

Organische Synthese im Computerzeitalter

Von James B. Hendrickson *

Die Planung von Synthesen mit dem Computer ist eine faszinierende Aufgabe. Lassen sich Kreativität, Phantasie und Intuition des Chemikers durch die Fähigkeiten des Computers ersetzen oder ergänzen? Jedes Programm zur Syntheseplanung mit dem Computer erfordert eine Strategie für das Bearbeiten großer Synthesebäume. Das hier beschriebene Programm SYNGEN basiert auf Vereinfachung und Systematisierung. Gesucht wird generell die „ideale konvergente Synthese“, bei der auch funktionelle Gruppen von Anfang an in die Planung einbezogen werden. Als wesentliche Kriterien dienen dabei die Zugänglichkeit und der Preis der Ausgangsverbindungen sowie die Anzahl der Reaktionsstufen.

1. Einleitung

Organische Synthese und Computeranwendungen haben traditionell kaum Berührungspunkte. Wenn wir nun unser Arbeitsgebiet an diesem 125. Geburtstag der BASF rückschauend betrachten, ist es sicher sinnvoll, den Gründen hierfür nachzugehen und sich zu fragen, was die Zukunft an gekoppelten Entwicklungen bringen mag. Der Einfluß von Computern in den Naturwissenschaften war zweifellos dramatisch; die Organische Synthese könnte von einer engen Zusammenarbeit ebenfalls außerordentlich stark profitieren.

Die folgende Übersicht faßt zusammen, was Computer zu leisten vermögen:

1. Komplizierte Berechnungen: Moleküleigenschaften, Reaktionsdynamik und Molecular Modeling.
2. Speicherung und Abruf von Informationen: Riesige gespeicherte Datenmengen; der Abruf erfordert effiziente Suchsysteme.
3. Künstliche Intelligenz zur Lösung intellektueller Probleme: Syntheseplanung und Strukturaufklärung.
4. Automatisierung von Laborexperimenten: Anwendung von Robotern zur Reaktionsoptimierung.

Diese Reihenfolge ist ungefähr historisch. Vor dreißig Jahren, als Naturstoffchemiker an der University of California at Los Angeles, lernte ich den Umgang mit Computern im Selbststudium, denn um die Möglichkeiten der Stereokontrolle bei der Sesquiterpensynthese zu verstehen, mußte ich die Konformationen von mittleren Ringen kennen. Die Lösung dieses Problems setzt die Berechnung der Molekülgestalt voraus. Am Beispiel von Cycloheptan wird deutlich, daß dabei sieben Bindungs- und sieben Diederwinkel (sowie die Bindungslängen) festgelegt werden müssen. Auf dieser Basis lassen sich dann alle 189 Abstände zwischen den nicht miteinander verbundenen Atomen errechnen, und zwar für jeden Einzelfall durch eine komplizierte Formel. Die Spannungsenergie, die der jeweiligen Geometrie entspricht, wird dann aus diesen Winkeln und Atomabständen berechnet, indem man individuelle Parameter für jede der drei Arten geometrischer Größen verwendet. Das Problem führt zu einer Suche nach dem Energieminimum und der dazugehörigen Geometrie, die bereits experimentell bekannt ist. Das wiederum erfordert die Berechnung einer Serie dieser Molekülgeometrien und ihrer Energien, um die richtige Mini-

mum-Energie-Konformation mit der richtigen Energie zu finden. Als weitere Schwierigkeit kommt allerdings noch hinzu, daß die gesamte Serie mit mehreren unterschiedlichen Abschätzungen der Energieparameter durchgeführt werden muß, die nur aus anderen physikalischen Messungen abgeleitet werden können.

Hier haben wir es mit einem typischen Problem zu tun, das sich nur mit Hilfe einer sehr großen Anzahl komplizierter iterativer Berechnungen lösen läßt. Ich machte mich also daran, diese Berechnungen für die Cycloalkane mit sechs bis zehn Ringgliedern durchzuführen; alles, was ich damals als Hilfsmittel hatte, waren ein Rechenschieber, die sin/cos-Tabellen sowie Bleistift und Papier! Es wurde schnell klar, daß dieses Problem systematisch nur mit einem Computer gelöst werden konnte. Deshalb nahm ich an einem eintägigen FORTRAN-Programmierkurs teil und erhielt Zugang zum damals einzigen Computer der UCLA, einer IBM 709 im Engineering Department. Die Prozedur bestand darin, jede Programmzeile mit einem aufdringlich lauten Lochkartenschreiber in eine IBM-Karte zu stanzen, den Stapel von Instruktionen dem Operateur zu übergeben und am nächsten Tag den gedruckten Output abzuholen. Das Problem wäre nie ohne Computer gelöst worden. So konnte ich die erste Publikation^[1] auf einem Gebiet veröffentlichen, das seither enorm gewachsen ist und nun als Molekülmechanik bezeichnet wird: eine intensive Zusammenarbeit von Computern und Organischer Chemie.

Durch die rasante Zunahme der Rechengeschwindigkeit und die Entwicklung der Computergraphik sind diese Berechnungen der Molekülgestalt zur Routine geworden; sie können dreidimensional am Bildschirm vorgeführt werden, das heißt, man kann der Energieminimierung zuschauen. Solche Rechnungen werden allgemein dazu benutzt, um die Gestalt von Molekülen mit einem Modell zu vergleichen. Die Vorführung ist graphisch so eindrucksvoll, daß es manchmal schwer fällt, sich der Schwachstelle der Methode bewußt zu bleiben; es ist dies die immer noch empirische Zuordnung der Parameter in den Gleichungen, welche die Energie des Moleküls und somit direkt die Vorzugskonformationen ergeben. Darüber hinaus fragen wir beim „modeling“ eigentlich nach der Wechselwirkung eines Moleküls mit einem anderen, z. B. einem Rezeptor. Durch das Berechnen der Gestalt lernen wir das Molekül als eine Art von neutralem Schwamm verstehen, der sich einer Durchdringung mit wachsendem Widerstand widersetzt, aber wir lernen nichts über seine elektrische „Farbe“: die intermolekularen Effekte lokaler Partialladungen und ihrer Felder, die π -Elektronen

[*] Prof. Dr. J. B. Hendrickson
Department of Chemistry, Brandeis University
Waltham, MA 02254-9440 (USA)

und die Orbitalwechselwirkungen. Kurz gesagt konzentrieren wir uns bei diesen Rechnungen auf die sterischen Wechselwirkungen und nicht auf die elektronischen.

Das Problem der molekularen Eigenschaften wurde zur selben Zeit auf einer physikalisch fundierteren Basis mit ab-initio- und semiempirischen Molekülorbitalberechnungen untersucht. In den dreißiger Jahren, als es noch keine Computer gab, war die Lösung des Problems des Wasserstoffmoleküls eine stolze Leistung. Dank der schnellen Zunahme der Geschwindigkeit iterativer Berechnungen mit dem Computer hat sich dieses Gebiet in den letzten dreißig Jahren enorm erweitert. Es sind nicht nur größere Moleküle für eine mehr oder weniger vollständige physikalische Beschreibung zugänglich, sondern es ist auch möglich, die Zuverlässigkeit einer Reihe von vereinfachenden Annahmen abzuschätzen und auf diese Weise sehr viel größere Moleküle auf einem weniger genauen theoretischen Niveau zu berechnen. Dennoch sind die Voraussagen nützlich. Diese Vorgehensweise ist die gleiche wie bei den oben besprochenen molekülmechanischen Berechnungen; Parameter und mathematische Details werden in zahlreichen individuellen Fällen so lange geändert, bis die Ergebnisse den experimentellen Ergebnissen nahekommen. Es ist den enormen Leistungen des Computers bei iterativen Berechnungen zu verdanken, daß diese Entwicklungen überhaupt möglich wurden.

2. Logische Grundlagen für die Syntheseplanung

Vor zwanzig Jahren wurde die Ära der künstlichen Intelligenz oder computergestützten Problemlösung auf dem Gebiet der Syntheseplanung durch die Corey-Wipke-Publikation^[2] des OCCS-Programms eröffnet. Die Autoren begannen damit, eine komfortable Kommunikation zwischen Mensch und Maschine zu schaffen, und entwickelten die Molekülgraphik, die eine Darstellung von Molekülen auf dem Bildschirm in Form der vertrauten Strukturformeln erlaubte. Dies ist kein triviales Problem; einige Einzelheiten der Implementierung werden im folgenden besprochen. Ohne Strukturformeln hätte der Synthesechemiker dem Computer vermutlich niemals große Aufmerksamkeit geschenkt, da die präparative Organische Chemie zunächst wenig mit mathematischen Berechnungen zu tun hat.

Coreys Ansatz für die computergestützte Syntheseplanung war einfach und geradlinig^[3]. Die Strukturänderung während einer Reaktion wird in Gegenrichtung als „Transform“ gesehen: vom Produkt zum Substrat. Dadurch können alle denkbaren Transforms auf das Zielmolekül angewendet werden, um die Zwischenstufen der vorangehenden Ebene (Synthesestufe) zu erzeugen. Diese wiederum führen zu Zwischenstufen, die noch eine Ebene weiter zurück liegen, usw. Der Prozeß wurde als „computergestützt“ bezeichnet, und der Benutzer sollte interaktiv mitwirken, indem er diesen Synthesebaum jeweils auf einige wenige Zwischenstufen auf jeder Ebene zusammenstutzte. Diese Prozedur hing wesentlich von einer Bibliothek allgemeiner Transforms ab, die als Datenbank im Computer gespeichert und immer angewendet wurde, wenn die Funktionalität des Zielmoleküls einem Transform aus der Bibliothek entsprach. Dies wurde als die Vorgehensweise bezeichnet, nach der ein Chemiker seine Synthesen plant, aber eigentlich schwer sich die Literatur über solche Planungsstrategien aus.

Das Problem der Syntheseplanung besteht einfach darin, daß die Anzahl der möglichen Synthesewege zu einem interessierenden Zielmolekül enorm groß ist, zumindest wesentlich größer als allgemein angenommen. Ein Blick auf den Synthesebaum in Abbildung 1 soll dies verdeutlichen. Die Punkte repräsentieren Verbindungen: das Zielmolekül auf der rechten Seite, die Zwischenstufen in der Mitte und die käuflichen Ausgangsverbindungen als große Punkte auf der linken Seite. Diese Ausgangsverbindungen werden kleiner, je weiter man nach links kommt, bis zu Verbindungen mit einem einzigen Kohlenstoffatom am linken Rand. Die Linien symbolisieren Reaktionen; sie zeigen nach rechts (oder die Transforms nach links), so daß einzelne Synthesewege herausgegriffen werden können, wie es hier in rot als Beispiel gezeigt ist. Die Zwischenstufen sind in Ebenen oder Reaktionsschritten angeordnet, die vom Zielmolekül aus rückwärts gezählt werden. Die Gesamtausbeute verringert sich dabei von Stufe zu Stufe; hier gehen wir von 80% Ausbeute pro Stufe aus.

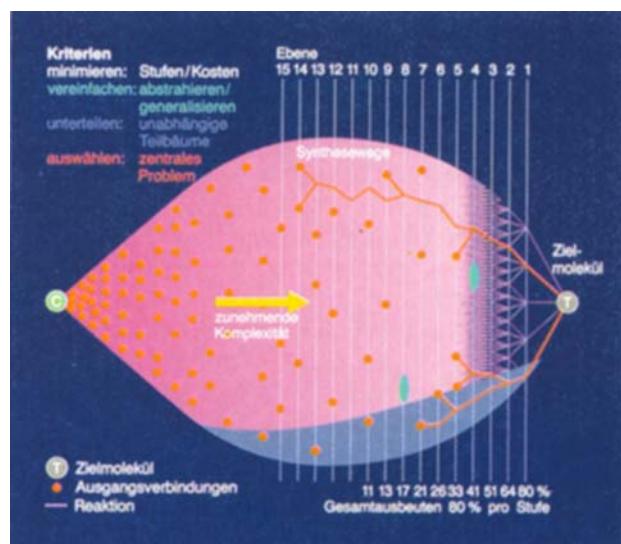


Abb. 1. Synthesebaum.

Für einen Computer ist es im Prinzip einfach, den gesamten Synthesebaum zu erzeugen; das wirkliche Problem ist es, die optimalen Synthesewege auszusuchen. Es ist sicher nicht unrealistisch anzunehmen, daß 30 Reaktionen existieren, die die Zwischenstufen der ersten Ebene in das Zielmolekül überführen können. Da jede dieser 30 Zwischenstufen von vergleichbarer Komplexität sein wird, kann man annehmen, daß jede von ihnen ebenfalls durch etwa 30 verschiedene Reaktionen aus den Zwischenstufen der Ebene 2 hergestellt werden kann. Geht man nur fünf Stufen zurück, so erzeugt man 30^5 oder 24 Millionen mögliche Synthesewege – erfahrungsgemäß sind viele Synthesewege noch wesentlich länger. Nun stellt sich die Frage: Welches ist der eine Weg, der im Laboratorium eingeschlagen werden soll?

Um einen Ansatz für die computergestützte Lösung des Problems zu finden^[4], können einige bekannte Prinzipien zur Suche in großen kombinatorischen Bäumen aus anderen Gebieten übernommen werden. Die Daten im Lösungsraum – hier verstehen wir darunter Strukturen und Reaktionen – müssen *systematisch* in digitalem Format beschrieben sein,

nicht nur für die schnelle Computerverarbeitung, sondern auch um sicherzustellen, daß alle Möglichkeiten als einfache mathematische Kombinationen abgeleitet werden können und daß keine Lösung verlorengeht. Der Synthesebaum ist so groß, daß eine *Vereinfachung* unumgänglich wird. Sie besteht darin, die Eintragungen zu verallgemeinern, indem man triviale Einzelheiten miteinander verschmilzt. Dadurch wird der Baum viel kleiner; er enthält weniger Eintragungen, die die ursprünglichen zusammenfassen. Die Algorithmen können so ausgelegt werden, daß diese Reduktion erleichtert wird. Die Vereinfachung ähnelt derjenigen von Landkarten: Bei gleichem Format zeigt die Karte eines kleinen Gebietes wesentlich mehr Einzelheiten als die Übersichtskarte eines großen Gebietes. In der Praxis benutzen wir die Übersichtskarte, um unseren Weg zu einem gewünschten Ort zu finden, und die andere, um Einzelheiten zu sehen, wenn wir uns dem Ort nähern. Das gleiche Prinzip – erst generalisieren und später bis zu den Details verfeinern – wird hier benutzt, um die optimalen Wege im Synthesebaum zu finden. Da der vereinfachte und verallgemeinerte Baum immer noch sehr groß ist, kann er leichter handhabbar gemacht werden, indem man ihn in unabhängige Unterbäume *aufteilt*, die nun einzeln und nacheinander durchforstet werden.

Schließlich bleibt das Problem, die wenigen besten Wege *auszuwählen*. Das erfordert die Definition von *Auswahlkriterien*, die streng genug formuliert sein müssen, um nur einige wenige Wege zuzulassen. In unserem Ansatz für die Syntheseplanung haben wir uns für das Kriterium der *Wirtschaftlichkeit* – der kürzesten Wege mit dem effizientesten Zusammenbau aus den billigsten Chemikalien – entschieden. Wir streben darüber hinaus an, daß alle diese besten Wege durch den Computer gefunden werden, und zwar durch wiederholtes Anwenden von Kriterien und nicht interaktiv und willkürlich durch den Benutzer.

Um diese Wege zu finden, teilen wir jede Struktur formal in das *Kohlenstoffskellett* oder -*gerüst* (später werden gegebenenfalls Heteroatome des Gerüsts berücksichtigt) und in die *Funktionalität* ein, d.h. Mehrfachbindungen im Gerüst und funktionelle Gruppen am Gerüst, die Heteroatome enthalten. Gemäß dieser Dichotomie betrachten wir *Aufbaureaktionen* (Verknüpfungsreaktionen), die CC- σ -Bindungen erzeugen und so das Gerüst zusammensetzen, und *Refunktionalisierungen*, die funktionelle Gruppen ändern, ohne das Gerüst zu beeinflussen. Aus dieser Sicht ist der Aufbau des Gerüsts elementar und wird zuerst betrachtet. Dies führt schon zu Anfang zu einer drastischen Vereinfachung des Synthesebaums – es werden nur die Grundgerüste berücksichtigt. Zum Vergleich: Während es vielleicht 50 000 mögliche Ausgangsverbindungen gibt, existieren nur 13 acyclische Gerüste mit sechs oder weniger Kohlenstoffatomen. Eine Überprüfung publizierter Synthesen zeigt, daß bei nichtaromatischen Zielmolekülen durchschnittlich nur *drei Kohlenstoffatome* der Ausgangsverbindung in das Zielgerüst eingebaut werden.

3. Das Molekülgerüst

Die elementare Bedeutung des Gerüsts und seines Zusammensetzes bei der Synthese hat eine philosophische Komponente: *Die Synthese selbst erfordert ein gerüstartiges Konzept*. Die chemische Synthese besteht darin, kleine Ausgangsver-

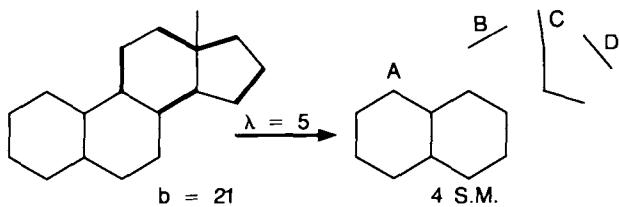
bindungen zu einem großen Zielgerüst zusammenzufügen. Deshalb sind die zentralen und wichtigsten Reaktionen die Aufbaureaktionen (Verknüpfungsreaktionen) für das Gerüst, die diese kleinen Stücke zusammensetzen. Im Prinzip werden Refunktionalisierungen überhaupt nicht benötigt, was im Gegensatz zu der Beobachtung steht, daß die durchschnittliche Synthese doppelt so viele Refunktionalisierungen wie Aufbaureaktionen umfaßt. Dies führt zu einem strengen Kriterium für die Wirtschaftlichkeit: nur nach *idealen Synthesen* zu suchen, nämlich solchen, die nur aus Aufbaureaktionen bestehen, da diese – obwohl selten – die kürzesten Wege liefern. Bei einer idealen Synthese haben die gewählten Ausgangsverbindungen die richtigen funktionellen Gruppen, um die Aufbaureaktionen einzuleiten, die nacheinander angewendet zum Zielgerüst mit allen funktionellen Gruppen in der korrekten Position als natürliche Konsequenz der Chemie der Reaktionssequenz führen.

Diese Überlegungen legen nahe, daß wir die Strukturen im Synthesebaum systematisieren müssen, um effizient – aber getrennt – mit Gerüst und Funktionalität umzugehen. Wir müssen bestimmen, welche der Gerüstbindungen des Zielmoleküls sich am besten zur Verknüpfung eignen und welche funktionellen Gruppen benötigt werden, um diese Bindungen so zu knüpfen, daß dabei die Funktionalität des Zielmoleküls direkt erhalten wird.

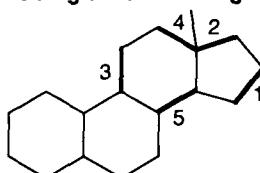
Obwohl die Beschränkung auf das Gerüst den Synthesebaum stark vereinfacht, bleibt dennoch eine große Zahl von möglichen Wegen, das Zielgerüst zusammenzusetzen. Das Problem beruht auf dem kombinatorischen Puzzle, welche Stücke auszuwählen und in welcher Reihenfolge sie zusammenzusetzen sind. Wir definieren einen *Satz von strategischen Bindungen* im Zielgerüst als einen Satz von λ Bindungen, die konstruiert werden sollen; Abbildung 2 zeigt dies für das Östrongerüst. Wenn wir alle strategischen Bindungen aus dem Gerüst entfernen, bleiben mehrere kleinere Gerüste von Ausgangsstoffen. Ein solcher Satz von strategischen Bindungen definiert eine Familie von Synthesen, die alle diese Gerüstbindungen knüpfen. Für jeden Satz von strategischen Bindungen definiert die Reihenfolge, in der die Bindungen geknüpft werden, einen verallgemeinernden Synthe-

Satz strategischer Bindungen

= Satz von λ Gerüstbindungen, die verknüpft werden



Geordneter Satz strategischer Bindungen:



Bauplan:

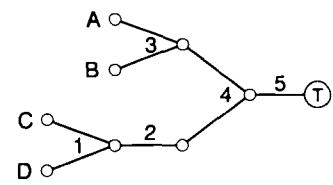


Abb. 2. Zersetzung und Zusammenbau des Östrongerüsts. λ = Anzahl strategischer Bindungen; b = Anzahl CC-Bindungen; S.M. = Anzahl von Ausgangsverbindungen.

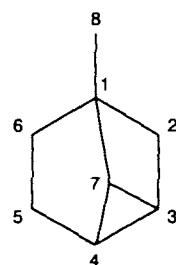
seweg, d. h. einen Bauplan für den Einbau der Startgerüste in das Zielgerüst. Der Bauplan in Abbildung 2 repräsentiert einen der Wege des Synthesebaums (Abb. 1), aber er setzt sich nur aus Aufbaureaktionen zusammen. In der Tat ist die einfachste allgemeine Beschreibung eines Syntheseweges dieser *geordnete Satz von strategischen Bindungen*, welcher die Gerüste der Ausgangsverbindungen und die Reihenfolge der zu knüpfenden Bindungen im Zielmolekül wiedergibt.

Die Anzahl der geordneten Sätze von strategischen Bindungen entspricht der Anzahl der möglichen Wege, das Gerüst zu zerschneiden: Für ein Zielmolekül mit b Bindungen ergibt dies $b!/(b-\lambda)!$ Wege. Beispielsweise erfordert der Aufbau des Östrongerüsts ($b = 21$ Bindungen) aus Bruchstücken mit durchschnittlich drei Kohlenstoffatomen einen Satz von $\lambda = 9$ strategischen Bindungen und somit 10^{11} Baupläne (geordnete Sätze von strategischen Bindungen), d. h. 100 Milliarden Wege, und das ohne jegliche Berücksichtigung chemischer Details. Drastisch vor Augen geführt werden kann diese kombinatorische Vielfalt durch die Synthese von Corticosteroiden aus Einkohlenstoff-Bausteinen. Hier ergeben sich 6×10^{23} verschiedene Baupläne, d. h. um ein mol Cortison zu synthetisieren, könnte man jedes Molekül auf einem anderen Weg herstellen.

Es gibt zwei wirkungsvolle Kriterien, die es erlauben, diese Anzahl von Möglichkeiten stark zu reduzieren. Das erste ist die Wahl einer *konvergenten* Synthese, die bezüglich der Gesamtmenge an Ausgangsstoffen wesentlich effektiver^[5] als eine lineare Synthese verläuft. Die konvergenten Synthesen können dadurch erzeugt werden, daß man das Zielgerüst in zwei Teile zerschneidet und jedes dieser Stücke in zwei weitere. Das zweite Kriterium ist die kommerzielle Verfügbarkeit der vier so abgeleiteten Startgerüste. Das bedeutet, daß ein Zielmolekül mit 20 Kohlenstoffatomen auf Ausgangsverbindungen zurückgeführt werden muß, die durchschnittlich fünf Gerüstkohlenstoffatome enthalten. Nach unseren Erfahrungen mit Chemikalienkatalogen fällt die Variationsbreite funktioneller Gruppen in Ausgangsstoffen mit mehr als fünf Kohlenstoffatomen scharf ab. Somit sollten wir durch Anwendung dieser Kriterien imstande sein, präparativ realistische Vorschläge für wenigstens einige dieser Pläne zu finden. Die konvergenten Synthesen sind nicht nur die kürzesten und effizientesten, sondern sie sind auch ziemlich selten. Damit gelingt auch die benötigte starke Reduktion in der Anzahl der Sätze von strategischen Bindungen, die untersucht werden müssen^[6].

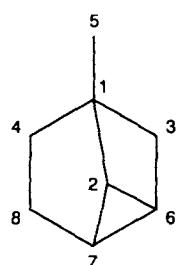
An dieser Stelle sollten wir auf das Problem der computergerechten Speicherung der Gerüste und das Problem des Vergleichs und der Feststellung von Übereinstimmungen mit bereits im Katalog abgespeicherten Strukturen eingehen^[7]. Abbildung 3 A zeigt ein Molekülergerüst aus acht Atomen, die willkürlich durchnumeriert sind, sowie eine Konnektivitätsmatrix der Bindungen, die die Struktur vollständig beschreibt. Diese Matrix kann in eine lineare binäre Zeichenfolge umgeformt werden, und dies ist nun die binäre Zahl, die als Nummer zur Identifizierung dieses Skeletts verwendet werden kann. Es gibt allerdings $8!$ Möglichkeiten, dieses Skelett durchzunumerieren, d. h. 40 320 Numerierungen, aus denen wir eine heraussuchen müssen, die das Gerüst eindeutig kennzeichnet. Das kann dadurch erreicht werden, daß man diejenige auswählt, für die sich die *größte binäre Zahl* ergibt. Das derart markierte Gerüst und die Binärzahlen sind in Abbildung 3 B gezeigt. Der Katalog besteht nun aus einer

A. Beispiel



	1	2	3	4	5	6	7	8
1	1	1	0	0	1	1	1	1
2		1	0	0	0	0	0	0
3			1	0	1	0	0	0
4				1	0	1	0	0
5					1	0	0	0
6						0	0	0
7							0	0
8								0

B. Maximal



	1	2	3	4	5	6	7	8
1	1	1	1	1	0	0	0	0
2		0	0	0	1	1	0	0
3			0	0	1	0	0	0
4				0	0	0	0	1
5					0	0	0	0
6						1	0	0
7							1	0
8								1

A: 10000111100000100101010100000

B: 111100000011000100001000101

Abb. 3. Kanonische Numerierung. A) Beispiel mit beliebiger kanonischer Numerierung; B) maximale kanonische Numerierung; unten: Binärzahlen für die Identifizierung der Molekülergerüste, die nach A und B mit Nummern versehen wurden.

Liste von größten binären Zahlen der Gerüste, die in aufsteigender Ordnung gespeichert werden. Dadurch reduziert sich die Suche eines generierten Gerüsts nach einem Gegenstück im Katalog zu einer einfachen binären Suche, die im Computer sehr schnell abläuft.

4. Systematisieren von Reaktionen

Um nun zur Chemie für die Synthesewege zu gelangen, müssen wir zuerst ein System schaffen, das die funktionellen Gruppen an den Gerüsten und ihre strukturelle Umwandlung durch Reaktionen beschreibt. Ein solches System muß zur Programmierung algorithmisierbar sein, eine einfache, aber rigorose Beschreibung der Chemie ermöglichen und die Fähigkeit zum Verallgemeinern und Vereinfachen durch Verschmelzen trivialer struktureller Details haben. In Abbildung 4 sind vier synthetisch wichtige Arten der Verknüpfung zusammengefaßt, die ein Gerüstkohlenstoffatom haben kann^[8]:

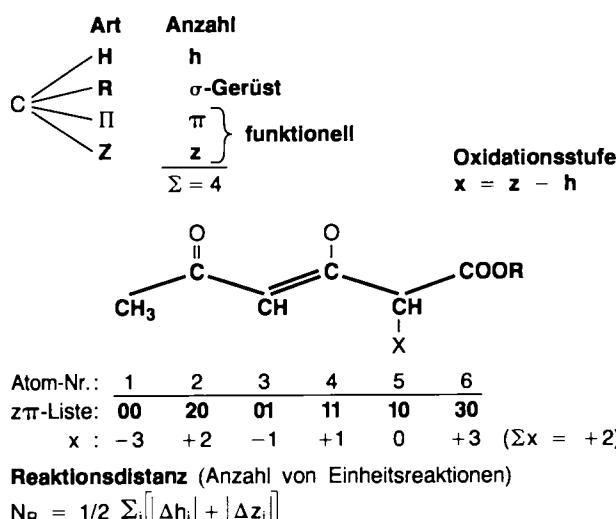
H bedeutet eine Bindung zu Wasserstoff, einem Elektronenpaar oder einem elektropositiven Atom,

R bedeutet eine σ -Bindung zu Kohlenstoff – d. h. die Verknüpfung im Grundgerüst^[9],

II bedeutet eine π -Bindung zu einem anderen Gerüst-Kohlenstoffatom,

Z bedeutet eine Bindung (σ oder π) zu einem elektronegativen Atom (O, N, X, S, P).

Die Anzahl der Bindungen zum Kohlenstoffatom für die vier Verknüpfungsarten wird dann durch h, σ , π bzw. z wiedergegeben. Die Summe beträgt jeweils vier. Die Oxidations-



16 Einheitsreaktionen (bezogen auf ein Kohlenstoffatom)

Oxidation	ZH
Reduktion	HZ
Substitution	HH, ZZ
Addition	HΠ, ZΠ, RΠ
Eliminierung	ΠH, ΠZ, ΠR
Aufbau	RH, RZ, RΠ
Fragmentierung	HR, ZR, ΠR

Abb. 4. Vier synthetisch wichtige Verknüpfungsarten eines Gerüstkohlenstoffatoms.

zahl des Kohlenstoffatoms beträgt $x = z - h$; die Änderung der Oxidationszahl bei einer Reaktion ergibt sich durch Aufsummieren aller Änderungen von x .

Die Funktionalität eines Kohlenstoffatoms kann nun als $z\pi$ -Wert codiert werden, der im Computer nur zwei Bits für z ($= 0 - 3$) und π ($= 0 - 2$) beansprucht. Die computergerechte Beschreibung eines Moleküls besteht dann aus der größten binären Zahl für sein Gerüst, gefolgt von der $z\pi$ -Liste der Funktionalität der durchnumerierten Kohlenstoffatome. Der σ -Wert eines jeden Kohlenstoffatoms stammt aus der Summe der entsprechenden Reihe der Konnektivitätsmatrix (Abb. 3), und der h -Wert ergibt sich durch Subtraktion von $(\sigma + z + \pi)$ von 4. Diese zusammengesetzte binäre Liste für das Gerüst und die Funktionalität einer Struktur kann dann mit den Ausgangsverbindungen im Katalog auf Übereinstimmung geprüft werden, entweder bezüglich des Gerüsts oder bezüglich der gesamten abstrahierten Struktur. Die Verschmelzung von Details geht aus Abbildung 4 hervor: Alle Carboxyl-Verbindungen erhalten einen $z\pi$ -Wert von 30.

Reaktionen lassen sich ebenfalls durch Verwendung des Systems in Abbildung 4 codieren. Unter Vernachlässigung des Mechanismus kann jede Reaktion durch die Netto-

Strukturänderung an jedem betroffenen Gerüst-Kohlenstoffatom vom Substrat zum Produkt beschrieben werden. Eine Einheitsreaktion an irgendeinem Kohlenstoffatom entspricht einem Austausch von Bindungspartnern; dafür gibt es $4 \times 4 = 16$ Möglichkeiten. Diese bezeichnen wir mit einem zweibuchstabigen Symbol. Der erste Buchstabe steht für die geknüpfte, der zweite für die gebrochene Bindung. Die möglichen Einheitsreaktionen an einem Kohlenstoffatom fallen in bekannte chemische Kategorien, wie in Abbildung 4 unten angegeben. Reaktionen wie Eliminierung, Addition und Verknüpfung erfordern Änderungen an mehr als einem (benachbarten) Kohlenstoffatom; jedem Kohlenstoffatom wird ein Symbol zugeordnet.

Der Wert eines solchen Systems besteht darin, daß es eine abstrakte Beschreibung jeder möglichen Änderung an einem Kohlenstoffatom liefert. Dadurch bildet es die Grundlage für eine strenge Organisation aller organisch-chemischen Reaktionen^[10]. Wie das Beilstein-System organische Verbindungen ordnen kann, läßt sich mit dem System ein Katalog organisch-chemischer Reaktionen aufstellen, in welchem jede mögliche Reaktion ihren Platz findet. Eine solche Organisation aller Einheitsreaktionen mit ein bis drei Kohlenstoffatomen ist in Abbildung 5 wiedergegeben; wie wir später sehen werden, umfaßt sie alle üblichen Reaktionen^[11].

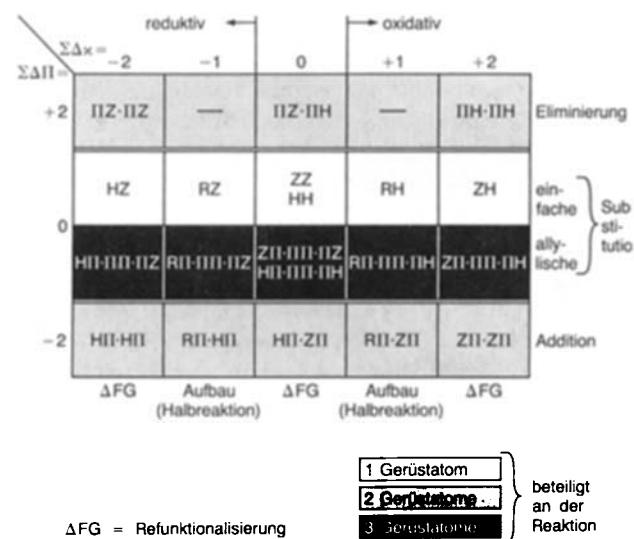


Abb. 5. Charakterisierung von Reaktionen mit ein bis drei Gerüstatomen.

Aufbau- oder Verknüpfungsreaktionen sind die Reaktionen, mit denen sich die Syntheseplanung besonders beschäftigt; sie sind auch die einzigen Reaktionen für ideale Synthesen. Eine Aufbaureaktion kann als Kombination zweier Halbreaktionen gesehen werden, üblicherweise einer oxidativen (nucleophil) und einer reduktiven (elektrophil). Die drei möglichen Halbreaktionen jeder Art sind in Abbildung 5 angegeben. Betrachtet man die Netto-Strukturänderungen, so stellt man fest, daß es sich um die üblichen bekannten Aufbau-Halbreaktionen handelt, z.B. um die Bildung von Carbanionen oder allylischen Carbanionen, Enolaten und Enaminen, um Alkylierungen und allylische Alkylierungen, Carbonyl-Additionen, Acylierungen und konjugierte Additionen. Mit dem obigen System können wir sehen, wie der Computer die Substrate aus den Produkten (oder umgekehrt) für jede gewünschte Verknüpfung generiert (Abb. 6).

Hier wird die zu knüpfende Bindung (fette Linie) durch den geordneten Satz von strategischen Bindungen festgelegt; maximal drei Kohlenstoffatome, von jedem Ende aus gezählt (markiert mit α , β , γ), enthalten die Funktionalität des Verknüpfungsproduktes. Für jeden $\alpha\beta\gamma$ -Strang (ausgehend von der zu knüpfenden Bindung) wird eine kombinierte $z\pi$ -Liste aufgerufen, und zu dieser wird ein $\Delta z\pi$ -Generator addiert,

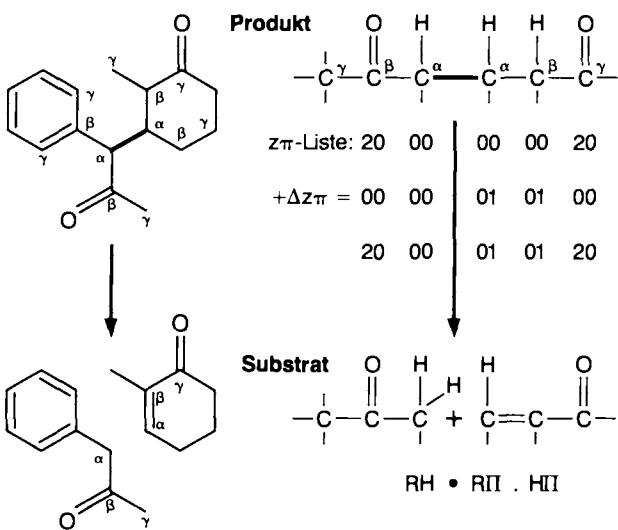


Abb. 6. Schnelle Generierung von Reaktionen mit dem SYNGEN-Programm.

der charakteristisch für die möglichen Kombinationen von nucleophilen und elektrophilen Halbreaktionen ist. Dieser Vorgang erzeugt direkt die $z\pi$ -Liste für die möglichen Substrate, wie Abbildung 6 rechts an einem Beispiel zeigt. Die Rechnungen laufen auf einem Computer sehr schnell.

5. Das SYNGEN-Programm

Die vorangehende Beschreibung einer Syntheseplanungsstrategie für Computer ist im SYNGEN-Programm realisiert, das in unserem Laboratorium während der letzten zehn Jahre entwickelt wurde^[12]. Das Programm wurde dazu entworfen, alle denkbaren kürzesten Synthesewege (konvergente ideale Synthesen) für beliebige Zielmoleküle zu lokalisieren und zu generieren. Das Programm ist nicht interaktiv und ist somit unabhängig von der Voreingenommenheit des Benutzers. Abbildung 7 zeigt die – zweistufige – Arbeitsweise im Überblick.

A) Zuerst wird nur das Zielgerüst geprüft und in zwei Ebenen aufgeteilt, um alle effizienten, konvergenten Wege für seinen Aufbau aus vier (oder weniger) Gerüsten von verfügbaren Ausgangsverbindungen zu finden, die im Katalog zusammengefaßt sind. (Der Katalog umfaßt zur Zeit mehr als 6000 Verbindungen.) Diese Wege entsprechen allen geordneten Sätzen von strategischen Bindungen; jeder dieser Sätze entspricht einem unabhängigen Unterbaum des gesamten Synthesebaums. In Abbildung 7 ist ein solcher Satz von strategischen Bindungen am Beispiel des Östrongerüsts aufgezeichnet. Durch Brechen von drei strategischen Bindungen über zwei Ebenen werden die drei Startgerüste B, C und D gefunden (siehe linke Seite).

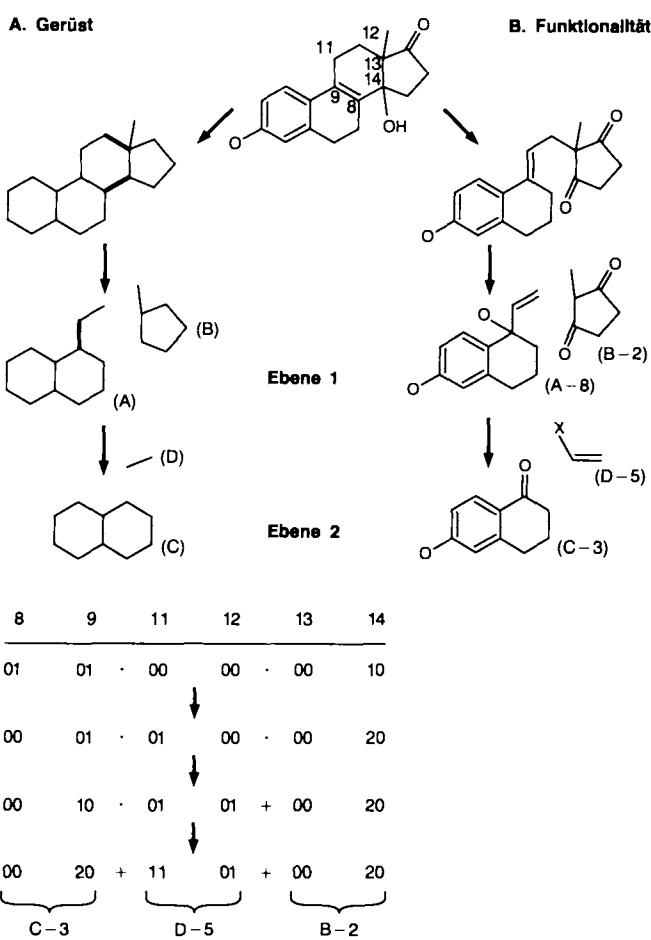


Abb. 7. Überblick über die Arbeitsweise des SYNGEN-Programms (siehe auch Abschnitt 6).

B) Danach generiert das Programm die Änderungen der Funktionalität, wie in Abbildung 6 gezeigt, rückwärts von der Funktionalität des Zielmoleküls. Wenn alle Bindungen in jedem Satz von strategischen Bindungen geknüpft sind (in retrosynthetischer Richtung), wird die benötigte Funktionalität jeder Ausgangsverbindung ermittelt, und diese Strukturen werden dann im Katalog gesucht. Nur solche Wege werden berücksichtigt, bei denen alle Ausgangsverbindungen im Katalog vorhanden sind. Die drei Startverbindungen in Abbildung 7 sind B-2, C-3 und D-5. Entsprechend der gezeigten Vereinfachung des Zielmoleküls erhält man so die kommerzielle (und ideale) Synthese von Östron. In Abbildung 7 unten sind die aufeinanderfolgenden $z\pi$ -Listen der reagierenden Kohlenstoffatome während der drei Verknüpfungsreaktionen aufgeführt.

Das Arbeiten mit dem SYNGEN-Programm ist sehr einfach. Mit einer einfachen Zeichenroutine gibt der Benutzer die rohe Struktur des Zielmoleküls ein, die vom Programm in ein einheitliches Format umgewandelt wird. Die Synthesewege werden dann ohne interaktives Eingreifen des Benutzers innerhalb von etwa 1–3 Minuten generiert. Das Ergebnis wird in einem Verzeichnis gespeichert und kann zu beliebiger Zeit zur Inspektion der Ergebnisse aufgerufen werden, die auf dem Bildschirm zusammengefaßt werden, wie es Abbildung 8 für die Synthese von 3'-Azido-3'-desoxythymidin (AZT) zeigt. Unten ist die Anzahl der Eintragungen pro Ebene für die folgenden vier Kategorien angefügt^[13]: Sätze von strategischen Bindungen, Ausgangsverbin-

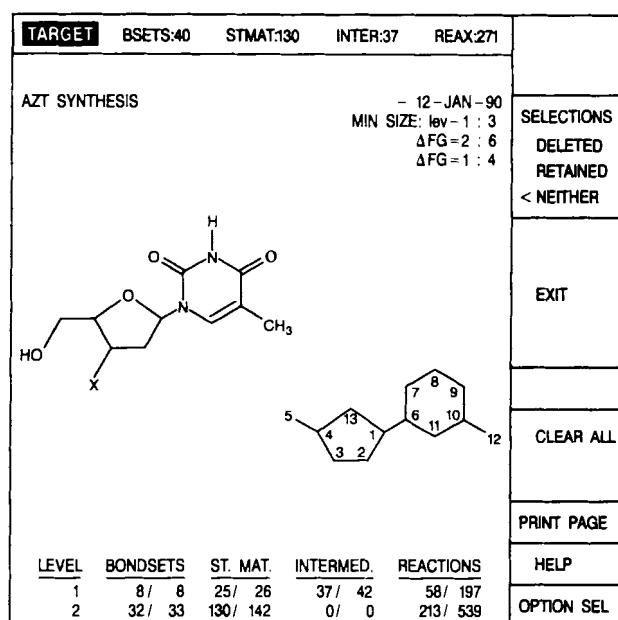


Abb. 8. Wiedergabe des Bildschirms bei Anwendung des SYNGEN-Programs. Beispiel: Synthese von AZT ($X = N_3$).

dungen, Zwischenstufen und Reaktionen. Jede Kategorie kann einzeln zur Überprüfung aufgerufen werden. Jede einzelne Eintragung in jeder der Kategorien kann zusammen mit den dazugehörigen Eintragungen in jeder anderen Kategorie beibehalten oder gelöscht werden. Wenn man demnach z. B. unerwünschte Sätze von strategischen Bindungen löscht, werden alle dazugehörigen Reaktionen und Verbindungen ebenfalls gelöscht. Die Gesamtsummen am oberen Bildschirmrand werden dabei automatisch angepaßt, so daß der Benutzer den Umfang der Lösungen seiner Wahl überblicken kann.

Ein Beispieldausdruck für Ausgangsverbindungen (Abb. 9) zeigt, daß einige Ausgangsverbindungen, die im Katalog nicht gefunden wurden, durch Refunktionalisierung aus an-

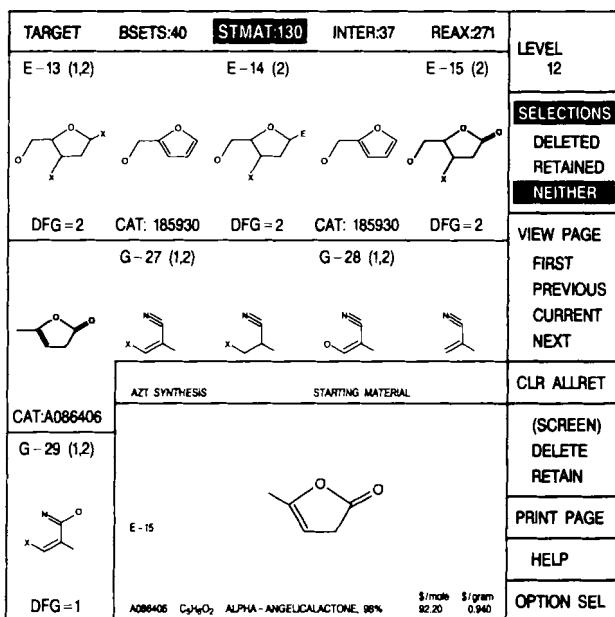


Abb. 9. Wiedergabe des Bildschirms bei Anwendung des SYNGEN-Programs. Beispiel: Ausgangsverbindungen für die Synthese von AZT.

deren hergestellt werden können. Diese fehlenden Verbindungen sind mit DFG gekennzeichnet; es folgt eine vorhandene Ausgangsverbindung mit ihrer Aldrich-Bestellnummer. DFG bedeutet eine „Reaktionsdistanz“ und gibt an, wieviele Stufen oder Refunktionalisierungen zwischen der generierten Verbindung und der ähnlichsten im Katalog liegen. Dieser Wert kann nach der Formel für N_k in Abbildung 4 berechnet werden. Die Einblendung in Abbildung 9 zeigt, daß eine Erweiterung der Angaben aufgerufen werden kann, in diesem Fall für die Verbindung E-15.

Im Beispieldausdruck für Reaktionen (Abb. 10) ist die zu synthetisierende Struktur oben links (das Zielmolekül auf Ebene 1, die Zwischenstufen auf Ebene 2) zu sehen. Die dazugehörigen Reaktionen erscheinen in abstrahierter Form mit abgekürzten Notationen der beteiligten Halbreaktionen für jede Verknüpfung. Wiederum läßt sich eine Erweiterung aller Eintragungen aufrufen, hier z. B. für Reaktion 3-34 des dritten Satzes von strategischen Bindungen. Betroffen davon sind auch die abgekürzten Reaktionsbezeichnungen.

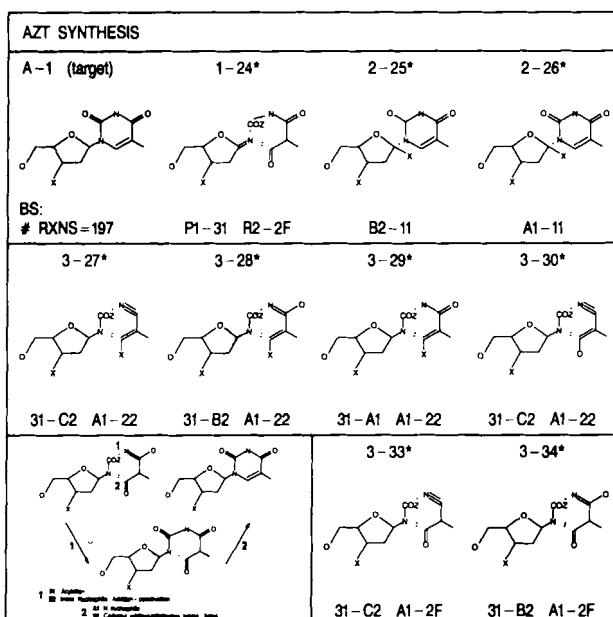


Abb. 10. Wiedergabe des Bildschirms bei Anwendung des SYNGEN-Programs. Beispiel: Reaktionen für die Synthese von AZT.

Auch wenn sehr viele Synthesewege gefunden werden, gibt es effektive Methoden zur Beschneidung des Synthesebaums, die es erlauben, die gewünschten Wege schnell einzuziehen^[14], z. B. die Möglichkeit, Eintragungen in jeder Kategorie zu löschen oder beizubehalten, oder die Möglichkeit, alle Reaktionen zurückzuverfolgen, die zu einem einzigen Satz von strategischen Bindungen, einer Ausgangsverbindung oder einer Zwischenstufe gehören. Ein anderes Werkzeug ist das Setzen von Randbedingungen bezüglich Verbindungen oder Reaktionen, wie aus Abbildung 11 hervorgeht. Diese Randbedingungen betreffen die Kosten von Ausgangsverbindungen oder die Notwendigkeit der Umwandlung ihrer funktionellen Gruppen und die Regiochemie oder andere Unsicherheitsfaktoren oder Inkompatibilitäten, die von den Reaktionen herrühren. Eine Auswahl dieser breiten Möglichkeiten kann die Gesamtzahl der erlaubten Wege stark verringern. Die im Auswahlmenü in Abbildung 11 ge-

REACTION / SELECTIONS	RXN-1-	RXN-2-	TOTAL REACTIONS	
UNCLEAR REGIO:KETONE	0	0	24	10
UNCLEAR REGIO:PI-BOND	10	0	MOVE CURSOR TO:	
MIXED MODE CONSTRUCTION	4	0	LEVEL -1-	
TWO-STEP ANNEL: ACID/BASE	0	0	LEVEL -2-	
TWO-STEP ANNEL: RED.CYC.	0	0	TEXT FOR 1&2	
BALDWIN RULES VIOLATION	24	0	ENTER:	
COMPATIBILITY: LEAVING GROUP	10	6	R FOR RETAIN	
COMPATIBILITY: KETONE	0	0	D FOR DELETE	
COMPATIBILITY: ALDEHYDE	0	0	C FOR CLEAR	
EQUIV: SAME CONCEPT	12	2	Clear All Flags	
EQUIV: SAME SUBSTRATES	4	2	HELP	
EQUIV: SAME REACTIONS	17	2	EXIT	
COMPOUND / SELECTIONS	SM-1-	SM-2-	RXN-1-	RXN-2-
COST LIMIT 2S(1) 1S(2)	6	14	24	10
SM NOT REFUNCTIONALIZED	0	8	0	10
SM REFUNCTIONALIZED	6	6	24	10
INTERMEDIATES	4	0	24	0

Abb. 11. Wiedergabe des Bildschirms bei Anwendung des SYNGEN-Programms. Auswahlmenü für Reaktionen.

setzten Randbedingungen senken die Anzahl der Reaktionen von 736 in Abbildung 8 auf nur 34.

Die Anwendung des Programms hat gezeigt, daß so auf einfacherem Weg alle denkbaren kürzesten Synthesen für viele Zielmoleküle gefunden werden. Das Programm findet bekannte Synthesen und schlägt viele andere interessante Wege vor. Es hat den entscheidenden Vorteil, daß der Benutzer sicher sein kann, keinen kurzen Syntheseweg zu seinem Zielmolekül übersehen zu haben.

6. Aktuelle Entwicklungen

Als Erweiterung für das SYNGEN-Programm haben wir nun ein Interface für die großen kommerziellen Reaktionsdatenbanken REACCS und SYNLIB geschaffen. Dies ermöglicht es, nahe verwandte Literaturbeispiele für die von SYNGEN generierten Reaktionen zu finden. Die Datenbanken müssen zunächst in das SYNGEN-Format umgewandelt werden, damit nicht Molekülstrukturen, sondern die *Netto-Strukturänderungen* bei der Reaktion erkannt werden. Dazu prüften wir die Edukt- und Produktstrukturen in der Datenbank, um Reaktionen zu finden, die CC-Bindungen verknüpfen, und listeten dann die π -Änderungen in den $\alpha\beta$ -Strängen für jede Halbreaktion auf. Diese werden dann zum schnellen Wiederfinden in eine Nucleophil \times Elektrophil-Matrix geordnet. Auf diese Weise lassen sich alle Übereinstimmungen mit den Reaktionen finden, die mit SYNGEN generiert wurden. Wenn es zu vielen Beispielen eines Reaktionstyps gibt, wird nach einer genaueren Übereinstim-

mung der Atomsorten und der σ , z , π -Werte an der Verknüpfungsstelle gesucht, bis die beste gefunden ist. Mit dieser Reaktionsklassifizierung haben wir eine Untermenge von etwa 26 000 Verknüpfungsreaktionen aus REACCS und SYNLIB erzeugt. Wir haben somit eine leistungsfähige Bibliothek von Literaturbeispielen für die ähnlichsten Strukturen zu jeder von SYNGEN generierten Reaktion; darüber hinaus versorgt das System den Chemiker mit Laborvorschriften, die er benötigt, um einen Syntheseweg in die Praxis umzusetzen.

Bei der Bearbeitung des Themas „effiziente Synthesen“ haben wir außerdem zwei Unterprogramme geschrieben, die nach speziellen Möglichkeiten für Mehrfachverknüpfungen in Zielmolekülen suchen, um ihre Synthese noch weiter abzukürzen^[15]. Des weiteren haben wir ein Programm aufgebaut, das *vorwärts* statt retrosynthetisch vorgeht, und zwar von den verfügbaren Ausgangsverbindungen zu nahe verwandten Derivaten des Zielmoleküls. Auf diese Weise sollen Wege gefunden werden, die einige Refunktionalisierungsreaktionen oder nur Verknüpfungsreaktionen des Gerüsts enthalten und so die Vielfalt der gefundenen Wege vergrößern.

Dieses FORWARD-Programm beginnt wie SYNGEN damit, zunächst die besten konvergenten Sätze von strategischen Bindungen und die daraus abgeleiteten Startgerüste zu definieren. Es kombiniert dann paarweise alle entsprechenden Ausgangsverbindungen aus dem Katalog und akzeptiert nur solche, für die die funktionellen Gruppen zum Zielmolekül hin konvergieren. Dies wird durch die Berechnung der „Reaktionsdistanz“ von den Zwischenstufen zum Zielmolekül möglich (Abb. 4); was sonst zu einer nicht mehr handhabbaren, kombinatorischen Explosion führen würde, wird radikal beschnitten. Dieses FORWARD-Programm produziert nicht nur das Zielmolekül, sondern alle nahe verwandten Strukturen, die leicht in das Zielmolekül überführt werden können, und liefert so eine viel größere Auswahl an kurzen, praktikablen Synthesewegen. Deshalb gilt die Synthese in Abbildung 7 nicht für Östron selbst (für das SYNGEN nur wenige und wenig praktikable Wege findet), sondern für ein Derivat. Im FORWARD-Programm werden alle Wege zu vielen nahe verwandten Östronderivaten gefunden, wenn Östron eingegeben wird, einschließlich des Weges von

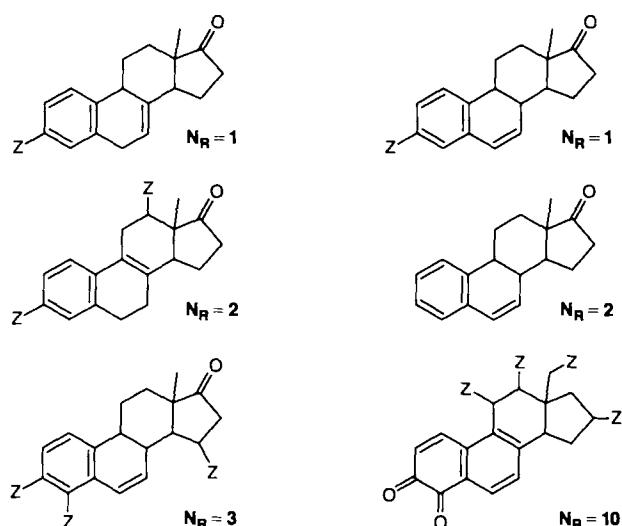


Abb. 12. Östronderivate aus dem FORWARD-Programm. N_R = Reaktionsdistanz (vgl. Abb. 4).

Abbildung 7. Andere Strukturen sind in Abbildung 12 gezeigt. Das Programm schließt allerdings solche aus, die mehr als zwei Stufen vom Zielmolekül entfernt sind.

Diese Programme basieren auf einer sich entwickelnden Strategie für die Syntheseplanung. Um Planungsstrategien zu erkennen, die in publizierten Synthesen enthalten sind, bauen wir derzeit eine Datenbank von publizierten Syntheserouten auf, speziell für wichtige Zielmoleküle, die auf vielen verschiedenen Wegen synthetisiert wurden. Diese Daten können nun im Computer verarbeitet werden, um numerische Größen für Komplexität und Ähnlichkeit zu erstellen, die einer multivarianten Analyse unterworfen werden. So könnte man feststellen, ob ein roter Faden allgemeiner Logik aus diesen Wegen extrahiert und ein Maß für die „Effizienz“ einer Synthese definiert werden kann.

7. Zukunftsperspektiven

Betrachten wir nun die zukünftigen Aussichten für das Verhältnis von Synthese und Computer, so können wir auf die in Abschnitt 1 genannten vier Gebiete zur Zusammenarbeit verweisen. Auf dem Gebiet der komplizierten Berechnungen ist das wichtigste Ziel die Echtzeit-Berechnung von Reaktionsgeschwindigkeiten. Bei der Generierung von Reaktionen benutzt SYNGEN verallgemeinerte abstrakte Beschreibungen. Diese müssen für die wirkliche Vorhersage im speziellen Fall reduziert werden; so muß der Einfluß von Lösungsmitteln, Basen, Gegenionen usw. auf die Reaktionsgeschwindigkeit festgelegt und mit dem bei möglichen, konkurrierenden Nebenreaktionen verglichen werden. Auf qualitativem mechanistischem Niveau dürfte sich die Vorhersage möglicher Nebenreaktionen mit dem CAMEO-Programm von Jorgensen bereits bald realisieren lassen^[16]. Aussagekräftige Berechnungen von kinetischen Daten setzen schnellere Computer voraus.

Auf diesem Gebiet können wir, wenn wir das historische Wachstum der Computerleistungen betrachten, auf einige erfolgversprechende Entwicklungen hoffen. In Abbildung 13 sehen wir die im Laufe der Zeit beachtlich gestiegene Re-

chengeschwindigkeit, gemessen in Fließkomma-Operationen pro Sekunde (FLOPS, Floating Point Operations per Second). Aufgrund dieser Daten kann man davon ausgehen, daß die gewünschten Berechnungen kinetischer Reaktionsparameter in naher Zukunft sicher in Echtzeit möglich sein werden. Darüber hinaus zeigt Abbildung 13, daß die Kosten solcher Rechnerleistungen auf derselben logarithmischen Skala fallen (MIPS = Million Instructions per Second) und dem „Normalchemiker“ zugänglich werden.

Aber Chemiker benutzen nicht nur Computer, sondern sie sind auch an ihrer Entwicklung beteiligt. Eine Möglichkeit, einen Computer schneller zu machen, besteht darin, ihn zu verkleinern. Die Größe der Transistoren hat sich im selben unglaublichen Maß verringert, wie die Schaltgeschwindigkeit gestiegen ist. Zur Zeit ist der kleinste kommerzielle Transistor ungefähr ein Mikron groß; die derzeitige Technologie hat wegen des Tunneleffekts ihre physikalische Obergrenze bei 0.1 µm. Wrighton, Whitesides et al. planen die Herstellung eines molekularen Schalters^[17]. Die Moleküle sollen kovalent an aufgedampfte metallische Leiterbahnen gebunden werden. Dadurch kann der Abstand zwischen ihnen in der Größenordnung von 0.01 µm liegen. Ein noch größeres Potential eröffnet sich, wenn man bedenkt, daß ein solches „Schaltmolekül“ mehr als zwei zugängliche Molekülzustände haben kann und so eine kompliziertere Schaltalgebra ausführen kann als die einfache „an/aus“-Logik.

Im Zeitalter der Informationsspeicherung versuchen wir verzweifelt, den immer stärker zunehmenden Verbrauch an Papier und Platz für Zeitschriften und Bücher in Bibliotheken einzudämmen. Wir benutzen den Computer bereits zur Suche nach Literaturstellen durch Eingabe von Autor, Stichwörtern oder Strukturen, aber wir benötigen immer noch die Bibliothek mit gedruckten Werken, um die Primärliteratur zu lesen. Was wir uns wünschen, wäre, von unserem Büroterminal eine Literaturstelle anzuwählen, sie auf dem Bildschirm zu lesen und dann auch alle folgenden zitierten Artikel aufzurufen, um einen schnellen Überblick über das interessierende Gebiet zu erhalten.

Dieses Quellenmaterial kann sowohl abgespeichert als auch wieder recherchiert werden, solange die Kapazität des Speichermediums ausreicht. Wie das Wachstum der Computerleistung (Abb. 13) schreitet auch die Entwicklung der Speicherkapazität für digitale Information voran; sie verdoppelte sich im letzten Jahrzehnt etwa jährlich. Eine optische Platte kann 2×10^{10} Bits speichern. Unsere Bibliothek an der Brandeis University enthält 1000 Meter an chemischen und biologischen Zeitschriften der letzten 15 Jahre. Wäre alles Text, hätten wir 10^{12} Bits an Information. Der jährliche Zuwachs an Zeitschriften würde dann auf vier gängigen Platten Platz finden; bei einer zwanzigfachen Steigerung der Speicherkapazität (vielleicht innerhalb von fünf Jahren) würde eine Platte die gesamte Bibliothek aufnehmen! Es wurde sogar bereits eine noch kleinere Speichereinheit untersucht^[18]. Es handelt sich um eine dreidimensionale molekulare Matrix zum optischen Lesen und Schreiben, die theoretisch imstande ist, mehr als 10^{10} Bits pro Kubikzentimeter zu speichern.

Das Problem, Abbildungen zusammen mit einem Text zu speichern, kann durch ein Kompressionsprogramm gelöst werden, das umgekehrt zum Lesen dekomprimiert. Zur Zeit kann eine Stunde Fernsehbilder (216 000 Bilder) auf eine einzige CD komprimiert werden. In Wirklichkeit haben che-

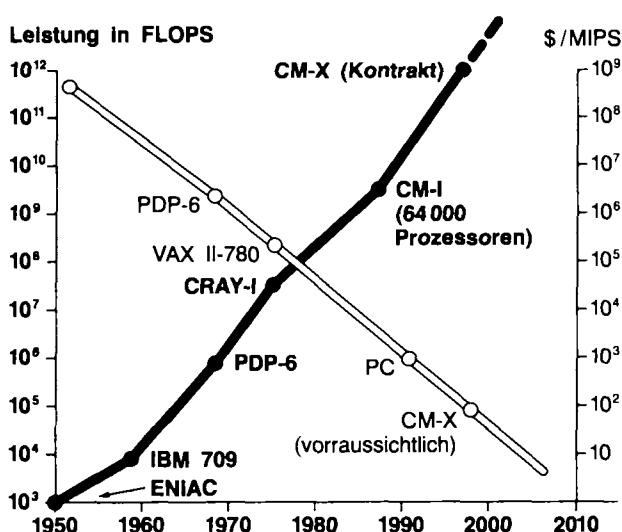


Abb. 13. Zunahme der Geschwindigkeit und Abnahme der Kosten von Computern seit 1950. FLOPS = Fließkomma-Operationen pro Sekunde; MIPS = 1×10^6 Rechengänge pro Sekunde.

mische Strukturen eine viel geringere Informationsdichte als Text. So benötigt unser Programm zur Speicherung von Strukturen (Abb. 3) nur 400 Bits, um die AZT-Struktur zu speichern, während dieselbe Fläche an gedrucktem Text in einer Zeitschrift 3200 Bits beansprucht. Demnach können wir uns ein System (Abb. 14) vorstellen, in dem der Inhalt von Zeitschriften vom Verlag zu einem Satelliten übertragen wird, der ihn dann kontinuierlich zu den Bibliotheken sendet, welche die Information auf einer Platte abspeichern (dasselbe System kann auch Kopien an einzelne Abonnenten senden). Mit einem solchen System könnte man von einem Büroterminal aus jede beliebige Zeitschrift aufrufen, sie lesen und eine Kopie davon machen, und man brauchte kein Papier zu sammeln außer dem, das wirklich gelesen wird.

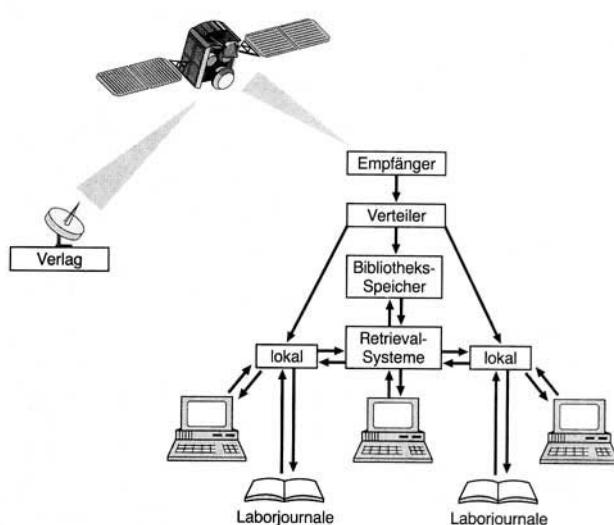


Abb. 14. Die computergesteuerte Bibliothek.

Das Problem des Wiederfindens einer Information (Retrieval) schneidet die Frage nach dem System der Dokumentation für eine solche Suche an. Während wir sinnvolle Algorithmen zur Suche nach Autoren, Literaturstellen, Stichwörtern und Strukturen haben, brauchen wir noch ein gutes System für die Suche nach Reaktionen in Reaktionsdatenbanken. Es sollte hierarchisch aufgebaut sein, wie das Beilstein-System für Strukturen, vom allgemeinen zum speziellen Fall, um jeder denkbaren Reaktion einen Platz zu geben, gleichgültig, ob sie bereits bekannt ist oder nicht. Unser Ansatz (Abb. 4) ausgehend von Einheitsreaktionen ist dementsprechend angelegt; wir prüfen zur Zeit mit beträchtlichem Erfolg Reaktionsdatenbanken, um sowohl Refunktionalisierungen als auch Verknüpfungsreaktionen zu suchen. Die Chemiker werden in Zukunft ein lebhafteres Interesse an der Schaffung solcher Dokumentationssysteme entwickeln müssen, wenn sie wirklich effiziente Methoden zur Suche nach Informationen zur Verfügung haben wollen.

Das dritte Gebiet der künstlichen Intelligenz kann sowohl auf die Synthese als auch auf die Analyse angewendet werden. In der Syntheseplanung kann das Programm SYNGEN alle kürzesten, verallgemeinerten Wege angeben und in Datenbanken die Literaturbeispiele für Reaktionen heraussuchen, welche die größte Übereinstimmung mit den generierten Reaktionen zeigen. Es benötigt allerdings noch eine

stereochemische Komponente. Im Zusammenspiel mit dem Programm CAMEO würden potentielle Nebenreaktionen entdeckt; eine Kopplung mit einem unabhängigen Programm zur Berechnung von Reaktionsgeschwindigkeiten könnte diese quantifizieren. In der Analytik benötigt der Chemiker ein übergeordnetes „black box“-Programm, um die Daten von den verschiedenen analytischen Geräten zu sammeln und sie im Sinne von Strukturen der Produkte zu interpretieren.

Schließlich kann der Computer dazu programmiert werden, selbst Experimente durchzuführen. Unter Verwendung eines computergesteuerten Zymark-Roboters demonstrierten Fuchs et al.^[19] ein integriertes System zur Durchführung einer Reihe von Parallelexperimenten, die dazu dienen sollten, die Bedingungen einer Reaktion im Maßstab von 0.1 mmol pro mL zu optimieren. Das System steuert einen Arm, der so programmiert ist, daß er durch vertikale und radiale Bewegungen Substrate und Lösungsmittel in einen Reaktor spritzt (und ihn anschließend wäscht!). Nach vorgegebener Zeit und Temperatur wird abfiltriert, extrahiert, chromatographisch aufgearbeitet, und es werden analytische Spektren vom Eluat aufgenommen. Mit acht Reaktoren und zwei Gaschromatographen kann das System so programmiert werden, daß es 400 Reaktionen am Tag ausführt. Am Beispiel einer komplizierten Reaktion ließ sich die Ausbeute von anfänglich nahezu Null auf über 70% steigern.

Diese rosigen (wenn nicht gefährlichen) Voraussagen für die Zukunft scheinen nahezulegen, daß die Chemiker selbst nur noch wenig mehr zu tun haben werden, als sich an ihre Terminals zu setzen! Obwohl dies sicher nicht im geringsten wahrscheinlich ist, wird klar, daß sie wesentlich vertrauter mit Computern werden müssen, als sie es jetzt sind, und daß sie die Computer intensiver nutzen müssen. Darüber hinaus werden sie der Bedeutung und dem Aufbau logischer Systeme zur Beschreibung und Organisation chemischer Information mehr Beachtung zollen müssen, und sie müssen aktiv an der Entwicklung klarer, nützlicher Software für chemische Zwecke mitarbeiten. Kurz gesagt müssen sie sich mehr mit Computern beschäftigen, als sie es jetzt tun.

Zum Schluß möchte ich die wundervolle Atmosphäre, den hingebungsvollen Eifer und die bemerkenswerten Computerkenntnisse meiner Kollegen anerkennen, welche die hier beschriebenen Arbeiten ermöglichten: Glenn Toczko, der mit mir seit den Anfängen von SYNGEN vor zehn Jahren zusammenarbeitet, David Grier, Zmira Bernstein, Ping Huang, Todd Miller und Camden Parks.

Eingegangen am 5. Juni 1990 [A 784]

- [1] J. B. Hendrickson, *J. Am. Chem. Soc.* 83 (1961) 4537.
- [2] E. J. Corey, W. T. Wipke, *Science (Washington, D.C.)* 166 (1969) 178.
- [3] E. J. Corey, A. K. Long, S. D. Rubenstein, *Science (Washington, D.C.)* 228 (1985) 408.
- [4] J. B. Hendrickson, *Acc. Chem. Res.* 19 (1986) 274.
- [5] J. B. Hendrickson, *J. Am. Chem. Soc.* 99 (1977) 5439.
- [6] Die so erhaltenen konvergenten Sätze von strategischen Bindungen lösen nicht mehr als zwei Bindungen pro Schnitt, d. h. sechs Bindungen ($\lambda = 6$) oder weniger. Für Östron ($b = 21$) gibt es insgesamt 39 Millionen geordnete Sätze von strategischen Bindungen; die Anzahl der völlig konvergenten Sätze, ausgehend von den Gerüsten kommerziell verfügbarer Verbindun-

- gen, ist aber kleiner als 900, und für die meisten von diesen wiederum gibt es keine akzeptablen Reaktionen für ideale Synthesen.
- [7] J. B. Hendrickson, A. G. Toczko, *J. Chem. Inf. Comput. Sci.* 23 (1983) 171.
 - [8] J. B. Hendrickson, *J. Am. Chem. Soc.* 93 (1971) 6847.
 - [9] Primäre Kohlenstoffatome haben $\sigma = 1$, sekundäre $\sigma = 2$, tertiäre $\sigma = 3$ und quartäre $\sigma = 4$.
 - [10] J. B. Hendrickson, *J. Chem. Inf. Comput. Sci.* 19 (1979) 129.
 - [11] Aus Gründen der Übersichtlichkeit sind in Abb. 5 keine Fragmentierungen gezeigt; sie können allerdings leicht als inverse Aufbaureaktionen eingefügt werden.
 - [12] J. B. Hendrickson, Z. Bernstein, T. M. Miller, A. G. Toczko in B. A. Hohne, T. H. Pierce (Hrsg.): „Expert System Applications in Chemistry,“ *ACS Symp. Ser. 408* (1989), Kap. 6.
 - [13] Für jede Kategorie und Ebene sind zwei Zahlen angegeben; bei der ersten Zahl sind alle äquivalenten Reaktionen ausgeschieden.
 - [14] J. B. Hendrickson, A. G. Toczko, *J. Chem. Inf. Comput. Sci.* 29 (1989) 137.
 - [15] J. B. Hendrickson, P. Huang, *J. Chem. Inf. Comput. Sci.* 29 (1989) 145.
 - [16] T. D. Salatin, W. L. Jorgensen, *J. Org. Chem.* 45 (1980) 2043; G. D. Paderes, W. L. Jorgensen, *ibid.* 54 (1989) 2058.
 - [17] P. E. Laibinis, J. J. Hickmann, M. S. Wrighton, G. M. Whitesides, *Science (Washington, D.C.)* 245 (1989) 845.
 - [18] D. A. Parthenopoulos, P. M. Rentzepis, *Science (Washington, D.C.)* 245 (1989) 843.
 - [19] A. R. Frisbee, M. H. Nantz, G. W. Kramer, P. L. Fuchs, *J. Am. Chem. Soc.* 106 (1984) 7143.
-