aufgebaut werden könnte, und die Suche nach Funktionalitäten in der Zielstruktur und deren Verteilung über das Gerüst. Diese werden nicht notwendigerweise sogleich nach Synthon-Gesichtspunkten eingeordnet, denn das Arsenal an Umpolungsmethoden ist mittlerweile sehr umfangreich, so daß eine falsche Polarität an einer Position kein unüberwindbares Hindernis mehr ist. Ebenfalls sehr früh werden weitreichende Beziehungen zu möglichen Ausgangsverbindungen und verwandten Verbindungen hergestellt.

Erst wenn grobe strukturelle Klassifizierungen von möglichen Zwischenstufen und Ausgangsmaterialien vorliegen, werden die Lücken mit konkreten Reaktionen und Synthonschemata überbrückt. Und erst bei diesem Überbrückungsschritt kristallisiert sich im Wechselspiel die genaue Konstitution der Zwischenstufen und Ausgangsmaterialien heraus. Wir haben es hier weder mit einer klassischen Bottom-Up- oder Top-Down-Suche noch mit einer bidirektionalen Suche, sondern mit einem komplexen Netz zwischen dem Synthesenziele und möglichen Ausgangsmaterialien zu tun, das vom Chemiker als sinnvoll erkannte Zwischenstufen sowie verfügbare Reaktionen und Reagentien umfaßt. Die Suche nach gangbaren Wegen erfolgt nicht in kleinsten Schritten von einem Punkt zum nächsten, sondern abwechselnd in kleinen und großen Sprüngen, in beiden Richtungen und unter ständiger Wandlung der exakten Struktur der angepeilten Zwischenstufen und Ausgangsmaterialien. Diese Vorgehensweise wird in Abbildung 2 anhand der Analyse dreier Synthesen des Antimalariamittels Qinghaosu 1 illustriert^[4,5a].

verbindung 1 sichergestellt worden war, galt es, die Synthese von 2 zu entwerfen. Hierbei suchten die drei Arbeitsgruppen nach leicht erhältlichen Ausgangsmaterialien, die wesentliche Strukturelemente der Synthesevorstufe 2 bereits enthielten. Mit R-(+)-Citronellol 3^[4,5b], (-)-Isopulegol 4^[4,5c] und (+)-Car-3-en 5^[4,5d] wurden sie fündig. Jeweils von einem dieser leicht verfügbaren Ausgangsstoffe ausgehend wurden dann in Vorwärtsrichtung die Synthesen entwickelt, immer als Ziel die Vorstufe 2 im Auge behaltend. Die Kombination einer Retrosynthese hin zu einer zentralen Synthesevorstufe mit der Entwicklung eines Verfahrens zur Umwandlung dieser Vorstufe in das Synthesenziele, dem Erkennen struktureller Ähnlichkeiten zwischen der Vorstufe und leicht erhältlichen Ausgangsstoffen sowie der wenn auch manchmal mühevollen Entwicklung eines Weges in Vorwärtsrichtung vom jeweils gewählten Ausgangsmaterial zur Synthesevorstufe führte bei all diesen Syntheseplanungen zum Erfolg. In keinem Fall wurde die Synthese von 2 durch systematische, stufenweise Retrosynthese geplant.

Ein weiterer Aspekt, mit dem sich die klassischen computergestützten Ansätze zur Syntheseplanung schwer tun, ist, daß die Zielstruktur häufig nur generisch bestimmt ist. Dies gilt insbesondere für den Pharmabereich, in dem die Zielstruktur in der Regel als Pharmakophor-Hypothese vorliegt und die genaue strukturelle Festlegung erst in einer eng an die Syntheseplanung gekoppelten Optimierung erfolgt. Dabei sind auch externe Parameter wie Ladungsverteilung und Raumerfüllung der Struktur zu berücksichtigen, bevor sich aus einer Leitstruktur das konkrete Zielmolekül herausschält. Damit wird das Synthesenziele selbst unscharf, und eine effiziente Synthese, die zu einer zur Modell-Zielstruktur funktionell äquivalenten, aber nicht strukturell identischen, sondern nur ähnlichen Verbindung führt, wird zu einer akzeptablen Problemlösung. Dies ist ein Gebiet, das mit traditionellen Transform-Methoden kaum zu bearbeiten ist, weil sie eine konkrete Zielstruktur als systematisch zu zer-gliederndes Objekt benötigen.

Um die Reaktionsstufen zwischen den Meilensteinen einer Synthese aufzufüllen, können Reaktionsdatenbanken sehr nützlich sein. Voll- oder Substruktursuchen in Chemikalienkatalogen sind gut geeignet, um konkrete Ausgangsmaterialien aufzufinden. Dies sind ausgereifte Verfahren, die auch vom präparativ arbeitenden Chemiker bereitwillig angenommen worden sind. Im Vergleich dazu fällt der Gebrauchswert von klassischen Syntheseplanungsprogrammen stark ab. Wenn systematische Suchen durchgeführt werden, entstehen meist enorm breite Synthesebaume aus Vorstufen und Einzelreaktionen, die kaum noch zu überblicken sind und überwiegend Sackgassen enthalten, in die ein pragmatisch urteilender Chemiker niemals hineingelangt wäre. Werden nur kurze Sequenzen untersucht, so werden brauchbare (aber natürlich häufig offensichtliche) Reaktionen erhalten, doch wegen der abstrahierten Form der zugrundeliegenden Transforms oder Synthonschemata ist dann noch keineswegs klar, ob Präzedenzfälle mit wirklich vergleichbaren Umständen existieren. Dies muß durch Konsultation einer Reaktionsdatenbank oder durch eine Literatursuche abgeklärt werden, was bei der Kürze der Sequenz auch direkt hätte geschehen können. Zudem entsprechen die Knoten im Synthesebaum vollständig spezifizierten Verbindungen, die durch konkrete Reaktionen verbunden sind. Kleinere Abwandlungen erfordern daher später erneute Programm läufe, oder es

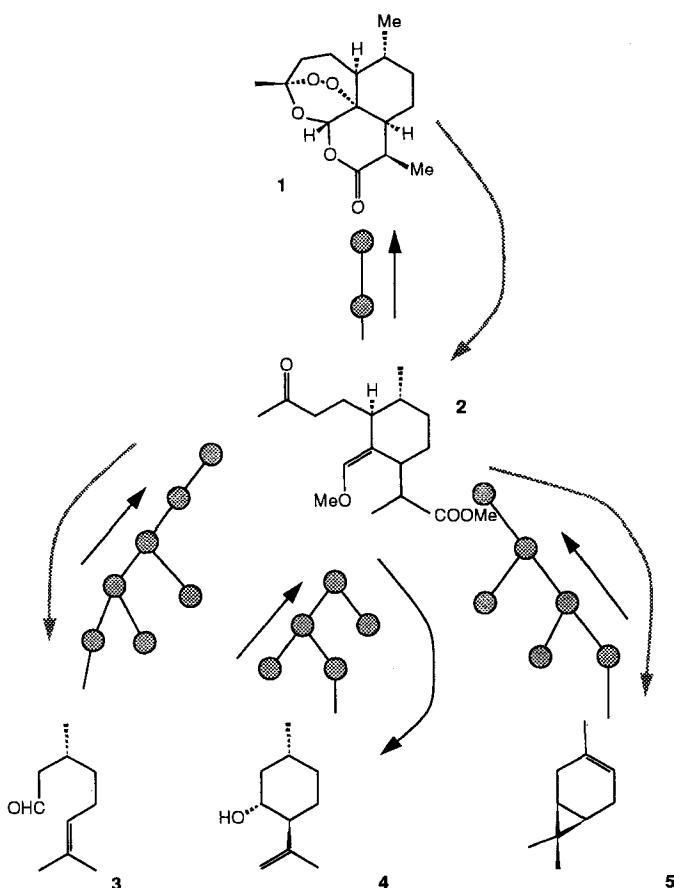


Abb. 2. Syntheseplan in der Realität: Es werden eine Reihe unterschiedlicher Methoden zur Suche und Planung – sowohl in retrosynthetischer, als auch in synthetischer Richtung – eingesetzt, um nicht den Synthesebaum in seiner ganzen Breite abarbeiten zu müssen. Für Einzelheiten siehe Text.

Ein erster wesentlicher Schritt bei der Entwicklung einer Synthese von 1 war die Idee, daß dessen ungewöhnlicher 1,2,4-Trioxan-Ring durch Photooxidation der Vorstufe 2 aufgebaut werden könnte. Diese Umwandlung von 2 in 1 konnte tatsächlich in zwei Stufen verwirklicht werden^[4,5b]. Nachdem mit 2 eine zentrale Zwischenstufe festgelegt und ihre Umwandlung in die Ziel-

ist, falls solche Variationen bereits im ersten Programmlauf erzeugt wurden, jede Variation ein eigener Subbaum, der die Ergebnisliste weiter aufblättert.

2.2. Neue Methoden zur Unterstützung des Synthesechemikers

Eine Aufgabe, bei der bisherige Systeme die Chemiker nicht unterstützen, ist das Erkennen von weitreichenden Beziehungen zwischen Zielverbindung, erhältlichen Ausgangsmaterialien und möglichen Zwischenstufen. Doch ist gerade dies ein wichtiger Bestandteil der Praxis. Für die Planung einer Synthese ist es hilfreich, von einem System auf möglicherweise synthetisch nutzbare Ähnlichkeiten zwischen dem Zielmolekül und der durch einen Menschen nicht mehr überschaubaren Vielfalt kommerziell erhältlicher Verbindungen hingewiesen zu werden. Die Bedeutung liegt hier auf *synthetisch nutzbarer* Ähnlichkeit, und dies ist eine Anforderung, der bekannte Ähnlichkeitsalgorithmen^[46] nur höchst unvollständig genügen.

Verfahren, bei denen die minimale chemische Distanz^[47], d.h. der minimale Satz an Bindungsumordnungen unter Erfüllung vorgegebener Restriktionen, zwischen zwei Verbindungen explizit berechnet wird, sind wertvoll, um mögliche Sequenzen von komplexen Bindungsumordnungen zu erkennen. Doch wegen des rechnerischen Aufwandes sind sie zur Zeit noch nicht zum direkten systematischen Abgleich einer Zielstruktur gegen Tausende von potentiellen Ausgangsmaterialien geeignet. Die gegenwärtigen Implementationen basieren noch ausschließlich auf topologischen Kriterien, d.h. sie kümmern sich noch nicht um eine Reaktivitätsabschätzung und damit um die praktische Realisierbarkeit eines Schemas an Bindungsumordnungen. Diese Methoden eignen sich speziell zum Erkennen von Umlagerungspfaden und werden möglicherweise in der Zukunft ein Werkzeug für Spezialfälle, aber keine komplette Problemlösung sein.

Wenn keine direkte Beziehung zwischen einer Zielverbindung und einer Ausgangsstufe herzustellen ist, sind Mechanismen zur systematischen Organisation einer Synthesestudie erforderlich, die über die Erzeugung eines Synthesebaums hinausgehen. Manuelle oder computergestützte Zerlegung der Zielverbindung an als strategisch erkannten Bindungen führt zu Meilensteinen, auf die dann weitere Such- und Planungsoperationen angewendet werden können: Ähnlichkeitssuche, Reaktionsplanung, Reaktionsvorhersage oder weitere Zerlegungsschritte. Die einzelnen Registrierpunkte einer solchen Studie sind dann nicht mehr nur einzelne Moleküle, sondern Fragmentierungsgruppen, Listen von Treffern aus Datenbanksuchen oder Reaktionsvorhersage-Ergebnisse. Dem Chemiker müssen Werkzeuge an die Hand gegeben werden, die ihn bei seiner Suche unterstützen und die Fortschritte protokollieren. Der Chemiker muß jederzeit frei sein, ein Werkzeug seiner Wahl anzuwenden, und dies an jedem Punkt der Studie, an dem es ihm sinnvoll scheint, und nach Gutdünken in synthetischer oder retrosynthetischer Richtung. Menschliche Urteilskraft ist zur Zeit jedem Syntheseplanungsprogramm überlegen, wenn es darum geht, uninteressante Regionen einer Studie großräumig auszugrenzen. Dies bedeutet keinesfalls, daß bei den neuen Systemen nicht mehr vom größten Vorteil vorurteilsfreier maschineller Berechnung profitiert

werden kann – dem Aufzeigen von Wegen, auf die der Chemiker, da er nicht erschöpfend eine Suche zu Ende führen will oder kann, nicht selbst gekommen wäre. Wenn eine Region einer Studie nicht offensichtlich oder nicht zufriedenstellend ist, steht es dem Benutzer frei, die Zügel in Stufen zu lockern. Die systematische Anwendung Tausender von Transforms über mehrere Stufen führt jedoch nach übereinstimmender Ansicht zu so viel Ausschuß, daß das Sichten der ausgegebenen Vorstufen und Reaktionsschritte der Suche nach einer Nadel im Heuhaufen gleicht und damit nicht zumutbar ist.

2.3. Strukturelle Ähnlichkeit

Methoden, um Ähnlichkeiten zwischen Strukturen zu erkennen, die in einer Synthese genutzt werden können, werden als mit am dringendsten nötig angesehen. Was man unter *synthetisch nutzbarer* Ähnlichkeit versteht, ist aber noch wenig definiert. Die Suche nach struktureller Ähnlichkeit wird zwar bei Untersuchungen zur biologischen Aktivität und bei der Aufstellung quantitativer Struktur-Wirkungs-Beziehungen (QSAR: Quantitative Structure-Activity Relationship) bereits mit Erfolg angewendet^[46], doch handelt es sich hierbei um Ähnlichkeitsmaße, die sich mit den Anforderungen der Syntheseplanung nur schwer zur Deckung bringen lassen. Ein populäres Verfahren ist das Abzählen einzelner Gruppierungen von Atomen in den zu vergleichenden Molekülen^[48]. Die Mengen der aufgefundenen Strukturelemente werden dann zu einem Ähnlichkeitsmaß verknüpft, das vielfach gut mit physiologischen Wirkungen korreliert. Das Problem hierbei ist die mangelnde Empfindlichkeit des Ähnlichkeitsmaßes gegenüber Gerüstveränderungen und die Überbewertung von Veränderungen an reaktiven Heteroatomzentren. So ist für physiologische Effekte die An- oder Abwesenheit einer zusätzlichen Methylgruppe häufig irrelevant, aber sie ist entscheidend dafür, ob eine Verbindung bei einer Synthese eingesetzt werden kann oder nicht, wenn das Synthesziel spezifisch ist oder Einflüsse auf die Reaktivität und die Raumerfüllung am Reaktionszentrum bestehen. Dagegen ist die spezifische Art eines reaktiven Heteroatomzentrums für eine Synthese relativ uninteressant, solange die prinzipielle Reaktivität und selektive Angreifbarkeit weiterhin gegeben sind und die Reaktionsbedingungen hierauf eingestellt werden können. Das Synthesepotential einer Verbindung hängt von vielen Faktoren ab, die bei der Syntheseplanung auf komplexe Weise zusammenhängen und für sie spezifisch sind. Ähnlichkeitsdefinitionen aus anderen Anwendungsgebieten sind nicht übertragbar.

Daher sind die Ähnlichkeitsmethoden aus dem QSAR-Gebiet nur sehr eingeschränkt zur Syntheseplanung einsetzbar. Klassische strukturorientierte Suchmethoden für Datenbanken wiederum bieten zu wenig Flexibilität. Suchen werden sehr ineffizient und unselektiv, wenn ein Kohlenstoffgerüst vorgegeben wird und alle offenen Valenzen z.B. als Wasserstoff- oder Heteroatome angegeben werden müssen. Obendrein sind in solche Suchmethoden keine Vorstellungen von Reaktivität, z.B. vom Einfluß der Umgebung auf die Reaktivität eines Wasserstoffatoms, eingebaut.

Das Pioniersystem, das auf elementare Art einen präparativ nutzbaren Zusammenhang zwischen einer Zielverbindung und einem Ausgangsmaterialkatalog herzustellen versucht, ist

SST^[49]. SST ist kein Syntheseplanungssystem, sondern bietet lediglich als Suchmethode drei Abstraktionsgrade für Substituenten am zentralen Kohlenstoffgerüst, die gegen eine mit denselben Methoden vortransformierte Datenbank gesucht wurden.

In unserer Gruppe wurde eine Reihe von Ähnlichkeitsdefinitionen entwickelt, die ganz gezielt auf strukturelle Beziehungen in einer Synthese gemünzt sind^[50a–c]. Einige stützen sich auf größere gemeinsame Teilstrukturen wie gleiche Ringsysteme oder Kohlenstoffgerüste in den zu vergleichenden Molekülen. Die überwiegende Mehrzahl aber verwendet Kriterien, die direkt aus chemischen Reaktionen abgeleitet werden. Dies sind zugleich die für die Planung von Synthesen besonders leistungsfähigen Verfahren. Algorithmen zur Bestimmung dieser neuartigen Ähnlichkeiten wurden entwickelt und in das WODCA-System integriert. Genaueres dazu wird in den folgenden Abschnitten noch zur Sprache kommen.

3. Das WODCA-System

Anfang der neunziger Jahre wurde in unserem Arbeitskreis das WODCA-System entwickelt^[50]. WODCA ist ein interaktives, graphikorientiertes System zur Unterstützung bei der Lösung der Probleme, mit denen ein in der Synthese arbeitender Chemiker konfrontiert ist. WODCA folgt den in den vorhergehenden Abschnitten entwickelten Anforderungen an eine neue Generation von Syntheseplanungssystemen. Es bietet eine Umgebung zum Testen und Weiterentwickeln der zugrundeliegenden Ideen und Methoden. Wie der Name bereits andeutet, ist WODCA nicht nur ein Programm zur Syntheseplanung, sondern erfüllt in unserem Arbeitskreis auch andere Aufgaben. In diesem Beitrag werden wir jedoch nur auf den Syntheseplanungsaspekt eingehen.

Schwerpunkte bei der Entwicklung dieses Systems waren die Implementierung neuartiger Ähnlichkeitsalgorithmen bei Datenbanksuchen, die Verwendung eines flexiblen Systems zur Organisation einer Studie und die dynamische Integration von Modulen, die intelligente Operationen mit und an Molekülen durchführen (Bewertung, Erkennung strategischer Bindungen). Von vornherein wurde auf den Einsatz einer klassischen Synthon- oder Transformdatenbasis konsequent verzichtet. Eine Reihe externer Programme kann von WODCA aus genutzt werden: die Programme der EROS-Serie^[20] (Reaktionsvorhersage, Reaktivitätsabschätzung) und PETRA (physikochemische Eigenschaften von Molekülen^[21–24]) sowie eine Reihe von Hilfs- und Visualisierungsprogrammen (Molekül- und Reaktionsdisplays, Planübersichten, Reportgenerator). Klassische Suchverfahren wie Vollstruktur- und Substruktursuche in Datenbanken mit Ausgangsmaterialien sowie einige bereits aus der Literatur bekannte Ähnlichkeitsalgorithmen sind integraler Bestandteil des Systems. Die Anbindung an eine Reaktionsdatenbank ist erst in Bearbeitung.

WODCA ist ein interaktives System mit ausgeprägter graphischer Komponente. Abbildung 3 zeigt ein typisches Beispiel für die über den Bildschirm übermittelte Information während einer WODCA-Sitzung. Dank der modernen Fenstertechnik (hier auf einer SUN-Workstation unter SunView, die Umstellung auf X-Windows und MS-Windows ist in Bearbeitung) können die

Ergebnisse mehrerer Analyse- und Suchmethoden gleichzeitig wiedergegeben werden. Damit wird in besonderem Maße das laterale Denken des Chemikers, der bei der Planung von Synthesen zwischen Gedankengängen und Fragestellungen hin- und herspringt, unterstützt. In Abbildung 3 sind gleichzeitig die Ergebnisse der Bestimmung der bestbewerteten strategischen Bindung, der Suche nach geeigneten Ausgangsmaterialien für die beiden Molekülfragmente und die Einträge im Chemikalienkatalog für eines der Ausgangsmaterialien zu sehen.

Mit diesem Bild soll auch deutlich werden, wie der Benutzer aktiv in eine Syntheseplanungsstudie eingebunden wird, wie er frei unter den in WODCA angebotenen Methoden wählen kann und wie er durch den Vergleich der Ergebnisse der von ihm gewählten Suchverfahren Anregungen erhält und damit rasch eine Syntheseplanungsstudie zu einem Ergebnis treiben kann, das seinen konkreten Anforderungen am besten entspricht.

Die Methodenkollektion in WODCA dient dazu, dem Chemiker sofortige Antwort auf „Was wäre, wenn ...“-Fragestellungen zu geben. Es läuft ein ständiges Wechselspiel zwischen der Kreativität des erfahrenen Chemikers und der Geschwindigkeit, Präzision und Gründlichkeit des Computers ab. Ein wichtiger Gesichtspunkt bei der Methodenentwicklung war die Ausführungsgeschwindigkeit, so daß im Normalfall eine Antwort auf einer Workstation innerhalb von Sekunden vorliegt oder zumindest bereits nach einer Teilaufgabe die Suche abgebrochen werden kann. Es liegt in der Verantwortung des Chemikers, aus dem zur Verfügung gestellten Werkzeugkasten das jeweils geeignete auszuwählen. In WODCA werden dem Benutzer keine Vorschriften über die Reihenfolge oder den Umfang der gewählten Operationen gemacht, sondern es wird auf seine Erfahrung und Intelligenz vertraut. Wenn das System Vorschläge macht, bei denen der Benutzer abweichender Meinung ist, kann er einen anderen Vorschlag manuell eingegeben oder die Systemantwort editieren, bevor diese Zwischenergebnisse an weitere Module weitergereicht werden. Ein Beispiel hierfür sind automatisch erkannte strategische Bindungen, bei denen der Benutzer unter Umständen eine andere Perspektive als das System hat. Strategische Bindungen können deshalb per Mausklick gelöscht, adiert oder mit anderen Attributen versehen werden.

Das ganze Arsenal der Methoden kann auf jedes sinnvolle Objekt innerhalb einer Studie angewendet werden und damit die Studie an jeder Stelle in neue Richtungen expandiert werden. Das System protokolliert alle Aktionen innerhalb einer Studie, d. h. nicht nur die Ergebnisse, sondern auch die durchgeführten Operationen und ihre Parameter werden in Dateien gespeichert und können jederzeit eingesehen werden. Die Beziehungen zwischen den Schritten werden graphisch dargestellt. Jeder Schritt wird zu einem Planelement. Wichtige Klassen von Planelementen sind Strukturen, Sätze von strategischen Bindungen, Fragmente, Ergebnisse von Datenbanksuchen (Match-Listen) und Ergebnisse von Reaktionsvorhersagen. Begonnene Studien können zu späteren Zeitpunkten erneut durchgesehen oder fortgeführt werden.

3.1. Kernoperationen des Systems

Die generelle Vorgehensweise bei der Studie eines Syntheseproblems mit WODCA zeigt Abbildung 4. In der derzeitigen

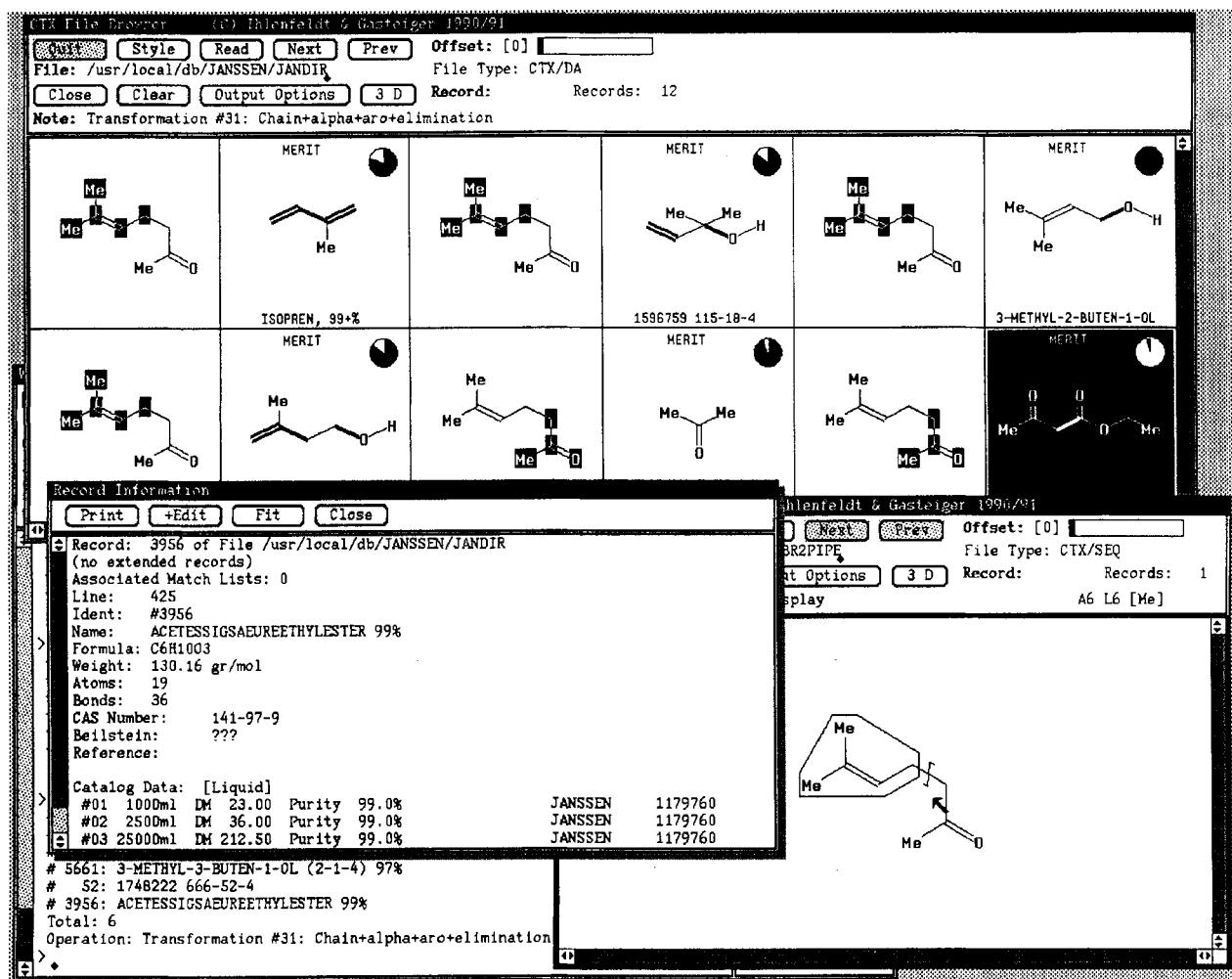


Abb. 3. Graphische Arbeitsumgebung von WODCA: Im rechten unteren Fenster ist das Syntheseziel zu sehen, zusammen mit der gerade bearbeiteten strategischen Bindung. Eines der Fragmente, das durch Bruch dieser strategischen Bindung erhalten wird, ist vom System durch Umrahmung hervorgehoben worden. Im oberen Bereich werden diejenigen Verbindungen aus einer Ausgangsmaterialdatei aufgeführt, die als zu einem der beiden Molekülfragmente ähnlich erkannt wurden. Jedem der Ausgangsmaterialien ist dabei links noch das Syntheseziel mit der Markierung des entsprechenden Fragments gegenübergestellt; das berechnete Ausmaß an struktureller Ähnlichkeit wird jeweils in Form eines Kuchendiagramms rechts oben visualisiert. Das vom Benutzer gewählte Ähnlichkeitskriterium (ganz unten links: Transformation Nr. 31: Haupt-Kohlenstoffgerüst + α -Heteroatome + alle Arene + Eliminierung) erfordert Übereinstimmung im Kohlenstoffgerüst und bei den Heteroatomen in der α -Sphäre. Arengerüste dürfen dabei nicht verändert werden, dagegen sind Strukturunterschiede, die sich durch Eliminierungsreaktionen beseitigen lassen, erlaubt. Der Benutzer hat ein Ausgangsmaterial, Acetessigester, das durch Schwärzung des Hintergrunds hervorgehoben wird, ausgewählt, um weitere Informationen darüber aus dem Ausgangsmaterialkatalog zu erhalten. Diese Daten wie Name, Summenformel, molare Masse, CAS-Registry-Nummer und Preisinformation werden im Fenster links unten aufgeführt.

Konfiguration lassen sich drei Kernoperationen identifizieren:

- Suche nach Ausgangsmaterialien
- Suche nach Synthesevorstufen
- Vorhersage von Reaktionen

Der Unterschied zwischen einer Synthesevorstufe und einem Ausgangsmaterial in diesem Kontext ist, daß ein Ausgangsmaterial ein in einem Chemikalienkatalog enthaltenes, vollständig spezifiziertes Molekül ist, während eine Vorstufe eine unscharf definierte, lose zwischen Ausgangsmaterial und Zielverbindung schwebende Modellstruktur ist, von der ausgehend die Lücken nach oben und unten im Syntheseplan schrittweise geschlossen werden.

Die vordringlichste Aufgabe in der Syntheseplanung ist die Rückführung des Syntheseziels auf geeignete Ausgangsmaterialien. Dementsprechend nehmen Datenbanken mit kommerziell erhältlichen Verbindungen eine zentrale Stellung im WODCA-System ein. Mit Hilfe variantenreicher Suchmethoden (siehe

Abschnitt 3.1.1 und Abb. 5) sollen für die Synthese geeignete Ausgangsmaterialien ermittelt werden (linker Zweig in Abb. 4). Bei komplexeren Zielmolekülen wird dies nicht gleich von Erfolg gekrönt sein, so daß das Molekül zuerst an als strategisch erkannten Bindungen in Vorstufen zerlegt werden muß. Für diese Synthesevorstufen sind dann geeignete Ausgangsmaterialien zu ermitteln (rechter Zweig in Abb. 4). Dieser Prozeß der Zerlegung und Ausgangsmaterialsuche wird so lange fortgesetzt, bis alle benötigten Vorstufen auf verfügbare Ausgangsmaterialien zurückgeführt sind. Aus den als geeignet erkannten Ausgangsmaterialien sind dann konkrete Verbindungen und Reagenzien auszuwählen, und mit Methoden der Reaktionsvorhersage ist zu überprüfen, ob die einzelnen Reaktionsschritte auch in der gewünschten Richtung ablaufen (unterer Teil in Abb. 4). Die einzelnen Vorstufen und Ausgangsmaterialien werden in einen Syntheseplan eingeordnet, der bei sehr grober Sichtweise einem Synthesebaum entspricht. Er ist allerdings bedeutend weniger umfangreich als in einer klassischen, systemati-

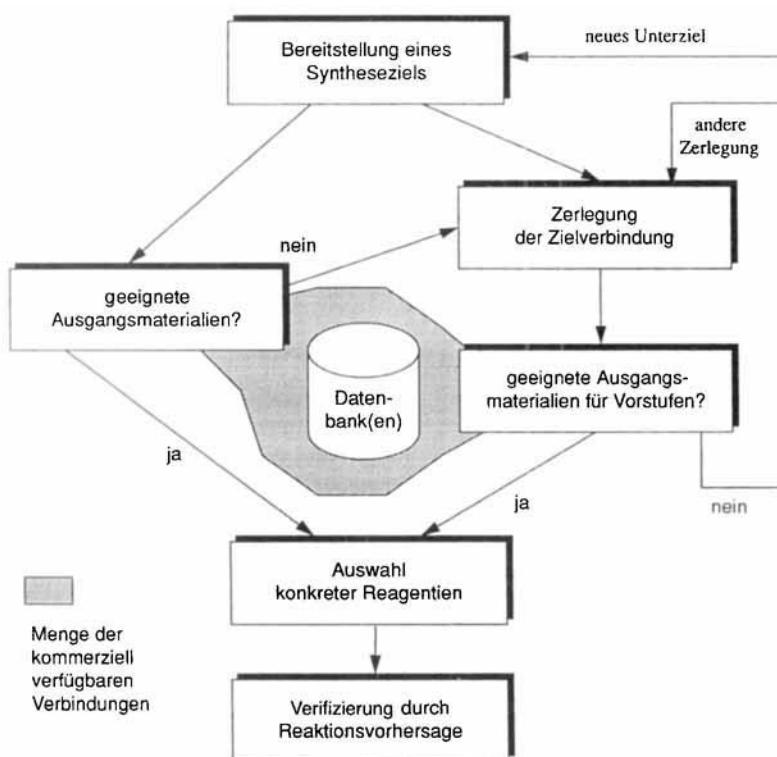


Abb. 4. Allgemeines Ablaufschema einer Syntheseplanungsstudie mit WODCA: Wechselbeziehungen zwischen der Suche nach Ausgangsmaterialien, der Ermittlung von Syntheseverstufen und der Bestätigung durch Reaktionsvorhersage.

schen retrosynthetischen Analyse. Der Informationsgehalt und die Komplexität (Klassifizierung, Historie, Zusammenhänge mit anderen Einträgen) eines solchen Eintrages ist dagegen sehr viel höher.

3.1.1. Die Suche nach Ausgangsmaterialien

Eines der Kernelemente beim Arbeiten mit WODCA ist es, Beziehungen zwischen einer Zielverbindung oder einer Vorstufe und einem Ausgangsmaterial herzustellen. An Ausgangsmaterialkatalogen stehen zur Zeit die Angebote mehrerer Feinchemikalienfirmen, spezialisierte, manuell kompilierte Kataloge sowie der Katalog des Programms CHIRON^[32] zur Verfügung; weitere Kataloge lassen sich leicht einfügen. Alle Operationen können mit jedem beliebigen Katalog durchgeführt werden; der Benutzer kann nach Gutdünken zwischen den Katalogen wechseln.

WODCA stellt zum einen klassische Methoden wie Substruktur- oder Vollstruktursuche (diese allerdings über ein neuartiges Hashcoding-Verfahren^[51]), fragmentorientierte Ähnlichkeitsmaße und allgemeine Datenbank-Suchverfahren wie numerische und textbezogene Suche nach Namen und Attributen bereit. Zum anderen wurden Methoden entwickelt, welche die Suche auf *synthetisch nutzbare* Ähnlichkeit fokussieren. Zwei der neu entwickelten Verfahren sind:

1. *Transformationssuchen*: Eine Anfragestruktur wird nach klar definierten Prinzipien, die allgemein gefassten Reaktionstypen entsprechen oder die be-

stimmte Strukturteile herausschälen, verändert und mit den entsprechend transformierten Molekülen der Ausgangsmaterialdatei verglichen. Auch hier gelangt das neue Hashcoding-Verfahren^[51] zum Einsatz, und daher liefern diese Suchen schnell Resultate (meist in weniger als einer Sekunde).

2. *Pfadlängencode-Suchen*: Substitutionsmuster und Strukturen von Kohlenstoffgerüsten werden unter Synthesegesichtspunkten verglichen. Anwendungsbeispiele finden sich in den Abschnitten 4.1 und 4.2. Eine Übersicht über das Methodenarsenal gibt Abbildung 5.

Da Transformationssuchen und Pfadlängencode-Suchen ausgesprochen leistungsfähig bei der Definition synthetisch nutzbarer Ähnlichkeit und damit bei der Suche nach geeigneten Ausgangsmaterialien sind, sollen sie im folgenden stärker betont werden. Eine genauere Erläuterung der Transformationssuche und Beispiele, die deren Bedeutung für die Syntheseplanung unterstreichen, finden sich auch in Lit. [50 a–c]. Die Transformationssuchen zerfallen in zwei Klassen: solche, die Ähnlichkeit zwischen zwei Strukturen anhand von Teilstrukturen bestimmen, und solche, die Ähnlichkeit feststellen, wenn sich zwei Strukturen durch verallgemeinerte Reaktionstypen wie Substitution, Hydrolyse oder Oxidation ineinander umwandeln lassen. Die Grenzen

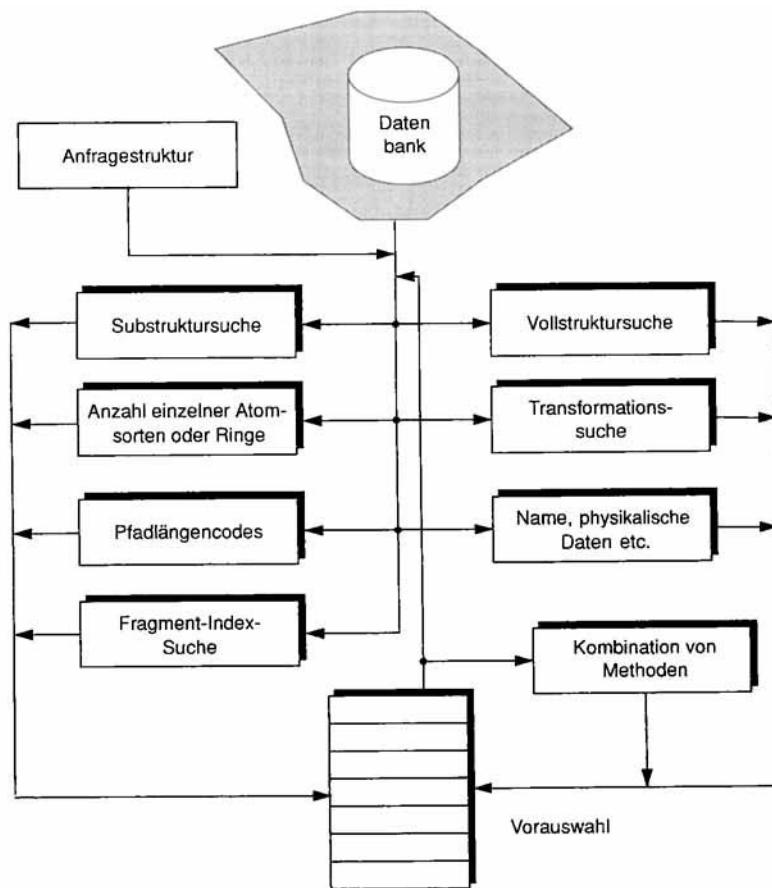


Abb. 5. Methodenarsenal zur Vorauswahl von Ausgangsmaterialien.

zwischen diesen beiden Klassen sind allerdings nicht scharf definiert, da auch bei Suchen der ersten Kategorie denjenigen Teilstrukturen Vorrang gegeben wird, die in chemischen Reaktionen intakt bleiben.

Abbildung 6 zeigt das Ergebnis einer Ähnlichkeitssuche mit 1,8-Diazabicyclo[5.4.0]undec-7-en **12** als Anfrage. Die Suche erfolgte hier, wie bei den meisten Beispielen in diesem Beitrag, im Katalog von Janssen Chimica (von 1988, die Firma heißt jetzt Acros Organica Division), der 8465 Verbindungen enthält, von 7849 strukturell verschiedenen sind. Als strukturell nicht verschiedenen werden nur abweichende Reinheitsgrade und Qualitätsstufen, Präparationen und isotopenmarkierte Verbindungen gewertet, Stereoisomere dagegen sind separate Verbindungen.

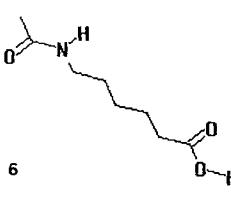
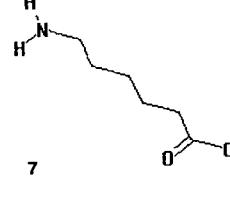
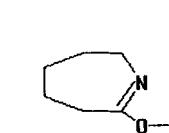
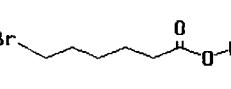
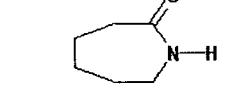
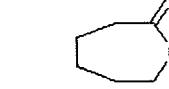
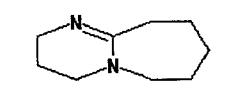
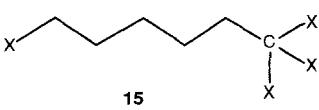
| | | |
|---|---|---|
| 28 JANSSEN 1865733  6 6-ACETAMIDOHEXANSÄURE | 370 JANSSEN 1033048  7 6-AMINOHEXANSÄURE 99+% | 804 JANSSEN 1051135  8 0-METHYLCAPROLACTIM 99% |
| 1332 JANSSEN 1656777  9 6-BROMHEXANSÄURE 98% | 1660 JANSSEN 1082356  10 EPSILON-CAPROLACTAM 99+ | 1661 JANSSEN 1734478  11 EPSILON-CAPROLACTON MON |
| 2653 JANSSEN 1606156  12 1,8-DIAZABICYCLO(5.4.0) | 7691 JANSSEN 1386995  13 1386995 7203-96-5 | 8443 JANSSEN 2087621  14 2087621 02556-73-2 |

Abb. 6. Verbindungen, die zu 1,8-Diaza-bicyclo[5.4.0]undec-7-en ähnlich sind, da sie das gleiche Kohlenstoffgerüst mit gleichartiger Substitution durch Heteroatome aufweisen.

Für diese Suche wurde ein Ähnlichkeitskriterium gewählt, das auf einer gemeinsamen Teilstruktur basiert; Moleküle werden als ähnlich erkannt, wenn sie das gleiche Kohlenstoffgerüst mit gleichartiger Substitution durch Heteroatome aufweisen. Letztlich wurden also alle Verbindungen auf die Teilstruktur **15** zurückgeführt. Dabei ist zu beachten, daß vom Zielmolekül **12** nur der Teil mit der größten Zahl an Kohlenstoffatomen für die

Ähnlichkeitssuche zugrundegelegt wurde. In der derzeitigen Implementation wird davon ausgegangen, daß dieses Fragment dasje-



nige mit der interessanten Funktionalität ist. Dies ist eine Vereinfachung, die sicherlich nicht unter allen Umständen korrekt ist, jedoch in den meisten Fällen als Faustregel gute Dienste leistet.

Konfrontiert mit den Verbindungen in Abbildung 6 ist es für den Chemiker leicht, aus seinem Wissensschatz Zwischenstufen und Reaktionen zu nennen, mit denen die Synthese von **12** durchgeführt werden sollte. Es dürfte klar sein, daß die Synthese über ϵ -Caprolactam **10** zu laufen hat. Sollten bezüglich dessen Umwandlung in **12** (Reaktion mit Acrylnitril und anschließende Hydrierung unter Ringschluß) noch Zweifel bestehen, so müssen weitere Ähnlichkeitssuchen mit dem kleineren Strukturelement, Tetrahydropyrimidin, durchgeführt werden. Möglichkeiten, ϵ -Caprolactam **10** zu synthetisieren, gehen klar aus den in Abbildung 6 enthaltenen Strukturen hervor: durch Ringschlußreaktionen aus **7**, desgleichen aus **6** nach Abspaltung der Schutzgruppe sowie aus den Verbindungen **9** und **11** nach Umsetzung mit Ammoniak. Die übrigen Verbindungen (**8**, **13** und **14**) wiederum lassen sich aus ϵ -Caprolactam **10** herstellen.

Dieses Beispiel zeigt, wie die Ergebnisse einer Ähnlichkeitssuche den Chemiker anregen können, für die konkrete Verwirklichung einer Synthese sein Wissen einzubringen: Chemiker und Programm bilden ein Team zur effizienten Lösung eines Syntheseproblems. Mit dem nächsten Beispiel (Abb. 7) wird ein Ähnlichkeitskriterium vorgestellt, das auf einem allgemeinen Reaktionstyp basiert: der Eliminierung. Alle in Abbildung 7 enthaltenen Verbindungen lassen sich über Eliminierungs- und/oder Additionsreaktionen ineinander umwandeln. Man beachte, daß dabei die Saytsev-Regel berücksichtigt wird. Solche an Reaktivitätsmustern orientierten Anfragen sind kaum über normale Substruktursuchmechanismen zu spezifizieren. Wiederum wird die Aufgabe, konkrete Reagentien und Reaktionsbedingungen für die tatsächliche Durchführung zu wählen (womit läßt sich Markovnikov- oder Anti-Markovnikov-Addition erzielen?), in der ersten Phase an den Chemiker delegiert. Die Absicherung durch eine Suche in einer Reaktionsdatenbank wird er nur in nichttrivialen Fällen vornehmen, und hierüber entscheidet er selbst.

Ein weiterer Ähnlichkeitsbegriff faßt Verbindungen zusammen, die durch Substitutionsreaktionen ineinander umgewandelt werden können. Es ist leicht einzusehen, daß dieser Ähnlichkeitsbegriff enger gefaßt ist, denn mit ihm würden von den Verbindungen in Abbildung 7 **16**, **17** und **19** in eine Gruppe und **18** und **20** in eine andere Gruppe als zueinander ähnlich eingeordnet. Der direkte Zusammenhang zwischen Verbindung **21** und den Verbindungen der beiden anderen Gruppen sowie die

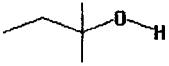
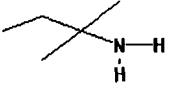
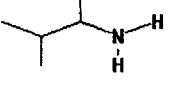
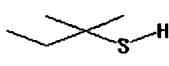
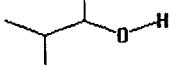
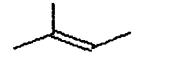
| | | |
|---|--|---|
| 714 JANSSEN 1666275  16 2-METHYL-2-BUTANOL 99% | 717 JANSSEN 1756104  17 2-METHYL-2-BUTYLAMIN 98 | 3607 JANSSEN 1836936  18 1836936 598-74-3 |
| 5653 JANSSEN 1586352  19 2-METHYL-2-BUTYL MERCAPT | 5657 JANSSEN 1499355  20 DL-3-METHYL-2-BUTANOL 9 | 5658 JANSSEN 1264939  21 1264939 513-35-9 |

Abb. 7. Verbindungen, die als ähnlich eingestuft werden, da sie durch Eliminierungsreaktionen in das gleiche Produkt umgewandelt werden.

über **21** führenden Möglichkeiten für Umwandlungen zwischen den Gruppen würden in jedem Fall verloren gehen. Einige der Ähnlichkeitsdefinitionen berücksichtigen auch die Konfiguration an Chiralitätszentren und stereogenen Bindungen.

Die Ähnlichkeitskriterien, die mit Strukturtransformationen arbeiten, lassen Ähnlichkeit feststellen – und spezifizieren, aufgrund welcher Strukturumwandlungen oder Reaktionen von dieser Ähnlichkeit Gebrauch gemacht werden kann –, sie geben aber kein quantitatives Maß für den Grad der Ähnlichkeit an. Um Ähnlichkeit quantitativ zu definieren, wurde ein anderer Ansatz gewählt: die Bestimmung von Pfadlängencodes. Ausgangspunkt war ein Vorschlag von Randic und Wilkins^[52], wonach in einem Molekül von jedem Atom ausgehend die Zahl der Pfade der Länge eins (eine Bindung oder ein Atompaar), zwei (eine Sequenz von drei aufeinanderfolgenden Atomen) usw. gezählt wird. Wir haben diesen einfachen Ansatz durch die Berücksichtigung von Heteroatomen, Ringen, Arenen, Mehrfachbindungen usw. erheblich erweitert. Abbildung 8 zeigt als Beispiel die Anfragestruktur **22** und zwei dazu aufgrund von Pfadlängencodes als ähnlich gefundenen Verbindungen, **23** und **24**, zusammen mit dem jeweiligen Ähnlichkeitsgrad. Die bestbewertete Verbindung **23** muß nur noch reduziert werden, um die Zielverbindung zu liefern, während Verbindung **24** erheblicher struktureller Umwandlungen bedarf, um sie in **22** überzuführen.

Mit diesem Beispiel wird vielleicht nochmals der große Vorteil der Ähnlichkeitsfeststellung auf der Basis von Transformationen deutlich: Sie geben meist direkte Hinweise darauf, mit welchen Reaktionen die Umwandlungen durchgeführt werden können, während die Pfadlängencodes nur aufgrund von zusammengezählten Teilstrukturen Ähnlichkeit erkennen lassen, ohne Aussagen über die anzuwendenden Reaktionen zu machen.

All diese Methoden dienen allerdings nur zur schnellen Auswahl der Kandidatenstrukturen und liefern deshalb nicht unbedingt das Endergebnis. WODCA enthält Mechanismen, um den Grad der Brauchbarkeit einer Verbindung als Ausgangsmaterial präzise und nachvollziehbar zu bestimmen. Zu

den leistungsfähigsten Verfahren zählt die Bewertung von Überlagerungen eines Ausgangsmaterials mit der Zielverbindung oder einer Zwischenstufe. Hierbei werden Faktoren wie stereochemische Übereinstimmung, Reaktivität der zu modifizierenden Zentren und die Möglichkeit zum selektiven Angriff abgeschätzt. Zugänglichkeit und Reaktivität fließen auch in die Bewertung von Stereozentren ein, die zu invertieren sind, so daß dies keineswegs nur eine einfache Projektion und Zählung von Modifikationen ist. WODCA verwendet zur Reaktivitätsabschätzung die gleichen physikochemischen Parameter, die bereits bei der Reaktionsmodellierung in den EROS-Reaktionsvorhersageprogrammen^[20–21, 25] ihre Relevanz bewiesen haben. Ein Beispiel für die Urteilskraft dieser Bewertungsfunktionen wird in Abschnitt 4.3 gegeben.

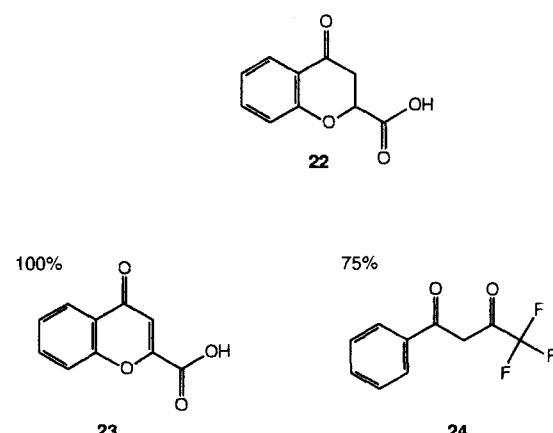


Abb. 8. Eine Zielverbindung und zwei dazu aufgrund von Pfadlängencodes als ähnlich bewertete Verbindungen; das Ähnlichkeitsmaß ist auf die erste Vorstufe normiert.

3.1.2. Die Suche nach Synthesevorstufen

Nur in einfachen oder günstig gelagerten Fällen wird direkt ein Bezug zwischen dem Syntheseziel und einem Ausgangsmaterial herstellbar sein. Daher enthält WODCA Mechanismen, um die Strukturen und damit das Problem zu zergliedern. Dazu gehören Funktionen, die strategische Bindungen entdecken und markieren. Dabei werden sowohl topologische als auch reaktivitätsorientierte Kriterien, die durch Analyse der physikochemischen Eigenschaftsmuster von Atomen und Bindungen gewonnen werden, eingesetzt. Dies stellt noch keine endgültige retrosynthetische Analyse dar, sondern legt erst einmal Schnittstellen fest. Die Atome und Gruppen, die dann tatsächlich verwendet werden, um aus den Fragmenten (Synthonen) konkrete Vorstufen oder Reagenzien zu erhalten, sind noch nicht oder nur tendenziell (elektronenziehende Gruppe) festgelegt.

Bindungen zwischen Kohlenstoff- und Heteroatomen sind wichtige Sollbruchstellen für die Zerlegung einer Verbindung in kleinere Fragmente. Die Art der Heteroatome und deren Elektronegativität, die Möglichkeit zur Stabilisierung von Ladungen, die nach heterolytischem Bruch auftreten, die Ringgröße usw. werden bei der Bewertung der einzelnen Bindungsbrüche berücksichtigt. In **25**, dem Ziel einer Arzneimittelsynthese, sind an derartig als strategisch erkannten Bindungen die ermittelten Bewertungen angegeben. Die in das automatisierte Verfahren eingebaute chemische Intelligenz erkennt in der tricyclischen Verbindung **25** korrekt erweise C-X-Bindungen als bevorzugt – in retrosynthetischer Richtung – zu brechende Bindungen. Im Imidazolthionring erhalten die Bindungen von den Stickstoffatomen zur Thiongruppe hervorragende Bewertungen, während die Bindungen von den Stickstoffatomen zum aromatischen Ring als ungünstige Bruchstellen erkannt werden. Die Verfahren berücksichtigen bei der Bewertung, daß letztere Bindungen durch nucleophile aromatische Substitutionen zu knüpfen wären, eine Reaktion, die nur unter drastischen Bedingungen abläuft und daher für Synthesen nicht sehr nützlich ist. Bemerkenswert ist auch der Unterschied in der Bewertung der beiden Bindungen zum dritten Stickstoffatom mit den skalierten Werten von 10 und 60. Der niedrige Wert von 10 führt daher, daß diese Bindung zu einem Chiralitätszentrum führt und daher ihre Knüpfung nach einer stereospezifischen Reaktion verlangen würde. Viel besser ist es, diese Bindung bereits im Ausgangsmaterial mitzubringen und dafür die Bindung zwischen dem Stickstoffatom und dem benzylichen C-Atom zu knüpfen. Diese Bindung erhält dementsprechend auch den viel höheren Beurteilungswert von 60.

Natürlich gibt es auch Verfahren, die strategische C-C-Bindungen erkennen. Da polare Reaktionen beim Aufbau des Kohlenstoffgerüsts organischer Verbindungen eine herausragende Rolle spielen, muß bei der Bestimmung strategischer Bindungen in einem Syntheseziel besonders darauf geachtet werden, wie beim Bruch einzelner C-C-Bindungen zu den Synthonen die Formalladungen verteilt werden sollen, um sie möglichst gut zu stabilisieren. Dabei spielt vor allem die Resonanzstabilisierung von Ladungen eine große Rolle. Für jede C-C-Bindung werden dazu beide heterolytischen Spaltungsmöglichkeiten untersucht. Je nach eingesetzter Detektionsfunktion können auch radikalische Prozesse in Betracht gezogen werden. Neben Überlegungen zur lokalen Stabilität fließen auch allgemeine topologische Faktoren wie Zentralität und vorhandene Ringe in die Endbewertung ein.

Abbildung 9 zeigt die Größe der Resonanzstabilisierung der bei der Heterolyse der C-C-Bindungen (außer der im aromatischen Ring) von 2-Methyl-2-phenylcyclopentan-1,3-dion **26** auftretenden Formalladungen. Die günstigsten Heterolyse führen zu Cyclopentan-2,5-dion-1-id, wobei die Abspaltung des Methyl-Kations günstiger sein sollte als die des Phenyl-Kations. Weiterhin wird noch eine relativ hohe Stabilisierung bei der Spaltung der C1-C2-Ringbildung gesehen, die zu einem Acyl-Kation und einem Enolat-Anion führt. Für die übrigen C-C-Ringbindungen wird nur eine relativ mäßige Stabilisierungs-

möglichkeit gesehen. Diese Abschätzungen zur Resonanzstabilisierung potentieller Ladungen bei den prinzipiell möglichen Heterolyse ergeben schließlich die beiden in Abbildung 9 b gezeigten strategischen Bindungen und die dazu aufgeführten Bewertungen. Die Abspaltung des Phenyl-Kations wird vom System noch zusätzlich „bestraft“, so daß die Bindung zum Benzolring nicht mehr als strategisch eingeordnet wird. Das Zielmolekül kann nun automatisch an den strategischen Bindungen zerlegt werden. Die so erhaltenen Fragmente

werden in Vorstufen umgewandelt. Mit ihnen können dann wieder Ähnlichkeitssuchen durchgeführt werden. Der Benutzer kann aber auch einzelne strategische Bindungen löschen oder neue hinzufügen und so die Analyse in die von ihm gewünschte Richtung leiten. Die Beziehung zwischen den Zerlegungen, die sich über mehrere Ebenen erstrecken können, und der Zielverbindung wird im Plan festgehalten. Alle Operationen, die für die Zielverbindung möglich sind, können auch auf diese Fragmente angewendet werden. Somit kann ein Pfad, der noch nicht vollständigen Reaktionen, sondern nur unscharf definierten Schritten entspricht, zu den Ausgangsmaterialien hin erarbeitet werden (siehe auch wieder Abb. 4). Die Lücken zwischen den Stationen werden dann z.B. durch Reaktionsvorhersage (Abschnitt 3.1.3) oder in einer künftigen Version durch Suche in einer Reaktionsdatenbank gefüllt, bis eine komplette Reaktionssequenz vorliegt.

3.1.3. Die Vorhersage von Reaktionen

Auch wenn WODCA noch keine Verbindung zu einer Reaktionsdatenbank enthält, die eine detaillierte, schrittweise retrosynthetische Planung mit konkreten Einzelreaktionen zulassen würde, ist es doch zumindest möglich, das EROS-System für Reaktionsvorhersagen und damit zur Verifizierung einer vorgeschlagenen Route zu nutzen. Mit der neuesten Version, EROS 6^[25], können automatisch aus einer Datenbasis an Beispielreaktionen Reaktionsregeln abgeleitet werden, die über den Trainingssatz hinaus Vorhersagekraft besitzen. Bei wirklich aussagekräftigen Trainingsdaten und gegenüber peripheren Einflüssen wie Temperatur, Ionenstärke und Lösungsmittel nicht zu erratisch sensiblen Reaktionen kann die Vorhersagbarkeit bis zu Reaktionsgeschwindigkeiten und auf wenige Prozent genauen Produktverhältnissen gehen^[25]. Reaktionssimulation ist immer dann das Mittel der Wahl, wenn keine wirklich sehr ähn-

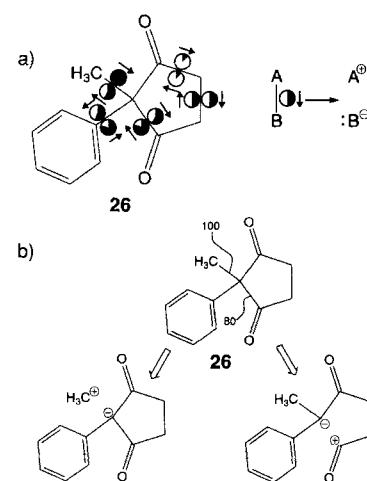
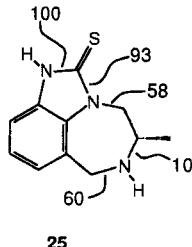


Abb. 9. a) Heterolytic Brüche von C-C-Bindungen in 2-Methyl-2-phenylcyclopentan-1,3-dion **26** und Ausmaß der Resonanzstabilisierung der dabei auftretenden Formalladungen. Die Pfeilrichtung gibt an, auf welches Atom das Elektronenpaar der Bindung verschoben wird, die Größe des gerasterten Kreissegments das Ausmaß der Resonanzstabilisierung. b) Aus der Analyse in (a) abgeleitete strategische Bindungen in **26** und Synthone.

lichen Reaktionen in der Literatur zu finden sind. EROS ist in der Lage, die Einflüsse veränderter Substituentenkonfigurationen zu modellieren und damit Aussagen zu liefern, deren Zuverlässigkeit größer ist als die von Datenbankabfragen. Die Diels-Alder-Reaktion unter Einbeziehung von Heteroatomen ist ein Beispiel für eine sehr gut modellierbare Reaktion. An der Amidhydrolyse wurde gezeigt, daß die Simulation bis zur Vorhersage absoluter Reaktionsgeschwindigkeiten getrieben werden kann^[27]. Bei Reaktionstypen, für die Modelle unter EROS existieren, ist es unter WODCA möglich, eine von WODCA vorgeschlagene oder eine manuell durch graphische Bindungsmarkierung erzeugte Zerlegung sofort auf ihre Realisierbarkeit zu untersuchen. Auch die Ergebnisse der Simulation werden in den Plan eingetragen und stehen für spätere Analysen zur Verfügung.

Ein Beispiel für eine Reaktionsvorhersage durch EROS 6 unter WODCA zeigt Abbildung 10. Hier wurden automatisch strategische Bindungen für eine Diels-Alder-Reaktion erkannt (kleines Fenster, Mitte links). Jedoch ist der Charakter der Reagenzien ungewöhnlich genug, um die Durchführbarkeit der Reaktion ungewiß zu machen. EROS 6 sagt nun aber ein einziges Produkt mit einer skalierten Bewertung von mehr als 95% (mittleres Fenster) und mit der richtigen Orientierung (unten Mitte) voraus^[53]. Die Vorhersage wird Teil des sich entwickelnden Plans (oben links).

3.2. Der Systemaufbau

Der Name Programmsystem drückt aus, daß es sich nicht um ein einzelnes Programm handelt, sondern um ein System von Programmen. Die meisten dieser Programme laufen parallel ab und tauschen in Echtzeit Daten aus. WODCA ist das Kernsystem, das von einer Schale aus peripheren Programmen umgeben ist. Diese strikte Modularisierung in eine Reihe von Methoden hat einige Vorteile:

- kleinere, leichter wartbare Programme
 - geringerer Speicherbedarf bei der Anwendung (der Benutzer wählt nur die Teilsysteme aus, die er für sein Projekt benötigt)
 - einfache parallele Entwicklung von Teilsystemen durch Programmiererteams
 - Wahlmöglichkeit zwischen mehreren Methoden, die eine unterschiedliche Bedienungsphilosophie haben, aber im wesentlichen dieselben Aussagen liefern
 - Verwendung einzelner Werkzeuge, losgelöst vom Gesamtsystem, für spezielle Einsatzgebiete

Neben den peripheren Programmen, die parallel zum Kernsystem laufen, gibt es noch sekundäre Programme, die von WODCA aus gestartet werden. Diese Programme werden entweder von WODCA aus aufgerufen und laufen dann unabhängig weiter, oder WODCA startet diese Programme und wartet, bis sie

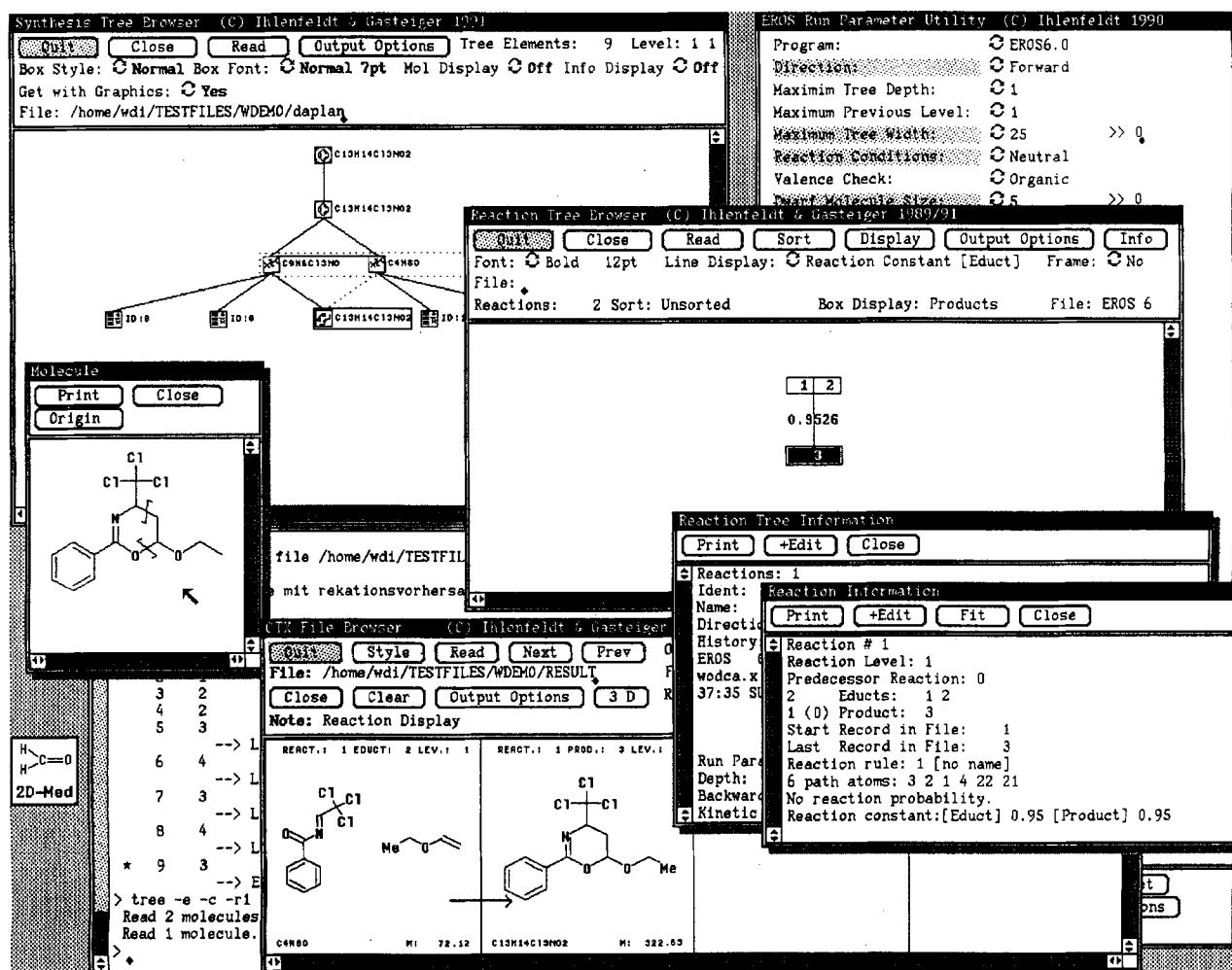


Abb. 10. Vorhersage der Regioselektivität einer Diels-Alder-Reaktion durch EROS zur Verifizierung einer durch WODCA entwickelten Zerlegung in Vorstufen. Zur Erklärung der Information in den einzelnen Fenstern siehe Text.

ihre Ergebnisse geliefert haben. Zwischen diesen Programmen kann nicht wie bei den peripheren beliebig hin und her gewechselt werden.

Alle Funktionen, welche die chemische Intelligenz des Systems repräsentieren (Bewertung von Vorstufen, Suche nach strategischen Bindungen, diverse Filterfunktionen), sind nicht fest im Programm codiert, sondern in externen, erst zur Laufzeit des Systems geladenen Dateien gespeichert. Damit ist gewährleistet, daß für eine bestimmte Aufgabe nicht nur eine generelle Funktion existiert, sondern jederzeit für Spezialfälle andere Funktionen herangezogen werden können. Der Benutzer hat also nicht nur die Kontrolle darüber, welche Methoden er anwendet, sondern auch darüber, welche Variante einer Methode zum Einsatz kommt und wie die Ergebnisse ausgewertet werden.

Abbildung 11 zeigt den Datenfluß innerhalb der interaktiven Teilprogramme. Unidirektionale und bidirektionale Datenströme sind durch die Pfeilspitzen an den Kanälen unterschieden.

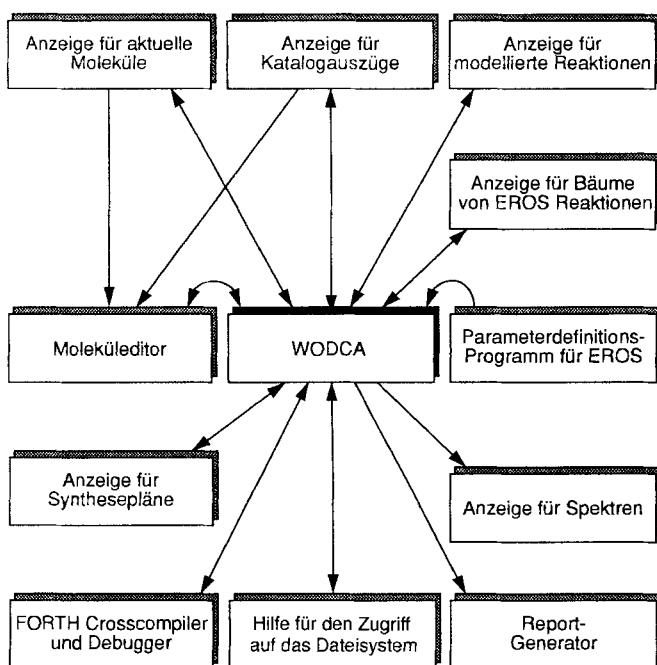


Abb. 11. Datenströme im WODCA-System.

Um keinen unnötigen Speicherplatz zu reservieren, kann der Benutzer das WODCA-System für seine Studie konfigurieren. Zunächst wählt er diejenigen Werkzeuge aus, die er zusätzlich zum Kernsystem benötigt. Sie werden synchron mit diesem gestartet und beendet. Zur Konfigurierung werden einfach die entsprechenden Symbole, die in Abbildung 12 zusammengestellt sind, markiert. In der derzeitigen Version muß der Satz an benötigten peripheren Programmen beim Start einer Studie ausgewählt werden.

Das System läuft in der beschriebenen Konfiguration zwar stabil, ist aber dennoch ein junges System, das direkt aus der Entwicklung kommt. Um es noch benutzerfreundlicher zu gestalten, muß noch einiges an Arbeit investiert werden.

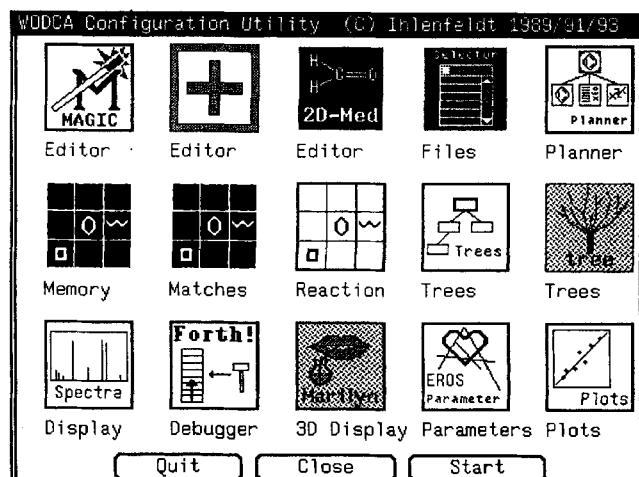


Abb. 12. Werkzeuge innerhalb des WODCA-Systems. Ausgewählte Programm-funktionen werden dunkel unterlegt. Es sind in diesem Beispiel für das Suchen in Katalogen ein graphischer Moleküleditor für die Struktureingabe, ein Visualisierungsprogramm für die graphische Molekülausgabe, ein Manager für das Dateisystem und ein Display für Listen von Suchergebnissen in Strukturdateien (Match-Listen).

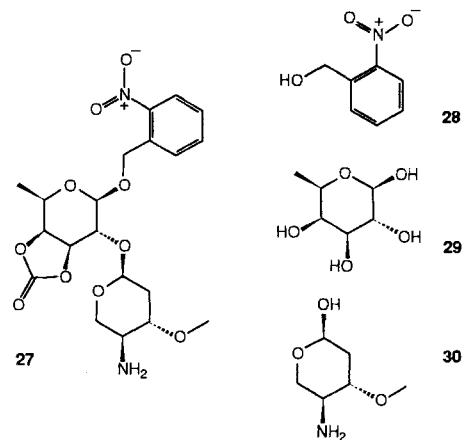
4. Beispiele für die Arbeit mit WODCA

Die folgenden drei Beispiele sollen einen allgemeinen und möglichst untechnischen Überblick über die Arbeit mit WODCA geben. Die Beispiele sind so ausgewählt, daß ein breites Spektrum von Kernelementen des Systems zum Einsatz kommt und dabei Wiederholungen vermieden werden. Aus Platz- und Übersichtlichkeitsgründen sind die entwickelten Synthesestudien und -pläne untypisch klein oder unvollständig gehalten. Die Beispiele sind somit als straff geführte Tour durch WODCA und weniger als komplettete Studien zu verstehen.

4.1. Calicheamicin

Das erste Beispiel soll das Entstehen eines Planes und einfache Datenbanksuchen nach Ausgangsmaterialien verdeutlichen. Es ist insofern etwas eingeschränkt und konstruiert, als die studierte Verbindung nur eine Zwischenstufe in einer größeren Synthese ist und ein Teil der Struktur nur eine Schutzgruppenfunktion hat.

Die Zielverbindung **27** ist eine Zwischenstufe bei der Synthese von Calicheamicin^[54].



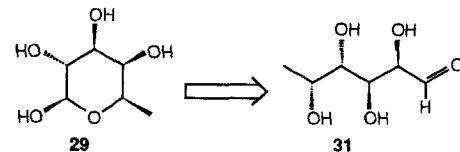
Die Struktur **27** lässt sich offensichtlich in die drei größeren Einheiten **28**, **29** und **30** zerlegen. Diese Zerlegung wird auch von WODCA aufgrund selbsttätig gefundener strategischer Bindungen vollautomatisch durchgeführt. Die kleine Kohlensäure-ester-Schutzgruppe wird dabei als strukturell anspruchslos erkannt und für die Grobzerlegung außer acht gelassen. WODCA stellt auch noch keine Vermutungen darüber an, von welcher Vorstufe die Sauerstoffatome an den Verknüpfungsstellen letztendlich stammen, sondern dupliziert sie zunächst einmal auf jedes Fragment, wo sie Platzhalterfunktionen für beliebige Funktionalitäten übernehmen. Diese Art der Zerlegung ist klar zu unterscheiden von einer Transform-orientierten Retrosynthese. Hier wird eben noch keine Aussage über die Reaktion, welche die Verknüpfung bewerkstelligen soll, getroffen, sondern strukturelle Teilelemente werden isoliert, ihr Bezug zur Gesamtstruktur festgelegt und sie dann getrennt weiter bearbeitet. Die Lücke zwischen den einzelnen verallgemeinerten Vorstufen durch spezifische Reaktionen aufzufüllen, ist eine davon vollständig getrennte Aufgabe.

Das Synthesenzieiel, der Satz an (hier automatisch gefundenen) Bindungen, an denen das Zielmolekül zerlegt wird, und der daraus erzeugte Satz an Vorstufen bilden die ersten Schritte im Plan, der in Abbildung 13 zu sehen ist. Dieser Plan verwendet nur einen einzigen Satz an strategischen Bindungen – hier Kohlenstoff-Heteroatom-Bindungen –, die auf eine einzige Art in konkrete Fragmente umgewandelt werden. Bei komplizierteren Studien können mehrere alternative Sätze an strategischen Bindungen verwendet werden, die z.B. durch variierende Kriterien zur Bestimmung strategischer Bindungen erhalten wurden, und die Bindungen eines jeden dieser Sätze können wiederum mit unterschiedlichen Methoden geschnitten werden. Parameter, die bei der Zerlegung eingestellt werden können, sind unter anderem die Zahl der Stufen und die Sequenz bei der Zerlegung sowie die Absättigung offener Valenzen durch Heteroatome oder die Eliminierung von Atomgruppen oder Ionen, um neutrale Vorstufen zu erhalten. Weiterhin kann nach dem Bruch der

strategischen Bindungen oder eines Teils davon eine genauere Bewertung sowohl der Fragmente als auch der noch nicht behandelten strategischen Bindungen nachgeschoben werden, da nun konkrete Vorstufen oder eine veränderte Umgebung der verbleibenden strategischen Bindungen vorliegen.

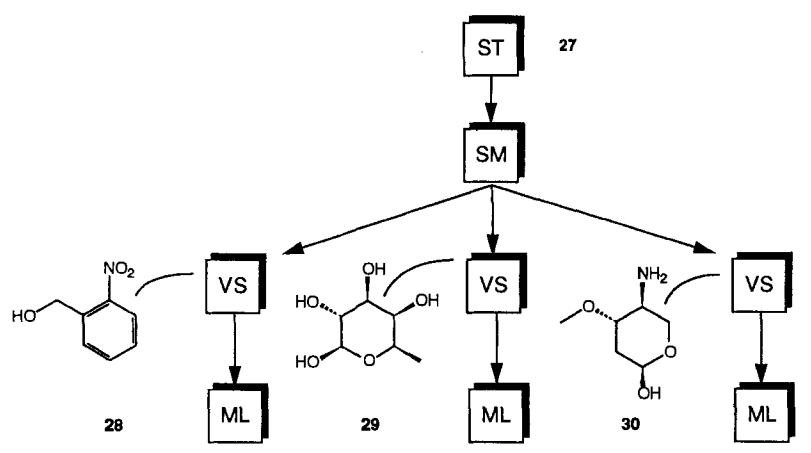
Im nächsten Schritt kann versucht werden, die Vorstufen als Vollstruktur in einem Katalog von Ausgangsmaterialien zu finden. Der größte Katalog, der WODCA zur Zeit zur Verfügung steht, ist der Katalog von Janssen Chimica mit fast 8000 verschiedenen Verbindungen. Die Suche lässt sich automatisch parallel für alle Vorstufen durchführen. Das Suchergebnis wird in logischer Verknüpfung mit der Anfragestruktur im Plan protokolliert, auch wenn nur eine leere Liste erhalten wurde, also kein Ausgangsmaterial gefunden wurde, so daß in jedem Fall alle Schritte nachvollziehbar sind. Nur *o*-Nitrobenzylalkohol **28** wird mit dieser klassischen Suchmethode direkt im Katalog gefunden. Auch im CHIRON-Katalog und in einem Auszug aus dem Aldrich-Katalog wird keine der restlichen beiden Verbindungen, **29** und **30**, direkt gefunden.

Lässt man die Zügel etwas lockerer und verwendet eine Ähnlichkeitssuche mit einem Transformations-Hashcoding, bei der das zentrale Kohlenstoffgerüst und die α -Heteroatome sowie natürlich die Stereochemie erhalten bleiben, aber keine Unterscheidung mehr zwischen Substituenten jenseits der ersten Sphäre sowie zwischen einer offenen und einer Ringform getroffen wird, so findet sich für Verbindung **29** als passendes Molekül D-Fucose **31** im Katalog von Aldrich.



Aldrich 85, 028-4

Es bleibt nun nur noch die dritte Vorstufe, der 2'-Desoxyaminozucker **30**, ohne direkten Bezug zu einem käuflichen Produkt. Diese Verbindung gibt es weder in offener noch in Ringform in irgendeinem der an WODCA angeschlossenen Kataloge. Hier muß also nach einer weiter gefaßten Ähnlichkeit gesucht werden. Verwendet man ein Ähnlichkeitsmaß, das auf das Substituentenmuster der Kohlenstoff-Hauptkette achtet, dabei aber die Art der Substituenten und die Stereochemie außer Acht lässt, so erhält man aus dem CHIRON-Katalog eine Liste von acht Verbindungen (Abb. 14). Die Suche wurde dabei über Pfadlängencodes durchgeführt, wobei aus den 2211 Verbindungen des CHIRON-Katalogs innerhalb weniger Sekunden diejenigen herausgeholt wurden, die ein Ähnlichkeitsmaß von über 10 % hatten. Diese Prozentzahlen spiegeln allerdings noch keine detaillierte Bewertung wider. WODCA ermöglicht es aber auch, eine solche vorgefilterte Menge an Kandidaten mit präziseren, zeitaufwendigeren Bewertungsfunktionen näher unter die Lupe zu nehmen,



Janssen-Katalog-Nr. 1285147

leere Liste

leere Liste

Abb. 13. Erste Stufe des Plans zur Synthese der Calicheamicin-Vorstufe **27**. Für das Synthesenzieiel (ST) werden zunächst mit einer allgemeinen Methode die strategischen Bindungen bestimmt, d.h. es wird in das Strategische Molekül (SM), das die im nächsten Schritt zu spaltenden Bindungen enthält, überführt. Die Zerlegung von **27** an diesen Bruchstellen liefert die drei Vorstufen (VS) **28**, **29** und **30**. Mit jeder dieser Vorstufen wird eine Vollstruktursuche im Janssen-Chimica-Katalog durchgeführt; nur *o*-Nitrobenzylalkohol **28** wird direkt als verfügbares Ausgangsmaterial gefunden, die anderen beiden Match-Listen (ML) sind leer, enthalten also keinen Struktureintrag.

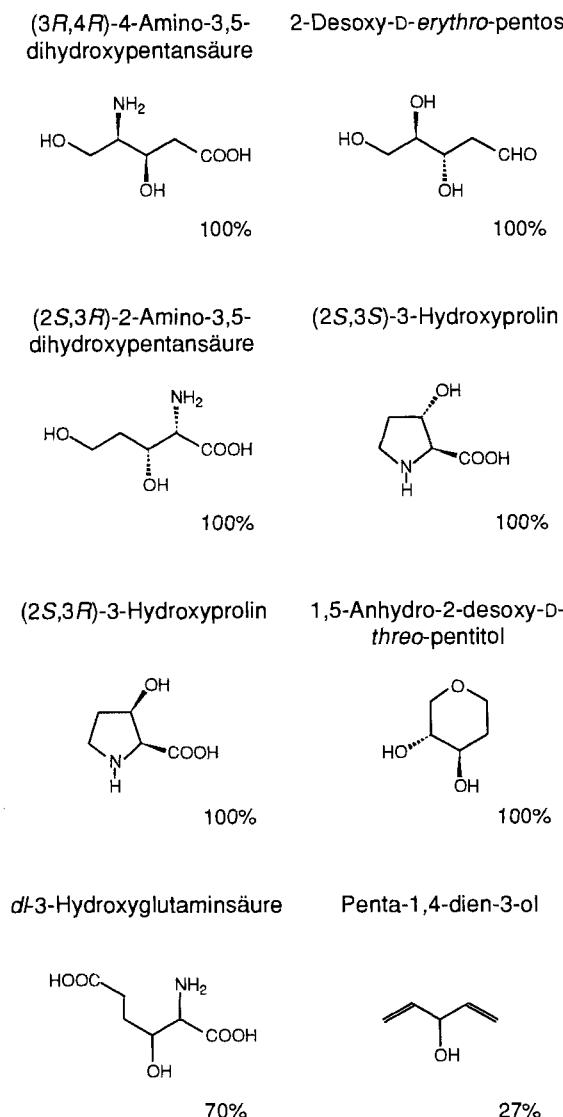


Abb. 14. Vorauswahl von acht potentiellen Ausgangsmaterialien für die Calicheamicin-Vorstufe **30** aufgrund einer Pfadlängencode-Suche. Die Prozentzahlen geben die mit dem inhärenten Maß dieses Suchverfahrens ermittelte Ähnlichkeit wieder; sie lassen Stereochemie und Praktikabilitätserwägungen außer Betracht.

wie in späteren Beispielen noch erläutert werden wird. Trotzdem ist diese grobe Vorauswahl schon sehr selektiv. Während Penta-1,4-dien-3-ol wohl eher uninteressant ist, sind einige der anderen Verbindungen einen zweiten Blick wert. Auf eine tiefgehende Analyse dieser Vorstufen und die Entwicklung von Reaktionssequenzen hin zur gewünschten Verbindung **30** wird hier verzichtet, denn solange keine experimentelle Verifizierung eines Syntheseplans vorgelegt werden kann, muß er Spekulation bleiben.

An diesem Punkt sind endlich alle Fragmente von **27** mit Ausgangsmaterialien assoziiert. Der nächste Schritt wäre nun die Suche nach konkreten Reaktionen zum Füllen der Lücken. Dieser Teil von WODCA befindet sich zur Zeit noch in Entwicklung.

Die publizierte Synthese verwendet tatsächlich *o*-Nitrobenzylalkohol **28** und D-Fucose **31**. Die dritte Vorstufe wurde als geschütztes Glycosylfluorid in die Reaktionssequenz eingeschleust. Dieses wurde aus dem Hydrochlorid des L-Serinmethylesters in zwölf Stufen (mit einer Gesamtausbeute von 14% bis

zur neunten Stufe) gewonnen. Es ist nicht ausgeschlossen, daß einige der hier vorgeschlagenen Verbindungen in weniger Stufen und besserer Ausbeute zu der eingesetzten Vorstufe führen können. Leider geben die Veröffentlichungen^[54] keine Auskunft darüber, wie die Ausgangsmaterialien und Reagentien ausgewählt wurden. Dies ist ein generelles Problem bei komplexen Synthesen: Während häufig die Reaktionen mit zahlreichen Literaturzitaten belegt werden und damit wahrscheinlich den Autoren bei der Planung ihrer Synthese bekannt waren, werden die Ausgangsmaterialien als selbstverständlich angesehen und höchstens mit einer Katalognummer des Herstellers dokumentiert. Dabei spielen gerade die Methoden, die zur Wahl der Ausgangsmaterialien führen, eine zentrale Rolle in der Syntheseplanung und wären es wert, genauer dokumentiert zu werden.

4.2. Sexualhormon des grünen Stinkkäfers

Die komplex aufgebauten Sexualhormone des grünen Stinkkäfers (Bisabolene) wurden erstmals 1989 durch Nicolaou et al. synthetisiert^[55]. Die einzelnen Pheromone unterscheiden sich in der Konfiguration der Anknüpfungspunkte der Alkylgruppen. Hier wird nur ein Beispiel behandelt: (+)- α -Bisabolene **32**.

Die automatische Analyse durch WODCA liefert nur zwei strategische C-C-Bindungen, und zwar die Doppelbindungen. Die mittlere wird nach allgemeinen taktischen Überlegungen (zentraler im Molekül gelegen und damit eher zu einem konvergenten Syntheseplan führend) höher bewertet. Läßt man WODCA unbeeinflußt fortfahren, so werden in der retrosynthetischen Studie zwei mögliche Zerlegungen in den Plan eingetragen (Abb. 15). Der rechte, von vornherein schlechter bewertete Zweig zur Vorstufe **35** und einer Isopropylverbindung ergibt sich als wenig vielversprechend. Das kleine Isopropyl-Fragment ist natürlich in Hunderten von Varianten erhältlich, aber zu dem größeren Fragment **35** läßt sich mit den in WODCA enthaltenen Substruktur- und Ähnlichkeitssuchen keine interessante direkte strukturelle Entsprechung in den vorhandenen Chemikalienkatalogen finden.

Bei der Analyse des linken Zweiges zu den Vorstufen **33** und **34** wird zunächst auf den CHIRON-Katalog umgeschaltet, um

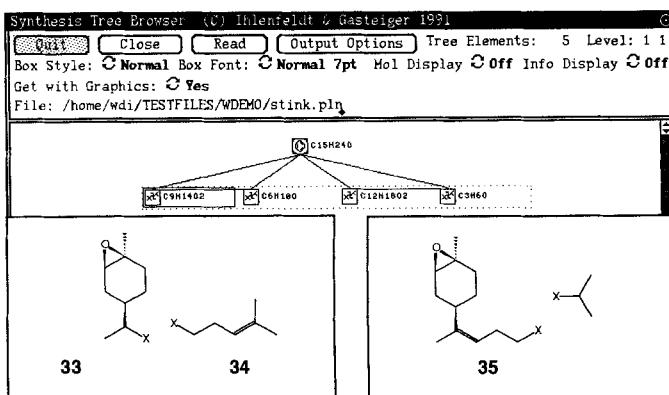
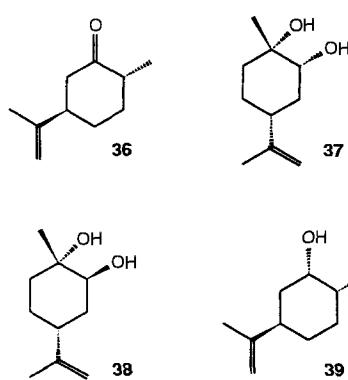


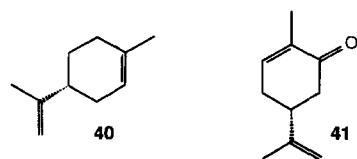
Abb. 15. Erste Stufen des Plans zur Synthese von (+)- α -Bisabololen **32**.

eine größere Auswahl an terpenoiden Strukturen mit definierter Konfiguration zur Verfügung zu haben. Schon durch einfache Substruktursuche (Einbettung des Fragments **33** in die Katalogstrukturen) unter Vernachlässigung der Stereochemie, aber mit der fokussierenden Nebenbedingung, daß die Gerüste der Katalogstrukturen um nicht mehr als ein Viertel größer als das der Anfragestruktur sein dürfen und daß mindestens ein Stereozentrum mit definierter Konfiguration irgendwo im Molekül vorhanden sein muß, werden 24 potentielle Vorstufen ermittelt. Hier wird nun die bereits im vorherigen Kapitel angesprochene Bewertungsfunktion eingesetzt. Dabei werden diese 24 Kandidaten mit dem Suchfragment **33** überlagert und die Komplexität der notwendigen Transformationen sowie die stereochemische Komplementarität abgeschätzt. Die nach der WODCA-Analyse



am besten geeigneten Kandidaten, **36**–**39**, sind nicht kommerziell erhältlich, sondern wurden für die Datenbank von CHIRON aus der Literatur entnommen. Die Datenbank enthält die Literaturstellen, die jederzeit für alle Kandidaten durch einen Mausklick abgerufen und auch ausgedruckt werden können. Das gemeinsame Merkmal der bestbewerteten Strukturen ist, daß das die Methylgruppe tragende Ringkohlenstoffatom bereits eine definierte Konfiguration hat.

Im Gegensatz dazu sind die bestbewerteten käuflichen Verbindungen ((*R*)-Limonen **40** und (*R*)-Carvon **41**) an dieser Position reaktiv (Doppelbindungen), aber noch nicht stereodifferenziert. Die stereochemische Bewertungsfunktion von WODCA versucht auch abzuschätzen, inwieweit andere Gruppen in der näheren Umgebung bei der Reaktion des zu manipulierenden Zentrums den stereochemischen Verlauf beeinflussen könnten. Bei **40** und **41** werden die Aussichten von WODCA allerdings pessimistisch eingeschätzt, und dies führt zu einer deutlichen Abwertung gegenüber den Verbindungen **36**–**39**.



Bemerkenswert ist, daß in allen von WODCA ausgewählten Verbindungen das Kohlenstoffgerüst größer ist als im daraus zu synthetisierenden Folgeprodukt **33**. Damit ist die normale Beziehung zwischen der Größe (und zum Teil auch der Komplexität) eines Syntheseziels und der eines Ausgangsmaterials umgekehrt: Das Ausgangsmaterial enthält mehr Kohlenstoffatome (und ist teilweise auch von komplexerer Struktur) als das daraus zu synthetisierende Produkt. Für klassische Methoden, die nach einer strukturellen Einbettung des Ausgangsmaterials in die Zielverbindung suchen, also die Ausgangsverbindung als Teilstruktur des Zielmoleküls erwarten, ist dies eine schwer zu überwindende Komplikation. Die hier eingesetzte Suchstrategie kommt damit aber problemlos zurecht.

In der Originalarbeit^[55] wurde racemisches, epoxidiertes Limonen (diese Verbindung wurde gekauft, ist aber in den an WODCA angeschlossenen Katalogen nicht aufgeführt) ozoniert und das entstandene Racemat getrennt. Während für Verbindungen aus der Literatur die Zitate abrufbar sind, stehen für kommerziell vertriebene Verbindungen die kompletten Katalogdaten bereit. Dies geht soweit, daß selektierte Kandidaten an einen externen Reportgenerator weitergeleitet werden können, der ein Bestellformular oder Tabellen ausdrucken kann.

Nachdem für die Synthese von **33** mit Carvon und Limonen offensichtlich brauchbare Verbindungen gefunden sind, gilt es, sich dem zweiten Bruchstück, **34**, zuzuwenden. Es sieht auf den ersten Blick nicht allzu kompliziert aus. Die Überraschung ist, daß es unter den mehr als 10 000 Verbindungen, die WODCA in diversen Katalogen zur Verfügung stehen, keine Verbindung gibt, die dieses Kohlenstoffgerüst mit einem entsprechenden Substitutionsmuster aufweist. Damit muß hier die neu entwickelte Ähnlichkeitssuche, die mit substitutionsorientierten Pfadlängencodes arbeitet, genutzt werden. Charakteristisch für diese Methode, die bereits in Abschnitt 3.1.1 kurz erwähnt wurde, ist, daß Verbindungen gefunden werden, deren Gerüste eine sehr weitgehende, wenn auch nicht absolute Ähnlichkeit zur Anfragestruktur aufweisen, wobei die Substituenten in einem Muster angeordnet sind, das eine etwas größere Variationsbreite aufweisen darf. Die Struktur der Substituenten wird ignoriert, und auch echte Substituenten und C-C-Doppelbindungen sind in Maßen austauschbar. Auch diese Suche findet keine auf den ersten Blick als ideal anzusehenden Verbindungen, aber die genauere Analyse der Treffer zeigt einige recht interessante Zusammenhänge auf, die beim Blättern durch Kataloge mit ziemlicher Wahrscheinlichkeit übersehen worden wären. Neun Strukturen, **42**–**50**, aus dem Konzentrat der dreißig bestbewerteten Verbindungen, in wenigen Sekunden aus mehreren tausend Kandidaten automatisch herausgeholt, zeigt Abbildung 16.

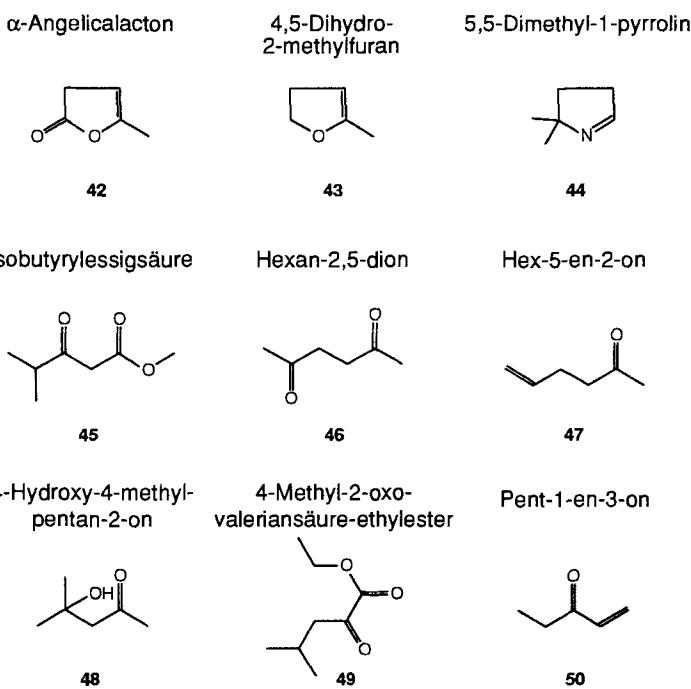
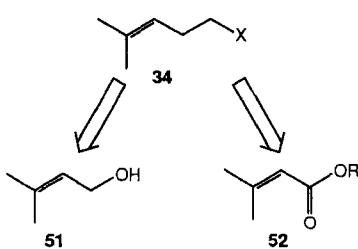


Abb. 16. Potentielle Ausgangsmaterialien für die Synthese der Verbindung **34**, erhalten mit einer sehr breit angelegten Suche.

Keines der dort tabellierten Moleküle wäre sofort einsetzbar, aber beim Betrachten dieser Auswahl kommen dem Chemiker doch neue Ideen: Der Ring des Pyrrolins **44** kann sicherlich geöffnet werden, und Funktionalität zur Einführung der Doppelbindung sowie am Terminus sind vorhanden. Auch das Angelicalacton **42** und das Dihydrofuranderivat **43** haben nach der Ringöffnung die Funktionalität an genau der richtigen Position, aber hier muß noch jeweils eine Methylgruppe eingeführt werden. In diesem Fall wurden die Verbindungen wegen der richtigen Plazierung der Funktionalität hoch bewertet, obwohl das zugrundeliegende Kohlenstoffgerüst unvollständig ist. Daneben finden sich auch Verbindungen wie Hex-5-en-2-on **47**, bei dem neben der Methylgruppe auch die terminale Funktionalität am Kohlenstoffgerüst eingeführt werden müßte – was allerdings nur durch Abbau (Doppelbindungsspaltung) gelänge. Andere Kandidaten wiederum erfordern eine Funktionalitätsverlagerung zum terminalen Atom, haben aber sonst, wie 4-Hydroxy-4-methyl-pentan-2-on **48**, die Grundstruktur einschließlich der inneren Doppelbindung bereits vorgebildet. Manches auf den ersten Blick rätselhaft erscheinende Molekül lohnt einen zweiten Blick und löst beim Chemiker eine Kaskade von Assoziationen aus. Natürlich ist es auch möglich, das Fragment **34** weiter zu zerlegen. WODCA findet dabei zwangsläufig 3-Methyl-but-2-en-1-ol **51** und 3,3-Dimethylacrylsäurederivate **52**.

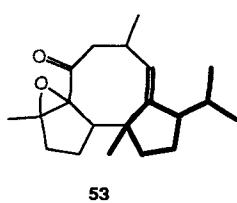


Diese Verbindungen werden auch in der direkten Suche als ähnlich gefunden, kommen jedoch wegen der geringeren Kohlenstoffzahl ziemlich spät in der nach Ähnlichkeit sortierten Liste.

In der Originalsynthese^[55] wurde das Fragment **34** in mühevoller Kleinarbeit aufgebaut. Dabei wurden in sechs Stufen drei kohlenstoffhaltige Verbindungen (zwei C₁-Moleküle und ein C₄-Bruchstück) an das Limonengerüst angehängt. Eine konvergente Synthese ist dies sicherlich nicht. Bei derart kleinen Bruchstücken kann man nicht mehr von Ähnlichkeit und erkennbarem Bezug zur Zielstruktur sprechen, so daß es nicht im Aufgabenbereich von WODCA liegt, solche winzigen Syntheseschritte in einer ersten Analyse zu assoziieren. Nur wenn in der ersten Analyse keine brauchbare Synthese gefunden wird, muß man sich in einem zweiten Anlauf einer mühevollen Detailanalyse zuwenden.

4.3. (+)-7,8-Epoxy-2-basmen-6-on

Als letztes Beispiel soll die Suche nach Ausgangsmaterialien für (+)-7,8-Epoxy-2-basmen-6-on **53** die Fähigkeiten von WODCA zeigen, über extrem große chemische Distanzen wertvolle Anhaltspunkte für die Planung zu liefern.



53, das in der Natur in sonnengedeckten Tabakblättern vorkommt, wurde 1991 von Paquette und Kang erstmals durch Totalsynthese hergestellt^[56]. Für diese Verbindung mit

ausgesprochen komplexer Stereochemie wurde wieder im Katalog von CHIRON gesucht. Bei Verbindungen dieser Komplexität ist es die natürliche Vorgehensweise, so viel Stereochemie wie möglich vom Ausgangsmaterial zu übernehmen, und es scheint sinnvoll, einen der Fünfringe bereits durch das Ausgangsmaterial mitzubringen. Daher wurde als Bedingung für die erste Vorwahl potentieller Kandidaten spezifiziert, daß das Ausgangsmaterial genau einen Fünfring enthalten muß, aber keinen Sechsring (im kleinsten Satz kleinsten Ringe; smallest set of smallest rings: SSSR) aufweisen darf^[57] und daß mindestens ein Stereozentrum mit definierter Konfiguration vorhanden sein muß. Diese Vorgaben lieferten 84 Kandidaten, deren Güte wieder mit der stereochemischen Bewertungsfunktion von WODCA bestimmt wurde. Die nach dem Bewertungsergebnis sortierte Liste wird dem Benutzer sowohl als Tabelle als auch graphisch in Form der Kandidatenstrukturen präsentiert.

Es wurde nicht vorgegeben, welcher der beiden Fünfringe von **53** als Ausgangspunkt dienen sollte. Jeder der 84 Verbindungen wurde in allen denkbaren Orientierungen^[58] in beide Fünfringe der Zielstruktur und auch in die anderen Strukturregionen eingepaßt. Besonders gute Übereinstimmungen wurden detailliert bewertet. Die Ergebnisse sind in Tabelle 2 und Abbildung 17 aufgeführt. Es ist deutlich, daß für den rechten Fünfring bessere Kandidaten gefunden wurden. Nur hier wurden überhaupt po-

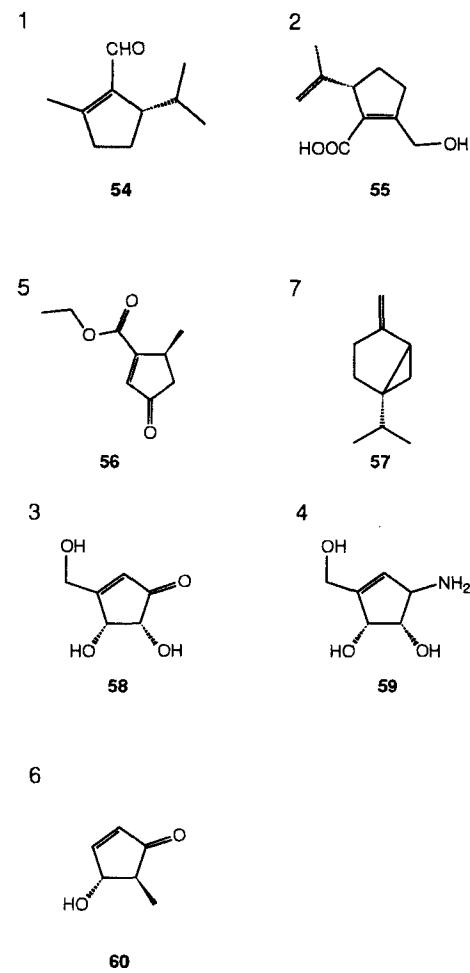
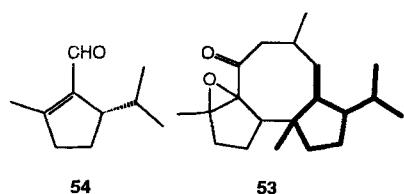


Abb. 17. Potentielle Vorstufen für den rechten (obere zwei Zeilen) und linken Fünfring von (+)-7,8-Epoxy-2-basmen-6-on **53** (untere zwei Zeilen).

Tabelle 2. Bewertung von Basmenon-Vorstufen, rechter Fünfring (1.–4. Eintrag) und linker Fünfring (5.–7. Eintrag).

| Rang | Bewertung | Verbindung |
|------|-----------|---|
| 1 | 979 | (5 <i>S</i>)-5-Isopropyl-2-methylcyclopentencarboxaldehyd 54 |
| 2 | 959 | DL-2-Hydroxymethyl-5-isopropenylcyclopent-1-encarbon-säure 55 |
| 5 | -148 | (5 <i>S</i>)-5-Methylcyclopent-1-en-3-oncarbonsäureethylester 56 |
| 7 | -194 | Sabinen 57 |
| 3 | -56 | (4 <i>R,5R</i>)-4,5-Dihydroxy-3-hydroxymethylcyclopent-2-enon 58 |
| 4 | -59 | (3 <i>R,4S,5R</i>)-3-Amino-4,5-dihydroxy-1-(hydroxymethyl)-cyclopenten 59 |
| 6 | -173 | (4 <i>R,5S</i>)- <i>trans</i> -4-Hydroxy-5-methylcyclopent-2-enon 60 |

sitive Bewertungen für die jeweils beste Einpassung erhalten. Die günstigste Einbettung der bestbewerteten Ausgangsverbindung 54 in die Zielstruktur 53 zeigt Schema 3.



Schema 3. Die bestbewertete Vorstufe **54** und ihre Einbettung in die Zielverbindung **53**.

Diese Einbettung scheint zunächst offensichtlich, ist aber bei weitem nicht die einzige untersuchte Orientierungsmöglichkeit. Die sechs bestbewerteten Orientierungen des Spitzenkandidaten aus einigen Dutzend näher untersuchter Möglichkeiten zeigt Abbildung 18. An diesem Beispiel kann man durch Vergleich der Bewertungsziffern und der Einbettung die Einflüsse bei der Bewertung verfolgen. Stereochemische Komplementarität, gewichtet nach Reaktivität der involvierten Zentren, Reaktivität

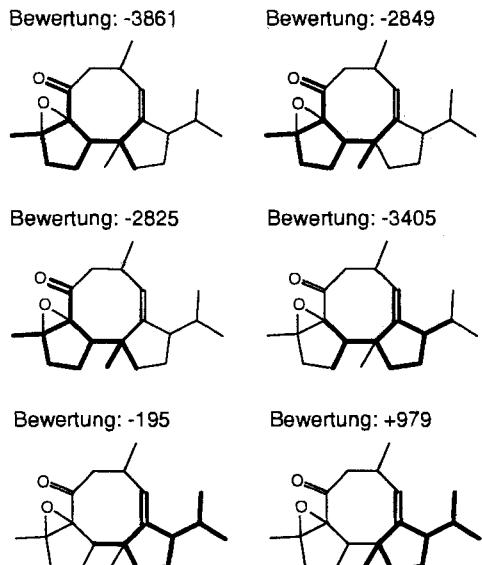


Abb. 18 Wichtige Einbettungen der besten Vorstufe 54 in die Ziельverbindung 53

der Anbindungsstellen beim Aufbau der Zielstruktur (vergleiche die Bewertungen der Einbettung unten rechts und unten links), und allgemeine topologische Faktoren sind die Kernelemente dieses Verfahrens. Es wurden nicht nur Einbettungen untersucht, bei denen das Ausgangsmaterial komplett, d. h. mit dem ganzen Kohlenstoffgerüst (Wasserstoffatome und Heteroatome werden natürlich ebenfalls mitgeführt), eingebettet vorliegt. Dutzende von Alternativen, in denen Teile des ursprünglichen Kohlenstoffgerüsts nicht mehr in der Zielstruktur vorhanden sind, wurden ebenfalls geprüft. Sie wurden allerdings (im Gegensatz zum Bisabolen-Beispiel) nicht mit konkurrenzfähigen Bewertungen versehen. Die stereochemische Komplementarität ist beim zweitbesten Kandidaten, **55**, fast ebenso gut, aber es sind mehr Umwandlungen notwendig. Dies beeinflußt natürlich die Bewertung – im vorliegenden Fall jedoch vergleichsweise wenig, weil die zu verändernden Stellen im Molekül **55** prinzipiell gut durch Reaktionen zugänglich sind. Im Gegensatz dazu steht die Bewertung des Enantiomers der bestbewerteten Verbindung, das auch in der Datenbank von CHIRON enthalten ist. Da es hier notwendig wäre, die Konfiguration an einem schwierigen und ziemlich unreaktiven Zentrum umzukehren, wird die Verbindung so weit abgewertet, daß sie nicht unter den besten 20 der 85 untersuchten Verbindungen erscheint.

Es wurden auch Verbindungen gefunden, bei denen der linke Fünfring von **53** bessere Ansatzmöglichkeiten für den Aufbau durch Synthese bietet. Die drei besten dieser Verbindungen, **58–60**, zeigt Abbildung 17 in den unteren zwei Zeilen, und die jeweils beste Einbettung zweier Kandidaten für diesen Ring zeigt Abbildung 19.

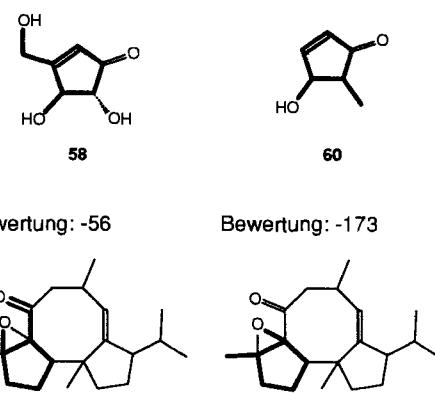


Abb. 19. Bestbewertete Einbettungen der Vorstufen **58** und **60** in den linken Fünfring von (+)-7,8-Epoxy-2-basmen-6-on **53**.

Bei der publizierten Synthese^[56] wurde tatsächlich (5S)-5-Isopropyl-2-methylcyclopentencarboxaldehyd **54** als Ausgangsmaterial herangezogen, was die Analyse von WODCA bestätigt. Die gesamte Synthese erstreckte sich hiervon ausgehend über 28 bis 30 Stufen, je nach Zählweise. Leider wurde wieder nicht erwähnt, welche Überlegungen zur Wahl der Ausgangsmaterialien geführt haben. WODCA war in diesem Fall in der Lage, über eine enorme chemische Distanz noch synthetisch verwertbare und überzeugende Ähnlichkeiten festzustellen. In diesem speziellen Fall konnte das jedoch wegen der Komplexität der Berechnungen nicht mehr interaktiv geschehen. Für die Bewertung der 85 Kandidaten, deren Vorauswahl in wenigen Sekunden geschehen war, mußten etwa 25 Minuten Rechenzeit auf

einer Sparc-II-Workstation aufgewendet werden. In dieser Zeit wurden einige hunderttausend Überlagerungen erzeugt und davon etwa 1000 genauer bewertet. Dieser Zeitbedarf muß mit dem Aufwand und der begrenzten Gründlichkeit des Menschen beim manuellen Durchgehen von Katalogen verglichen werden und ist somit immer noch vernachlässigbar. Zudem sind technisch aktuellere und im Bereich des Molecular Modeling routinemäßig verwendete Workstations mittlerweile mehr als zehnmal so leistungsfähig, so daß ein derartig komplexes Problem in absehbarer Zukunft auch interaktiv bearbeitet werden kann.

5. Zusammenfassung und Ausblick

WODCA ist der Prototyp einer neuen Generation von Syntheseplanungsprogrammen, die sich mehr an der Denkweise des präparativ arbeitenden Chemikers orientieren. WODCA enthält ein breites Spektrum von Methoden, die den Chemiker bei seiner von ihm selbst geleiteten Suche nach Synthesen für organische Zielverbindungen unterstützen. Einige der Werkzeuge, die sich als besonders wichtig herausgestellt haben, wurden speziell für die Anwendung in der Syntheseplanung entwickelt und zusammen mit bekannten Suchverfahren zu einem Programm gebündelt. Dieses ist mittlerweile so ausgereift, daß konkrete Denkanstöße für Synthesen gegeben werden konnten. Es bleibt zu hoffen, daß WODCA und andere Systeme, die sich vom starren Synthon- und Transformschema der Syntheseplanungsprogramme der ersten Generation befreien, breite Akzeptanz finden und sich zu einem ganz normalen Planungshilfsmittel entwickeln werden, wie es die Bibliothek seit Jahrhunderten ist und Reaktionsdatenbanken im letzten Jahrzehnt geworden sind. Es ist zu früh, das Requiem für die computergestützte Syntheseplanung anzustimmen. Die Fehler und Versäumnisse der ersten Generation dieser Programme sind analysiert und verstanden worden. Die zweite Generation wartet jetzt auf ihr Signal zum Einsatz.

Die Entwicklung des WODCA-Programmsystems wurde durch ein Promotionsstipendium der Studienstiftung des deutschen Volkes für W.-D. Ihlenfeldt und in der Fortsetzung durch DFG-Fördermittel des Projektes Ga 208/17-1 ermöglicht. Wir danken der Shell-Forschung (jetzt: Cyanamid-Forschung) und hierbei vor allem Herrn Prof. Dr. R. Mengel für die Unterstützung unserer Arbeiten. Wir sind Prof. S. Hanessian, Universität Montreal, für die großzügige Überlassung des CHIRON-Katalogs verpflichtet. Janssen Chimica (jetzt: Acros Organica Division) hat uns den Katalog an Ausgangsmaterialien dankenswerterweise in computerlesbarer Form zur Verfügung gestellt. Herr Dipl.-Chem. R. Fick entwickelte einige der in dieser Übersicht vorgestellten Beispiele.

Eingegangen am 10. Mai 1994,
veränderte Fassung am 6. April 1995 [A 65]

Erklärung der Akronyme

AIPHOS: Artificial Intelligence for Planning and Handling Organic Synthesis
CASP: Computer Assisted Synthesis Planning
CAMEO: Computer Assisted Mechanistic Evaluation of Organic reactions

CHIRON: CHIRal synthON (und Name des weisesten der Zentauren)

COMPASS: COMPuter ASSisted organic synthesis

EROS: Evaluation of Reactions for Organic Synthesis

IGOR: Interactive Generation of Organic Reactions

LHASA: Logics and Heuristics Applied to Synthetic Analysis

LILITH: kein Akronym, Name einer Dämonin in Spinnen-
gestalt

OCSS: Organic Chemistry Synthesis Strategies

RAIN: Reactions And Intermediates Network

SECS: Simulation and Evaluation of Chemical Synthesis

SST: Starting material STrategies

SYNCHEM: SYNthetic CHEMistry

TRESOR: TRacing and Evaluation of Syntheses in Organic Reactions

WODCA: Workbench for the Organization of Data for Chemical Applications

- [1] E. J. Corey, *Pure Appl. Chem.* **1969**, *14*, 19–37.
- [2] E. J. Corey, *Angew. Chem.* **1991**, *103*, 469–479; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 455–465.
- [3] S. Warren, *Organic Synthesis – The Disconnection Approach*, Wiley, Chichester, 1981.
- [4] E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, 1989.
- [5] Übersichten zur computergestützten Syntheseplanung sind zu finden in: a) M. Bersohn, *Chem. Rev.* **1976**, *76*, 269–282; b) J. Gasteiger in *Computer in der Chemie* (Hrsg.: E. Ziegler), 2. Aufl., Springer, Berlin, **1985**, S. 207–257; c) R. Barone, M. Chanon in *Computer Aids to Chemistry* (Hrsg.: G. Vermilye, M. Chanon), Ellis Horwood, Chichester, **1986**, S. 19–102; d) G. Loftus, *Chemical Information Systems – Beyond the Structure Diagram*, Ellis Horwood, New York, **1991**, S. 222–262; e) M. A. Ott, J. H. Noordik, *Recl. Trav. Chim. Pays Bas* **1992**, *111*, 239–246. Die Zeitschrift *Recueil des Travaux Chimiques des Pays-Bas* widmete ein ganzes Heft (Juni 1992, Band 111) diesem Thema und brachte Beiträge von zehn Forschungsgruppen, die Programmsysteme zur computergestützten Syntheseplanung entwickeln. Der erste Beitrag gibt hierzu eine Einführung.
- [6] a) E. J. Corey, G. A. Petersson, *J. Am. Chem. Soc.* **1972**, *94*, 460–465; b) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, G. Petersson, *ibid.* **1975**, *97*, 6116–6124; c) E. J. Corey, A. K. Long, J. Mulzer, H. W. Orf, P. Johnson, A. P. W. Hewett, *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 221–230; d) E. J. Corey, A. P. Johnson, A. K. Long, *J. Org. Chem.* **1980**, *45*, 2051–2057; e) E. J. Corey, A. K. Long, G. I. Lotto, S. D. Rubenstein, *Recl. Trav. Chim. Pays Bas* **1992**, *111*, 304–309; f) A. P. Johnson, C. Marshall, P. N. Hudson, *Recl. Trav. Chim. Pays Bas* **1992**, *111*, 310–316.
- [7] E. J. Corey, W. T. Wipke, *Science* **1969**, *176*, 178–186.
- [8] W. T. Wipke, G. I. Ouchi, S. Krishnan, *Artif. Intell.* **1978**, *11*, 173–193.
- [9] H. L. Gelernter, A. F. Sanders, D. L. Larsen, K. K. Agarwal, R. H. Boivie, G. A. Spritzer, J. E. Searleman, *Science* **1977**, *197*, 1041–1049.
- [10] J. B. Hendrickson, *Angew. Chem.* **1990**, *102*, 1328–1338; *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 1286–1296.
- [11] Z. S. Hippe, G. Fic, M. Marcus, *Recl. Trav. Chim. Pays Bas* **1992**, *111*, 255–261.
- [12] E. V. Gordeeva, D. E. Lushnikov, N. S. Zevirov, *Tetrahedron Comput. Methodol.* **1990**, *3*, 445–459.
- [13] I. Dohgane, T. Takabatake, M. Bersohn, *Recl. Trav. Chim. Pays Bas* **1992**, *111*, 291–296.
- [14] K. Funatsu, S. Sasaki, *Tetrahedron Comput. Methodol.* **1988**, *1*, 27–38; *AI-PHOS konpyuuta ni yoru yuukigouseikeirotansaku*, Kyouritsushuppan, Tokio, 1994.
- [15] G. Sello, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 120–129.
- [16] R. Moll, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 117–119.
- [17] A. Weise, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 490–491.
- [18] M. Yanaka, K. Nakamura, A. Kurumisawa, W. T. Wipke, *Tetrahedron Comput. Methodol.* **1990**, *3*, 359–375.
- [19] E. S. Blurock, *Tetrahedron Comput. Methodol.* **1989**, *2*, 207–222.
- [20] a) J. Gasteiger, C. Jochum, *Top. Curr. Chem.* **1978**, *74*, 93–128; b) J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki, *ibid.* **1987**, *137*, 19–73; c) J. Gasteiger, M. Marsili, M. G. Hutchings, H. Saller, P. Löw, P. Röse, K. Rafeiner, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 467–476.
- [21] J. Gasteiger, *Tetrahedron* **1979**, *35*, 1419–1426.

- [22] a) J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219–3228; b) J. Gasteiger, H. Saller, *Angew. Chem.* **1985**, *97*, 699–701; *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 687–689.
- [23] M. G. Hutchings, J. Gasteiger, *Tetrahedron Lett.* **1983**, *24*, 2541–2544.
- [24] J. Gasteiger, M. G. Hutchings, *J. Chem. Soc. Perkin Trans. 2* **1984**, 559–564.
- [25] a) P. Röse, J. Gasteiger, *Anal. Chim. Acta* **1990**, *235*, 163–168; b) in *Software-Development in Chemistry 4* (Hrsg.: J. Gasteiger), Springer, Heidelberg, **1990**, S. 275–288.
- [26] V. Simon, J. Gasteiger, J. Zupan, *J. Am. Chem. Soc.* **1993**, *115*, 9148–9159.
- [27] J. Gasteiger, U. Hondelmann, P. Röse, W. Witzenbichler, *J. Chem. Soc. Perkin Trans. 2* **1995**, 193–204.
- [28] a) T. D. Salatin, W. L. Jorgensen, *J. Org. Chem.* **1980**, *45*, 2043–2051; b) W. L. Jorgensen, E. R. Laird, A. J. Gishorst, J. M. Fleischer, S. A. Gothe, H. E. Helson, G. D. Paderes, S. Sinclair, *Pure Appl. Chem.* **1990**, *62*, 1921–1932; E. R. Laird, W. L. Jorgensen, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 458–466.
- [29] a) J. Dugundji, I. Ugi, *Top. Curr. Chem.* **1973**, *39*, 19; b) I. Ugi, J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, R. Herges, M. Knauer, K. Reitsam, N. Stein, *Angew. Chem.* **1993**, *105*, 210–239; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 201–227.
- [30] a) J. Bauer, I. Ugi, *J. Chem. Res. Synop.* **1982**, 298; *J. Chem. Res. Miniprint* **1982**, 3101; b) J. Bauer, R. Herges, E. Fontain, I. Ugi, *Chimia* **1985**, *39*, 43–53; c) J. Bauer, *Tetrahedron Comput. Methodol.* **1989**, *2*, 269–293.
- [31] a) E. Fontain, J. Bauer, I. Ugi, *Chem. Lett.* **1987**, *37*; E. Fontain, J. Bauer, I. Ugi, *Z. Naturforsch. B* **1987**, *42*, 297; b) E. Fontain, *Tetrahedron Comput. Methodol.* **1990**, *3*, 469; E. Fontain, K. Reitsam, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 96–101.
- [32] a) S. Hanessian, J. Franco, B. Larouche, *Pure Appl. Chem.* **1990**, *62*, 1887–1910; b) S. Hanessian, J. Franco, G. Gagnon, D. Laramée, B. Larouche, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 413–425.
- [33] E. Zass, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 360–372.
- [34] J. L. Blake, R. C. Dana, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 394–399.
- [35] a) A. Parlow, C. Weiske, J. Gasteiger, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 400–402; b) J. Gasteiger, C. Weiske, *Nachr. Chem. Tech. Lab.* **1992**, *40*, 1114–1120.
- [36] D. F. Chodosh, J. Hill, L. Shpiisky, W. L. Mendelson, *Recl. Trav. Chim. Pays Bas* **1992**, *111*, 247–254.
- [37] G. A. Hopkinson, T. P. Cook, I. P. Buchan in *Chemical Information Systems – Beyond the Structure Diagram* (Hrsg.: D. Bawden, E. Mitchell), Ellis Horwood, New York, **1990**, S. 83–91.
- [38] T. E. Moock, J. G. Nourse, D. Grier, W. D. Hounshell in *Chemical Structures: The International Language of Chemistry* (Hrsg.: W. A. Warr), Springer, Berlin, **1988**, S. 303–313.
- [39] C. Jochum, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 71–73.
- [40] A. Barth, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 384–393.
- [41] a) H. L. Gelernter, J. R. Rose, C. Chen, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492–504; b) J. R. Rose, J. Gasteiger, *ibid.* **1994**, *34*, 74–90; c) E. S. Blurock, *ibid.* **1990**, *30*, 505–510; d) L. Chen, J. Gasteiger, J. R. Rose, *J. Org. Chem.*, im Druck.
- [42] H. Satoh, K. Funatsu, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 34–49.
- [43] Zusammenfassung mehrerer Antworten auf eine Anfrage nach Benutzererfahrungen mit Syntheseplanungssystemen durch K. M. Huss, Posting auf der OSC Chemistry Mailing List chemistry@osc.edu, 26. Juli 1993. Die Originalmitteilung, auf die wir uns hier beziehen, stammt von J. W. Boiten.
- [44] Einer der Gutachter dieses Beitrags war der Auffassung, daß hier Äpfel mit Birnen verglichen werden. Eine Suche in einer Reaktionsdatenbank diene häufig nur dazu, die Durchführbarkeit einer vom Chemiker selbst geplanten Reaktion zu überprüfen. Nach Meinung der Autoren ist dies aber gerade ein Symptom für die mangelnde Akzeptanz der Syntheseplanungsprogramme im Vergleich zu den Reaktionsdatenbanken: Im Idealfall sollte eine Planung integriert aus einem Planungssystem heraus verlaufen, also eine Reaktionsrecherche aus einem Syntheseplanungsprogramm zur Verifizierung aufgerufen werden.
- [45] a) W. S. Zhou, X. X. Xu, *Acc. Chem. Res.* **1994**, *27*, 211–216; b) X. X. Xu, J. Zhu, D. Z. Huang, W. S. Zhou, *Tetrahedron* **1986**, *42*, 819; c) G. Schmid, W. Hofsteinz, *J. Am. Chem. Soc.* **1983**, *105*, 624–625; d) T. Ravindranathan, M. A. Kumar, R. B. Memon, S. V. Hiremath, *Tetrahedron Lett.* **1990**, *31*, 755.
- [46] a) P. Willett, *Similarity and Clustering in Chemical Information Systems*, Wiley, New York, **1987**; b) M. A. Johnson, G. M. Maggiore, *Concepts and Applications of Molecular Similarity*, Wiley, New York, **1990**.
- [47] M. Wochner, J. Brandt, A. von Scholley, I. Ugi, *Chimia* **1988**, *42*, 217–225.
- [48] P. Willett, V. Winterman, *Ouant. Struct. Activ. Relat.* **1986**, *5*, 18–25.
- [49] W. T. Wipke, D. Rogers, *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 71–78.
- [50] a) J. Gasteiger, W. D. Ihlenfeldt, *Software Dev. Chem. 4 Proc. Workshop Comput. Chem. 4th* **1990**, 57–65; b) W. D. Ihlenfeldt, J. Gasteiger, *Software Dev. Chem. 5 Proc. Workshop Comput. Chem. 5th* **1991**, 59–67; c) J. Gasteiger, W. D. Ihlenfeldt, R. Fick, J. R. Rose, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 700–712; d) J. Gasteiger, W. D. Ihlenfeldt, P. Röse, R. Wanke, *Anal. Chim. Acta* **1990**, *235*, 65–75; e) J. Gasteiger, W. D. Ihlenfeldt, P. Röse, *Recl. Trav. Chim. Pays Bas* **1992**, *111*, 270–290; f) R. Fick, W.-D. Ihlenfeldt, J. Gasteiger, *Heterocycles* **1995**, *40*, 993–1007.
- [51] W. D. Ihlenfeldt, J. Gasteiger, *J. Comput. Chem.* **1994**, *15*, 793–813.
- [52] M. Randic, C. L. Wilkins, *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 31–37.
- [53] A. D. Sinita, B. S. Drach, *J. Org. Chem. USSR* **1973**, *9*, 706–709.
- [54] a) K. C. Nicolaou, R. D. Groneberg, N. A. Stylianides, T. Miyazaki, *J. Chem. Soc. Chem. Commun.* **1990**, 1275–1277; b) K. C. Nicolaou, E. P. Schreiner, W. Stahl, *Angew. Chem.* **1991**, *103*, 566–568; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 585–587.
- [55] B. E. Marron, K. C. Nicolaou, *Synthesis* **1989**, *55*, 537–539.
- [56] L. A. Paquette, H.-J. Kang, *J. Am. Chem. Soc.* **1991**, *113*, 2610–2621.
- [57] Im Fall des Kandidaten Sabinen 57 ist der Sechsring ein umhüllender Ring, dessen Elementarringe der Dreiring und der Fünfring sind. Daher ist der Sechsring hier zulässig.
- [58] Diese Aussage ist nicht ganz exakt. Die erschöpfende Überlagerung zweier Graphen in allen Orientierungen ist ziemlich rechenaufwendig. Daher nutzt WODCA Heuristiken, die den Kreis der Überlagerungen unter Synthesegesichtspunkten sinnvoll eingrenzen. Die maximale gemeinsame Substruktur ist nur dann unter den untersuchten Überlagerungen, wenn zumindest in einem Teil der Grenzgebiete zwischen der gemeinsamen Substruktur und den Zielverbindungen reaktive Zentren vorhanden sind.