

DAS SYNTHESEPLANUNGSPROGRAMM E R O S

J.Gasteiger, C.Jochum, M.Marsili und J.Thoma

Organisch-Chemisches Institut, Technische Universität München,
8046 Garching, W.Germany

Zusammenfassung

Es wird über ein Computerprogramm berichtet, das ganze Folgen chemischer Reaktionen erzeugt. Reaktionen werden als Bindungs- und Elektronen-umordnungsschemata aufgefaßt und dementsprechend aufgrund eines kombinatorischen Prozesses erzeugt. Dadurch werden sämtliche denkbaren Reaktionsmöglichkeiten erhalten, gleichgültig ob die Reaktion bereits bekannt ist oder ob sie völlig neuartig ist. Um aus den mathematisch möglichen Reaktionen die chemisch tatsächlich ablaufenden auszuwählen werden verschiedene Bewertungen vorgenommen. Hier soll etwas näher auf die Berechnung der Reaktionsenthalpie eingegangen werden. Verschiedene Beispiele demonstrieren den Einsatz des Systems, sowohl zum Studium der Folgeprodukte vorgegebener Verbindungen, als auch zur Planung der Synthese organischer Verbindungen.

Einleitung

Aufgaben, von denen früher angenommen wurde, daß sie nur mit Hilfe menschlicher Intelligenz zu lösen sind, werden immer mehr auch an Computer delegiert. Artificial Intelligence hat sich zu einem wichtigen Forschungsgebiet entwickelt. Wir erstellen Computerprogramme, die Methoden der Artificial Intelligence zur Lösung chemischer Probleme einsetzen. Von den verschiedenen Anwendungsgebieten wurde vorrangig die Planung der Synthese organischer Verbindungen gewählt. Dies sowohl wegen der enormen Bedeutung der Synthese für die Arbeit eines Chemikers, als auch wegen der Herausforderung, die durch die Komplexität der Problemstellung gegeben ist.

geführt. Die Bindungen sind in einer Dreizeilenmatrix enthalten. Die ersten beiden Zeilen geben die Nummern der aneinander gebundenen Atome an, in der dritten Zeile steht die entsprechende Bindungsordnung. So bedeutet demnach die erste Spalte, daß Atom 1 an Atom 3 mit einer Doppelbindung gebunden ist. Diese Darstellung besitzt die gleichen algebraischen Eigenschaften wie eine BE-Matrix², insbesondere lassen sich Valenzkonfigurationen leicht abprüfen.

Ein Großteil der Vorträge dieser Tagung ist graphentheoretischen Problemen gewidmet. Auch wir mußten bei der Entwicklung unseres Programmsystems Aufgaben lösen, die durch die topologische Darstellung von Molekülen gegeben sind. So mußte eine eindeutige Repräsentation eines Moleküls definiert werden. Dies gelang durch eindeutige Bezifferung der Atome eines Moleküls³. Diese Bezifferung kann auch als Grundlage für die Behandlung der Stereochemie von Molekülen und Reaktionen dienen⁴. Desweiteren wurde ein Programm erstellt, das einen eindeutigen Satz an Ringen automatisch aus einer topologischen Darstellung heraus ermittelt: Der kleinste Satz kleinster Ringe⁵.

Auf diese Problemlösungen soll aber hier in der Kürze der Zeit nicht weiter eingegangen werden. Die Repräsentation von Molekülen ist nur eines der Probleme bei der Behandlung chemischer Information mittels Computer. Genauso wichtig ist die Lösung der Aufgabe chemische Reaktionen zu repräsentieren. Hierzu wählten wir einen formalen Ansatz: Reaktionen werden als Bindungs- und Elektronen-verschiebungsschemata behandelt. Die den R-Matrizen entsprechenden Reaktionsgeneratoren lassen sich in verschiedene Kategorien einteilen. In Abb. 2 sind Typen enthalten, die eine Bindung brechen und bei denen nur gepaarte Elektronen auftreten.

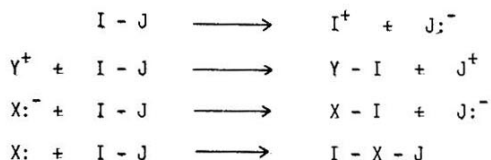
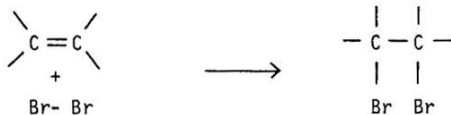


Abb. 2: Reaktionskategorien, die eine Bindung brechen

Es befindet sich darunter der heterolytische Bindungsbruch, der Prototyp einer elektrophilen und einer nucleophilen Substitution, sowie eine Insertionsreaktion. Die verschiedenen Reaktionskategorien haben ganz unter-

schiedliche Bedeutung. Während die in Abb. 2 aufgezeigten Kategorien hauptsächlich mechanistische Schritte repräsentieren, gibt es andere, welche wichtige synthetische Reaktionen umfassen. In Abb. 3 ist eine Reaktionskategorie dargestellt, die einem Großteil aller synthetischen Reaktionen zugrunde liegt.



	I	J	K	L
I		-1	+1	
J	-1			+1
K	+1			-1
L		+1	-1	

Abb. 3: Reaktionskategorie, welche zwei Bindungen bricht und zwei Bindungen knüpft

Reaktionen, welche diesem Schema gehorchen sind Additionen an Mehrfachbindungen, Eliminierungen und Substitutionsreaktionen. In Abb. 3 unten sind die von Null verschiedenen Elemente der entsprechenden R-Matrix aufgeführt. Die diesen Eintragungen zugeordneten Bindungsbrechungen und -knüpfungen werden nun direkt auf der Bindungsliste der reagierenden Stoffe ausgelöst. Mit Abb. 4 soll die Anwendung obiger Reaktionskategorie erläutert werden. Links oben in Abb. 4 wird ein Ausschnitt einer Bindungsliste gegeben, welche die Partialstruktur des Eduktes repräsentiert. Die Bindungsbrechungen werden direkt auf der Bindungsliste ausgeführt, indem die entsprechenden

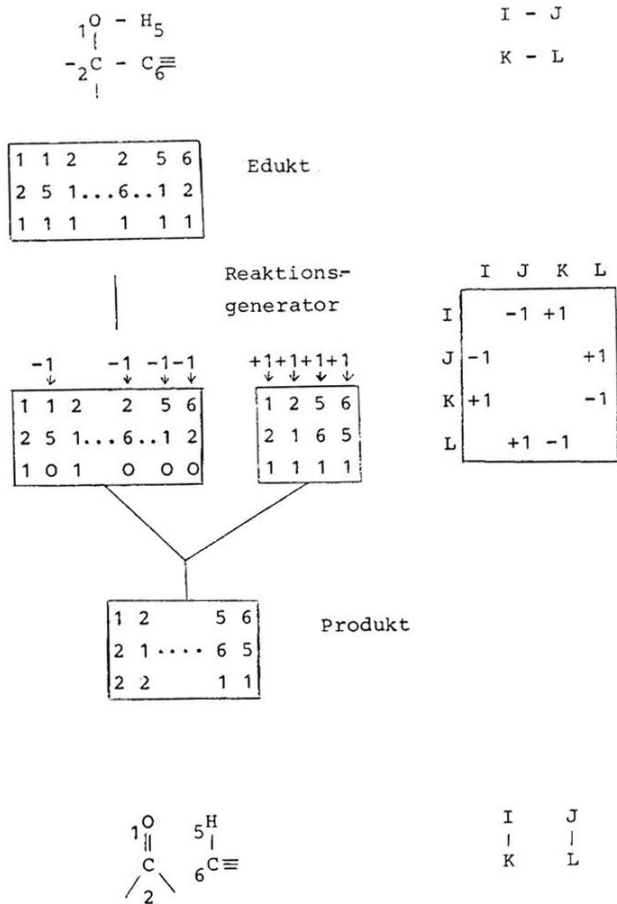


Abb. 4: Erzeugung von Reaktionen

Bindungsordnungen um eine Einheit herabgesetzt werden. Wenn vorher eine Einfachbindung bestand, so erscheint dann eine Bindungsordnung von Null. Die neu gemachten Bindungen werden in einer separaten Matrix geführt. Kombination der veränderten Bindungsliste mit der Matrix der geknüpften Bindungen, wobei Bindungen mit der Ordnung Null eliminiert werden, führt bereits zur Bindungsliste des Reaktionsproduktes, hier also zu der Doppel-

bindung zwischen Atom 1 und 2 und der Einfachbindung zwischen den Atomen 5 und 6. Es ist nun wichtig festzuhalten, daß das Reaktionsprodukt allein durch Anwendung der formalen Bindungsumordnungen erhalten wurde. Wir verwendeten keinerlei Information über die Bekanntheit oder Durchführbarkeit der erzeugten Reaktion. Aufgrund dieses Ansatzes erhält man demnach alle denkbaren Reaktionen. Darin können auch völlig neuartige, präzedenzlose Reaktionen vorkommen; sie werden damit sozusagen zur Entdeckung vorgeschlagen.

Kurz seien einige Worte zur Kommunikation zwischen Mensch und Maschine gesagt. Moleküle werden in Form von Strukturformeln auf dem Zeilendrucker ausgegeben. Da die internen Bindungslisten als topologische Darstellungen jeglicher geometrischer Information entbehren, muß das zweidimensionale Bild von einem eigenen Programmsystem erzeugt werden. In den Strukturformeln symbolisieren Sternchen Einfachbindungen, Pluszeichen Doppelbindungen (siehe Abb. 5).

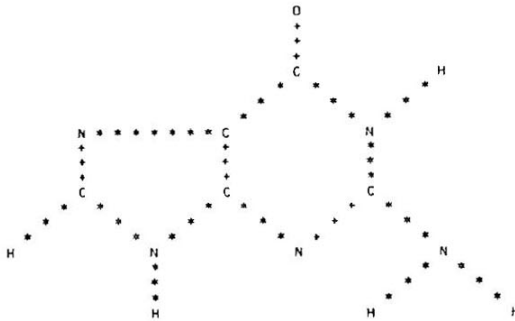


Abb. 5: Strukturformel von Guanin

Die Güte der Strukturformeln ist begrenzt, da auf dem Zeilendrucker nur ein Punkteraster zur Verfügung steht und weiterhin die Zeichenkette nur einen recht beschränkten Zeichenvorrat hat. Wir wählten aber den Zei-

lendrucker zunächst als schnelles Ausgabemedium, da damit keine speziellen hardware-Anforderungen gestellt werden mußten, denn ein Zeilendrucker ist an jeder Rechnerkonfiguration vorhanden.

Seit einiger Zeit haben wir einen graphischen Bildschirm zur Verfügung und entwickeln darauf ein Programm, das natürlich ästhetisch viel befriedigendere Bilder erzeugen kann. Einige Beispiele dieser so erhaltenen Strukturen zeigen Abb. 6 und 7. Es wird ein dreidimensionales Modell eines Moleküls aufgebaut, das dann auf einen Bildschirm projiziert wird. Mit Hilfe eines Lichtgriffels kann das Modell vergrößert oder verkleinert und in allen drei Raumrichtungen rotiert werden.

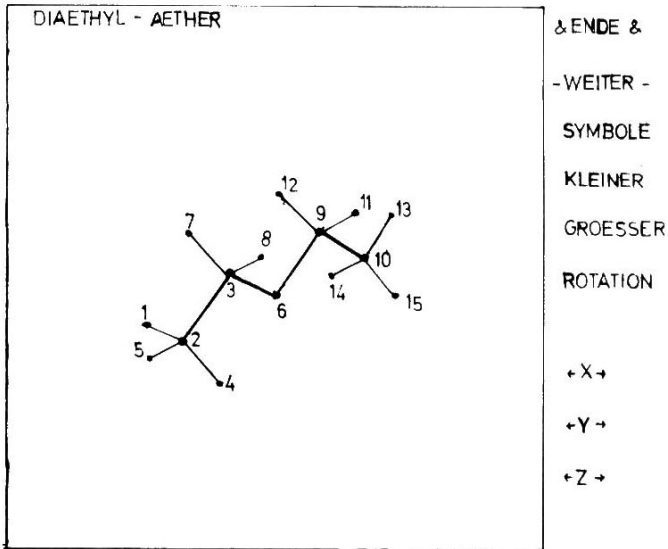


Abb. 6: Molekülmodell von Diäthyläther

Man muß sich dabei vor Augen halten, daß die dreidimensionalen Bilder wiederum nur aus der Bindungsliste heraus erzeugt wurden; die dreidimensionale Struktur muß also vom Programm selbst aufgebaut werden. Dies wird durch Arbeiten mit monozentrischen Modellen erreicht, aus denen durch geeignetes Aneinandersetzen sowohl Ketten- als auch Ringstrukturen erhalten

werden.

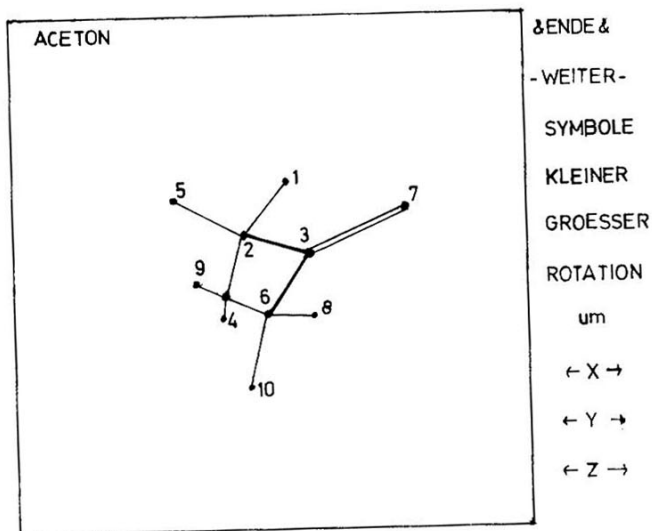
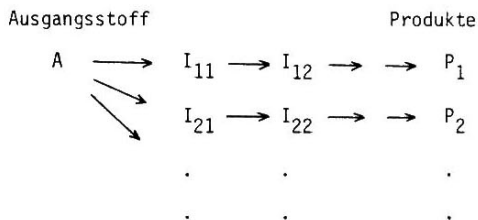


Abb. 7: Molekülmodell von Aceton

Bewertung und Auswahl

Mit EROS liegt zunächst nur ein Programmsystem vor, das chemische Reaktionen erzeugt und damit eingegebene Verbindungen in andere umwandelt. Werden die so erhaltenen Moleküle wiederum den Reaktionsgeneratoren unterworfen, so erhält man ganze Folgen von Reaktionen. Es sei nun betont, daß EROS zwei grundsätzlich verschiedene Aufgabenstellungen bewältigen kann. Diese unterscheiden sich darin wie die erzeugten Reaktionen aufgefaßt werden. Wird angenommen, daß die Reaktionen in der Abfolgerichtung, in der sie vom Programm erhalten werden, auch ablaufen, so macht man sich eine Vorwärtsstrategie zu eigen. Man kann dann mit EROS z.B. Probleme wie die Suche nach der Verwendbarkeit eines industriellen Abfallproduktes oder die zu erwartenden Folgeprodukte von Chemikalien in der Umwelt studieren.

Vorwärtsrichtung



Retrosynthese

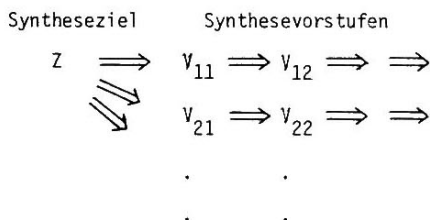


Abb. 8: Grundstrategien für den Einsatz von EROS

Die vom Programm erzeugten Reaktionen können aber auch als retrosynthetische Schritte aufgefaßt werden; die Reaktionen selbst laufen dann umgekehrt zur Richtung der Erzeugung ab. Diese Auffassung macht man sich in der Syntheseplanung zu eigen. Vom Syntheseeziel kommt man über Synthesezwischenstufen schließlich zu leicht erhältlichen Ausgangsstoffen.

Welche Problemart man nun mit EROS studiert hängt von den Spezifikationen zu Beginn des Programmlaufes und vom unterschiedlichen Einsatz der Bewertungskriterien ab.

Wie schon erwähnt, können die zunächst durch Anwendung der Reaktionsgeneratoren erzeugten Moleküle erneut in diesen Prozess eingeschleust werden. Man erhält dann eine ausfächernde Abfolge von Reaktionen und Molekülen, wofür in der Syntheseplanung der Ausdruck Synthesebaum geprägt wurde (Abb. 9). Der absteigende Übergang von Niveau zu Niveau entspricht einer Retroreaktion. Eine solche Retroreaktion kann dabei auch zu zwei Verbindungen führen, d.h. zur Synthese der Zielverbindung sind dann zwei

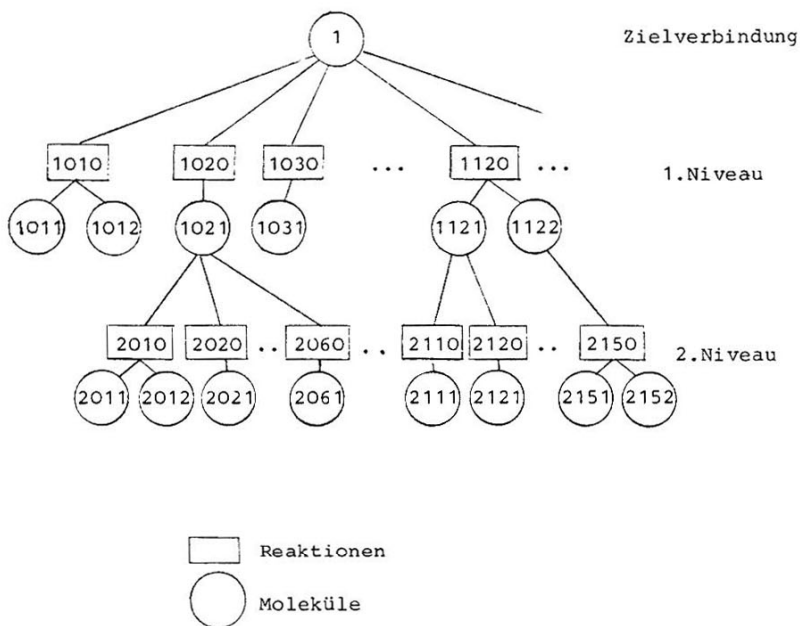


Abb. 9: Synthesebaum

Moleküle nötig. Damit sind im Synthesebaum auch Querverweise von Bedeutung, denn die Gesamtsynthese hat nur dann Erfolg, wenn auch beide Verbindungen ihrerseits erzeugt werden können. Beide müssen also ebenfalls den Reaktionsgeneratoren unterworfen werden. Eine vollständige Synthese umfaßt dann sich verzweigende Pfade, wobei die Endpunkte zugängliche Verbindungen sein müssen. Die Erzeugung und Abarbeitung des Synthesebaumes mit all seinen strategischen Beziehungen wird in EROS vollautomatisch über eine beliebige Anzahl an Niveaus hinweg vorgenommen, wobei komplexe Entscheidungen zu fällen sind.

In Abb. 9 werden von den Vorstufen des ersten Niveaus nur einige erneut den Reaktionsgeneratoren unterworfen. Zu diesem Zwecke ist vorher eine Bewertung und Auswahl durchgeführt worden. Der Auswahl kommt in unserem Ansatz zentrale Bedeutung zu, da wir prinzipiell alle denkbaren Reaktionen erzeugen können. Aus dieser großen Zahl mathematisch möglicher

Reaktionen sind die chemisch sinnvollen zu ermitteln. Von den verschiedenen zu diesem Zwecke vorgenommenen Bewertungen soll hier nur auf die Berechnung der Reaktionsenthalpie näher eingegangen werden. Das Verfahren⁶ zur Berechnung der Reaktionsenthalpie stützt sich auf thermochemische Parameter, die analog dem Schema von Allen⁷ zur Abschätzung von Atomisierungswärmen zusammengesetzt werden.

Beispiel:	Alkane	4 Parameter	
		B (CC)	78.64 (alle Werte in kcal/mol)
		B (CH)	99.30
		G (CCC)	2.58
		D (CCC)	-0.55

n-Butan

$ \begin{array}{cccc} \text{H} & \text{H} & \text{H} & \text{H} \\ & & & \\ \text{H}-\text{C}-\text{C}-\text{C}-\text{C}-\text{H} \\ & & & \\ \text{H} & \text{H} & \text{H} & \text{H} \end{array} $	10	B (CH)	993.0	
	3	B (CC)	236.52	
	2	G (CCC)	5.16	
	ber.	1234.68	exp.	1234.96
		Fehler		-0.28

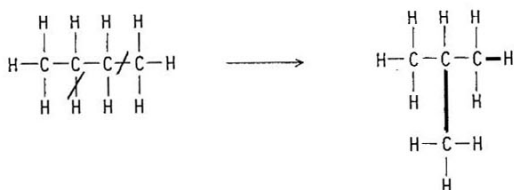
i-Butan

$ \begin{array}{cccc} \text{H} & \text{H} & \text{H} & \\ & & & \\ \text{H}-\text{C}-\text{C}-\text{C}-\text{H} \\ & & & \\ \text{H} & & \text{H} & \\ & & & \\ \text{H}-\text{C}-\text{H} \\ \\ \text{H} \end{array} $	10	B (CH)	993.0	
	3	B (CC)	236.52	
	3	G (CCC)	7.74	
	1	D (CCC)	-0.55	
	ber.	1236.71	exp.	1237.01
Differenz	ber.	2.03	exp.	2.05

Abb. 10: Allen-Schema zur Berechnung von Atomisierungswärmen

Für Alkane benötigt man dabei vier Parameter, je einen Parameter für eine CC- und CH- Bindung, einen weiteren Parameter, den G(CCC) Parameter, für die Anordnung von drei aneinander gebundenen C-Atomen, sowie den D(CCC) Parameter, der einzusetzen ist, wenn drei C-Atome an ein zentrales C-Atom gebunden sind (siehe Abb. 10). Aufaddieren der einzelnen Parameterwerte ergibt dann die Atomisierungswärme der entsprechenden Verbindung. Die Fehler sind im allgemeinen recht gering; bei dem in Abb. 10 angegebenen Beispiel des n- und i-Butans jeweils etwa 0.30 kcal/mol. Der Unterschied in den Bildungswärmen der beiden isomeren Verbindungen wird damit sogar exakt wiedergegeben.

Wie kann nun dieses Schema zur Berechnung von Reaktionsenthalpien herangezogen werden? Wir haben hierzu ein Verfahren entwickelt, bei dem nicht die gesamten Atomisierungswärmen von Ausgangsstoffen und Reaktionsprodukten ermittelt werden müssen. Vielmehr wird die Enthalpie des Reaktionsprozesses direkt abgeschätzt⁶. Dies sei am Beispiel der Isomerisierungsreaktion von n-Butan zu i-Butan erläutert (Abb. 11).



$$\begin{aligned}
 \Delta H_{\text{rea}} = \Delta \Delta H_{\text{a}} &= 1 \text{ B (CH)} + 1 \text{ G (CCC)} + 1 \text{ B (CC)} + 1 \text{ G (CCC)} \\
 &\quad - [1 \text{ B (CH)} + 1 \text{ B (CC)} + 3 \text{ G (CCC)} + 1 \text{ D (CCC)}] = \\
 &= -1 \text{ G (CCC)} - 1 \text{ D (CCC)} \\
 &= -2.58 \quad + \quad 0.55 \quad = \quad -2.03 \text{ kcal/mol} \\
 &\qquad \qquad \qquad \text{exp.} \quad -2.05 \text{ kcal/mol}
 \end{aligned}$$

Abb. 11: Berechnung der Reaktionsenthalpie

An den gebrochenen Bindungen findet man Substrukturen, die Zugang zu den jeweiligen Parametern liefern; an den geknüpften Bindungen werden entsprechende Parameter erhalten. Die Reaktionsenthalpie ergibt sich als

Differenz der beiden Sätze an Parametern. Die fast vollständige Übereinstimmung zwischen berechnetem und experimentellem Wert ist hier etwas glücklich, aber der Fehler liegt generell meist unter 0.4 kcal/mol. Wichtig ist, daß zur Ermittlung der Reaktionsenthalpie nur die direkte Umgebung des Reaktionsortes abzusuchen ist. Die Rechenzeiten sind damit praktisch unabhängig von der Größe der an der Reaktion beteiligten Moleküle. (Die Rechenzeiten betragen etwa 0.01 - 0.02 sec auf einer AMDAHL 470 V/6) Dies ist ein großer Vorteil gegenüber anderen Verfahren, bei denen die Rechenzeiten mit steigender Molekülgröße meist enorm zunehmen.

In Abb. 12 ist eine ausführlichere Studie zur C_4H_{10} -Potentialfläche enthalten. Es sind eine Reihe von Isomerisierungsprodukte von n-Butan zusammen mit den entsprechenden Werten der Reaktionsenthalpie aufgeführt. Diese überstreichen einen Bereich von 90 kcal/mol; die Abweichungen zwischen berechneten und experimentellen Werten betragen meist weniger als 0.4 kcal/mol und sind damit unter der experimentellen Fehlergrenze. Nur bei den stark gespannten Verbindungen Methylcyclopropen und Bicyclobutan ist ein größerer Unterschied festzustellen. Unser Verfahren ist auch dann anwendbar, wenn keine experimentellen Daten zugänglich sind. Dies ist z.B. bei Methylcyclopropan der Fall, wo wir dennoch eine Berechnung vornehmen können.

Der Einsatz der Reaktionsenthalpien bei der Auswahl unter verschiedenen Reaktionen sei anhand von Abb. 13 erläutert. Darin ist eine Reihe möglicher, vom Programm formulierter Reaktionsprodukte aus n-Butan und Chlor aufgeführt, zusammen mit den entsprechenden Reaktionsenthalpien. Nimmt man thermodynamische Produktkontrolle an, und vernachlässigt Entropie-effekte, so ist diejenige Reaktion zu erwarten, die am stärksten exotherm ist, also den negativsten Enthalpiewert hat. Im vorliegenden Falle würde dann als Reaktionsprodukt 2-Chlorbutan zu erwarten sein, vielleicht noch zusammen mit etwas 1-Chlorbutan. Und dies ist ja auch tatsächlich das beobachtete Ergebnis.

Die Verwendung der Reaktionsenthalpie zur Auswahl unter den Retroreaktionen bei der Synthesepaltung ist nun nicht so einfach wie bei Reaktionen in der Vorwärtsrichtung. Der geeignete Einsatz der Reaktionsenthalpie als Auswahlkriterium sei am Beispiel einer vielstufigen Synthese erläutert (siehe Abb. 14). Nähme man jeweils die exothermste Reaktion, so würde man bei immer stabileren Vorstufen anlangen (Pfad C). Die Überführung dieser Verbindungen zum gewünschten Zielmolekül müßte dann mit so großer Energie-

	Reaktionsprodukte	Reaktionsenthalpie (in kcal/mol)	
		ber.	exp.
$ \begin{array}{cccc} & \text{H} & \text{H} & \text{H} & \text{H} \\ & & & & \\ \text{H} & -\text{C} & -\text{C} & -\text{C} & -\text{C}-\text{H} \\ & & & & \\ & \text{H} & \text{H} & \text{H} & \text{H} \end{array} $	$ \begin{array}{c} \text{CH}_3-\text{CH}-\text{CH}_3 \\ \\ \text{CH}_3 \end{array} $	-2.03	-2.05
	$ \text{CH}_3-\text{CH}_2-\text{CH}=\text{CH}_2 + \text{H}_2 $	+29.92	+30.16
	$ \text{CH}_3-\text{CH}=\text{CH}-\text{CH}_3 + \text{H}_2 $	+27.15	+27.37
	$ \text{CH}_3-\text{CH}=\text{CH}_2 + \text{CH}_4 $	+16.93	+17.35
	$ \text{CH}_3-\text{CH}_3 + \text{CH}_2=\text{CH}_2 $	+22.29	+22.57
	$ \text{CH}_2=\text{CH}_2 + \text{CH}_2=\text{CH}_2 + \text{H}_2 $	+54.98	+55.26
	$ \begin{array}{c} \text{CH}_2-\text{CH}_2 \\ \diagdown \quad / \\ \text{CH}_2 \end{array} + \text{CH}_4 $	+25.03	+25.20
	$ \begin{array}{c} \text{CH}_2-\text{CH}-\text{CH}_3 \\ \diagdown \quad / \\ \text{CH}_2 \end{array} + \text{H}_2 $	+35.96	-----
	$ \begin{array}{c} \text{CH}_2-\text{C}-\text{CH}_3 \\ \diagdown \quad // \\ \text{CH} \end{array} + 2\text{H}_2 $	+87.15	+88.57
	$ \begin{array}{c} \text{CH}_2-\text{CH}_2 \\ \quad \\ \text{CH}_2-\text{CH}_2 \end{array} + \text{H}_2 $	+36.60	+37.14
	$ \begin{array}{c} \text{CH}-\text{CH}_2 \\ \quad \\ \text{CH}-\text{CH}_2 \end{array} + 2 \text{H}_2 $	+67.35	+67.81
	$ \begin{array}{c} \text{CH} \\ / \quad \quad \backslash \\ \text{H}_2\text{C} \quad \text{CH}_2 \\ \backslash \quad \quad / \\ \text{CH} \end{array} + 2 \text{H}_2 $	+85.14	+82.26

Abb. 12: Isomerisierungen von n-Butan

		Reaktionsprodukte	Reaktionsenthalpie (in kcal/mol)	
			ber.	exp.
$ \begin{array}{cccc} \text{H} & \text{H} & \text{H} & \text{H} \\ & & & \\ \text{H}-\text{C}-\text{C}-\text{C}-\text{C}-\text{H} \\ & & & \\ \text{H} & \text{H} & \text{H} & \text{H} \end{array} + \text{Cl}_2 $		$\text{CH}_3-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{Cl} + \text{HCl}$	-27.86	-26.8 (± 2.0)
		$\text{CH}_3-\text{CH}_2-\underset{\text{Cl}}{\text{CH}}-\text{CH}_3 + \text{HCl}$	-31.07	-30.3 (± 2.0)
		$\text{CH}_3-\text{CH}_2-\text{CH}_2-\text{Cl} + \text{CH}_3\text{Cl}$	-21.41	-21.19
		$\text{CH}_3-\text{CH}_2-\text{CH}=\text{CH}_2 + 2 \text{HCl}$	-14.17	-13.96
		$\text{CH}_3-\text{CH}_2-\underset{\text{Cl}}{\text{CH}}-\text{CH}_2-\text{Cl} + \text{H}_2$	-14.83	-----
		$\text{CH}_3-\text{CH}_2-\text{Cl} + \text{CH}_3-\text{CH}_2-\text{Cl}$	-22.04	-23.84
		$\text{CH}_3-\text{CH}_2-\text{Cl} + \text{CH}_2=\text{CH}_2 + \text{HCl}$	-5.58	-5.35
		$\text{CH}_3-\text{CH}_3 + \text{Cl}-\text{CH}_2-\text{CH}_2-\text{Cl}$	-22.04	-20.58

Abb. 13: Reaktion von n-Butan mit Chlor

zufuhr ablaufen, daß sie nicht mehr realisierbar ist. Andererseits darf auch nicht jeweils die endothermste Reaktion ausgewählt werden. Die synthetische Reaktion selbst, die ja in umgekehrter Richtung abläuft, wäre dann zwar energetisch günstig, aber bei der Restrostudie gelangte man dann zu so energiereichen Vorstufen, daß sie synthetisch nicht mehr zugänglich sind (Pfad A). Für optimale Synthesen über mehrere Niveaus scheint vielmehr nur ein energetische Fenster zur Verfügung zu stehen, innerhalb dessen die Reaktionsenthalpie liegen muß. Wir nehmen an, daß dabei die Reaktionsenthalpie im leicht endothermen Bereich zu liegen hat, die synthetische Reaktion selbst also etwas exotherm sein soll (Pfad B).

Diese Betrachtungen seien an einem Beispiel illustriert (siehe Abb. 15). Synthesziel war β -Lactam, ein wichtiger Baustein der Penicillin- und Cephalosporin-Antibiotika. β -Chlorpropionsäureamid ist so stabil, daß es nicht mehr direkt in β -Lactam umgewandelt werden kann. Dagegen ist eine Reihe von

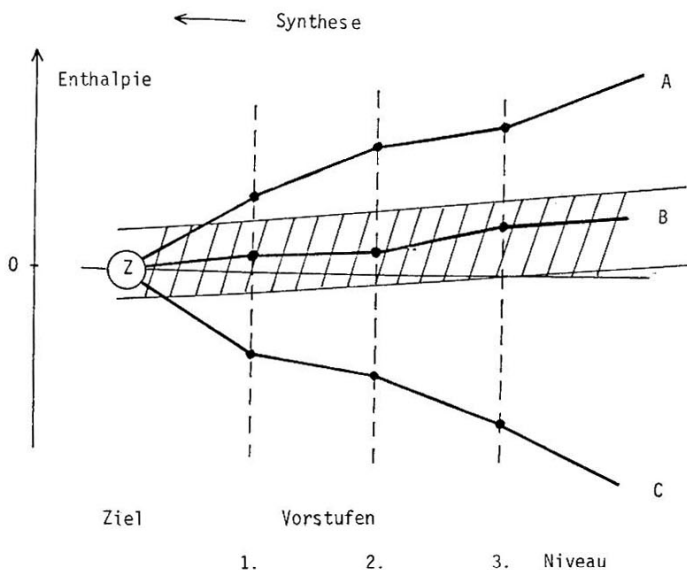


Abb. 14: Energetische Verhältnisse bei Synthesen

Synthesen bekannt, die von Olefinen und Isocyanaten ausgehen. In diesem Falle schätzen wir, im Einklang mit den obigen Betrachtungen, die synthetische Reaktion als leicht exotherm ab. Die Verbindungen in der letzten Zeile sind alle energiereicher als β -Lactam, sollten sich also leicht in das Zielmolekül umwandeln. Verbindungen 5 und 7 sind aber so energiereich, daß sie kaum noch zugänglich erscheinen (vgl. Abb. 14, Pfad A). Verbindungen 6 und 8 sind isomer zu den vorigen und ihr Energieinhalt ist nicht gar so hoch. Damit könnten sie noch zugänglich sein und sollten sich aber dann auch leicht in β -Lactam umwandeln. Und in der Tat wurden 1971 von Wasserman⁸ jeweils von Vertretern des Typs 6 und 8 Synthesen zu β -Lactam gefunden. Diese Reaktionen sind wegen der dabei ablaufenden Gerüstumwandlung recht bemerkenswert. Es sei nochmals betont, daß diese synthetischen Reaktionen vom Programm gefunden wurden, ohne daß irgendeine Information über diese Reaktionen in den Computer eingegeben wurde. Sie wurden alleinig mit Hilfe der formalen Reaktionsgeneratoren erhalten. Wären diese Reaktionen noch nicht bekannt gewesen, so hätte ein Studium der Reaktionsenthalpien genügend Grund liefern müssen

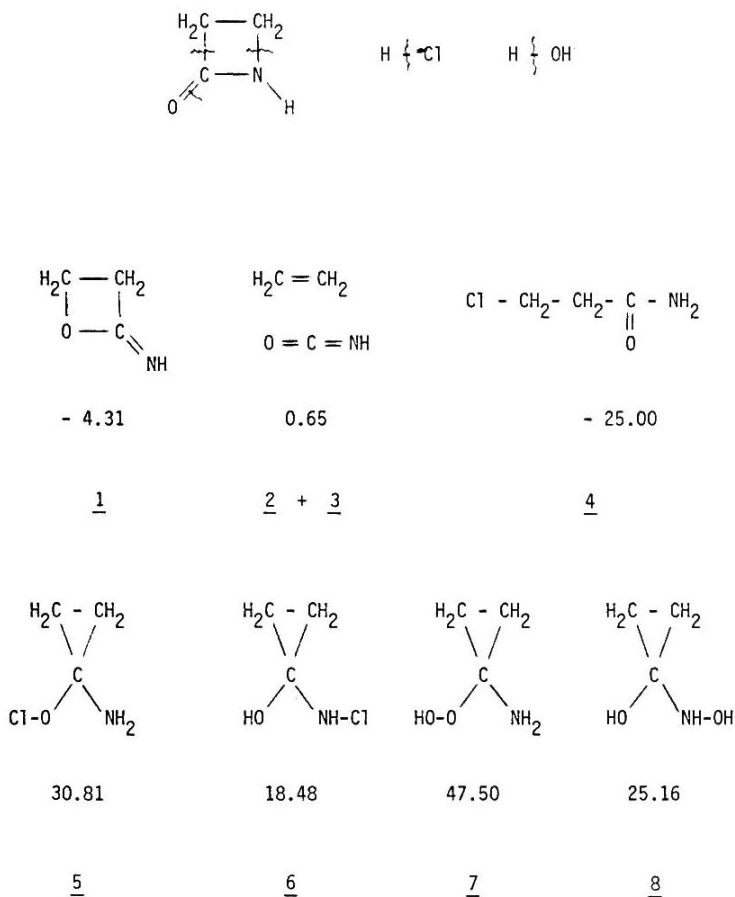


Abb. 15: β -Lactam Synthesen (Reaktionsenthalpien in kcal/mol)

für den Versuch diese Reaktionen im Labor zu verifizieren. Dieses Beispiel soll also auch illustrieren, wie durch Studium der Ergebnisse eines Programmlaufs neue synthetische Reaktionen aufgefunden werden könnten.

Nachdem etwas ausführlicher auf die Berechnung von Reaktionsenthalpien und deren Verwendung bei der Auswahl und Bewertung von Reaktionen eingegangen wurde, soll kurz eine neuere Entwicklungsrichtung aufgezeigt werden. Der Ablauf chemischer Reaktionen ist nicht nur von thermodynamischen Faktoren

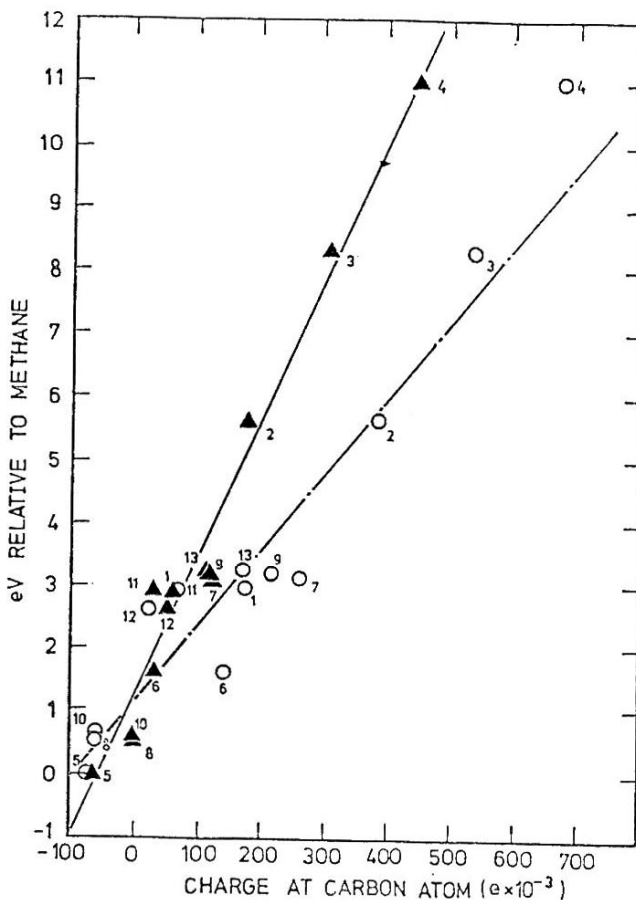


Abb. 16: Korrelation von C1s-ESCA shifts mit Atomladungen; ▲ Ladungen nach der PEOE-Methode (Lit. 9), ○ Ladungen aufgrund von ab initio-Rechnungen (Lit. 10)

- 1) CH_3F , 2) CH_2F_2 , 3) CHF_3 , 4) CF_4 , 5) CH_4 , 6) CH_3OH ,
 7) $\text{CH}_3\overset{\ominus}{\text{C}}\text{COCH}_3$, 8) $\text{CH}_3\overset{\oplus}{\text{C}}\text{COCH}_3$, 9) $\text{CH}_3\overset{\oplus}{\text{C}}\text{CHO}$, 10) $\overset{\oplus}{\text{C}}\text{H}_3\text{CHO}$, 11) $\overset{\oplus}{\text{C}}\text{H}_3\text{CN}$,
 12) $\text{CH}_3\overset{\oplus}{\text{C}}\text{N}$, 13) H_2CO

Die eingezeichneten Geraden entstammen einer Korrelationsrechnung

(wie z.B. ausgedrückt durch die Reaktionsenthalpien) geprägt. Eine wichtige Rolle bei der Diskussion von Reaktionsmechanismen spielen elektronische Effekte. Ladungsverteilungen sind durch quantenmechanische Rechnungen zugänglich. Für unsere Zwecke erschienen uns diese Verfahren zu aufwendig (hohe Rechenzeiten) und auf zu kleine Systeme beschränkt. Darüberhinaus sind in der am häufigsten verwendeten Mulliken'schen Populationsanalyse prinzipielle Unzulänglichkeiten enthalten, die manchmal zu unrealistischen Atomladungen führen. Wir entwickelten ein eigenes Verfahren zur Berechnung von Atomladungen, das auf einem Modell über den partiellen Ausgleich von Orbitalelektronegativitäten basiert (PEOE-Methode)⁹. Um eine Beurteilung der so erhaltenen Ladungswerte vornehmen zu können, korrelierten wir sie mit C-1s ESCA shifts, die bekanntermaßen von Atomladungen abhängen.

In Abb. 16 ist der entsprechende Vergleich durchgeführt. Die Korrelation ist ausgezeichnet; sie ist auch besser als diejenige, die mit Atomladungen, erhalten aus einer Mulliken'schen Populationsanalyse mit ab initio-Wellenfunktionen¹⁰, resultiert. Die Zuverlässigkeit unserer Ladungswerte ermutigt uns ihre Verwendung zur Vorhersage chemischer Reaktivitäten zu untersuchen.

Anwendungsbeispiele zur Syntheseplanung

Schon bei der Diskussion im vorigen Kapitel wurden einige Beispiele der Anwendung von EROS erwähnt. Hier sei der Einsatz von EROS zur Planung der Synthese organischer Verbindungen an zwei biochemisch wichtigen Verbindungen illustriert.

In Abb. 17 ist ein Ausschnitt einer Suche nach einer Synthese von Guanin enthalten. Der Pfad A entspricht in den letzten beiden Stufen der Biosynthese von Guanin. Die letzten beiden Schritte des Syntheseweges B laufen analog ab wie die Traube'sche Harnsäuresynthese. Die Darstellung des 2.5.6-Triamino-4-hydroxy-pyrimidins aus Cyanamid und Glycinester ist unserer Kenntnis nach neuartig, sie erscheint aber durchaus plausibel. Die Synthesewege C und D sind unseres Wissens noch nicht durchgeführt worden; es spricht aber nichts dagegen, daß sie nicht ablaufen können. Bei einigen Schritten sind vielleicht mehrere Angriffsorte gegeben, so daß zur geeigneten Reaktionslenkung die Blockierung unerwünschter reaktiver Stellen oder eine zusätzliche Aktivierung der reagierenden Gruppen nötig wird. In allen Fällen gehen die Synthesewege von wohlfeilen Ausgangsmaterialien aus und sind attraktiv genug für eine experimentelle Nachprüfung.

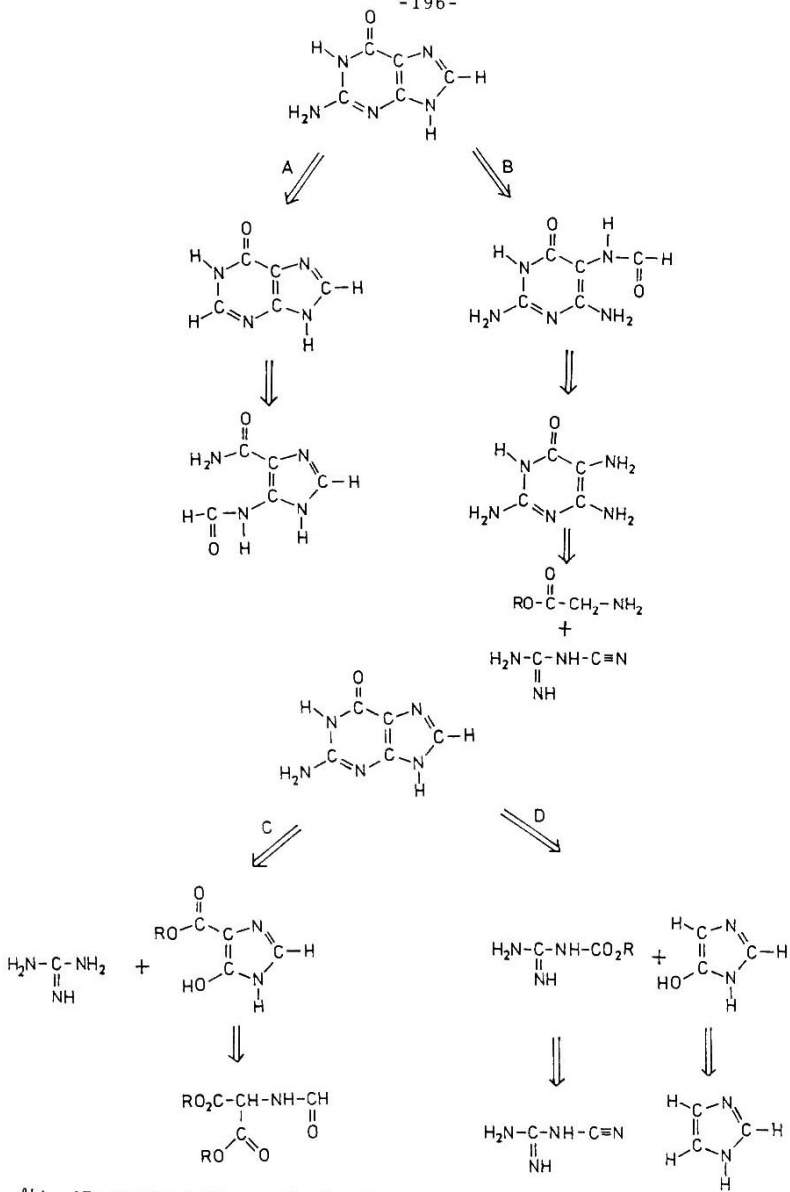


Abb. 17: Syntheseplan für Guanin

Während im vorigen Beispiel in den Pfaden B, C und D die Entwicklung von Retroreaktionen bis zu leicht zugänglichen Ausgangsmaterialien getrieben wurde, sollen in der folgenden Studie (Abb. 18) nur Strukturen des ersten Niveaus des Synthesebaumes aufgeführt werden. Die Vielfalt dieser Strukturen soll die Originalität der Ergebnisse eines Programmlaufes von EROS demonstrieren. Synthesziel war der sog. Corey-Aldehyd, eine wichtige Zwischenverbindung bei der Synthese von Prostaglandinen. Die Verbindungen umfassen die verschiedensten Ringsysteme und sollen als Anregungen für eine breit angelegte Betrachtung des Syntheseproblems dienen. Sie stellen die ersten Meilensteine in der Erarbeitung einer Synthese der Zielverbindung dar. Weitere Analyse der Struktur dieser Vorstufen muß dann ihre Zugänglichkeit klären und wird schließlich zu endgültigen Synthesevorschlägen führen. Hier soll aber anhand der ersten Vorstufen illustriert werden welche Vorteile die Unvoreingenommenheit eines formalen Reaktionsgenerators liefert. Es wird eine Reihe überraschender Strukturen erzeugt, von denen einige bei näherer Betrachtung doch Anregungen zu interessanten Synthesewegen darstellen dürften. Immerhin sind Synthesen des Corey-Aldehyds von Verbindungen mit dem Gerüst 2, 3, 4 und 9 bekannt.

Zusammenfassung

Mit der Entwicklung von EROS konnte nachgewiesen werden, daß die Erzeugung von Reaktionen durch formale Generatoren mit einem endlichen und sogar ziemlich kleinen Aufwand an Rechenzeit und Speicherplatz durchgeführt werden kann. Aufgrund dieses Ansatzes lassen sich alle denkbaren Reaktionen erhalten - bekannte und neuartige. Die Notwendigkeit der Auswahl der chemisch tatsächlich ablaufenden Reaktionen aus den mathematisch möglichen hat zur Entwicklung von Modellen geführt, die neue Einblicke in die Triebkraft chemischer Reaktionen gewähren. Mit Hilfe von EROS lassen sich eine Vielzahl von Problemen behandeln, die von unmittelbarem Interesse für den Chemiker sind: Die Folgeprodukte vorgegebener Verbindungen, der Metabolismus biochemischer Stoffe, industrielle Prozesse und die Planung der Synthese organischer Verbindungen.

Die Deutsche Forschungsgemeinschaft und die Stiftung Volkswagenwerk trugen durch finanzielle Unterstützung zu diesem Projekt bei. Oliver Dammer, Erika Diener-Weselsky und Helmut Ibl sei für Mithilfe bei den Programmierarbeiten gedankt.

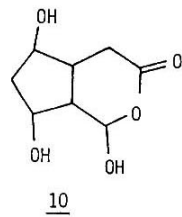
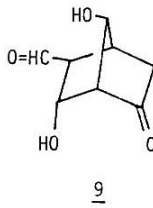
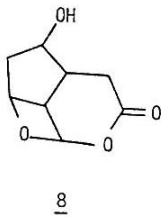
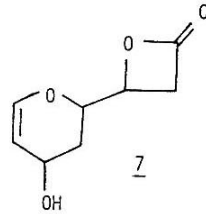
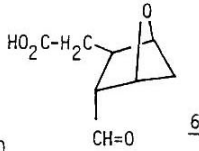
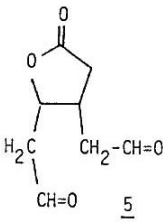
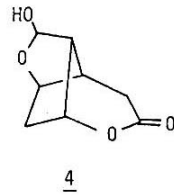
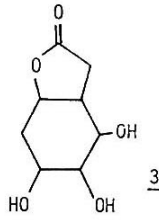
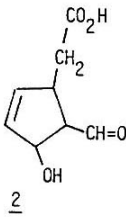
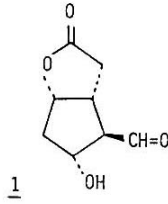


Abb. 18: Synthesestudie zum Corey-Aldehyd

Literatur

- 1) J.Gasteiger, C.Jochum, Topics Curr. Chem., 74, 93 (1978)
- 2) J.Dugundji, I.Ugi, Topics Curr. Chem., 39, 19 (1973)
- 3) C.Jochum, J.Gasteiger, J.Chem.Inf.Comput.Sci.,17, 113 (1977)
- 4) J.Blair, J.Gasteiger, C.Gillespie, P.Gillespie, I.Ugi, Tetrahedron, 30, 1845 (1974)
- 5) J.Gasteiger, C.Jochum, J.Chem.Inf.Comput.Sci., im Druck
- 6) J.Gasteiger, Comput.Chem., im Druck
- 7) T.L.Allen, J.Chem.Phys.,31, 1039 (1959)
- 8) H.H.Wasserman, H.W.Adickes, O.F.Ochoa, J.Amer.Chem.Soc.,93, 5586 (1971)
- 9) J.Gasteiger, M.Marsili, Tetrahedron Letters, 1978, 3181
- 10) W.J.Hehre, J.A.Pople, J.Amer.Chem.Soc., 92, 2191 (1970)