

Institut für Organische Chemie
der Technischen Universität München
Lehrstuhl I

Die bilaterale Generierung von Reaktionsnetzwerken

Eric Fontain

Vollständiger Abdruck der von der Fakultät für Chemie, Biologie und Geowissenschaften
der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender : o.Prof. Dr. H.-D. Belitz
1.Prüfer : o.Prof. Dr. I. Ugi
2.Prüfer : o.Prof. Dr. W. Nitsch
3.Prüfer : o.Prof. Dr. H.-P. Fritz

Die Dissertation wurde am 03.06.87 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie, Biologie und Geowissenschaften am 12.06.87 angenommen.

Meinem sehr verehrten Lehrer,

Herrn o.Prof. Dr. Ivar Ugi

danke ich herzlichst für sein immerwährendes Interesse und die großzügige Unterstützung dieser Arbeit.

Meinem Kollegen *Herrn Dr. Johannes Bauer* gilt mein besonderer Dank für viele nützliche Diskussionen und vor allem für die gründliche Einweisung in die Geheimnisse der Programmierkunst.

Frau Dipl.Inf. FH Mechthild Smoliniski danke ich für die Erstellung eines Netzwerkeditors.

Frau cand. chem. Andrea Wasserloos danke ich für die Durchführung von Testläufen mit dem Programm *RAIN*.

Frau Dr. Micaela Wochner danke ich für die Ermöglichung einiger PEMCD-Rechnungen.

Für Marina und Fabian

Inhalt

1	Einleitung	1
2	Das mathematische Modell der konstitutionellen Chemie	3
2.1	Matrixbeschreibung von Molekülen und Reaktionen	3
2.2	T-Matrizen	8
3	Reaktionsnetzwerke	13
3.1	Eigenschaften der Reaktionsnetzwerke	13
3.2	Erzeugung der Reaktionsnetzwerke	16
3.3	Bilaterale Erzeugung von Reaktionsnetzwerken	19
4	Der chemische Ereignishorizont	24
4.1	Definition der globalen Gesamtreaktion	24
4.2	Elemente und Pseudoelemente	24
4.3	Minimale Bindungsordnung	25
4.4	Atom-auf-Atom-Abbildung von Edukt und Produkt	26
4.5	Valenzschemata	28
4.6	Konvertierbarkeit der Valenzschemata	30
4.7	Definition des Reaktionskerns	32
4.7.1	Automatische Bestimmung des Reaktionskerns	32
4.7.2	Heuristische Bestimmung des Reaktionskerns	33
4.8	Grenzwerte für R- und T-Matrizen	35
4.8.1	Maximaler Rang einer irreduziblen R-Matrix	35
4.8.2	Maximale Anzahl von außerdiagonalen Einträgen in einem R-Dreieck ...	35
4.8.3	Maximaler Rang einer irreduziblen T-Matrix	35
4.8.4	Maximale Anzahl von außerdiagonalen Einträgen in einem T-Dreieck ...	36
4.8.5	Maximaler Betrag eines außerdiagonalen Eintrags in einer R-Matrix	36
4.8.6	Maximaler Betrag eines außerdiagonalen Eintrags in einer T-Matrix	37
4.8.7	Maximale Anzahl von außerdiagonalen Einträgen in einer R-Zeile	37
4.8.8	Maximale Anzahl von außerdiagonalen Einträgen in einer T-Zeile	38

II

4.8.9	Maximaler Betrag einer T-Zeilensumme	38
4.8.10	Maximaler Betrag einer T-Matrixgesamtsumme	39
4.9	Einfluß der R- und T-Grenzwerte auf die Valenzschemakonvertierbarkeit	40
4.10	Konstitutionelle Randbedingungen	42
4.10.1	Gespannte Ringsysteme	42
4.10.2	Bredt'sche Regel	45
4.10.3	Kumulierte Mehrfachbindungen in Ringen	46
4.10.4	Dreifachbindungen in Ringen	47
4.10.5	Maximale Länge von Heteroatomketten	47
4.10.6	Maximale Anzahl geladener Atome	48
4.10.7	Benachbarte gleichnamige Ladungen	48
5	Reaktionserzeugung	49
5.1	Das reaktionsgenerierende Verfahren	52
5.2	Eindeutige Darstellung der Moleküle	63
6	Zielgerichtete Entwicklung der Reaktionsbäume	66
6.1	Reaktionspfade entlang der minimalen chemischen Distanz	67
6.2	Obere Schranke für die MCD zur Zielverbindung	73
7	Anwendungsbeispiele	76
7.1	Protonenkatalysierte Spaltung von Amiden	77
7.2	Bildung von Diazoniumsalzen	83
7.3	Elektrocyclische Tandemreaktion von Benzocyclobutenen	88
7.4	Streith-Reaktion	93
7.5	Praebiotische Adeninsynthese	99
8	Zusammenfassung und Ausblick	110
9	Literatur	112

1 Einleitung

Die enorme Geschwindigkeit der Innovation in allen Wissensgebieten, der sich ein Naturwissenschaftler heute gegenüber sieht, macht den Einsatz moderner Methoden der Informationsverarbeitung in steigendem Maße erforderlich. Auf dem Gebiet der Chemie wurden in den letzten Jahrzehnten große Anstrengungen gemacht, das bekannte Wissen zu strukturieren, zu formalisieren und damit den Methoden der elektronischen Datenverarbeitung zugänglich zu machen. So ist es in relativ kurzer Zeit nahezu vollständig gelungen, dem Chemiker die neuere Primärliteratur über Terminals "on-line" zur Verfügung zu stellen^{1,2}. Die Inkonsistenz chemischer Sachverhalte macht den Aufbau stoffbezogener Faktendatenbanken ungleich schwieriger. Dies zeigt sich in der Tatsache, daß eine umfassende Datenbank dieses Typs bis heute keine Anwendungsreife erlangt hat. Diese Entwicklung vollzieht sich neben einer zunehmenden Perfektionierung rein numerischer Verfahren in der physikalischen, theoretischen und technischen Chemie. Hieraus ziehen vor allem quantenmechanische Rechnungen und Simulationstechniken ihren Nutzen.

Doch nicht nur zur Verarbeitung und Präsentation bereits vorliegender Informationen und in der Verwendung als reine Rechenautomaten höchster Geschwindigkeit und Präzision, sondern auch als Werkzeuge zur kreativen Neuschöpfung von Ideen und Methoden sollen Computerprogramme Hilfestellung leisten. Auch wenn die anfängliche Euphorie über die Mächtigkeit und schnelle Verfügbarkeit solcher Systeme, besonders in der chemischen Industrie, mittlerweile etwas kleiner geworden ist, sind in den letzten Jahren eine Reihe interessanter Arbeiten über die "Denkzeuge" in der chemischen Forschung entstanden³.

Ausschlaggebend hierfür war die Schaffung eines streng mathematischen Fundaments für die chemische Konstitution und deren Transformationen. Die Anwendung graphentheoretischer Verfahren stellt ein geeignetes Instrumentarium zur Verfügung, um chemische Strukturen und Reaktionen "computergerecht" in den Griff zu bekommen. Dies hat zu grundlegenden Arbeiten zur computerunterstützten Syntheseplanung⁴, Reaktionsdokumentation⁵, Reaktionserzeugung⁶⁻⁸ u.v.m. geführt.

Das mathematische Modell der konstitutionellen Chemie bildet auch die Ausgangsbasis für die vorliegende Arbeit, die sich mit dem Problem der Aufstellung von Reaktionsnetzwerken befaßt. Unter Reaktionsnetzwerken sollen alle möglichen Kombinationen von Reaktionspfaden verstanden werden, die zwei Moleküle miteinander verbinden. Solche Reaktionspfade sind bekannt z.B. aus der Biochemie in Form der metabolischen Auf- und Abbauewege, aus der Syntheseplanung als Sequenzen von Syntheseschritten und aus der chemischen Mechanistik in Form der Abfolge elementarer Einzelschritte einer komplexen Reaktion. Der Erarbeitung solcher Netzwerke von Reaktionspfaden gehen oftmals äußerst umfangreiche mechanistische und analytische Studien voraus, um z.B. Übergangszustände und Zwischenstufen zu erkennen und Entscheidungen über Alternativwege zu treffen. Ein System, das in der Lage ist, auf der Grundlage formaler Regeln, unter Beachtung von definierten Randbedingungen, vollständige Reaktionsnetzwerke mit allen kombinatorisch möglichen Alternativen der Reaktionspfade zu generieren, stellt hierbei eine große Hilfe dar. Insbesondere kann das Augenmerk auf neue Alternativen gelenkt werden, und es gibt die Sicherheit, keine wichtigen denkbaren Reaktionspfade übersehen zu haben.

Mit dieser Arbeit wird das formale System *RAIN* (*Reaction And Intermediate Network*)^{9,10} vorgestellt, das in der Lage ist, Reaktionsnetzwerke jeglicher Art zu erzeugen und somit als vielseitig verwendbares Werkzeug in der biochemischen Forschung, in der Syntheseplanung und vor allem bei der Aufklärung von Reaktionsmechanismen seinen Einsatz finden kann. Die Netzwerke zwischen Edukt und Produkt einer chemischen Reaktion oder Reaktionsfolge werden von *RAIN* aufgebaut, indem unter Beachtung von Regeln für die Reaktionsschritte und die chemische Konstitution, alle kombinatorisch möglichen Reaktionssequenzen, die jeweils vom Edukt und Produkt ausgehen, erzeugt werden. Dies geschieht stufenweise, es werden also zuerst die unmittelbaren Nachfolgemoleküle z.B. des Edukts, dann deren Nachfolger usw. erzeugt. Finden sich in den Reaktionssequenzen des Edukts und des Produkts einander entsprechende Moleküle, so existieren über diese Verbindungen Reaktionspfade zwischen Edukt und Produkt, und das Reaktionsnetzwerk ist geschlossen.

2 *Das mathematische Modell der konstitutionellen Chemie*

Das von *Ugi* und *Dugundji* entwickelte Modell der konstitutionellen Chemie^{11,12}, die Algebra der *BE*- und *R*-Matrizen, ist Grundlage einer Vielfalt von Computerprogrammen zur Lösung komplizierter Problemstellungen in der organischen Chemie. Es stellt einen wesentlichen Beitrag bei dem Versuch dar, das komplexe Gebiet der organischen Chemie, mit seinen vielfältigen Strukturen und Transformationen, durch ein strenges mathematisches Konzept zu formalisieren und damit computergerecht darzustellen. Es ist eine der Zielsetzungen des Modells durch Erarbeitung von Regeln auf sehr einfachem Niveau zu Systemen zu gelangen, die unter gegebenen Randbedingungen einen kombinatorisch vollständigen Satz von chemischen Aussagen erzeugen können und dabei ohne katalogisiertes Hintergrundwissen auskommen. Systeme dieser Art sind in der Lage, die Kreativität und den Erfahrungshorizont der sie anwendenden Chemiker in großem Maße zu erweitern und zu bereichern.

2.1 *Matrixbeschreibung von Molekülen und Reaktionen*

Das Modell definiert ein Ensemble von Molekülen *EM* als Menge von Atomrümpfen, die durch eine bestimmte Zahl von Valenzelektronen verbunden sind. Diese Elektronen befinden sich einerseits in Bindungen zwischen Atomen und andererseits an den Atomrümpfen, in Form freier Valenzen. Das Bindungsmodell entspricht dem der Valenzbindungstheorie mit ganzzahligen Bindungsordnungen. Eine Umverteilung der Elektronen (dies entspricht einer chemischen Reaktion) transformiert ein *EM* in ein anderes, isomeres *EM*. Die Gesamtheit der so generierbaren *EM* bildet die Familie isomerer *EM* (*FIEM*). Sie umfaßt alle kombinatorisch möglichen Strukturen, die aus einer Menge von Atomen und Elektronen unter Beachtung der Valenzregeln gebildet werden können.

Ein *EM* von *n* Atomen kann durch eine *n*×*n* *BE*-Matrix (*B*indungs- und *E*lektronenmatrix) wiedergegeben werden, wobei jedem Atomrumpf eine Zeile/Spalte dieser

Matrix zugeordnet ist. Ein Matrixeintrag in der Spalte j der Zeile i entspricht der formalen Ordnung einer kovalenten Bindung zwischen den Atomen i und j . Die Einträge auf der Hauptdiagonalen der Matrix stehen für die Anzahl von freien Elektronen der entsprechenden Atome. Die mathematischen Eigenschaften der BE -Matrizen sind eingehend untersucht und beschrieben worden¹¹⁻¹⁷.

Eine chemische Reaktion $EM(B) \rightarrow EM(E)$ wird durch eine Matrixtransformation $B \rightarrow E$ dargestellt, wobei diese Transformation durch die Addition einer sogenannten R -Matrix zu B nach der Fundamentalgleichung

$$B + R = E$$

ausgeführt wird. R -Matrizen beschreiben die Elektronenverschiebung von Reaktionen. Negative Einträge stehen für Abzug, positive Einträge für Hinzugewinn von Elektronen in Bindungen (bei außerdiagonalen Einträgen) oder an Atomrümpfen (bei diagonalen Einträgen). Eine irreduzible R -Matrix wird erhalten, indem aus einer vollständigen R -Matrix alle Nullzeilen/spalten, also die an der Transformation unbeteiligten Atome, entfernt werden. Sie beschreibt den eigentlichen Reaktionstyp, d.h. die individuelle Art des Elektronenflußmusters.

Abbildung (2-1) zeigt einige BE -Matrizen mit den sie ineinander transformierenden R -Matrizen der $FIEM$ aus den Atomen C, O, O, H und H mit 18 Elektronen.

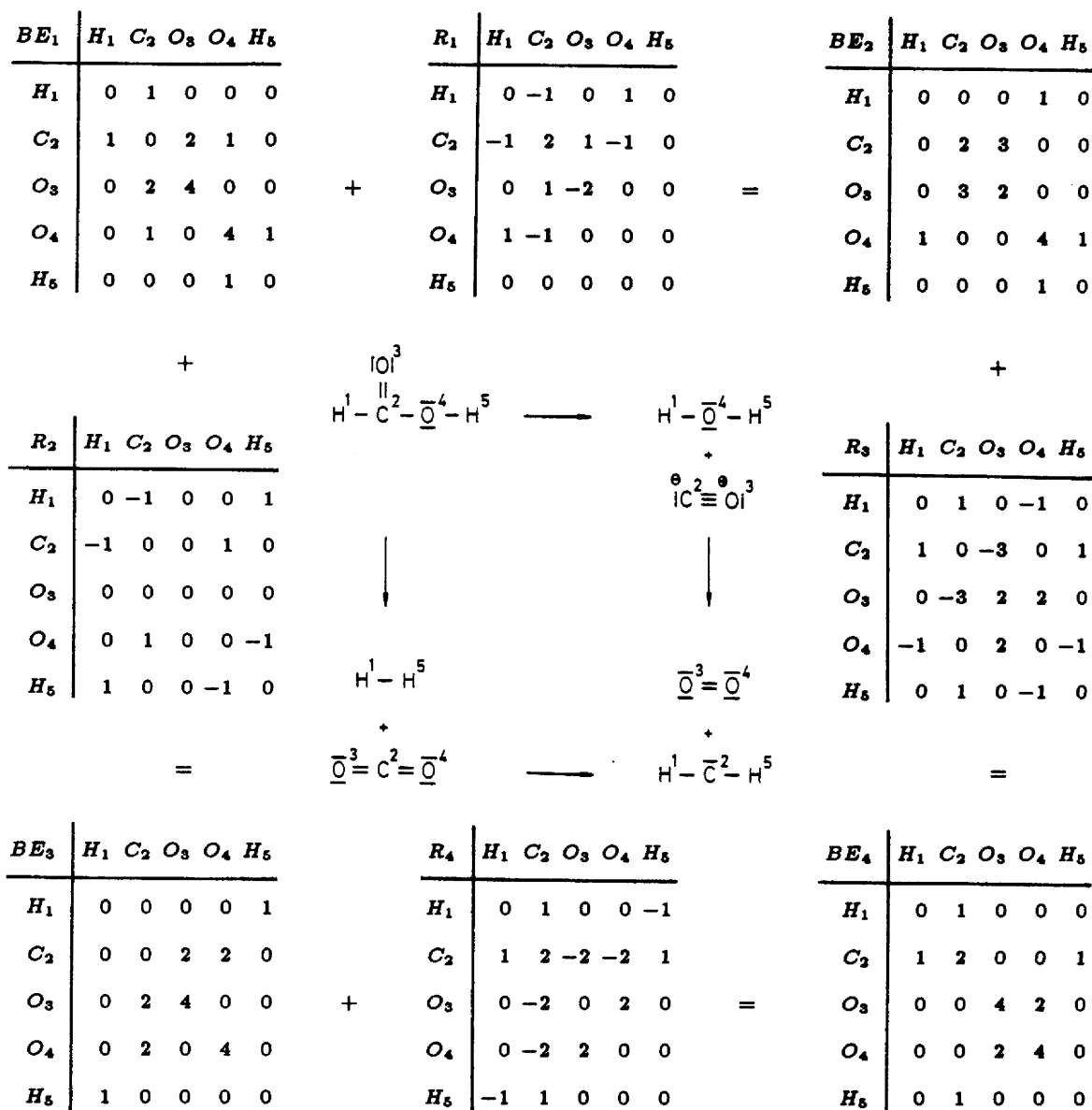


Abb. (2-1) Ausschnitt aus einer FIEM.

In ihrer Eigenschaft als Operatoren auf *BE*-Matrizen werden irreduzible *R*-Matrizen in einigen Synthesepungs- bzw. Reaktionsprädiktionsprogrammen als Reaktionsgeneratoren verwendet. Hierzu werden *R*-Matrizen aus einem fest vorgegebenen Katalog von sehr häufig beobachteten^{19–22} (z.B. EROS^{23,24}, SCANMAT²⁵) oder elementaren (ASSOR^{26,27}) Reaktionsschritten herangezogen. Diese begrenzte Auswahl führt zu einigen Nachteilen; so können in EROS und SCANMAT nur Reaktionsschritte erzeugt werden, deren *R*-Matrizen zu den (wenigen) gehören, die als Reaktionsgeneratoren zugelassen sind. ASSOR kann zwar alle der möglichen Reaktionsschritte generieren, jedoch führt die elementare Einfachheit der verwendeten *R*-Matrizen zu einer starken Auffächerung der Reaktionsbäume. Auch eine kleine Entfernung vom Ausgangsmolekül*, in "realen" Reaktionsschritten gerechnet, führt zu Problemen, da die Reaktionsschritte ja aus einer Aneinanderreihung sehr vieler dieser Basistransformationen bestehen.

Um diese Nachteile zu beseitigen, werden in *RAIN* keine vordefinierten *R*-Matrizen verwendet, sondern stattdessen globale Randbedingungen für *R*-Matrizen definiert. Im Rahmen dieser Randbedingungen sind zur Reaktionsgenerierung alle möglichen *R*-Matrizen erlaubt. Die Randbedingungen beschreiben die maximale Anzahl beteiligter Atome, Bindungen und die Komplexität der Elektronenflußmuster.

Dies reicht jedoch zur Beschreibung eines sehr allgemeinen Reaktionsgenerators dieses Typs noch nicht aus. Vielmehr muß auch die Anwendbarkeit der möglichen *R*-Matrizen auf unterschiedliche Molekülgerüste limitierend beschrieben werden. Dies ist notwendig, weil die Einträge der von null verschiedenen Einträge der *R*-Matrizen nur Netto-Elektronenverluste oder -gewinne protokollieren. Die Information über die Ordnung einer Bindung in Ausgangs- und Endmatrix einer Reaktion ist in ihnen nicht enthalten. So kann der Eintrag einer außerdiagonalen -1 in einer *R*-Matrix z.B. bedeuten, daß eine Doppelbindung zur Einfachbindung erniedrigt, oder daß eine Einfachbindung gebrochen wird. Der Bindungsbruch ist jedoch ein wesentlich schwerwiegenderer Ein-

* Im folgenden wird anstatt des Ausdrucks "*EM*" auch der Ausdruck "Molekül" verwendet. Dies geschieht aus Gründen der besseren Lesbarkeit. Ein Molekül im Sinne eines "*EM*" kann also auch aus mehreren, individuellen Molekülen bestehen.

griff in die molekulare Struktur als die Erniedrigung der Bindungsordnung, bei der das σ -Gerüst erhalten bleibt. Er ist aus diesem Grund sowohl aus energetischer Sicht, als auch im Sinne eines synthetisch-strategischen Vorgehens von wesentlich größerer Bedeutung.

Eine Möglichkeit, die Anwendbarkeit von *R*-Matrizen auf bestimmte Molekülgerüste zu limitieren, ist die sogenannte "intakte *BE*-Matrix"⁵. Sie beschreibt das nach der Addition der negativen Anteile einer *R*-Matrix verbleibende Bindungsgerüst. Dies ist jedoch eine zu spezielle, von Eduktmolekül zu Eduktmolekül neu zu definierende Beschreibung, für die es keine global festlegbaren Randbedingungen gibt. In einem netzwerkgenerierenden Programm sind globale Randbedingungen für die Reaktionserzeugung jedoch unbedingt nötig, da die potentiell entstehenden Zwischenstufenstrukturen a priori nicht bekannt sind. Um diese Problematik in den Griff zu bekommen, werden die *T*-Matrizen definiert.

2.2 *T*-Matrizen

T-Matrizen beschreiben die Nettoveränderungen von Molekülgerüsten. Sie protokollieren, zwischen welchen Atomen neue Bindungen geknüpft, oder bestehende Bindungen gebrochen werden und stellen hiermit eine Beschreibung der Veränderungen der kovalenten Nachbarschaftsverhältnisse einer Reaktion dar.

T-Matrizen werden wie folgt definiert :

Die *T*-Matrix einer chemischen Reaktion ist gleich der Differenz der Adjazenzmatrizen von Edukt (B_{ad}) und Produkt (E_{ad}).

$$T := E_{ad} - B_{ad}$$

oder

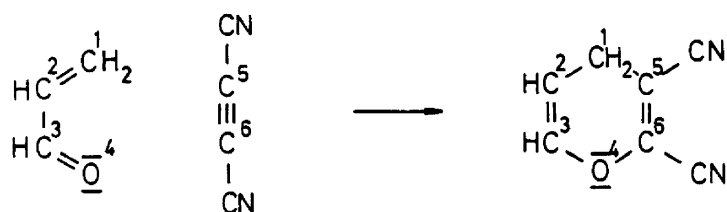
$$B_{ad} + T = E_{ad}$$

mit

$$B_{ad} := \begin{cases} b_{ij} = 0 & \text{wenn Atom } i \text{ und } j \text{ nicht verbunden sind} \\ b_{ij} = 1 & \text{wenn Atom } i \text{ und } j \text{ verbunden sind} \end{cases}$$

analog für E_{ad} .

Abbildung (2-2) zeigt eine Beispielreaktion mit *BE*-, *R*-, Adjazenz- und *T*-Matrizen.



<i>B</i>	<i>C</i> ₁	<i>C</i> ₂	<i>C</i> ₃	<i>O</i> ₄	<i>C</i> ₅	<i>C</i> ₆
<i>C</i> ₁	0	2	0	0	0	0
<i>C</i> ₂	2	0	1	0	0	0
<i>C</i> ₃	0	1	0	2	0	0
<i>O</i> ₄	0	0	2	4	0	0
<i>C</i> ₅	0	0	0	0	0	3
<i>C</i> ₆	0	0	0	0	3	0

+

<i>R</i>	<i>C</i> ₁	<i>C</i> ₂	<i>C</i> ₃	<i>O</i> ₄	<i>C</i> ₅	<i>C</i> ₆
<i>C</i> ₁	0	-1	0	0	1	0
<i>C</i> ₂	-1	0	1	0	0	0
<i>C</i> ₃	0	1	0	-1	0	0
<i>O</i> ₄	0	0	-1	0	0	1
<i>C</i> ₅	1	0	0	0	0	-1
<i>C</i> ₆	0	0	0	1	-1	0

=

<i>E</i>	<i>C</i> ₁	<i>C</i> ₂	<i>C</i> ₃	<i>O</i> ₄	<i>C</i> ₅	<i>C</i> ₆
<i>C</i> ₁	0	1	0	0	1	0
<i>C</i> ₂	1	0	2	0	0	0
<i>C</i> ₃	0	2	0	1	0	0
<i>O</i> ₄	0	0	1	0	0	1
<i>C</i> ₅	1	0	0	0	0	2
<i>C</i> ₆	0	0	0	1	2	0

<i>B_{ad}</i>	<i>C</i> ₁	<i>C</i> ₂	<i>C</i> ₃	<i>O</i> ₄	<i>C</i> ₅	<i>C</i> ₆
<i>C</i> ₁	0	1	0	0	0	0
<i>C</i> ₂	1	0	1	0	0	0
<i>C</i> ₃	0	1	0	1	0	0
<i>O</i> ₄	0	0	1	1	0	0
<i>C</i> ₅	0	0	0	0	0	1
<i>C</i> ₆	0	0	0	0	1	0

+

<i>T</i>	<i>C</i> ₁	<i>C</i> ₂	<i>C</i> ₃	<i>O</i> ₄	<i>C</i> ₅	<i>C</i> ₆
<i>C</i> ₁	0	0	0	0	1	0
<i>C</i> ₂	0	0	0	0	0	0
<i>C</i> ₃	0	0	0	0	0	0
<i>O</i> ₄	0	0	0	0	0	1
<i>C</i> ₅	1	0	0	0	0	0
<i>C</i> ₆	0	0	0	1	0	0

=

<i>E_{ad}</i>	<i>C</i> ₁	<i>C</i> ₂	<i>C</i> ₃	<i>O</i> ₄	<i>C</i> ₅	<i>C</i> ₆
<i>C</i> ₁	0	1	0	0	1	0
<i>C</i> ₂	1	0	1	0	0	0
<i>C</i> ₃	0	1	0	1	0	0
<i>O</i> ₄	0	0	1	0	0	1
<i>C</i> ₅	1	0	0	0	0	1
<i>C</i> ₆	0	0	0	1	1	0

Abbildung (2-2) Reaktion mit *B*-, *R*-, *E*-, *B_{ad}*-, *T*- und *E_{ad}*-Matrizen.

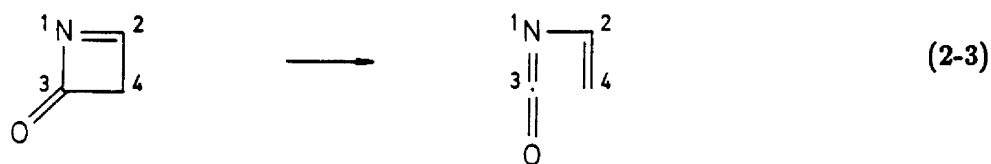
T -Matrizen besitzen die folgenden Eigenschaften :

- grundsätzlich gilt : $|t_{ij}| \leq 1$.
- sie sind symmetrisch, d.h. es gilt : $t_{ij} = t_{ji}$.
- die Matrixsumme $S_T = \sum_{ij} t$ ist nicht notwendigerweise gleich null, wie dies bei den R -Matrizen der Fall ist. $S_T > 0$ zeigt an, daß mehr Bindungen geknüpft als gebrochen werden, das Molekülgerüst topologisch also komplexer wird. $S_T < 0$ zeigt an, daß mehr Bindungen gebrochen als geknüpft werden, was einer topologischen Vereinfachung des Molekülgerüsts gleichkommt. $S_T = 0$ bedeutet, daß die topologische Komplexität des Moleküls erhalten bleibt.

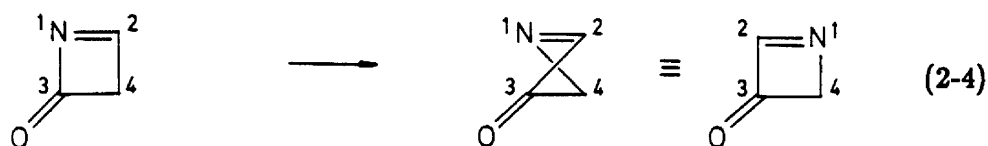
Diese Definition der T -Matrizen eröffnet die Möglichkeit, in Analogie zur Beschränkung der R -Matrizen, auch Grenzwerte für die maximalen topologischen Veränderungen in einem Reaktionsschritt festzulegen.

Das folgende Beispiel möge dies erläutern :

Die beiden Reaktionen



und



verlaufen nach den irreduziblen R -Matrizen :

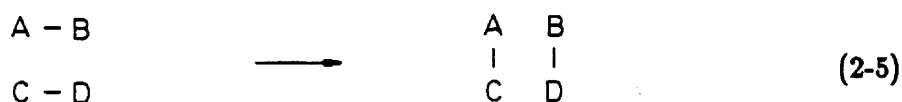
R_1	N_1	C_2	C_3	C_4
N_1	0	-1	1	0
C_2	-1	0	0	1
C_3	1	0	0	-1
C_4	0	1	-1	0

(2-3a)

R_2	N_1	C_2	C_3	C_4
N_1	0	0	-1	1
C_2	0	0	1	-1
C_3	-1	1	0	0
C_4	1	-1	0	0

(2-4a)

Diese beiden R -Matrizen sind äquivalent, d.h. sie beschreiben beide dasselbe Elektronenflußmuster (2-5).



Ein nur auf der Limitierung von R -Matrizen beruhendes reaktionsgenerierendes System würde sowohl (2-3), als auch die weniger wahrscheinliche Reaktion (2-4) erzeugen. Unterscheidbar werden die Reaktionen erst durch den Einsatz der T -Matrizen.

T_1	N_1	C_2	C_3	C_4
N_1	0	0	0	0
C_2	0	0	0	0
C_3	0	0	0	-1
C_4	0	0	-1	0

(2-3b)

T_2	N_1	C_2	C_3	C_4
N_1	0	0	-1	1
C_2	0	0	1	-1
C_3	-1	1	0	0
C_4	1	-1	0	0

(2-4b)

Die T -Matrix (2-3b) betrifft lediglich 2 Atome und 1 Bindung, während in (2-4b) 4 Atome und 4 Bindungen beteiligt sind. Eine Begrenzung der maximal beteiligten Bindungszahl in T -Matrizen würde die Reaktion (2-4) ausschließen, Reaktion (2-3) jedoch zulassen.

Die T -Matrizen können auch zur Klassifizierung und zum Ordnen von Basisreaktionen in reaktionsgenerierenden Systemen wie IGOR mit einer hierarchischen Reaktionsklassifizierung herangezogen werden⁸. So können z.B. die Basisreaktionen der R -Matrix $R_{8,6}^c$ nach steigender Anzahl von beteiligten Atomen A_t und Bindungen B_t in den zugehörigen T -Matrizen klassifiziert werden (siehe Abbildung (2-6)).

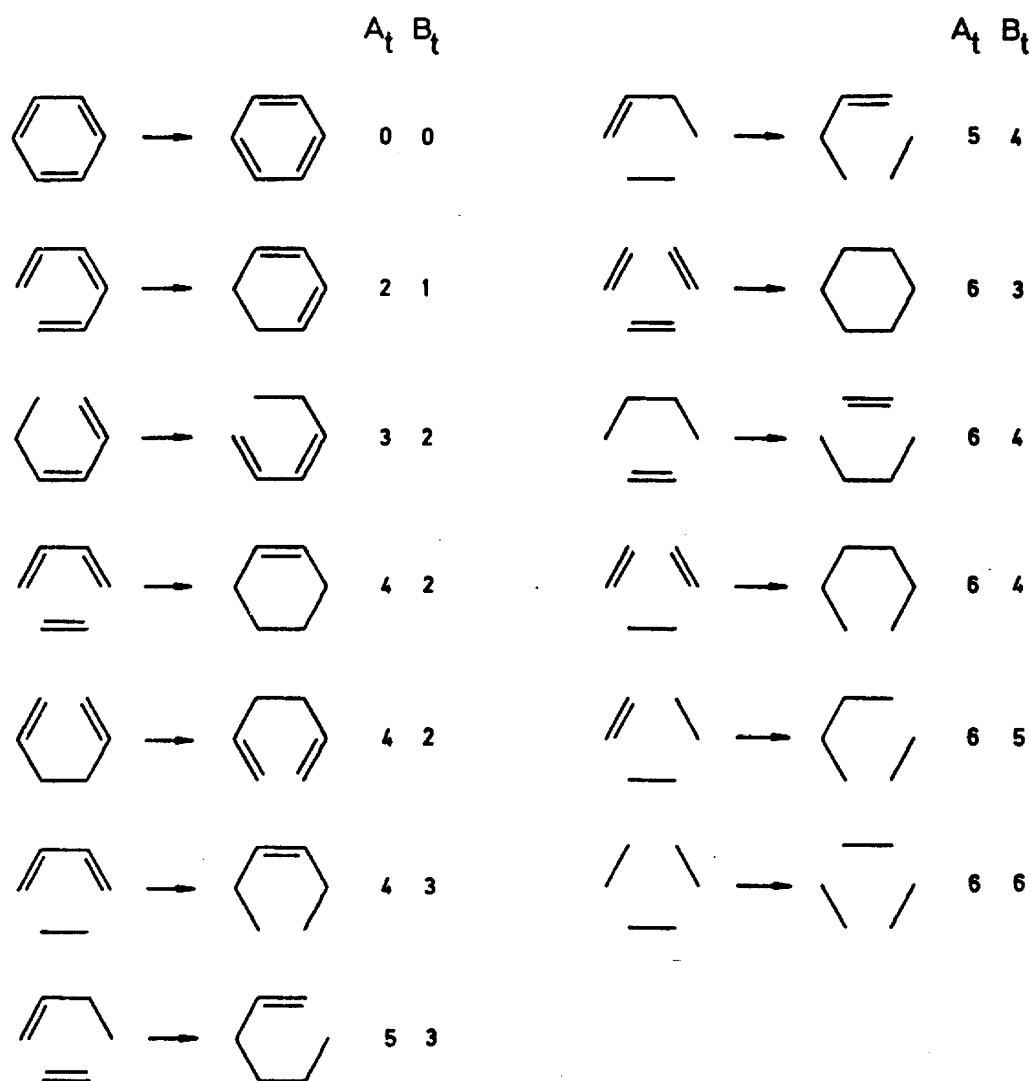


Abbildung (2-6) Basisreaktionen der R -Matrix $R_{8,6}^c$

3 Reaktionsnetzwerke

3.1 Eigenschaften der Reaktionsnetzwerke

Das Programmsystem *RAIN* erzeugt Reaktionsnetzwerke zwischen Edukten *E* und Produkten *P*. Diese Netzwerke repräsentieren die Kombination aller rechnerisch möglichen Reaktionspfade zwischen *E* und *P*, die gegebenen Randbedingungen gehorchen. Die folgende Abbildung (3-1) zeigt schematisch ein solches Netzwerk mit Reaktionswegen und Zwischenstufen.

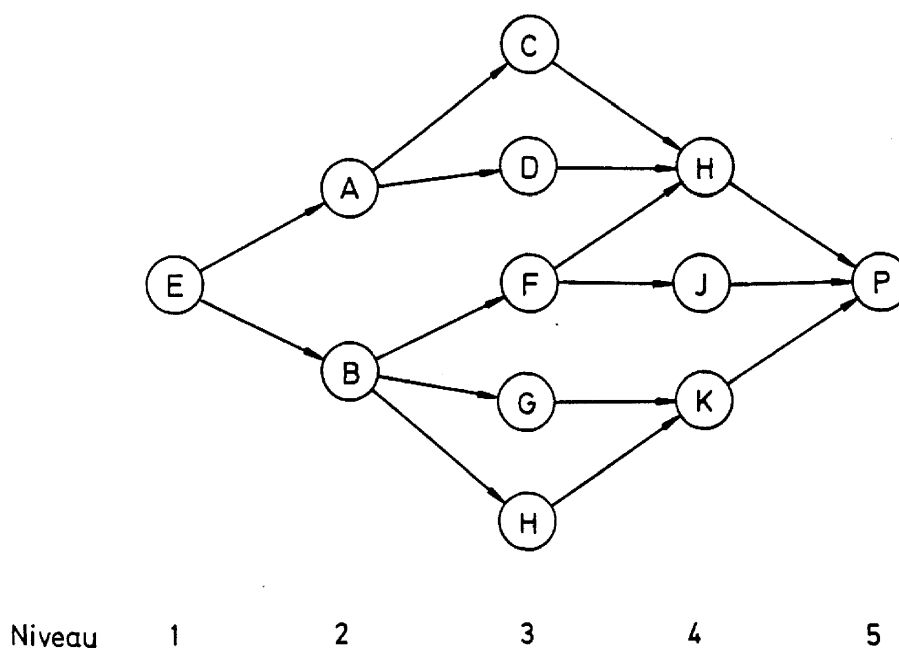


Abb. (3-1) Schematisches Reaktionsnetzwerk.

Diese Reaktionsnetzwerke besitzen definitionsgemäß folgende Eigenschaften :

- Ein Reaktionsnetzwerk ist in Niveaus eingeteilt. Alle Moleküle eines Niveaus haben denselben graphentheoretischen Abstand (gezählt in Reaktionsschritten) vom Edukt E oder vom Produkt P .
- Jeder Reaktionsschritt verbindet ein Molekül mit einem Molekül im benachbarten, höheren Niveau. Damit ist eine Orientierung des Netzwerks und somit eine globale Reaktionsrichtung definiert. Es gibt keine Verknüpfungen innerhalb eines Niveaus oder über mehrere Niveaus hinweg.
- Ein bestimmtes Molekül existiert in einem Niveau nur einmal.
- Ein bestimmtes Molekül kann in verschiedenen Niveaus mehrfach auftreten. In jedem dieser Niveaus wird das Molekül jeweils vollkommen eigenständig behandelt.
- In einem Reaktionsnetzwerk mit N Niveaus sind alle Reaktionspfade zwischen E und P genau $N - 1$ Reaktionsschritte lang, wenn Automerisationsreaktionen* als Reaktionsschritte verboten werden.
- Werden Automerisationsreaktionen als Reaktionsschritte zugelassen, so finden sich in einem Reaktionsnetzwerk mit N Niveaus alle Reaktionspfade mit einer Länge von bis zu $N - 1$ Reaktionsschritten.
- Von jedem Molekül eines Niveaus im Reaktionsnetzwerk führt mindestens ein Pfad zum Edukt E (entgegen der globalen Reaktionsrichtung) und ein Pfad zum Produkt P , d.h. das Netzwerk ist frei von "ins Leere" Wegen.
- Ein von RA/N erzeugtes Netzwerk stellt unter Beachtung gegebener Randbedingungen die Gesamtheit aller formal möglichen Reaktionspfade zwischen dem Edukt E und dem Produkt P dar.

* Automerisationsreaktionen überführen ein Molekül in sich selber. Formal ist hier auch die "Nicht"-Reaktion miteinbezogen.

Automerisationsreaktionen stellen Schleifen dar, die über null Zwischenstufen verlaufen. Schleifen über eine Zwischenstufe, d.h. daß auf einen Reaktionsschritt die unmittelbare Umkehrung dieses Schritts folgt, können von *RAIN* erkannt und optional zugelassen oder verboten werden. Hingegen ist es nicht möglich, Schleifen über mehr als eine Zwischenstufe zu erkennen. Da sich die zu erzeugenden Netzwerke jedoch in den meisten Fällen über eine möglichst geringe Anzahl von Zwischenstufen erstrecken sollen, tritt dieses Problem nicht auf, da für Reaktionspfade, die solche Schleifen enthalten, immer ein schleifenfreies und damit kürzeres Pendant existiert.

3.2 Erzeugung der Reaktionsnetzwerke

Die bekannten retrosynthetischen Syntheseprogramme oder Reaktionssimulationssysteme arbeiten nach dem Prinzip der sukzessiven Anwendung von reaktionsgenerierenden Verfahren⁴. Diese Technik basiert entweder auf katalogisierten, bekannten Reaktionen²⁸⁻³³ (Transforms), oder auf formalen Prinzipien^{13-15,23-27,34-36}, die in der Lage sind, auch aus der bekannten Chemie hinauszuführen.

RAIN ist mit einem global definierten, formalen Reaktionsgenerator ausgestattet. Dieser erlaubt es, von einem Molekül alle potentiellen Nachfolger (nächste Zwischenstufen) auf formalem Weg zu erzeugen, ohne die Transformationen spezifisch festzulegen (siehe 5.1). Dieses Verfahren ist symmetrisch bezüglich der Reaktionsrichtung, d.h. dieser Reaktionsgenerator kann ebenso zur Erzeugung aller potentiellen Vorgänger eines Moleküls herangezogen werden.

Eine Möglichkeit, das Netzwerk aller Reaktionspfade zwischen einem Edukt *E* und einem Produkt *P* zu generieren, ist eine einseitige Entwicklung aller möglichen Reaktionswege zu einem Reaktionsbaum*:

- a) Von einem Ausgangsmolekül (Edukt *E* oder ein Zwischenprodukt) werden mittels des reaktionsgenerierenden Prinzips alle Nachfolgemoleküle (Zwischenstufen des nächsten Niveaus) erzeugt.
- b) Schritt a) wird sukzessive auf alle Moleküle des nächsten Niveaus angewandt, bis das zu erreichende Produkt als Nachfolger mindestens einer Zwischenstufe erzeugt wird.
- c) Die rekursive Streichung aller Pfade, die nicht zum Produkt führen, erzeugt das geschlossene Netzwerk (siehe Abbildung (3-2)).

* Der Begriff "Baum" wird in dieser Arbeit nicht im streng graphentheoretischen Sinn eines gerichteten "Baums" verwendet, in welchem jeder Knoten (Molekül) nur einen Vorgänger haben darf, sondern eher im Sinne eines "monopolar gerichteten Netzwerks". "Reaktionsnetzwerke" stellen damit "bipolar gerichtete Netzwerke" dar.

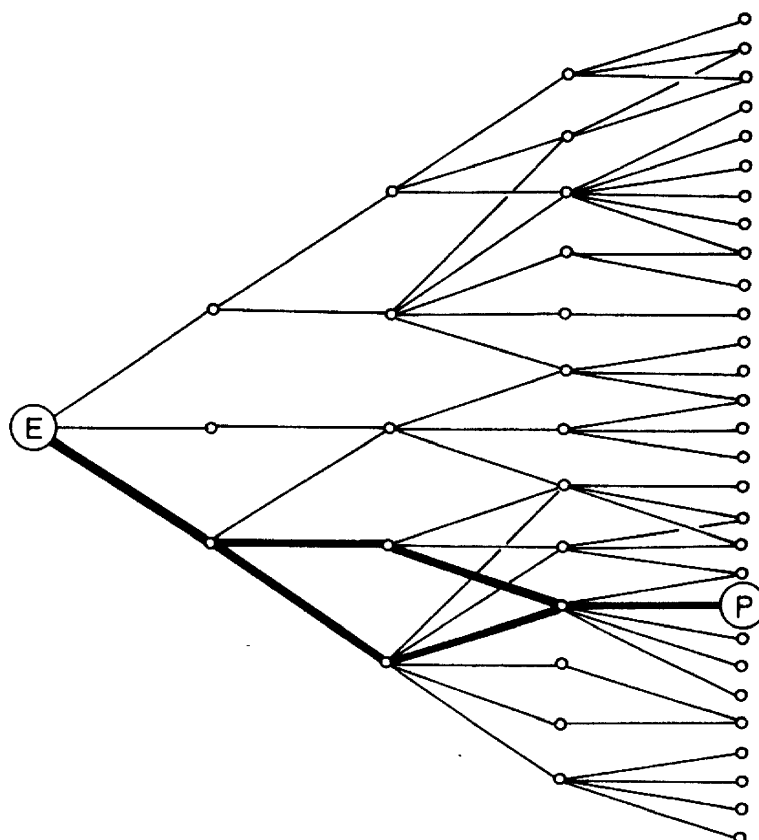


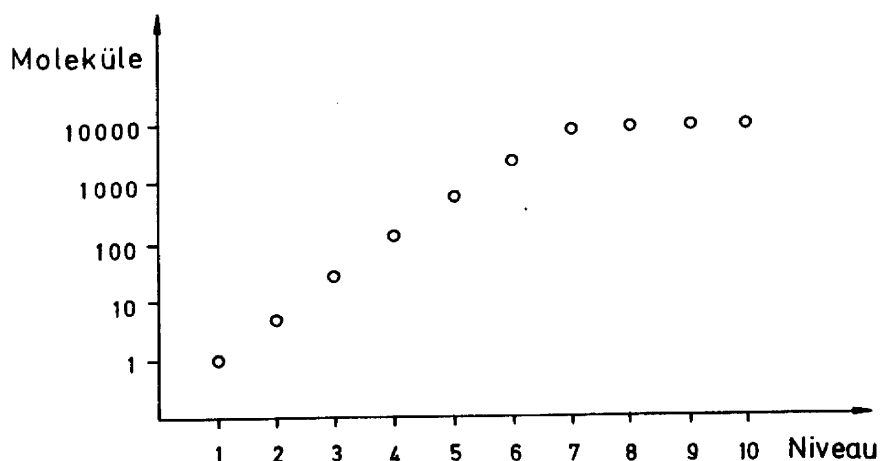
Abb. (3-2) Einseitig entwickeltes Reaktionsnetzwerk.

Ein Nachteil dieses einseitigen (monolateralen) Vorgehens ist die sehr große Auffächerung des Reaktionsbaumes bei anwachsender Zahl der Reaktionsschritte. Das folgende Beispiel eines hypothetischen Reaktionsbaumes verdeutlicht dies :

Annahmen :

- Der Reaktionsgenerator erzeuge aus einem Molekül immer 5 Nachfolgemoleküle im nächsthöheren Niveau ($N_f = 5$).
- Die Gesamtzahl aller Strukturen N_{ges} , die kombinatorisch möglich sind und die Randbedingungen erfüllen sei 10000.
- Die Wahrscheinlichkeit, daß ein erzeugtes Nachfolgemolekül noch nicht im nächsten Niveau (mit bereits N_v Molekülen) enthalten ist sei $P = 1 - \frac{N_v}{N_{ges}}$.

Abbildung (3-3) zeigt die Abhängigkeit der Niveaumächtigkeit von dem Abstand des Niveaus vom Niveau 1 (Niveau des Ausgangsmoleküls).



Niveau i :	1	2	3	4	5	6	7	8	9	10
Moleküle M_i :	1	5	25	124	601	2596	7296	9736	9923	9930

Abb. (3-3) Niveaumächtigkeit von einseitig erzeugten Netzwerken.

Um ein Produkt im zehnten Niveau zu finden, sind $\sum_{i=1}^9 M_i = 30280$ Moleküle zu entwickeln. Von dieser großen Anzahl von Molekülen sind alle Nachfolger zu ermitteln und in die jeweils nächsthöheren Niveaus einzutragen.

3.3 *Bilaterale Erzeugung von Reaktionsnetzwerken*

Eine wirksame Reduktion der Anzahl der zu erzeugenden Intermediate kann durch Ausnutzung der Symmetrie der Reaktionserzeugung erlangt werden. Statt einen Reaktionsbaum in einer Richtung vom Edukt aus bis hin zum Produkt zu entwickeln, baut *RAIN* zwei Reaktionsbäume auf, einer vom Edukt, der andere vom Produkt ausgehend. Die Entwicklung dieser beiden Bäume geschieht unabhängig voneinander. Die Erstellung des Produktbaums unterscheidet sich von der des Eduktbaums lediglich dadurch, daß hier die Matrizen der Valenzschemakonvertierbarkeit (siehe 4.5 und 4.6) von den Spalten her (also an der Hauptdiagonalen gespiegelt) interpretiert werden. Sind die Matrizen für alle Elemente symmetrisch aufgefüllt, so verschwindet auch dieser Unterschied.

Ist ein Niveau eines Baums komplett entwickelt, d.h. daß alle Nachfolger (bzw. Vorgänger) der Moleküle im höchsten Niveau erzeugt sind, so können diese Moleküle des Nachfolge-Niveaus im höchsten Niveau des jeweils anderen Reaktionsbaums gesucht werden. Werden keine äquivalenten Moleküle entdeckt, muß einer der beiden Bäume um ein Niveau weiterentwickelt werden. Dieses Verfahren setzt sich so lange fort, bis jeweils gleiche Moleküle in den Bäumen auftauchen, oder eine entsprechende Benutzerinteraktion den Abbruch erzwingt.

Sobald in den am weitesten entwickelten Niveaus der beiden Reaktionsbäume übereinstimmende Moleküle gefunden werden, existieren Pfade zwischen Edukt und Produkt der Gesamtreaktion. Alle diese Reaktionspfade verlaufen über diese gemeinsamen Moleküle im Edukt- und Produktbaum. Abbildung (3-4) zeigt schematisch ein bilateral erzeugtes Reaktionsnetzwerk.

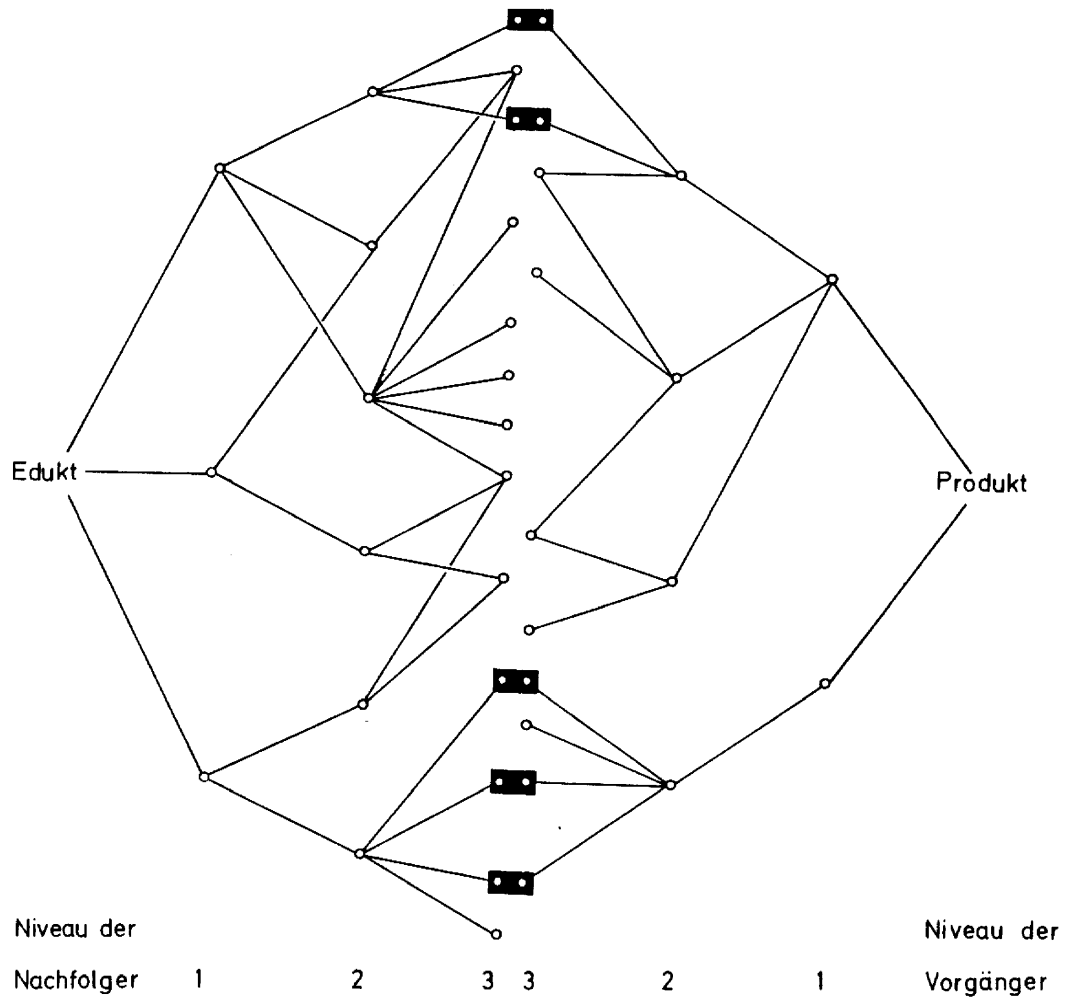


Abb. (3-4) Bilateral entwickeltes Reaktionsnetzwerk.

Nach Streichung terminaler Äste in beiden Reaktionsbäumen bleibt ein Netzwerk übrig, das dem gleich ist, das durch einseitige Entwicklung, entweder des Eduktbaums bis hin zum Produkt, oder des Produktbaums bis hin zum Edukt, erhalten worden wäre. Abbildung (3-5) zeigt das verbleibende, geschlossene Netzwerk von Abbildung (3-4).

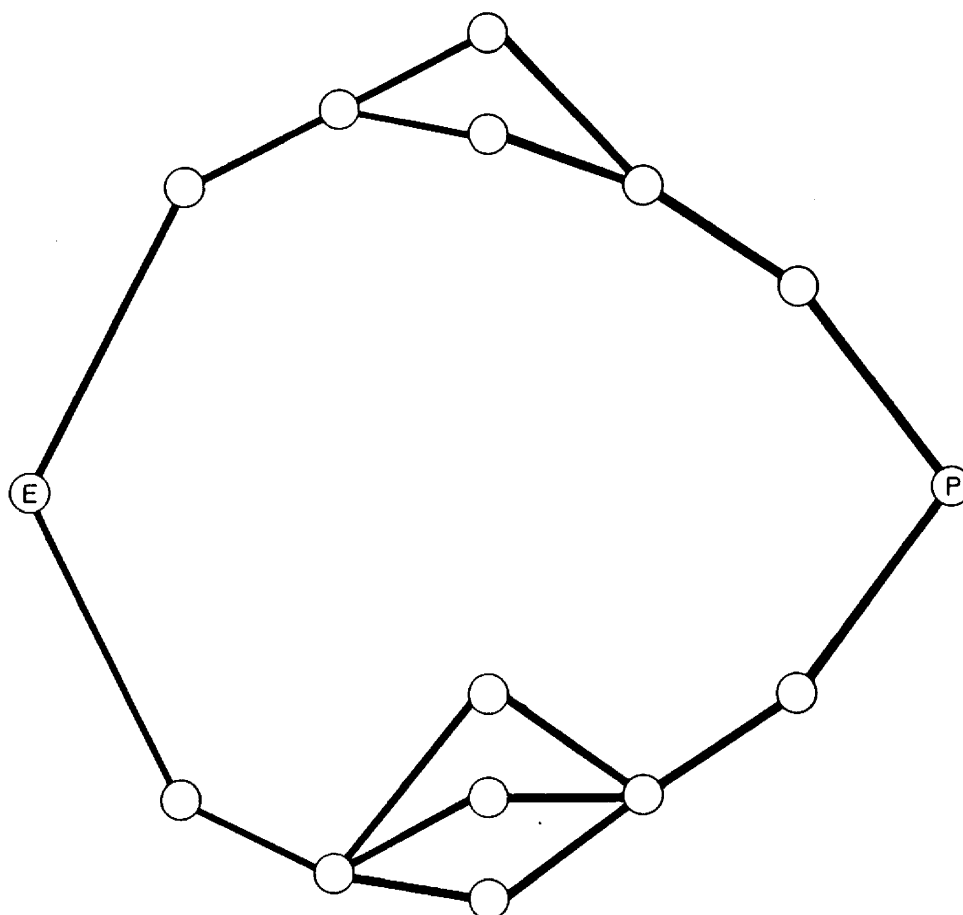


Abb. (3-5) Geschlossenes Netzwerk.

Dem Anwender bleibt es jedoch unbenommen, auch jetzt noch die Reaktionsbäume weiter zu entwickeln, um Netzwerke zu erhalten, die mehr als die minimal nötige Anzahl von Reaktionsschritten repräsentieren.

Bei der bilateralen Vorgehensweise zur Generierung eines N -Niveau Netzwerkes sind im allgemeinen in jedem der beiden Bäume nur $\frac{N}{2} + 1$ Niveaus zu entwickeln. Die dadurch erzielte Verringerung der Anzahl der zu behandelnden Zwischenstufen ist beträchtlich. Abbildung (3-6) verdeutlicht dies (Annahmen wie bei Abbildung (3-3)). Um in diesem Fall das Netzwerk zu schließen, sind nun nicht 30280, sondern lediglich $2 \cdot \sum_{i=1}^5 M_i = 1512$ Moleküle zu entwickeln.

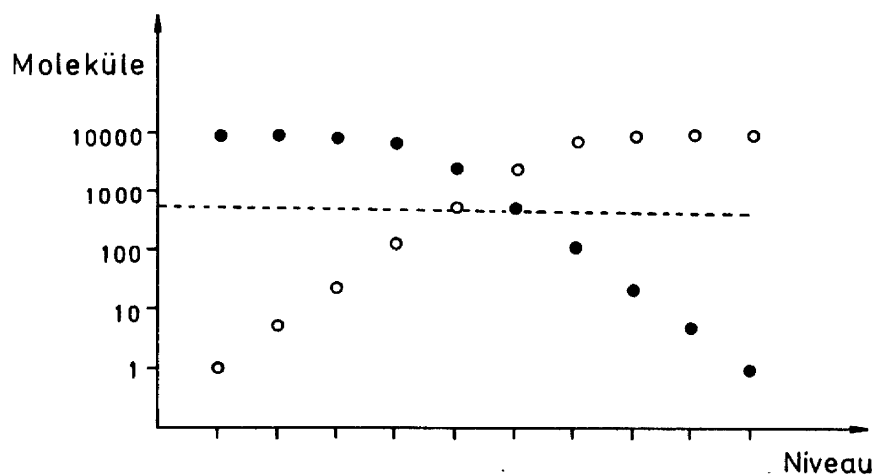


Abb. (3-6) Niveaumächtigkeit von bilateralen erzeugten Netzwerken.

Statt beide Reaktionsbäume bis hin zur Mitte des Netzwerks zu generieren, kann auch einer der beiden Bäume, in asymmetrischer Weise, weiter entwickelt werden als der andere. Eine Indikation für dieses Vorgehen ist dann gegeben, wenn, bedingt durch strukturelle Voraussetzungen oder starke Asymmetrie der Valenzschemakonvertierbarkeitsmatrizen, einer der beiden Bäume wesentlich schneller anwächst als der andere. In diesem Fall kann es von Vorteil sein, den langsam anwachsenden Reaktionsbaum einige Niveaus weiter zu entwickeln, als den stark wachsenden. Die Verknüpfungstelle befindet sich dann nicht "auf halber Strecke" im Netzwerk, sondern ist in Richtung auf die stärker wachsende Seite hin verschoben. Auf die Struktur und den Inhalt des entstehenden Netzwerks hat dies jedoch keinerlei Auswirkung, denn alle Generierungsoperationen sind symmetrisch.

4 *Der chemische Ereignishorizont*

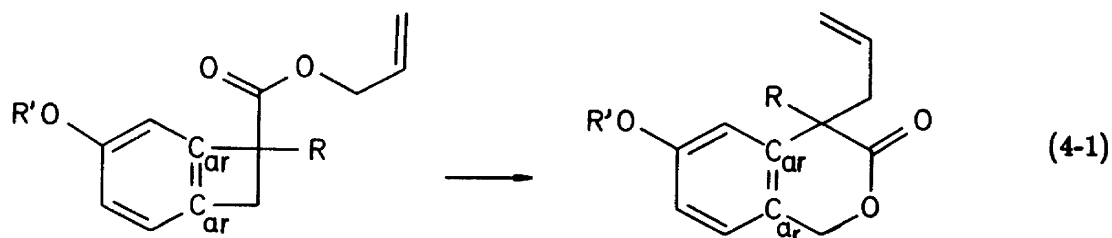
4.1 *Definition der globalen Gesamtreaktion*

Das Edukt-Produkt-Paar einer Reaktion oder Reaktionsfolge legt die äußersten Randknoten des von *RAIN* aufzuspannenden Reaktionsnetzwerks fest. Im Prinzip kann jede beliebige Reaktion mit eindeutigen Bindungsordnungen in Edukt und Produkt dafür dienen. Die Gesamtreaktion zwischen Edukt und Produkt muß lediglich stöchiometrisch ausgeglichen sein, d.h. die Anzahl der Atome einer Elementsorte müssen einander in Edukt und Produkt entsprechen (Gesetz der Erhaltung der Atomzahl). Dasselbe gilt auch für die Zahl der Elektronen in Bindungen und freien Elektronen (Gesetz der Ladungserhaltung).

4.2 *Elemente und Pseudoelemente*

Die in Edukt und Produkt vorkommenden chemischen Elemente können beliebiger Art sein. Sie werden durch ein Elementsymbol bezeichnet, das sowohl das Symbol eines Elements aus dem Periodensystem sein kann, als auch eine vom Anwender beliebig gewählte Zeichenfolge.

Alle Elementsorten, die zur Anwendung kommen, müssen jeweils durch Valenzschemata und eine Matrix der Valenzschemakonvertierbarkeit (siehe 4.5 und 4.6) charakterisiert werden. Für spezielle Anwendungen können auf diese Weise auch Pseudoelemente mit definierten Eigenschaften kreiert werden. So werden im Beispiel (4-1) die Atome C_{ar} grundsätzlich, d.h. in allen Reaktionsintermediaten, sowie in Edukt und Produkt der Gesamtreaktion, als sp^2 -hybridisiert angenommen.

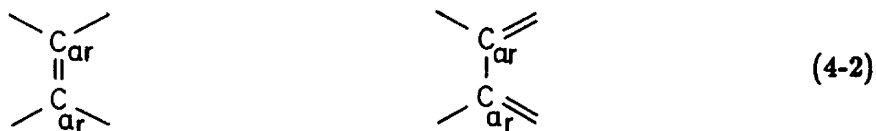


Sie stellen eine eigene Atomsorte dar, deren Eigenschaften (die möglichen Valenzschemata und deren Konvertierbarkeit) stark von denen der anderen C -Atome abweichen.

4.3 Minimale Bindungsordnung

Um gewisse gerüstspezifische Eigenheiten des reagierenden Systems vorgeben zu können, ist es möglich, für bestimmte Bindungen im Edukt und Produkt eine minimale Bindungsordnung anzugeben. Ist diese ungleich null, so werden die durch sie verbundenen Atome in keinem Reaktionsschritt getrennt. Dies ist insbesondere nützlich, um die Aromatizität in Ringen oder anderen konjugierten Systemen zu erhalten.

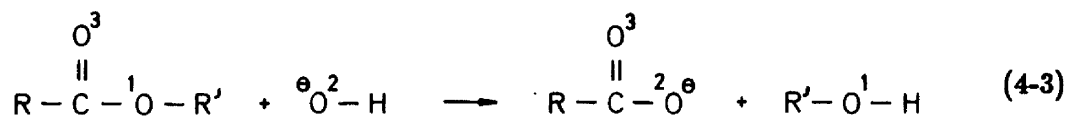
Wird z.B. die Bindung zwischen den Atomen C_{ar} in (4-1) mit der Minimalordnung 1 versehen, so bleibt der Benzolring in allen Intermediaten intakt (unter der Voraussetzung, daß die anderen Atome des Rings als an der Reaktion unbeteiligt markiert werden). Zusammen mit der Forderung, daß C_{ar} grundsätzlich sp^2 -hybridisierte Atome sein sollen, bleiben für diese Gruppierung in allen Reaktionszwischenstufen nur die Alternativen in (4-2).



4.4 Atom-auf-Atom-Abbildung von Edukt und Produkt

Grundsätzlich ist es für *RAIN* nicht nötig, daß die Atome des Edukts individuell den Atomen des Produkts zugeordnet werden. Diese Zuordnung ist, insbesondere bei komplexeren mechanistischen Studien oder biochemischen Auf- oder Abbauwegen, auch nur selten vorgegeben. Sind jedoch, z.B. aus Markierungsexperimenten, individuelle Atomzuordnungen bekannt, so kann dies mit in die Beschreibung der Reaktion aufgenommen werden. Dies ist jedoch nur dann sinnvoll, wenn die Zuordnung nicht schon dadurch eindeutig gegeben ist, daß diese Atome jeweils die einzigen Vertreter ihrer Elementsorte sind.

Atome im Edukt, zu denen die Korrespondenzatome im Produkt bekannt sind, werden dazu, sowohl im Edukt als auch im Produkt, mit einer eindeutigen Nummer versehen. Im Beispiel (4-3) sind die Sauerstoffatome auf diese Weise markiert und somit individuell unterscheidbar³⁷.



Eine mögliche automatische, vollständige Zuordnung der Atome im Edukt und Produkt, unter dem Gesichtspunkt der geringsten Anzahl der zu verschiebenden Elektronen, erhält man durch die Anwendung des Prinzips der minimalen chemischen Distanz PMCD^{38,39} (siehe hierzu auch 7.1).

Die Einführung einer partiellen oder vollständigen Atom-auf-Atom-Abbildung kann zu einer Reduktion des resultierenden Netzwerks führen, da nur spezielle Reaktionssequenzen die markierten Atome im Edukt an die geforderten Positionen im Produkt bringen. Diese Reduktion trifft jedoch nicht zu für die beiden unabhängigen Edukt- und Produkt-Reaktionsbäume, die *RAIN* zur Aufspannung des Netzwerks erzeugt. Da die markierten Atome individualisiert sind, kommt es im Gegenteil zu einer weiteren Aufspaltung der Reaktionsbäume. Jede unterscheidbare Reihenfolge der markierten Atome in sonst identischen Molekülen führt zu einem individuellen, neuen Eintrag in dem entsprechenden Baum (siehe 5.2).

Für den Fall einer vollständigen Zuordnung aller Atome ist ein anderes Verfahren zur Erzeugung eines Reaktionsnetzwerks denkbar als das, welches mit *RAIN* hier vorgestellt wird. Dieses Verfahren analysiert die durch die vollständige Zuordnung der Atome resultierende *R*-Matrix der Gesamtreaktion. Diese wird in Matrixsummanden zerlegt, die dann auf alle Weisen kombiniert werden, die nach den valenzchemischen Randbedingungen erlaubt sind. Hierbei können die Null-Einträge der *R*-Matrix entweder grundsätzlich beibehalten werden, oder in Einträge gleichen Betrags, von gegengesetztem Vorzeichen, aufgelöst werden. Im ersten Fall bekommt man nur Reaktionspfade mit der gleichen Anzahl von bewegten Elektronen, wie sie die chemische Distanz von Edukt und Produkt vorgibt, bzw. gar keine Lösungen, wenn aus Gründen der Valenzchemie oder konstitutioneller Randbedingungen bestimmte — in diesem Fall notwendige — Schlüsselreaktionsschritte nicht erlaubt sind. Im zweiten Fall können auch "Umwege" begangen werden, die die Reaktionsbäume ebenfalls wieder stark erweitern, in vielen Fällen jedoch zu interessanten und plausiblen Reaktionspfaden führen*.

In der Praxis hat es sich jedoch herausgestellt, daß der hier mit *RAIN* vorgestellte Lösungsansatz des Netzwerksproblems sehr gut und praktikabel ist und mit weitaus weniger Beschränkungen arbeitet als ein System mit den eben angeführten Eigenschaften.

* Ein Programm dieser Art ist Gegenstand weiterer Untersuchungen.

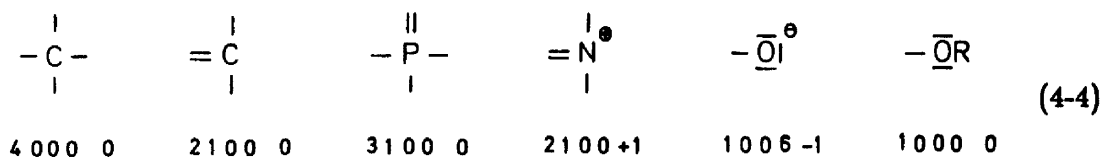
4.5 Valenzschemata

Valenzschemata dienen in *RAIN* zur Definition der Eigenschaften der an einer Reaktion beteiligten chemischen Elemente oder Pseudoelemente. Ein Valenzschema beschreibt, wie die einzelnen Valenzelektronen eines Atoms auf verschiedene Bindungen verschiedener Ordnung und auf freie Elektronen verteilt sind. Das Valenzschema eines Atoms wird erhalten, in dem das Atom mit seinen Bindungen und freien Valenzen aus dem Molekülverband herausgelöst wird, ohne die gebundenen Nachbarn in ihrer Identität zu berücksichtigen.

Die numerische Beschreibung eines Valenzschemas erfolgt in Form eines Zahlenquintupels (v_s, v_d, v_t, v_f, v_l) mit

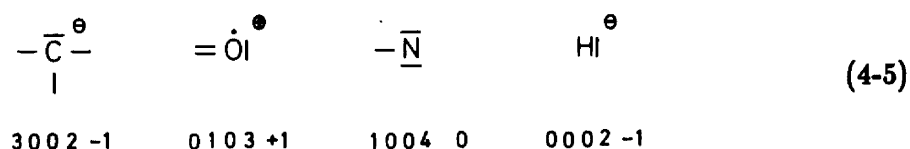
v_s	:	Anzahl der Einfachbindungen des Atoms
v_d	:	Anzahl der Doppelbindungen des Atoms
v_t	:	Anzahl der Dreifachbindungen des Atoms
v_f	:	Anzahl der freien Elektronen des Atoms
v_l	:	Formalladung des Atoms

Beispiele :



Diese Definition und Darstellung der Valenzschemata unterscheiden sich von denen, wie sie in früheren Publikationen gegeben wurden^{14,15}. Die Reihenfolge der numerischen Beschreibung ist verändert und die Formalladung wird als fünfter Zahlenwert aufgenommen. Die Formalladung könnte zwar aus den Bindungszahlen, den freien Elektronen und der Ordnungszahl errechnet werden. Da jedoch für Pseudoelemente keine Ordnungszahlen definiert sind, wird sie mit in das die lokalen Eigenschaften eines Elements beschreibende Valenzschema aufgenommen.

Für alle Elemente oder Pseudoelemente, die an den Reaktionsschritten beteiligt sind (Reaktionskern), wird ein Satz von Valenzschemata angegeben, in denen sie auftreten können. Damit wird in einer sehr umfassenden Weise der chemische Ereignishorizont beschrieben, innerhalb dessen das Reaktionsnetzwerk aufgespannt werden kann. Die Wahl der zulässigen Valenzschemata beeinflusst die Natur der möglichen Intermediate. So erlaubt z.B. die Aufnahme geladener, radikalischer oder elektronendefizienter Valenzschemata, wie sie in Beispiel (4-5) gezeigt werden, die Studie von Reaktionsmechanismen mit reaktiven Zwischenstufen.



4.6 Konvertierbarkeit der Valenzschemata

Um dem Anwender auch die Möglichkeit zu geben, das dynamische Verhalten der Elemente zu beeinflussen, wird zusätzlich zu der Angabe der möglichen Valenzschemata eines Elements eine Matrix der Konvertierbarkeit K_V definiert. Diese Matrix legt fest, in welche Valenzschemata jedes Valenzschema eines Elements innerhalb eines Reaktionsschritts übergehen darf. Hierzu bilden die einem Element zugeordneten Valenzschemata die Zeilen i und Spalten j der quadratischen Matrix K_V .

Es gilt dann folgendes :

$$K_{V,i,j} = \begin{cases} + & \text{Übergang zwischen den Valenzschemata } i \text{ und } j \text{ ist erlaubt} \\ - & \text{Übergang zwischen den Valenzschemata } i \text{ und } j \text{ ist verboten} \end{cases}$$

Diese Matrix ist nicht notwendigerweise symmetrisch. Somit ist es möglich, das "Schicksal" von Valenzschemata entlang der Reaktionspfade innerhalb der definierten Reaktionsrichtung vom Edukt zum Produkt festzulegen.

Abbildung (4-6) zeigt eine Matrix K_V für das Element O.

O	$\rightarrow \bar{O} -$	$\rightarrow \bar{O} =$	$\rightarrow O^+ -$	$\rightarrow O^+ \equiv$	$\rightarrow \bar{O} \vdash$	$\rightarrow \dot{O}^+ =$
$\rightarrow \bar{O} -$	+	+	+	-	+	-
$\rightarrow \bar{O} =$	+	+	+	+	+	-
$\rightarrow O^+ -$	-	+	-	-	-	+
$\rightarrow O^+ \equiv$	-	+	+	-	-	-
$\rightarrow \bar{O} \vdash$	+	+	-	-	+	-
$\rightarrow \dot{O}^+ =$	-	-	+	-	-	-

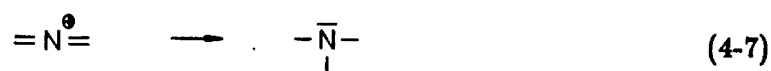
Abb. (4-6) Matrix der Valenzschemakonvertierbarkeit.

Die Stabilität eines Valenzschemas i findet ihren Ausdruck in dem diagonalen Eintrag $K_{V,i,i}$:

$$K_{V,i,i} = \begin{cases} + & \text{das Valenzschema } i \text{ ist stabil (kann erhalten bleiben)} \\ - & \text{das Valenzschema } i \text{ ist instabil (muß abreagieren)} \end{cases}$$

Dies gibt die Möglichkeit, hochreaktive Valenzschemas wie die der Carbene, Nitrene oder Radikale zu definieren, die, sobald sie in einem Molekül auftauchen, im nächsten Reaktionsschritt sofort in ein (stabileres) Valenzschema umgewandelt werden.

Die Matrixeinträge in K_V werden vom Anwender vorgegeben. Es kann jedoch vorkommen, daß Einschränkungen für R - oder T -Matrizen gegeben werden, die bestimmte, vom Benutzer erlaubte Valenzschemakonversionen verbieten. So erfordert z.B. der Übergang der Valenzschemas in (4-7) mindestens drei außerdiagonale Einträge in einer R -Matrixzeile.



Dies könnte ein gesetztes Limit für die entsprechende Randbedingung überschreiten. In diesem Fall wird von *RAIN* die Matrix der Konvertierbarkeit für das betreffende Element automatisch so verändert, daß diese den Randbedingungen der R - und T -Matrizen genügen (siehe 4.9). Auf diese Weise wird von vornherein vermieden, daß, im Zuge der Reaktionsgenerierung, Valenzschemakonversionen in Betracht gezogen werden, die in jedem Fall zu ungültigen R - oder T -Matrizen führen würden.

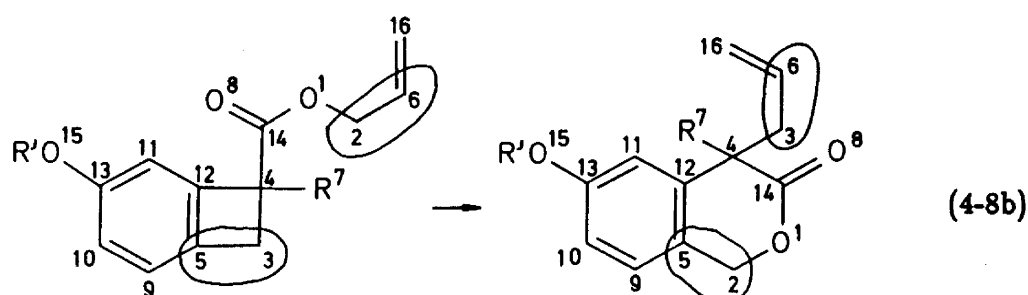
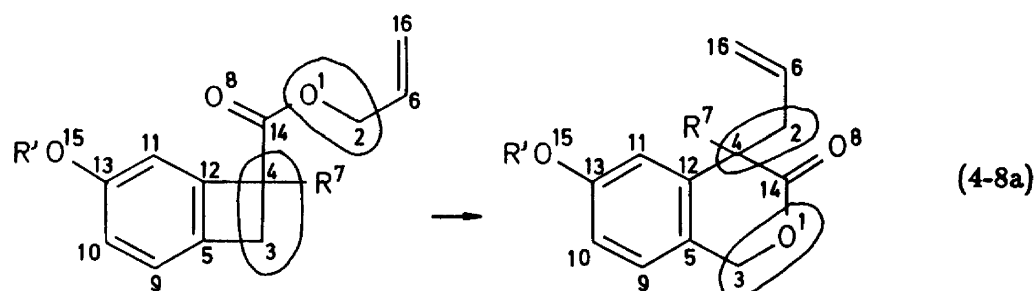
4.7 *Definition des Reaktionskerns*

Die Anzahl der formalen Reaktionsmöglichkeiten eines Moleküls hängt in starkem Maße von der Anzahl der reaktiven Atome im Molekül ab. Die Grenze der von heutigen Rechenanlagen mit kombinatorischen Verfahren vollständig bewältigbaren Molekülgrößen liegt ungefähr bei 20 bis 30 Atomen. Da jedoch in den meisten Fällen nur Teile der Moleküle an den Reaktionsschritten einer Reaktionsfolge beteiligt sind, während andere Teile über die gesamte Reaktion hinweg unverändert erhalten bleiben, ist es möglich, auch Reaktionsnetzwerke größerer Moleküle aufzuspannen, ohne eine "kombinatorische Explosion" auszulösen.

Die Auswahl des Reaktionskerns, d.h. der Menge von Atomen, deren Bindungen und freie Valenzen sich im Zuge einer Reaktionsfolge ändern, kann auf zwei Arten geschehen.

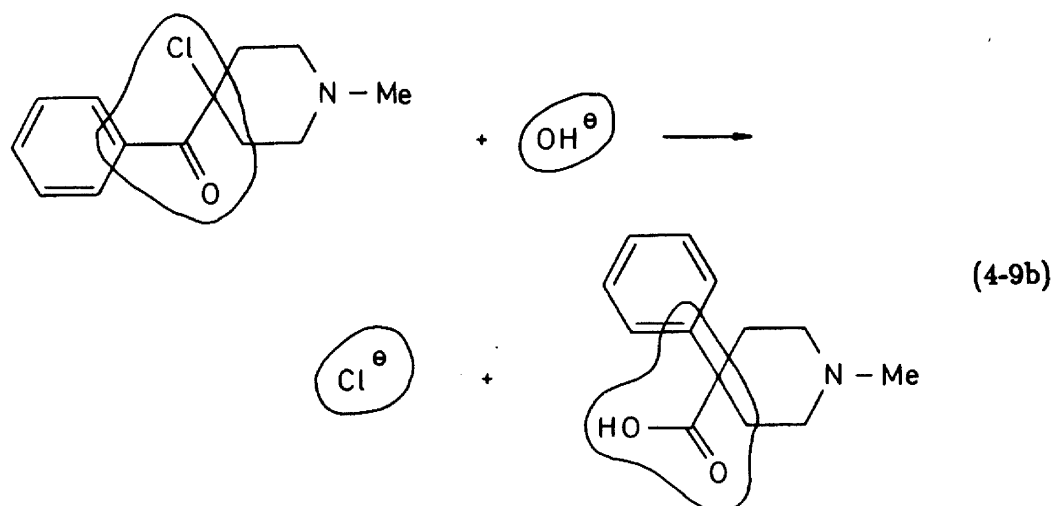
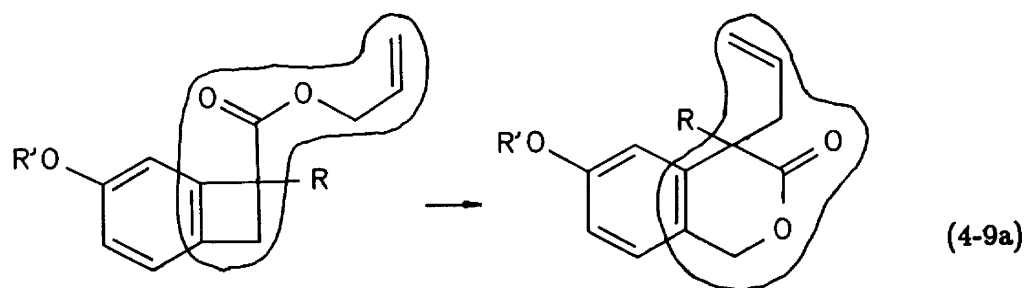
4.7.1 *Automatische Bestimmung des Reaktionskerns*

Das Programm PEMCD³⁹ bildet die Atome des Edukts auf die Atome des Produkts hinsichtlich minimaler Anzahl der zu bewegenden Elektronen ab. Der Reaktionskern wird hier also derart bestimmt, daß die Menge der unveränderten Strukturteile möglichst groß ist. Oftmals gibt es mehr als eine Zuordnungsmöglichkeit unter Beachtung dieses Prinzips. Die Beispiele (4-8a und 4-8b) zeigen diese Zuordnungsmöglichkeiten und (jeweils hervorgehoben) die entsprechenden Reaktionskerne. Diese Nichteindeutigkeit der Abbildungsvorschrift führt zu einem schwerwiegenden Nachteil (siehe 6.1).



4.7.2 Heuristische Bestimmung des Reaktionskerns

RAIN bietet dem Anwender die Möglichkeit, jede beliebige Menge von Atomen im Edukt und Produkt als Reaktionskern zu definieren. Es müssen nur die Atome markiert werden, die an den Reaktionsschritten beteiligt sein sollen. Hierbei ist jedoch wichtig, daß sowohl die Anzahl der Atome einer Elementsorte im Reaktionskern auf Edukt- und Produktseite, als auch die der Elektronen in Bindungen und freien Valenzen dieser Atome jeweils gleich sind. Das Beispiel (4-9a und 4-9b) zeigt Reaktionen mit vom Anwender heuristisch definierten Reaktionskernen.



Mit der heuristischen Definition des Reaktionskerns ist es möglich, auch Atom-auf-Atom-Abbildungen zuzulassen, die nicht dem PMCD gehorchen, was in vielen Fällen zu plausibleren Reaktionsfolgen führt. Durch Einbeziehung von eigentlich unveränderten Atomen ist es oft erreichbar, mechanistisch interessante Zwischenstufen zu erhalten, die in Reaktionswegen mit der minimalen chemischen Distanz nicht auftauchen.

4.8 Grenzwerte für *R*- und *T*-Matrizen

Die Definition der maximalen Komplexität eines einzelnen Reaktionsschritts ist von zentraler Bedeutung bei der Reaktionsgenerierung in *RAIN*. Sie wird dadurch erreicht, daß für die Transformationsmatrizen vom Typ *R* und *T* Grenzwerte hinsichtlich Rang, Anzahl außerdiagonaler Einträge u.ä. gesetzt werden können.

4.8.1 Maximaler Rang einer irreduziblen *R*-Matrix

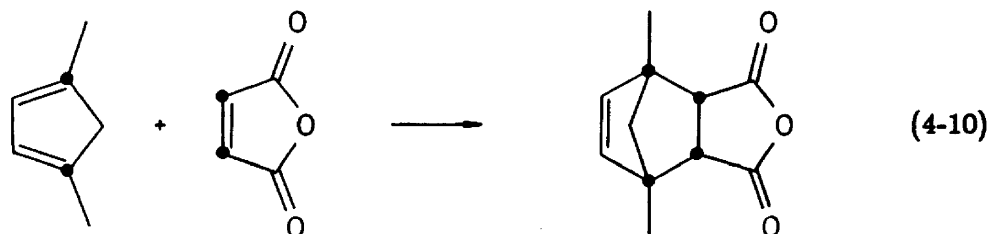
Diese Randbedingung bestimmt die maximale Anzahl von Atomen, die an einem Reaktionsschritt beteiligt sein dürfen. Diese Atome können bestehende Bindungen brechen, neue Bindungen knüpfen, Bindungsordnungen ändern, oder freie Elektronen hinzugewinnen oder verlieren. Um z.B. auch Reaktionsschritte vom *Diels-Alder*-Typ (4-10) zu erhalten, müssen bis zu 6 Atome in einer irreduziblen *R*-Matrix vorkommen dürfen.

4.8.2 Maximale Anzahl von außerdiagonalen Einträgen in einem *R*-Dreieck

Diese Randbedingung bestimmt die maximale Anzahl von Bindungen, die an einem Reaktionsschritt beteiligt sein dürfen. Diese Bindungen können sich lediglich in ihrer formalen Ordnung ändern, oder aber auch neu gebildet oder ganz gebrochen werden.

4.8.3 Maximaler Rang einer irreduziblen *T*-Matrix

Diese Randbedingung bestimmt die maximale Anzahl von Atomen, die im Verlauf eines Reaktionsschritts Bindungen zu neuen Nachbarn knüpfen bzw. bestehende Bindungen zu alten Nachbarn aufgeben dürfen. Hiermit wird also (in Kombination mit 4.8.4, 4.8.8, 4.8.9 und 4.8.10) festgelegt, wie groß die topologischen Gerüstveränderungen maximal sein dürfen. Die *Diels-Alder*-Reaktion z.B. beteiligt 4 Atome in der irreduziblen *T*-Matrix. Sie sind im Beispiel (4-10) markiert.

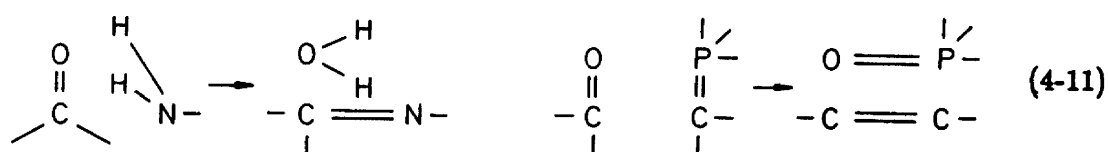


4.8.4 Maximale Anzahl von außerdiagonalen Einträgen in einem *T*-Dreieck

Diese Randbedingung bestimmt die maximale Anzahl von Bindungen, die in einem Reaktionsschritt gebrochen oder neu geknüpft werden dürfen. Dieser Wert ist der wichtigste bei der Limitierung der topologischen Komplexität eines Reaktionsschrittes, da er die maximalen Gerüstveränderungen exakt festlegt. Die *Diels-Alder*-Reaktion in (4-10) beteiligt zwei Bindungen in der *T*-Matrix.

4.8.5 Maximaler Betrag eines außerdiagonalen Eintrags in einer *R*-Matrix

Diese Randbedingung bestimmt die maximale Änderung einer Bindungsordnung im Verlauf eines Reaktionsschritts. Der Wert 1 an dieser Stelle bewirkt, daß Reaktionen, die in einem Schritt Doppelbindungen brechen oder knüpfen, nicht erlaubt sind. Liegt das Augenmerk jedoch nicht auf der Erzeugung eines mechanistischen Netzwerks, sondern ist der Anwender an der Einbeziehung komplexerer synthetischer Reaktionsschritte interessiert, ist hier ein Wert ≥ 2 anzugeben. Hiermit werden Reaktionsschritte mit einbezogen, die selbst wieder in mehrere mechanistische Einzelschritte zerlegt werden können. Beispiel (4-11) zeigt solche Reaktionen.

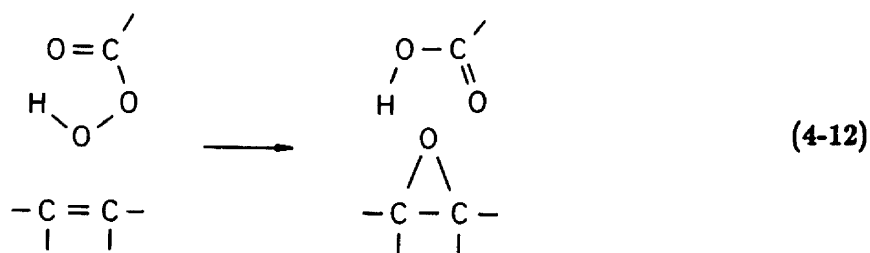


4.8.6 Maximaler Betrag eines außerdiagonalen Eintrags in einer T-Matrix

T-Matrizen haben gemäß ihrer Definition außerdiagonale Einträge, die vom Betrag nie größer sind als 1.

4.8.7 Maximale Anzahl von außerdiagonalen Einträgen in einer R-Zeile

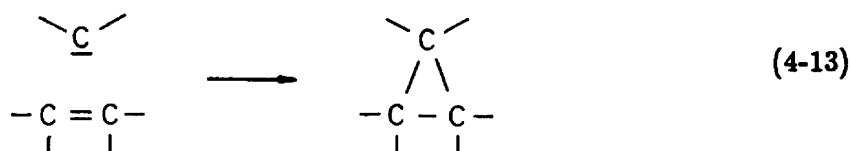
Diese Randbedingung bestimmt die maximale Anzahl von Bindungen an einem Atom, die innerhalb eines Reaktionsschritts in ihrer Ordnung verändert werden dürfen. Reaktionsschritte mit mehr als 2 veränderten Bindungen an einem Atom sind selten. Ein Beispiel ist die Epoxidierung mit Persäuren (4-12), doch kann auch diese Reaktion wieder in mehrere, einfachere Reaktionschritte zerlegt werden.



Sollen auch Reaktionen dieses Typs als gültige Reaktionschritte zugelassen werden, so ist die maximale Anzahl von außerdiagonalen *R*-Zeileneinträgen z.B. auf 4 zu setzen.

4.8.8 Maximale Anzahl von außerdiagonalen Einträgen in einer *T*-Zeile

Diese Randbedingung bestimmt die maximale Anzahl von Bindungen an einem Atom, die innerhalb eines Reaktionsschritts neu geknüpft oder gebrochen werden dürfen. Beispiel (4-13) zeigt eine Reaktion, in der ein Carben-C-Atom zwei neue Bindungen knüpft. Für das zentrale O-Atom im Beispiel (4-12) beträgt dieser Wert 4.

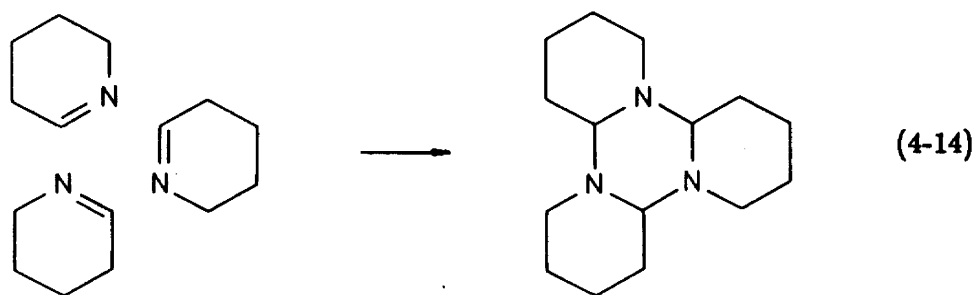


4.8.9 Maximaler Betrag einer *T*-Zeilensumme

Diese Randbedingung bestimmt die größten topologischen Veränderungen an einem Atom, die innerhalb eines Reaktionsschritts geschehen dürfen. Im Gegensatz zu (4.8.8) werden hier jedoch die Bindungsbrüche und -neuknüpfungen gegeneinander aufgerechnet. So hat die *T*-Zeilensumme für das Carben-C-Atom in Beispiel (4-13) den Wert 2, während sie für das zentrale O-Atom des Beispiels (4-12) den Wert 0 annimmt, da hier 2 Bindungen gebrochen und 2 Bindungen neu geknüpft werden.

4.8.10 Maximaler Betrag einer T-Matrixgesamtsumme

Diese Randbedingung bestimmt die maximale Nettoveränderung der topologischen Komplexität eines einzelnen Reaktionsschritts. Werden in einem Molekülgerüst in einem Schritt sehr viele zusätzliche Bindungen eingeführt, so ist dies aus der Sicht der Entropie meist sehr ungünstig. Trotzdem sind Reaktionen bekannt, in denen die Zeilen/Spaltensumme der zugehörigen T-Matrix einen Wert von 3 annimmt. Beispiel (4-14) zeigt einen Vertreter solcher Reaktionen. Auch hier stellt sich die Frage, ob Reaktionen dieser Art wirklich konzertiert ablaufen, oder sich in einzelne Reaktionsschritte gliedern lassen.



4.9 *Einfluß der R- und T-Grenzwerte auf die Valenzschemakonvertierbarkeit*

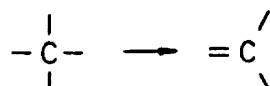
Die maximale Anzahl der außerdiagonalen Einträge in R - und T -Matrixzeilen beeinflussen unmittelbar die möglichen Belegungen in den Matrizen der Valenzschemakonvertierbarkeit (siehe 4.6). Die Überprüfung der vom Anwender angegebenen Möglichkeiten der Konversion eines Valenzschemas i in ein Valenzschema j auf die Einhaltung der Grenzwerte für R - und T -Matrizen geschieht auf folgende Weise :

- a) Aus den Valenzschemata i und j werden die charakteristischen Vektoren C_V der Länge n gebildet⁷. Die charakteristischen Vektoren stellen jeweils eine ausgewählte BE -Zeile dar, die aus dem Valenzschema erstellt werden kann.
- b) Aus der Differenz der beiden charakteristischen Vektoren wird eine R -Zeile gebildet und diese auf die Anzahl und den Betrag der von null verschiedenen Einträge hin überprüft und mit den Randbedingungen für R -Matrixzeilen verglichen.
- c) Aus den charakteristischen Vektoren werden "Adjazenzzeilen" gebildet, in dem jeder von null verschiedene Eintrag durch eine 1 ausgetauscht wird.
- d) Aus der Differenz der in c) erhaltenen Zeilen wird eine T -Zeile gebildet und diese auf die Anzahl und den Betrag der von null verschiedenen Einträge hin überprüft und mit den Randbedingungen für T -Matrixzeilen verglichen.
- e) Nur Valenzübergänge, die in b) und d) den Randbedingungen gehorchen, werden zugelassen.

Beispiel mit folgenden Valenzübergängen :



(4-15a)



(4-15b)

Es sei u.a. folgende Randbedingung gegeben :

Maximale Anzahl von außerdiagonalen Einträgen in R -Zeilen : 2

Valenzschemaübergang	:	(4-15a)	(4-15b)
Valenzschemata i	:	4000 0	4000 0
Valenzschemata j	:	0200 0	2100 0
Char. Vektor i	:	1111	1111
Char. Vektor j	:	2200	1120
R -Zeile	:	11-1-1	001-1
T -Zeile	:	00-1-1	000-1

Der Valenzschemaübergang (4-15a) aktiviert 4 außerdiagonale R -Zeileneinträge, ist also nach den gegebenen Randbedingungen verboten, während (4-15b) mit nur 2 Einträgen erlaubt ist.

4.10 Konstitutionelle Randbedingungen

Die von *RAIN* erzeugten Intermediate enthalten unter Umständen komplexe, gespannte Ringsysteme, lange Ketten von Heteroatomen o.ä.. Um diese konstitutionellen Eigenschaften zu beeinflussen, können strukturelle Randbedingungen gesetzt werden. Diese sind größtenteils heuristischer Natur und erlauben es dem anwendenden Chemiker auf eine sehr wirksame Weise, das Reaktionsnetzwerk von erfahrungsgemäß unmöglichen Strukturen zu befreien und damit oft stark zu vereinfachen. Bei der Auswahl der Grenzparameter liegt das Augenmerk stets darauf, dem Anwender eine große Transparenz zu bieten und die Auswahlmechanismen nicht vollständig zu automatisieren, d.h. abschaltbar zu machen. Dies bedeutet, daß auch der an einer kombinatorisch kompletten Lösung, ohne heuristische Auswahlregeln, interessierte Anwender die Möglichkeit hat, mit *RAIN* einen Satz von Lösungen zu erarbeiten.

4.10.1 Gespannte Ringsysteme

Der in *RAIN* zur Anwendung kommende Algorithmus zur Erkennung der Ringe in einem Molekül liefert stets einen kompletten Satz von Ringen. Somit sind bekannte Verfahren zur Berechnung der Ringspannungsenergien nach einem Additivitätsschema⁴⁰ nicht anwendbar, da diese mit einem reduzierten Satz von Ringen arbeiten⁴¹.

Der komplette Satz von Ringen wird durch vollständiges Aufspannen eines Molekülbaumes, ausgehend von einem beliebig gewählten Wurzelatom, gewonnen. Das Beispiel (4-16) zeigt ein Molekülgerüst mit dem vollständig aufgespannten Baum. Die eingekreisten Nummern verweisen auf einen Ringschluß zu einem Atom im Pfad zur Wurzel des Molekülbaums⁴².

Die Ringspannung S_R wird nun so berechnet, daß die Spannungsenergie von kondensierten kleinen Ringen mit größerem Gewicht eingeht, als dies bei den bekannten Additivitätsschemata der Fall ist. Dies geschieht, um die hochgespannten System in ihrem Ringspannungswert besser von den weniger gespannten Ringverknüpfungen trennbar zu machen. Ein Molekül dessen Wert für S_R ein angebbares Limit übersteigt, wird nicht in das Netzwerk aufgenommen.

Die Ringspannung S_R wird wie folgt berechnet :

$$S_R = \sum_i s_i \quad \text{mit} \quad s_i = \sum_{j=3}^5 k_i \cdot q_{ij} \cdot a_{ij} \quad \text{mit} \quad k_i = \sum_{j=3}^5 a_{ij}$$

mit

s_i Anteil des Atoms i and der Gesamtringspannung.

a_{ij} Anzahl der Ringe mit der Größe j , in denen das Atom i enthalten ist.

q_{ij} Ringsspannungsanteil des Atoms i an einem Ring der Größe j in $\frac{1}{10} kcal/mol$.

Diese Werte entsprechen den Ringspannungsenergien⁴³ geteilt durch die zugehörigen Ringgrößen.

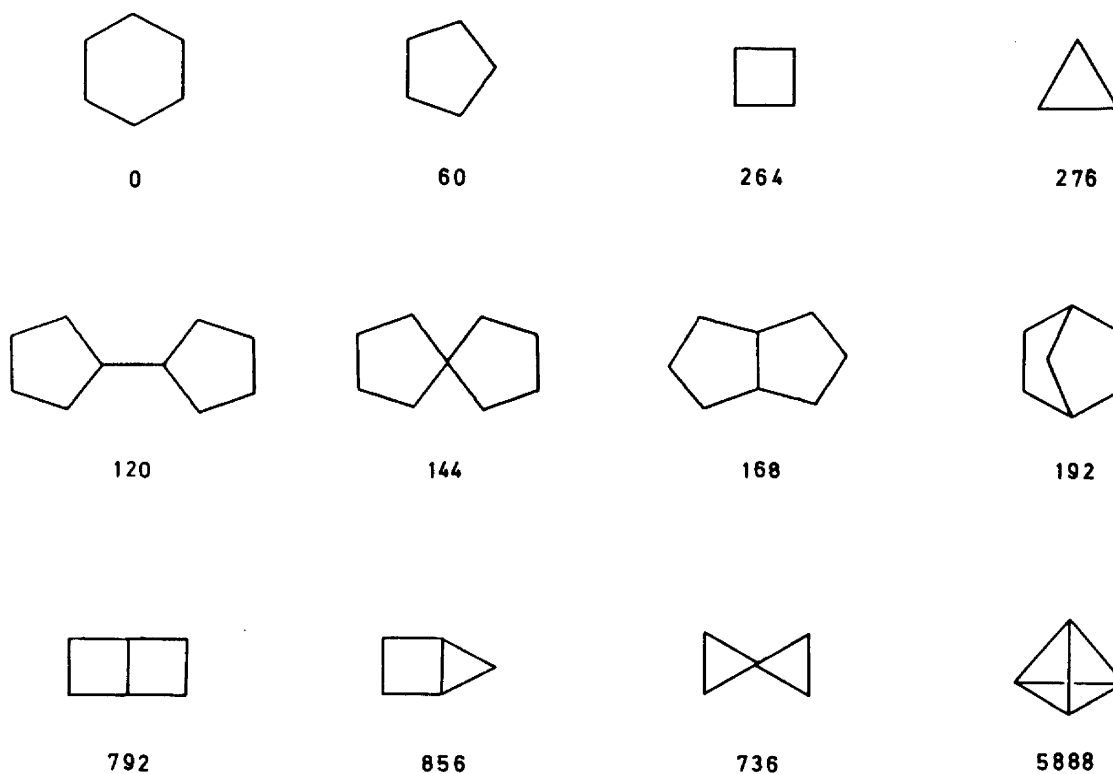
$$q_{i3} := 92$$

$$q_{i4} := 66$$

$$q_{i5} := 12$$

k_i Wichtungsfaktor des Atoms i für Einbettung in komplexe Ringsysteme.

Abbildung (4-18) zeigt einige Strukturen mit den zugehörigen Ringspannungen.

Abb. (4-18) Ringspannungen S_R .

4.10.2 Bredt'sche Regel

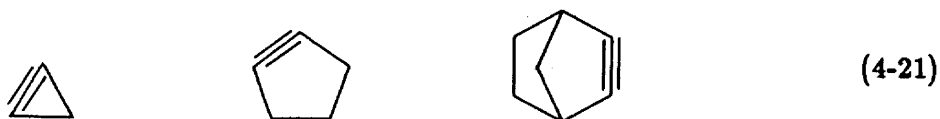
Molekülstrukturen, die nicht der Bredt'schen Regel⁴⁴ gehorchen, können aus dem Netzwerk ausgeschlossen werden. Hierbei gilt folgendes als definiert :

- Ein Atom, das mindestens 2 Ringen angehört, die mehr als 2 Atome gemeinsam haben, ist ein Brückenkopfatom.
- Brückenkopfatome dürfen keine Mehrfachbindungen ausbilden.

Die entsprechenden Ringgrößen werden nicht beachtet. Das Beispiel (4-19) zeigt einige nicht existenten Bredt-Verbindungen.

4.10.4 Dreifachbindungen in Ringen

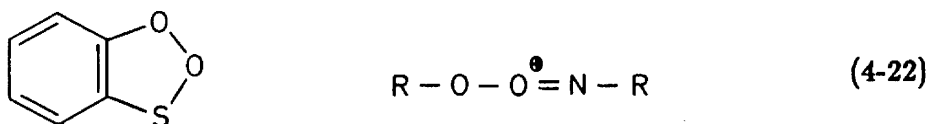
Tritt an einem Ringatom eine Dreifachbindung auf, so führt auch dies zu Strukturen mit sehr großer Ringspannung, die ebenfalls erst ab einer bestimmten Ringgröße von ca. 7 bis 8 Atomen in den Bereich der Stabilität geraten. Beispiel (4-21) zeigt einige hochgespannte Strukturen mit Dreifachbindungen.



Auch hier stellt die kleinste Ringgröße, ab der Dreifachbindungen in Ringen auftreten dürfen, den Grenzparameter in *RAIN* für dieses Strukturelement dar.

4.10.5 Maximale Länge von Heteroatomketten

Jedes der Elemente oder Pseudoelemente, die vom Anwender durch Symbol, Valenzschemata und deren Konvertierbarkeit definiert werden, kann als Heteroatom markiert werden. Um nun Verbindungen wie in Beispiel (4-22) angegeben, ausschließen zu können, wird eine größte Anzahl (maximal 3) gebundener Heteroatome definiert.



Bei der Erzeugung der *BE*-Zeilen (siehe 5.1) für ein Heteroatom wird jede *BE*-Zeile daraufhin überprüft, wieviele Bindungen zu anderen Heteroatomen sie knüpft. *BE*-Zeilen die zu viele Heteroatome verbinden, werden eliminiert.

4.10.6 Maximale Anzahl geladener Atome

Jedes Valenzschema der definierten Elemente und Pseudoelemente beinhaltet als fünften Zahlenwert die Information über die zugehörige Formalladung. Die Angabe einer maximalen Anzahl geladener Atome in einem *EM* führt bereits zu einem sehr frühen Zeitpunkt innerhalb der Reaktionsgenerierung (siehe 5.1) zu einem Abbruch der Entwicklung von Valenzschemakombinationen. Auf diese Weise werden Zwischenstufen mit zuviel Ladungen gar nicht erst erzeugt.

4.10.7 Benachbarte gleichnamige Ladungen

Bei der Entwicklung der *BE*-Zeilen kann, unter Zuhilfenahme der Information über die Formalladung der zugehörigen Valenzschemata, bereits darauf geachtet werden, daß es keine Bindungen zwischen Atomen gibt, deren Valenzschemata von null verschiedene Ladungen mit gleichem Vorzeichen aufweisen.

5 Reaktionserzeugung

Ein reaktionserzeugendes System generiert aus einem gegebenen Molekül (Edukt*) alle Moleküle (Produkte), die potentiell aus diesem Molekül in einem Reaktionsschritt erzeugt werden können. Ein symmetrisch ausgelegtes System kann auch alle Moleküle erzeugen, aus denen das gegebene Molekül hervorgehen kann.

Bei der Entwicklung von *RAIN* wurden an das reaktionsgenerierende Prinzip folgende Forderungen gestellt :

- 1) Die Reaktionsgenerierung sollte nicht auf der Basis von Individualreaktionen arbeiten. Ein derartiges System kann nur Reaktionen anwenden, die in einer Reaktionsdatei gespeichert sind. Umfangreiche Reaktionsdateien sind teuer und müssen ständig erweitert werden, um neuen Erkenntnissen gerecht zu werden. Die Erzeugung der Reaktionen sollte vielmehr nach rein formalen Gesichtspunkten, unter Beachtung von gut parametrisierbaren und verallgemeinerungsfähigen Randbedingungen, erfolgen.
- 2) Die Begrenzung der Reaktionsvielfalt sollte nicht durch Auswahl einiger *R*-Matrizen geschehen, wie es in bekannten formalen Reaktionsgeneratoren geschieht^{13,23-27}, sondern durch einen, die Kreativität des Systems viel weniger einschränkenden Satz von Randbedingungen für *R*-Matrizen.
- 3) Die erforderlichen Randbedingungen für die Reaktionsschritte sollten für den Anwender in ihrer potentiellen Auswirkung möglichst transparent sein.
- 4) Das reaktionserzeugende Prinzip sollte unabhängig von der Reaktionsrichtung sein, d.h. es sollten sowohl die Nachfolger, als auch die Vorgänger eines Moleküls erzeugt werden können. Für formale Reaktionsgeneratoren auf der Basis der Algebra

* Innerhalb dieses Kapitels werden die Begriffe "Edukt" und "Produkt" nicht im Sinne von Edukt und Produkt der Gesamtreaktion oder der Reaktionsfolge verwendet, sondern stehen für die Anfangs- und Endmoleküle eines einzelnen Reaktionsschritts im Netzwerk.

der *BE*- und *R*-Matrizen stellt diese Forderung i.a. keine Schwierigkeit dar, da die Anwendung der *R*-Matrizen nicht richtungsbezogen ist. Sobald jedoch Reaktivitätsanalysen von Bindungen²⁴ durchgeführt werden, wird eine Erzeugung der Vorstufen problematischer. Hier müssen erst alle möglichen Vorstufen erzeugt werden, von denen dann diejenigen verworfen werden, bei denen Bindungen beteiligt sind, für die das Programm eine niedrige Reaktivität vorhersagt.

- 5) Trotz der Forderungen 1) bis 4) sollte der Reaktionserzeugungsalgorithmus ein gutes Rechenzeitverhalten zeigen. Dies kann durch eine Konstruktion erreicht werden, die es erlaubt, jedwede Abbruchkriterien für kombinatorische Verfahren zum frühest möglichen Zeitpunkt wirksam werden zu lassen.
- 6) Der Reaktionsgenerator sollte als Modul in *RAIN* integriert und somit beliebig gegen Alternativen austauschbar sein.

Für ein formales reaktionserzeugendes Verfahren mit globalen Randbedingungen, das diesen Forderungen gehorcht, werden zwei prinzipiell verschiedene Algorithmen in Erwägung gezogen.

Algorithmus A :

- 1) Erzeuge nach den gegebenen Regeln für Valenzschemata, deren Konvertierbarkeit und unter Beachtung der konstitutionellen Randbedingungen die Gesamtheit aller Moleküle, die als Lösungsmenge in Frage kommen.
- 2) Berechne die *R*-Matrizen, die das Eduktmolekül in die unter 1) erzeugten Produktmoleküle transformieren und prüfe diese auf die Einhaltung der Regeln für *R*-Matrizen. Moleküle, deren *R*-Matrizen nicht alle Randbedingungen einhalten, werden verworfen.

Algorithmus B :

- 1) Erzeuge aus den Randbedingungen für die Reaktionsschritte alle möglichen *R*-Matrizen.

- 2) Wende diese *R*-Matrizen auf alle möglichen Weisen auf das Eduktmolekül an.
- 3) Prüfe die dadurch entstehenden Moleküle auf die Einhaltung der Regeln für Valenzschemata, deren Konvertierbarkeit und konstitutionelle Randbedingungen. Moleküle, die nicht alle Regeln erfüllen, werden verworfen.

Eine Entscheidung, welcher dieser Algorithmen der praktikablere sein wird, ist nur schwer zu treffen. Wahrscheinlich müßten beide vollständig implementiert werden, um dann vergleichende Aussagen über das Zeitverhalten machen zu können.

Die vorliegende Programmversion von *RAIN* arbeitet auf der Grundlage von Algorithmus A*. Allerdings werden bereits beim Aufbau der Moleküle die Regeln für die *R*- und *T*-Matrizen, so weit dies möglich ist, beachtet. Abbildung (5-1) zeigt eine Übersicht über das Ablaufschema des reaktionsgenerierenden Verfahrens von *RAIN*.

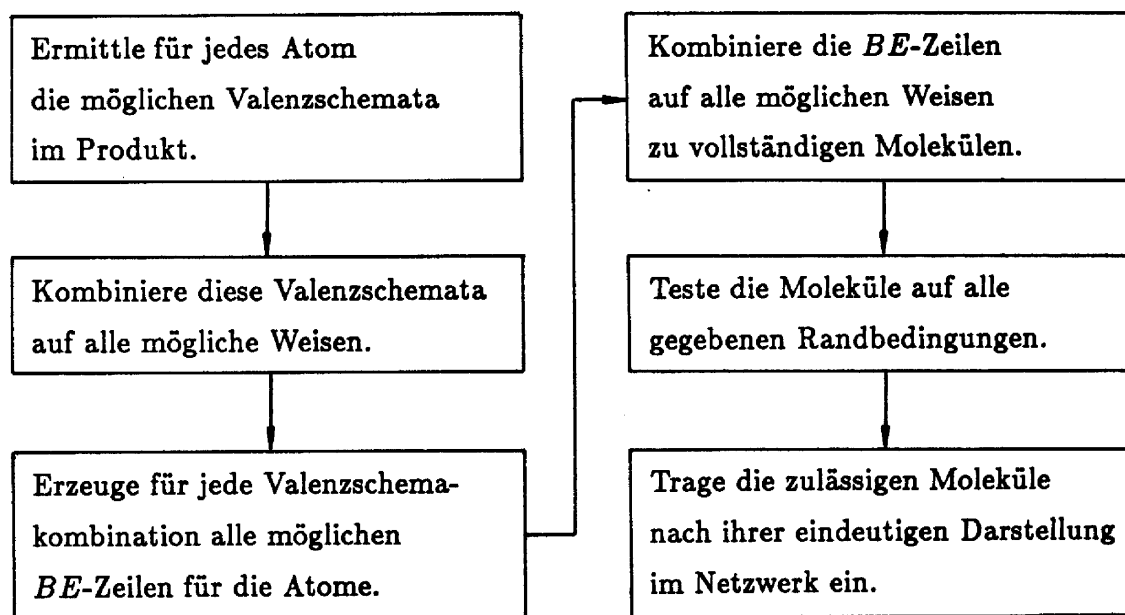


Abb. (5-1) Ablaufschema des Reaktionsgenerators.

* Die Alternative B ist, insbesondere im Zusammenhang mit der minimalen chemischen Distanz (siehe 6.1), Gegenstand weiterer Untersuchungen.

C	$\begin{array}{c} \\ -C- \\ \end{array}$	$\begin{array}{c} // \\ C- \\ / \end{array}$	$\equiv C-$	$\begin{array}{c} \backslash \\ C^+- \\ / \end{array}$
$\begin{array}{c} \\ -C- \\ \end{array}$	+	+	-	+
$\begin{array}{c} // \\ C- \\ / \end{array}$	+	+	+	+
$\equiv C-$	-	+	+	-
$\begin{array}{c} \backslash \\ C^+- \\ / \end{array}$	+	+	-	-

Nummer	Produkt- Valenz- schema	
1	4000	0
2	1200	0 (5-3)
3	1001	0
4	3000	+1

Soll der Reaktionsgenerator Vorstufen erzeugen, so müssen an dieser Stelle die Valenzschemata berechnet werden, aus denen das Eduktvalenzschema des Atoms hervorgehen kann.

3) Kombination der Produktvalenzschemata :

Die verschiedenen Produktvalenzschemata aller Atome werden auf alle möglichen Weisen miteinander mit der Zielsetzung kombiniert, hieraus unter Beachtung aller Randbedingungen Produktmoleküle zusammensetzen zu können. Hierbei werden Verfahren angewandt, die sowohl redundante Kombinationen weitestgehend ausschließen, als auch innerhalb der Kombinationen die sich akkumulierenden Größen ständig überwachen und bei Überschreitung der Grenzwerte das Kombinationsverfahren so steuern, daß mit diesen angefangenen Kombinationen nicht weitergerechnet wird.

3a) Konstitutionelle Symmetrie :

Existieren im Edukt Untermengen konstitutionell äquivalenter Atome*, so werden für die mächtigste dieser Untermengen – und nur für diese – lediglich Kombinationen von Valenzschemata erzeugt, deren interne Numerierung monoton mit den Atomnummern steigt. Valenzschemakombinationen, die dieser Bedingung nicht gehorchen, sind auf jeden Fall schon erzeugt worden und, bedingt durch die Symmetrie des Eduktmoleküls, redundant.

3b) Anzahl geladener Atome :

Wird für ein Atom i ein formal geladenes Valenzschema in die Kombination mit aufgenommen und ist damit die maximale Anzahl von geladenen Atomen in einem Molekül überschritten (siehe 4.10.6), so wird für dieses Atom sofort das nächste Valenzschema genommen, d.h. die Kombinationen der Valenzschemata für die Atome $> i$ nicht erzeugt.

3c) Anzahl von Reaktionszentren :

Wird für ein Atom i ein Valenzschema in die Kombination mit aufgenommen, das verschieden vom Eduktvalenzschema dieses Atoms ist, so ist das Atom i auf jeden Fall an der Reaktion beteiligt. Ist damit die maximale Anzahl von reagierenden Atomen (siehe 4.8.1) überschritten, so wird wie in 3b) verfahren.

3d) Anzahl beteiligter Elektronen :

Für die an der Kombination beteiligten Valenzschemata wird die Anzahl von Elektronen bestimmt, die diese Valenzschemata zum Reaktionskern beisteuern. Im Laufe des Kombinationsverfahrens darf die Gesamtzahl dieser Elektronen die Gesamtzahl der Elektronen in den Eduktvalenzschemata nicht überschreiten. Ist dies bei einem Atom i der Fall, so wird wie bei 3b) verfahren. Ist eine Kombination von n Valenzschemata komplett, so muß die Gesamtzahl von Elektronen gleich der im Edukt sein. Damit ist auf jeden Fall gewährleistet,

* Die Erkennung der konstitutionellen Äquivalenz von Atomen geschieht mittels des Verfahrens zur eindeutigen Darstellung (siehe 5.2).

daß jede entstehende *R*-Matrix die Nullsumme besitzt, d.h. daß das Gesetz der Ladungs- und Massenerhaltung beachtet wird.

3e) Ordnungsspezifische Anzahl der Bindungen :

Für eine komplette Kombination von *n* Valenzschemata muß für jede Bindungsordnung die Summe der Bindungen dieser Ordnung, die von den Valenzschemata beigesteuert werden, gerade sein. Damit wird gewährleistet, daß eine paarweise Absättigung der Valenzen potentiell möglich ist. Dies ist eine notwendige, aber nicht hinreichende Bedingung dafür, daß aus einer Valenzschemakombination auch wirklich ein Molekül zusammengesetzt werden kann.

Für jede nach 3a) bis 3e) gültige Kombination von Valenzschemata wird Schritt 4) durchgeführt.

4) Erzeugung der *BE*-Zeilen :

Das im folgenden beschriebene Verfahren zur Erzeugung aller *BE*-Zeilen zu einem Valenzschema bei gegebenem *n* zieht seinen Vorteil aus der Tatsache, daß i.a. die meisten *BE*-Zeileneinträge aus Nullen bestehen. Ein reines Permutationsverfahren müßte also sehr viele Redundanzen eliminieren. Das Einlesen extern erzeugter, katalogisierter Permutationen, wie es in einer Reihe von Programmen geschieht^{7,39}, kommt hier nicht in Betracht, da mit *RAIN* angestrebt wird, eine Anzahl von mindestens 15 beteiligter Atome zuzulassen. Die erforderlichen Speicherkapazitäten wären viel zu groß, um alle Permutationen aufzunehmen. Aus diesem Grund werden alle Permutationen "vor Ort" in einem zweistufigen Verfahren erzeugt.

Eine im Schritt 3) erzeugte Kombination von Valenzschemata unterscheidet sich bis zu einem Atom *i* nicht von der vorhergehenden Kombination*. Für die Atome 1 bis *i* können aus diesem Grund die schon erzeugten *BE*-Zeilen unverändert übernommen werden, während für die Atome *i* + 1 bis *n* die Schritte 4a) bis 5g) durchgeführt werden.

* Für die erste Kombination gilt : *i* = 0.

- 4a) Aus den Valenzschemata werden die reduzierten charakteristischen Vektoren C'_V erzeugt. Diese Vektoren enthalten, im Gegensatz zu den von *Bauer*⁷ definierten, in steigendem Betrag nur die von Null verschiedenen Einträge einer *BE*-Zeile, die das zugehörige Valenzschema repräsentiert. Hierbei werden nur die Bindungen, also nur die ersten drei Zahlen in den Valenzschemata interpretiert. Bindungen zu Atomen, die als nicht an der Reaktion beteiligt markiert sind, werden vorher abgezogen. Beispiel (5-4) zeigt, wie aus Valenzschemata die reduzierten charakteristischen Vektoren erzeugt werden.

Valenzschema	:	1010 0	2100 0	3100 0	4000 0	(5-4)
C'_V	:	13	112	1112	1111	

Die Länge der C'_V entspricht der Koordinationszahl des Atoms vermindert um die Bindungen zu "nicht reaktiven" Atomen. Für ein Valenzschema ohne Bindungen (Koordinationszahl 0) wird ein C'_V der Länge n erzeugt, der nur aus Nullen besteht.

- 4b) Von den reduzierten charakteristischen Vektoren C'_V werden alle verschiedenen Permutationen $P(C'_V)$ erzeugt. Dies geschieht, indem nach einem publizierten Verfahren⁴⁶⁻⁴⁸, zunächst alle Permutationen von C'_V erzeugt werden und die redundanten Permutationen dann eliminiert werden. Diese redundanten Permutationen sind bedingt durch identische Einträge in den C'_V (z.B. bei $C'_V = 112$). Das Beispiel (5-5) zeigt einige reduzierte charakteristische Vektoren mit jeweils allen möglichen Permutationen.

C'_V	:	13	112	1112	1111	(5-5)
$P(C'_V)$	:	13, 31	112, 121, 211	1112, 1121, 1211, 2111	1111	

- 4c) Für den reduzierten charakteristischen Vektor des Atoms i mit der Länge m werden nun alle möglichen Verteilungen von $n - m - 1$ Nullen auf $n - 1$ Plätzen erzeugt. Beispiel (5-6) zeigt für $n = 6$ und $C'_V = 12 (\Rightarrow m = 2)$ die möglichen Verteilungen von $n - m - 1 = 3$ Nullen auf $n - 1 = 5$ Plätzen.

..000	.0.00	.00.0	
.000.	0..00	0.0.0	(5-6)
0.00.	00..0	00.0.	
000..			

- 4d) Die *BE*-Zeilen für ein Atom i werden aus den in 4c) erhaltenen Null-Zeilen dadurch erreicht, daß in jeder Null-Zeile die nicht mit Nullen besetzten Positionen von links her mit allen unterscheidbaren Permutationen des reduzierten charakteristischen Vektors aufgefüllt werden. Gleichzeitig wird jeweils an der Position i die Anzahl der freien Elektronen eingefügt (späterer Hauptdiagonaleintrag). Beispiel (5-7) zeigt dies an den Null-Zeilen und dem C'_V aus (5-6) für ein Atom $i = 2$ mit 2 freien Elektronen.

Null-Zeile	mit $P(C_V')$	BE-Zeile
..000	12000	122000
	21000	221000
.0.00	10200	120200
	20100	220100
.00.0	10020	120020
	20010	220010
.000.	10002	120002
	20001	220001
0..00	01200	021200
	02100	022100
0.0.0	01020	021020
	02010	022010
0.00.	01002	021002
	02001	022001
00..0	00120	020120
	00210	020210
00.0.	00102	020102
	00201	020201
000..	00012	020012
	00021	020021

(5-7)

Für jede auf diese Weise erhaltene *BE*-Zeile wird Schritt 5) durchgeführt.

5) Überprüfung der *BE*-Zeilen auf Gültigkeit :

Jede *BE*-Zeile für ein Atom i muß den folgenden Forderungen genügen, um in die Liste der gültigen *BE*-Zeilen für das Atom i aufgenommen zu werden.

- 5a) Aus der zu prüfenden *BE*-Zeile und der *BE*-Zeile des Atoms i im Edukt werden die Adjazenzzeilen und deren Differenz, die zugehörige *T*-Zeile gebildet. Für die Adjazenzzeilen wird jeder von Null verschiedene Eintrag einer *BE*-Zeile auf den Wert 1 gesetzt. Die erhaltene *T*-Zeile wird nun in Bezug auf Anzahl der außerdiagonalen Einträge und Betrag der Summe aller Einträge auf die Einhaltung der Randbedingungen geprüft (siehe 4.8.8 und 4.8.9).
- 5b) Aus der zu prüfenden *BE*-Zeile und der *BE*-Zeile des Atoms i im Edukt wird die Differenz, die zugehörige *R*-Zeile gebildet. Diese erhaltene *R*-Zeile wird nun in Bezug auf Anzahl und Betrag der außerdiagonalen Einträge auf die Einhaltung der Randbedingungen geprüft (siehe 4.8.5 und 4.8.7).
- 5c) Trägt das Atom i eine Ladung, wird die *BE*-Zeile daraufhin überprüft, ob sie eine Bindung zu einem ebenfalls geladenen Atom $j < i$ ausbildet, dessen Ladung dasselbe Vorzeichen hat wie die des Atoms i . Benachbarte gleichnamige Ladungen sind grundsätzlich ausgeschlossen (siehe 4.10.7).
- 5d) Ist das Atom i als Heteroatom markiert, so wird die Anzahl der Heteroatome $j < i$ gezählt, zu denen die *BE*-Zeile eine Bindung ausbildet. Hierdurch können Ketten von Heteroatomen bis zur Länge 3 erkannt und gegebenenfalls ausgeschlossen werden (siehe 4.10.5).
- 5e) Die *BE*-Zeile des Atoms i wird daraufhin überprüft, ob sie die geforderten Mindestbindungsordnungen zu den Atomen $j < i$ einhält (siehe 4.3).
- 5f) Jeder Eintrag (auch die Null-Einträge) der *BE*-Zeile des Atoms i an einer Position $j < i$ muß in mindestens einer gültigen *BE*-Zeile eines Atoms j an

der Spaltenposition i vorkommen. Dies gewährleistet, daß zumindest eine Möglichkeit der Kombination der erhaltenen *BE*-Zeilen zu einem kompletten Molekül existiert.

- 5g) Jeder von Null verschiedene Eintrag der *BE*-Zeile des Atoms i an einer Position $j > i$ muß als Bindung der entsprechenden Ordnung in mindestens einem der möglichen Produktvalenzschemata des Atoms j vorkommen.

Die Forderung 5g) ist wesentlich weniger weitreichend als 5f), jedoch betrifft sie als einzige die Bindungen zu Atomen höherer Nummer als das aktuelle Atom i . Diese Atome können in einer folgenden Kombination mit anderen Valenzschemata versehen sein und somit andere Bindungsverhältnisse eingehen. Da die *BE*-Zeilen der Atome $\leq i$ aber auch in diesen Kombinationen ihre Gültigkeit behalten sollen, dürfen explizite Tests, die Nachbarschaft von Atomen betreffend (5c, 5d, 5e und 5f), nur mit Bindungen zu Atomen $j < i$ durchgeführt werden.

Eine *BE*-Zeile, die die Forderungen 5a) bis 5g) erfüllt, wird in die Liste der *BE*-Zeilen für das Atom i aufgenommen. Diese Liste dient auch zur Wiederherstellung der *BE*-Zeileninformation von Atomen, für die keine neue *BE*-Zeilengenerierung durchgeführt wird, da sich ihr Valenzschema gegenüber dem vorherigen Durchlauf der *BE*-Zeilengenerierung nicht verändert hat (siehe 4).

Erfüllt für ein Atom i keine der erzeugten *BE*-Zeilen diese Forderungen, so wird das Verfahren der *BE*-Zeilenerzeugung für die aktuelle Valenzschemakombination sofort abgebrochen und für das Atom i unmittelbar das nächste Valenzschema in die Kombination eingesetzt. Die noch folgenden Kombinationen der Atome $j > i$ brauchen nicht durchgeführt zu werden.

Sind für alle Atome die möglichen *BE*-Zeilen erzeugt, so können diese vor dem *BE*-Zeilenkombinationsverfahren im Schritt 6) noch reduziert werden.

6) Reduktion der *BE*-Zeilen :

Das Reduktionsverfahren der Schritte 6a) und 6b) wird so lange durchgeführt, bis sich keine weitere Verringerung der Anzahl der *BE*-Zeilen mehr ergibt. Reduziert

sich für ein Atom i diese Anzahl auf Null, so wird das *BE*-Zeilenerzeugungsverfahren unmittelbar abgebrochen. Es wird dann jedoch nicht wie unter 5) beschrieben verfahren, sondern es wird lediglich für das Atom n das nächste Valenzschema in die Kombination aufgenommen.

6a) Atome, die unter ihren gültigen *BE*-Zeilen keine Nullzeile haben, sind auf jeden Fall an der Reaktion beteiligt. Übersteigt die Anzahl dieser Atome das entsprechende Limit (siehe 4.8.1), so kann auch hier das *BE*-Zeilenerzeugungsverfahren abgebrochen werden. Entspricht diese Zahl genau dem Wert des Limits, so müssen für alle potentiell unveränderten Atome die *BE*-Zeilen bis auf die jeweilige Nullzeile reduziert werden.

6b) Wie 5f), jedoch für alle Positionen $j \neq i$.

7) Kombination der *BE*-Zeilen :

Die nach dem Schritt 6) für alle Atome verbleibenden *BE*-Zeilen werden nun auf alle möglichen Weisen zu vollständigen Molekülen kombiniert. Eine Zeile für ein Atom i kann nur dann in die Kombination aufgenommen werden, wenn ihre Einträge an den Positionen $j < i$ in der bis dahin aufgebauten *BE*-Matrix mit den Spaltenpositionen i aller Atome j übereinstimmen. Hierbei werden die bezüglich der Edukt-*BE*-Matrix veränderten Atome und Bindungen laufend mitgezählt. Wird einer der gegebenen Grenzwerte dafür bei einem Atom i überschritten, kann der für die Atome $j > i$ verbleibende Rest der *BE*-Zeilen dieser Teilkombination verworfen werden und für das Atom i kann unmittelbar die nächste passende *BE*-Zeile in die Kombination eingesetzt werden.

8) Test auf konstitutionelle Randbedingungen :

Jede aus Schritt 7) resultierende komplette *BE*-Matrix, d.h. jedes neue Produktmolekül wird den Tests auf konstitutionelle Randbedingungen unterworfen (siehe 4.10). Zu diesem Zweck wird zuerst eine Ringerkennung durchgeführt. Anschließend wird auf die Ringspannung S_R , kumulierte Mehrfachbindungen in Rin-

gen, Dreifachbindungen in Ringen und Einhaltung der Bredt'schen Regel geprüft. Die weiteren Randbedingungen für die chemische Konstitution werden bereits in den davorliegenden Kombinationsverfahren beachtet.

9) Kanonisierung der Moleküle :

Jedes zulässige Produktmolekül wird nun einem Kanonisierungsverfahren unterworfen (siehe 5.2), das eine eindeutige Darstellung der chemischen Konstitution erzeugt.

5.2 Eindeutige Darstellung der Moleküle

Im Verlauf der Erzeugung von Reaktionsnetzwerken tritt für *RAIN* an zwei Punkten das Problem auf, Moleküle, deren Atome auf verschiedene Weisen beziffert sein können, auf Identität hin zu überprüfen :

- Jede als Vorgänger oder Nachfolger eines Moleküls durch den Reaktionsgenerator erzeugte Struktur wird in dem Niveau, in das es eingetragen werden soll, gesucht. Nur konstitutionell neue Strukturen werden in das Niveau aufgenommen. Bei bereits vorhandenen Strukturen werden lediglich zusätzliche Verkettungsverweise eingetragen. Wird aus ein und demselben Molekül ein Produktmolekül auf verschiedene Weisen mehrfach erzeugt, so werden diese zusätzlichen Verkettungsverweise nicht noch einmal eingetragen.
- Um Verbindungen zwischen dem Eduktbaum und dem Produktbaum aufbauen zu können, werden die Moleküle des einen Baums im anderen Baum gesucht.

Um zwei Moleküle mit willkürlicher Numerierung auf strukturelle Identität zu prüfen, wird folgendermaßen vorgegangen :

- a) Beide Moleküle werden einem Kanonisierungsverfahren unterzogen, das eine eindeutige, nur von der Konstitution abhängige, Numerierung erzeugt^{16,49-57}.
- b) Aus beiden Molekülen wird, nach der kanonischen Bezifferung, eine komprimierte Darstellung der Atom- und Bindungslisten erzeugt.
- c) Die komprimierten Darstellungen werden direkt miteinander verglichen und auf Gleichheit überprüft.

Als Kanonisierungsverfahren wird der in unserem Arbeitskreis entwickelte Algorithmus CANON⁵²⁻⁵⁴ in einer modifizierten Form herangezogen. Im folgenden seien die Modifikationen ausführlich, die dem publizierten Verfahren gleichen Schritte des Algorithmus nur kurz dargestellt :

- 1) Den Atomen des zu kanonisierenden Moleküls werden jeweils Atomdeskriptoren zugewiesen, die folgende Informationen in der angegebenen Reihenfolge enthalten :
 - a) interne Ordnungszahl (Nummer des Element- oder Pseudoelementsymbols in der Liste aller vom Anwender definierten Elementsorten).
 - b) Anzahl der Nachbarn des Atoms.
 - c) Anzahl der freien Elektronen des Atoms.
 - d) Numerische Markierung für Atome, die dadurch im Edukt und Produkt einander zugeordnet sind (siehe 4.4). Atome ohne Zuordnung erhalten hier alle den gleichen Wert Null und werden dadurch nicht individualisiert.
 - e) Kanonische Atomnummern und Bindungsordnungen der Nachbarn des Atoms, die als nicht zum Reaktionskern gehörend markiert sind. Die entsprechenden kanonischen Atomnummern des unveränderten Rests der globalen Gesamtreaktion werden erhalten, indem zu Beginn des Netzwerkaufbaus einmal der unveränderliche Teil des Eduktmoleküls für sich dem Kanonisierungsverfahren unterworfen wird. Definitionsgemäß muß dieser Teil genau dem des unveränderlichen Teil des Produktmoleküls entsprechen.
 - f) Ringdeskriptor des Atoms. Dieser besteht aus den Anzahlen der m -Ringe, an denen das Atom beteiligt ist (für $m = 3$ bis n).
- 2) Die Atome werden lexikographisch nach ihren Deskriptoren geordnet.
- 3) Atome mit gleichen Deskriptoren werden zu Äquivalenzklassen zusammengefaßt (mit entsprechenden Äquivalenzklassenindizes versehen), die aufsteigend numeriert werden. Ist die Anzahl der verschiedenen Äquivalenzklassen gleich der Anzahl der Atome n , so entspricht die in 2) erhaltene Reihenfolge der "CANONischen" Bezifferung und das Verfahren bricht ab.
 Ist die Anzahl der verschiedenen Äquivalenzklassen gleich der im vorigen Durchlauf erzielten, so wird der Äquivalenzklassenindex eines Atoms, dessen Äquivalenzklasse mehrere Atome enthält, auf einen bis dahin noch nicht vorkommenden Wert gesetzt.

Geschieht dies zum ersten Mal, so enthalten die bis dahin gewonnenen Äquivalenzklassenindizes die Information über die konstitutionelle Symmetrie des Moleküls. Atome mit gleichen Äquivalenzklassenindizes sind konstitutionell äquivalent. Diese Information ist wichtig, um bei der Reaktionsgenerierung von vorneherein die Redundanz der Erzeugungsoperationen zu minimieren (siehe 5.1).

- 4) Den Atomen werden jeweils Atomdeskriptoren zugewiesen, die folgende Informationen enthalten :
 - a) Äquivalenzklassenindex des Atoms.
 - b) Äquivalenzklassenindizes der Nachbarn des Atoms in aufsteigender Reihenfolge.
- 6) Sprung nach 2).

6 Zielgerichtete Entwicklung der Reaktionsbäume

Bei einer vollkommen voneinander unabhängigen Entwicklung des Edukt- und des Produkt-Reaktionsbaums ist es nicht vermeidbar, daß mitunter sehr viele Pfade erzeugt werden, die innerhalb einer bestimmten Anzahl von Reaktionsschritten keine Verbindung zum jeweils anderen Reaktionsbaum finden und somit im resultierenden Netzwerk nicht auftauchen. Dies ist insbesondere von Bedeutung, wenn bis zur Auffindung einer Verknüpfungsstelle der Reaktionsbäume sehr viele Niveaus zu entwickeln sind, d.h. die Gesamtreaktion über eine große Anzahl von Zwischenstufen verläuft (siehe 3.2).

Ein zielgerichtetes Vorgehen bei der Entwicklung der Reaktionsbäume kann viele unnötige Pfadentwicklungen vermeiden. Dies bedeutet, daß in jedem Niveau nur diejenigen Moleküle behandelt werden, die auf einem relativ "kurzen" Weg vom Edukt zum Produkt liegen. Die Dimension von "kurz" ist ungefähr in der Größenordnung der Mindestanzahl von Reaktionsschritten vom Edukt zum Produkt unter den gegebenen Randbedingungen anzusetzen. Für die Lösung dieses Problems werden zwei Möglichkeiten erörtert :

- a) Forderung der größtmöglichen strukturellen Ähnlichkeit einer Zwischenstufe zum Produkt (für den Edukt-Reaktionsbaum) oder zum Edukt (für den Produkt-Reaktionsbaum).
- b) Verbot einer zu großen strukturellen Verschiedenheit einer Zwischenstufe zum Produkt und zum Edukt der Gesamtreaktion.

In beiden Fällen dient die chemische Distanz, bzw. die durch Permutation erzielbare minimale chemische Distanz zweier *EM* zur Abschätzung der strukturellen Ähnlichkeit.

6.1 Reaktionspfade entlang der minimalen chemischen Distanz

Ein Maß für die strukturelle Ähnlichkeit zweier Moleküle ist die größte Menge größter gemeinsamer Substrukturen mit einer minimalen Anzahl der Elektronen, die zu bewegen sind, um eine Struktur in die andere zu transformieren, also die minimale chemische Distanz *MCD*. Da in *RAIN* Grenzparameter für die Komplexität eines einzelnen Reaktionsschritts implementiert sind, ist die *MCD* ein grobes Maß zur Abschätzung der Mindestanzahl an Reaktionsschritten, die nötig sind, um das eine Molekül in das andere überzuführen. Die Reaktionsnetzwerke mit den wenigsten Reaktionsschritten werden erhalten, wenn entlang der Reaktionspfade vom Edukt zum Produkt die Zwischenstufen dem Produkt strukturell immer ähnlicher werden. Die strikteste Anwendung dieses Prinzips führt zu einer sehr zielstrebigem Vorgehensweise, die jedoch mit einigen Nachteilen behaftet ist (siehe unten) :

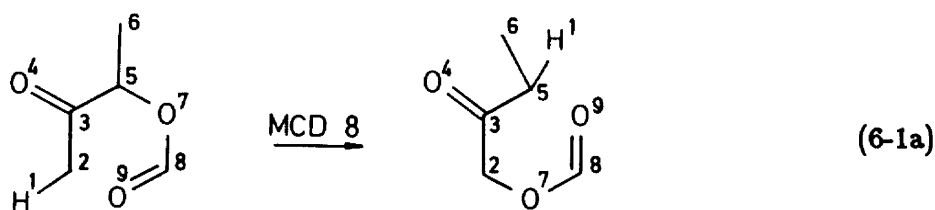
- 1) Ordne die Atome des Eduktmoleküls so um, daß sie der Zuordnung der *MCD* zum Produkt entsprechen.
- 2) Führe das Reaktions- bzw. Niveaugenerierungsverfahren wie beschrieben durch, aber
- 3) entwickle in jedem Niveau nur die Nachfolger/Vorgänger der Moleküle mit der kleinsten chemischen Distanz* *CD* zum Produkt/Edukt. Die chemische Distanz errechnet sich aus $CD = \sum_{ij} |e_{ij} - b_{ij}|$ unter Beibehaltung der Numerierung.

Da hier mit einer festen Zuordnung der Atome gearbeitet wird, muß auf eine Kanonisierung ($\cong$ Umordnung) der Moleküle bei der Generierung (siehe 5.2) verzichtet werden. Dieses Verfahren erzeugt auf jeden Fall die Zuordnung nach dem Prinzip der *MCD* und die kleinste Anzahl der unter dieser Bedingung nötigen Reaktionsschritte.

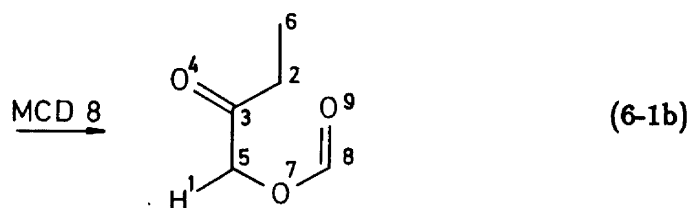
Als direktes Mittel zur Atomzuordnung vom Edukt auf das Produkt einer chemischen Reaktionsfolge ist die *MCD* jedoch nicht immer geeignet, da für viele Reaktio-

* Nicht die minimale chemische Distanz *MCD* !

nen** sehr plausible Reaktionswege generierbar sind, die diesem Prinzip nicht gehorchen. Das Beispiel (6-1) zeigt eine hypothetische Reaktion dieser Art. Die Atomzuordnung nach der *MCD* erfordert eine komplexe Gerüstumstrukturierung (6-1a) oder eine Wanderung eines CH_3 -Restes (6-1b).

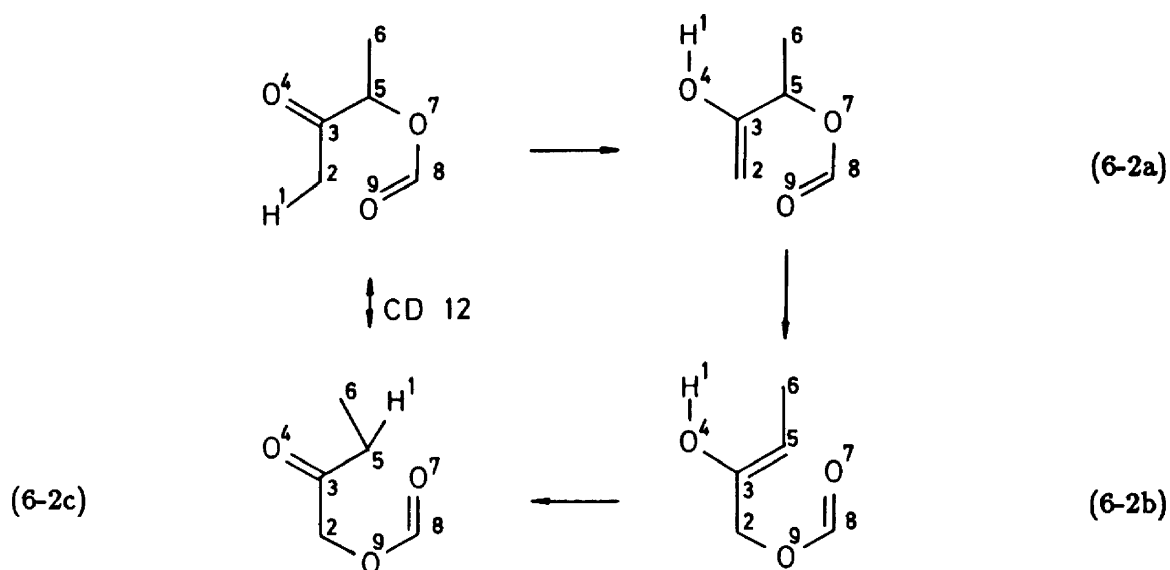


III



Eine zweimalige Keto-Enol-Tautomerie (6-2a und 6-2c) mit dazwischenliegender *Cope*-Umlagerung (6-2b) führt über zwei Zwischenstufen ebenfalls zum Ziel, diesmal jedoch mit einer chemischen Distanz von 12 statt dem *MCD*-Wert von 8.

** z.B. electrocyclische Reaktionen mit großen Strukturveränderungen.



Der Reaktionspfad in (6-2) ist wesentlich einleuchtender, als die unter der Abbildungsvorschrift von (6-1) erzeugbaren.

Ein weiterer Nachteil der Voraussetzung einer Atom-auf-Atom-Abbildung nach dem Prinzip der *MCD* liegt in der oftmaligen Nichteindeutigkeit der Abbildungsvorschrift, d.h. in der Existenz mehrerer Möglichkeiten, die Atome des Edukts denen des Produkts unter einer Verschiebung einer Mindestanzahl von Elektronen zuzuordnen (siehe Beispiel (4-8) und (6-1)). Für jede der Zuordnungsmöglichkeiten unter Wahrung des Prinzips der *MCD* müßte also ein separater Netzwerkgenerierungslauf angesetzt werden.

Wesentlich besser als in der Form einer zwingenden Abbildungsvorgabe, eignet sich die *MCD* als Entscheidungshilfe, um bei der Entwicklung der Reaktionsbäume Wege zu vermeiden, die weit vom entsprechenden Ziel abweichen. Um ein Niveau z.B. des Eduktbaums unter Ausnutzung der *MCD*-Information zu entwickeln, wird wie folgt vorgegangen :

- 1) Berechne für alle Moleküle des Niveaus die *MCD* zum Produkt*.
- 2) Sortiere die Moleküle nach steigender *MCD* und fasse Moleküle mit gleicher *MCD* zu Gruppen zusammen.
- 3) Entwickle lediglich die *k* ersten Gruppen weiter zum nächsten Niveau, d.h. generiere nur für die Moleküle mit den *k* niedrigsten *MCD*s die Nachfolger. Der Parameter *k* ist eine vom Anwender definierbare Randbedingung in *RAIN*.

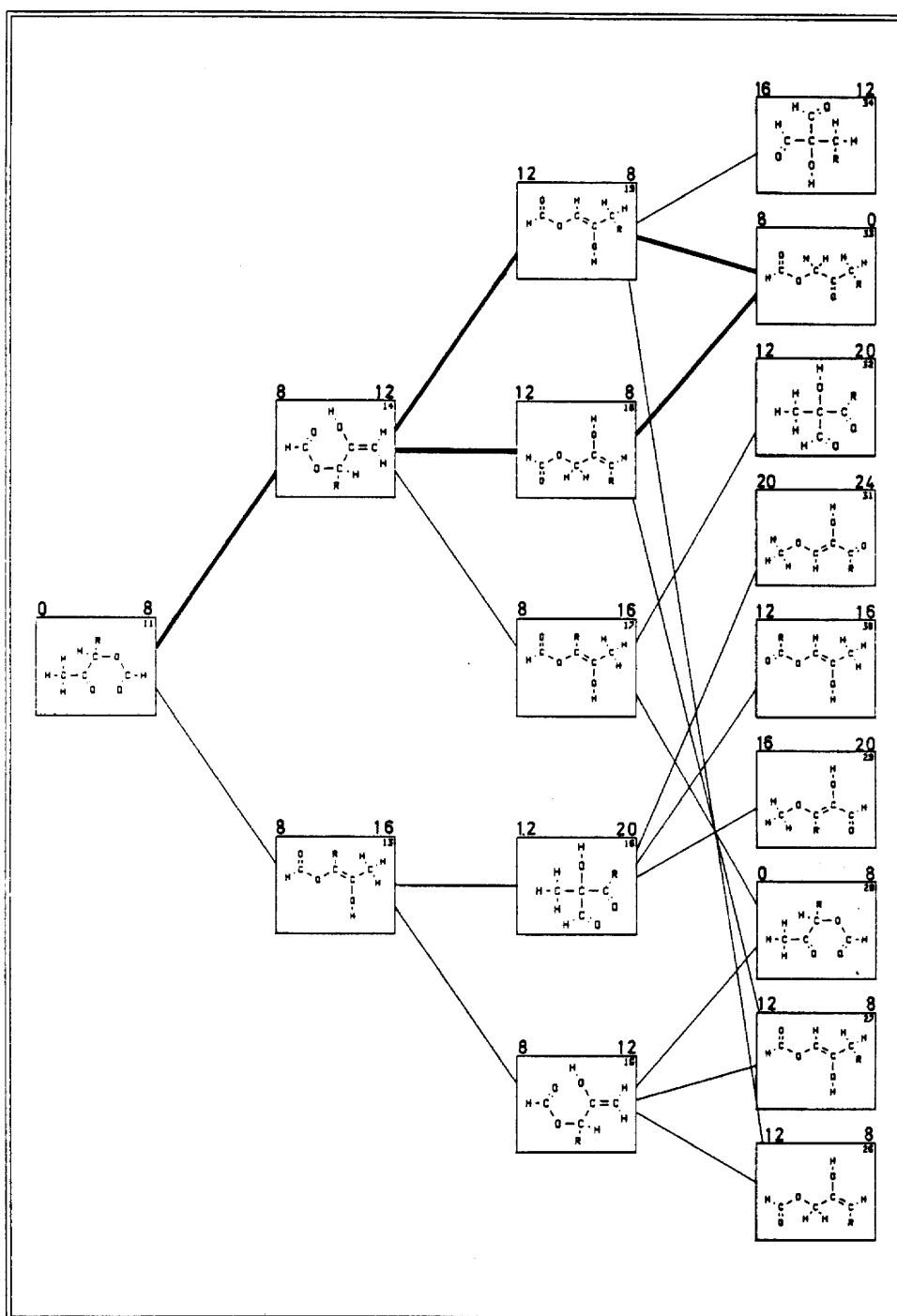
Dieses Verfahren führt zu einem, je nach dem Wert von *k*, mehr oder weniger zielstrebigem Vorgehen bei der Entwicklung der Reaktionsbäume. Die so erhaltenen Ergebnisse beinhalten eine Atom-auf-Atom-Abbildung, die folgendem Prinzip gehorcht :

RAIN erzeugt die Atomzuordnung in Edukt und Produkt, die durch eine kleinste Anzahl einfacher Reaktionsschritte zu erhalten ist.

Diese Atomzuordnung kann — und in vielen Fällen trifft dies auch zu — gleich sein der, die dem Prinzip der minimalen chemischen Distanz gehorcht.

Abbildung (6-3) zeigt einen Edukt-Reaktionsbaum für das Beispiel (6-1 und 2).

* Für den Produktreaktionsbaum die *MCD* zum Edukt.

Abb. (6-3) Reaktionsbaum mit *MCDs*.

Dieser Baum ist bis zum Produkt (Molekül Nummer 33) hin entwickelt. Die resultierenden Netzwerkspfade sind hervorgehoben. Der Pfad über Molekül 12 und 19 führt zu einer Atomzuordnung mit der minimalen chemischen Distanz (6-1), der Pfad über 12 und 18 repräsentiert die in (6-2) gezeigte Variante. Die über den jeweiligen Molekülen angegebenen Zahlen stehen für die *MCD* zu Edukt (links) und zum Produkt (rechts). Es wird deutlich, daß die Moleküle 12, 18, und 19 jeweils die niedrigste Produkt-*MCD* innerhalb des jeweiligen Niveaus haben. Die Anwendung des beschriebenen Prinzips würde (z.B. mit $k = 2$) also die Moleküle 16 und 20 nicht mehr für eine Weiterentwicklung freigeben und somit den Reaktionsbaum verkleinern.

6.2 Obere Schranke für die *MCD* zur Zielverbindung

Eine andere Möglichkeit, zu weit vom Ziel abweichende Pfade abzuschneiden, liegt darin, für die erlaubten Zwischenstufen maximale Werte der minimalen chemischen Distanz *MCD* zum Edukt und zum Produkt zu anzugeben. Dies bedeutet, daß ein im Edukt-Reaktionsbaum in ein Niveau einzutragendes Molekül keine größere *MCD* zum Edukt und zum Produkt haben darf, als zwei vom Anwender festzulegende Grenzparameter hierfür angeben.

Diese Methode hat gegenüber der in (6.1) beschriebenen den Vorteil, daß "vom Ziel zu weit entfernte" Moleküle gar nicht erst in die Reaktionsbäume eingetragen werden müssen. Bei der Anwendung von (6.1) muß zuerst ein komplettes Niveau aufgebaut werden, um die Entscheidung über eine eventuelle Weiterentwicklung der Moleküle treffen zu können.

Verdeutlicht werden kann die Limitierung der *MCD* zu den Wurzeln der Reaktionsbäume mit der Einhüllung dieser Wurzeln im $(\frac{n*(n+1)}{2})$ -dimensionalen Hyperraum, dem Darstellungsraum der Ensembles mit n Atomen¹⁵. Die Familie isomerer *EM* (*FIEM*) mit n Atomen wird repräsentiert durch eine Menge von Punkten auf der Oberfläche einer Hyperkugel, die in einen Raum eingebettet ist, der für jedes Atom i und für jede potentielle Bindungsmöglichkeit (d.h. alle möglichen Paare i, j der n Atome) eine Koordinate aufweist. Die Punkte, die durch die *FIEM* belegt werden, liegen in dem "Oktanten" der Hyperkugeloberfläche, dessen Koordinaten alle ≥ 0 sind und haben alle eine L_1 -Entfernung vom Ursprung, die der Gesamtzahl der Elektronen in der *FIEM* entspricht.

Die Maximalwerte der *MCD* zum Edukt und Produkt definieren L_1 -Kreise um die Edukt- bzw. Produktabbildungspunkt auf dieser nicht-euklidischen Hyperfläche⁵⁸. Die beiden Kreisflächen gemeinsame Schnittfläche legt den L_1 -Ereignishorizont des Netzwerks. Nur Moleküle, deren Abbildungspunkte im $(\frac{n*(n+1)}{2})$ -dimensionalen Hyperraum innerhalb dieser Fläche liegen, werden in die Reaktionsbäume aufgenommen. Abbildung (6-4) zeigt den Versuch einer zweidimensionalen Darstellung dieser diskreten Hyperfläche.

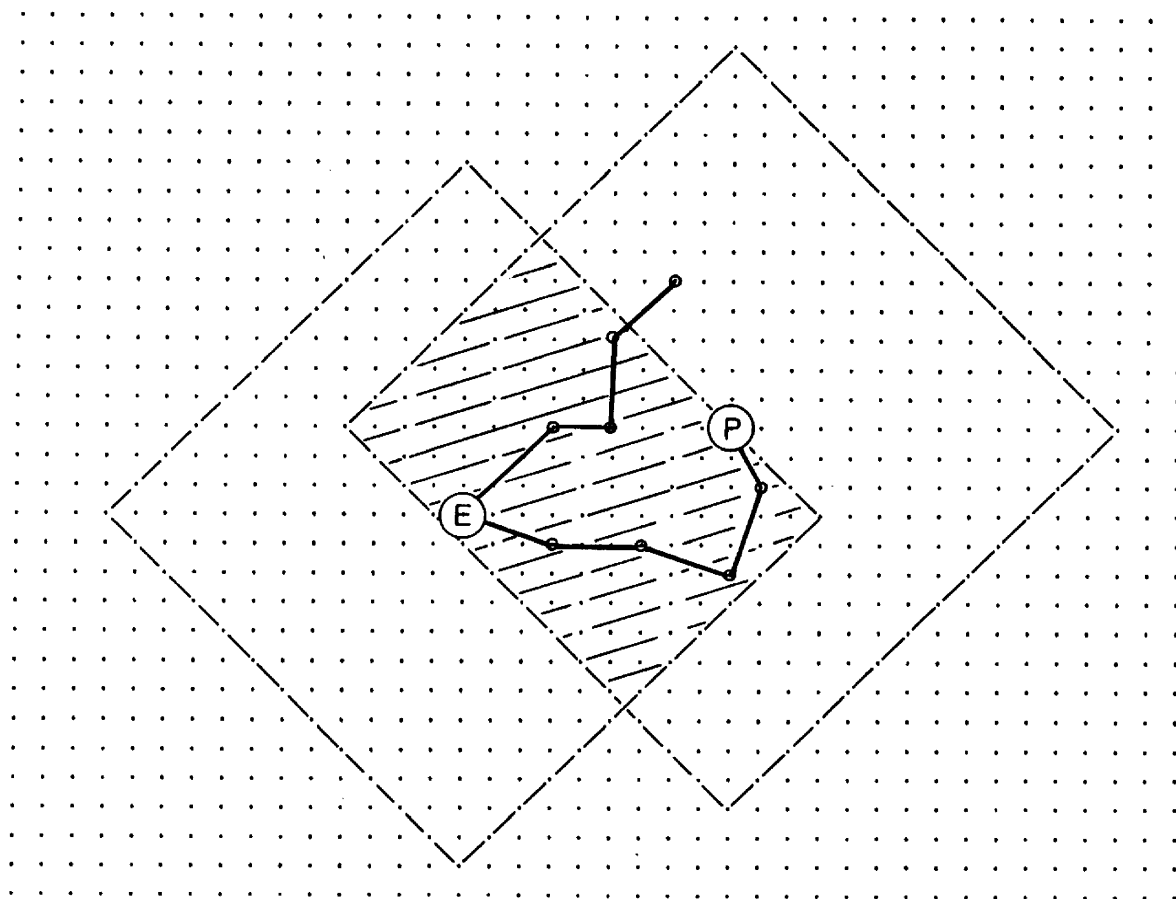


Abb. (6-4) Edukt und Produkt auf der Hyperfläche mit max. *MCDs*.

Die einzelnen Punkte repräsentieren Moleküle des gezeigten Ausschnitts der *FIEM*. Das Edukt- und das Produkt-*EM* sind hervorgehoben. Es sei jedoch bemerkt, daß eigentlich jedes individuelle Molekül durch eine Menge von Punkten, bedingt durch die unterschiedliche Numerierbarkeit, repräsentiert ist. Die Abbildung (6-4) zeigt also lediglich eine Repräsentationsmöglichkeit von vielen.

Die L_1 -Kreise erscheinen in der Abbildung als euklidquadratähnliche Gebilde. Sie markieren die *EM*, die die maximal erlaubte *MCD* zum Edukt oder zum Produkt besitzen. In der schraffierten Schnittfläche liegen die zugelassenen Moleküle der Reaktionsbäume. Es ist ein Weg gezeigt, der vom Edukt zu Produkt führt, also im Reaktionsnetzwerk auftaucht, und ein anderer Weg, der nach einigen Reaktionsschritten den erlaubten Bereich verläßt und somit bei der Generierung des Eduktbaums nicht weiter verfolgt wird.

7 *Anwendungsbeispiele*

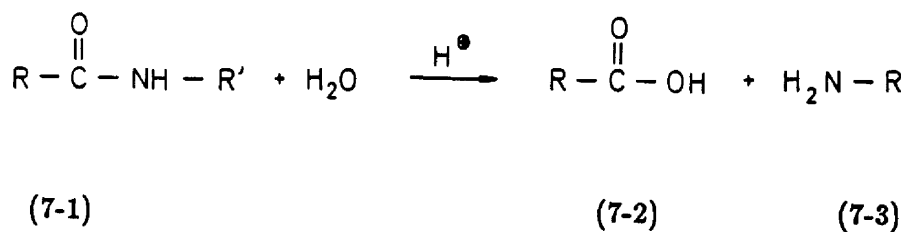
Das Programmsystem *RAIN* ist unter dem Betriebssystem MS-DOS 3.10 in der Programmiersprache FORTRAN auf einem Personalcomputer COMPAQ Deskpro 386 implementiert. Es besteht aus mehreren Programmodulen zur Struktur- und Parametereingabe, zur Erzeugung, Bearbeitung und zur grafischen Ausgabe von Reaktionsnetzwerken.

Auf eine Darlegung des Programmtextes mit Flußdiagrammen, sowie auf eine Beschreibung der Dateiformate und anderer programmtechnischer Einzelheiten wird verzichtet, da diese handwerklichen Details nicht in den wissenschaftlichen Rahmen dieser Arbeit passen würden.

Bei der Auswahl der Beispielreaktionen lag das Augenmerk darauf, zu Netzwerken zu kommen, die im Rahmen dieser Arbeit darstellbar sind. Außerdem war wichtig, daß zumindest für einige Beispiele die wirklich eingeschlagenen Reaktionssequenzen bekannt sind. Aus diesem Grund sind die gewählten Beispiele nicht unbedingt repräsentativ für das gesamte Leistungsspektrum von *RAIN*.

7.1 Protonenkatalysierte Spaltung von Amid

Mit dieser Reaktion soll die Fähigkeit von *RA/N* demonstriert werden, detaillierte Mechanismenstudien zu erleichtern. Betrachtet wird eine sauer katalysierte Verseifung eines Carbonsäureamids (7-1) zur entsprechenden Carbonsäure (7-2) und zum Amin (7-3). Als Reaktionsschritte sollen nur solche in Frage kommen, die nicht mehr als eine Bindung knüpfen oder brechen, um eine möglichst genaue Erkenntnis über die Reaktionspfade gewinnen zu können.



Für die Analyse dieser Spaltungsreaktion mit *RA/N* werden die folgenden Randbedingungen gesetzt :

1) Reaktionskern :

Alle Atome außer den Resten *R* dürfen sich an den Reaktionsschritten beteiligen.

2) Valenzschemata :

C	$\begin{array}{c} \\ -C- \\ \end{array}$	$\begin{array}{c} // \\ C- \\ / \end{array}$	$\begin{array}{c} \backslash \\ C^+- \\ / \end{array}$
$\begin{array}{c} \\ -C- \\ \end{array}$	+	+	+
$\begin{array}{c} // \\ C- \\ / \end{array}$	+	+	+
$\begin{array}{c} \backslash \\ C^+- \\ / \end{array}$	+	+	-

H	H-	H ⁺
H-	+	+
H ⁺	+	+

N	$\begin{array}{c} : \\ \backslash N- \\ / \end{array}$	$\begin{array}{c} \equiv \\ N- \end{array}$	$\begin{array}{c} // \\ N^+- \\ / \end{array}$	$\begin{array}{c} \backslash \\ -N^+- \\ / \end{array}$
$\begin{array}{c} : \\ \backslash N- \\ / \end{array}$	+	+	+	+
$\begin{array}{c} \equiv \\ N- \end{array}$	+	+	+	-
$\begin{array}{c} // \\ N^+- \\ / \end{array}$	+	+	+	+
$\begin{array}{c} \backslash \\ -N^+- \\ / \end{array}$	+	-	+	+

O	$\begin{array}{c} \backslash \\ -\ddot{O}- \\ / \end{array}$	$\begin{array}{c} : \\ \ddot{O}= \end{array}$	$\begin{array}{c} \equiv \\ O^+- \end{array}$	$\begin{array}{c} \backslash \\ :O^+- \\ / \end{array}$
$\begin{array}{c} \backslash \\ -\ddot{O}- \\ / \end{array}$	+	+	+	+
$\begin{array}{c} : \\ \ddot{O}= \end{array}$	+	+	+	-
$\begin{array}{c} \equiv \\ O^+- \end{array}$	+	+	-	+
$\begin{array}{c} \backslash \\ :O^+- \\ / \end{array}$	+	-	+	-

Die Elemente O und N sind als Heteroatome markiert.

3) Grenzwerte für *R*- und *T*-Matrizen :

Maximaler Rang einer irreduziblen <i>R</i> -Matrix	3
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>R</i> -Dreieck	2
Maximaler Rang einer irreduziblen <i>T</i> -Matrix	2
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>T</i> -Dreieck	1
Maximaler Betrag eines außerdiagonalen Eintrags in einer <i>R</i> -Matrix .	1
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>R</i> -Zeile ...	2
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>T</i> -Zeile ...	2
Maximaler Betrag einer <i>T</i> -Zeilensumme	∞
Maximaler Betrag einer <i>T</i> -Matrixgesamtsumme	∞

4) Konstitutionelle Randbedingungen :

Maximale Ringspannung	0
Bredt'sche Regel beachten	ja
Kumulierte Mehrfachbindungen in Ringen erlaubt ab Ringgröße	10
Dreifachbindungen in Ringen erlaubt ab Ringgröße	10
Maximale Länge von Heteroatomketten	1
Maximale Anzahl geladener Atome	1
Maximale chemische Distanz zum Edukt	∞
Maximale chemische Distanz zum Produkt	∞
Anzahl weiter zu entwickelnder MCD-Gruppen	∞

5) Sonstiges :

- Es werden keine minimalen Bindungsordnungen gefordert.
- Automerisationsreaktionen sind als Reaktionsschritte nicht erlaubt.
- Unmittelbare Rückreaktionen sind nicht erlaubt.

Abbildung (7-4a mit 7-4b) zeigt das von *RAIN* unter diesen Randbedingungen für die Amidverseifungsreaktion erzeugte Reaktionsnetzwerk.

Amidspaltung

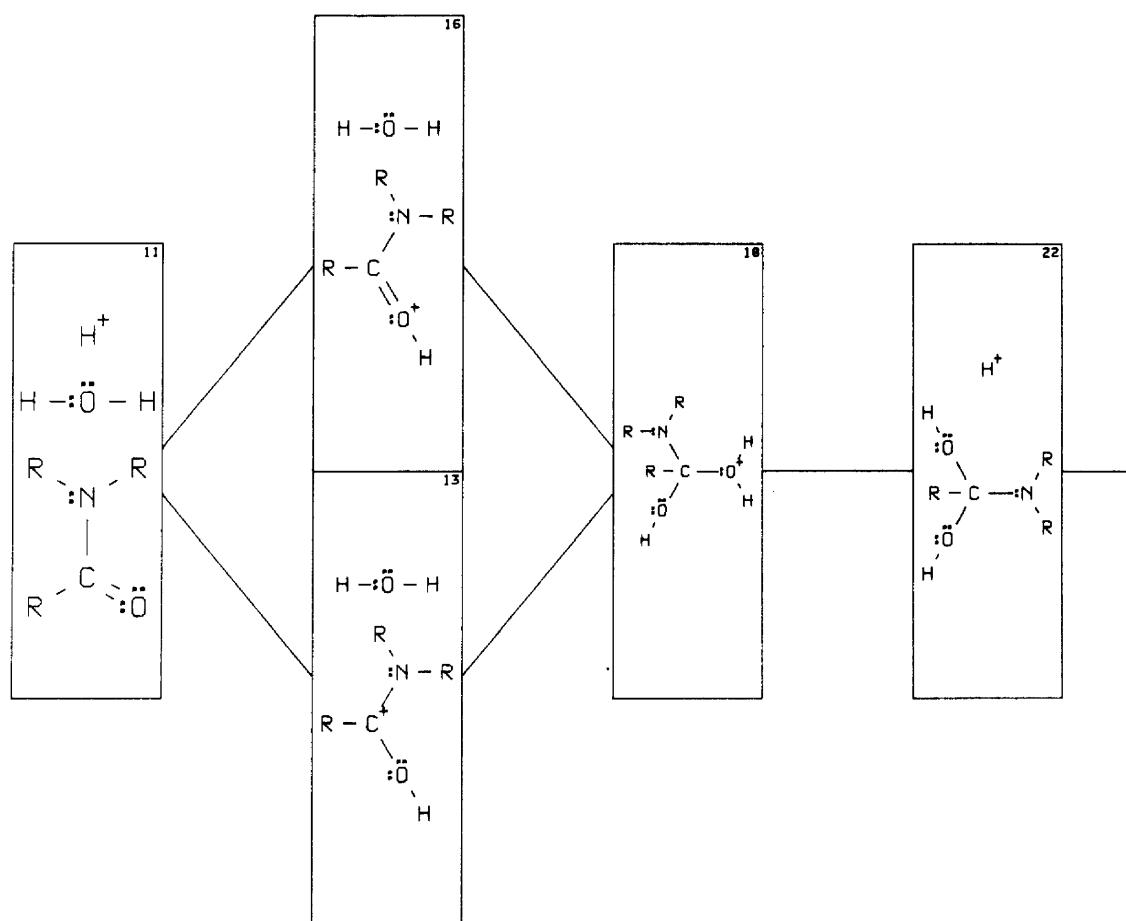


Abb. (7-4a) Reaktionsnetzwerk der sauren Amidspaltung (Teil a).

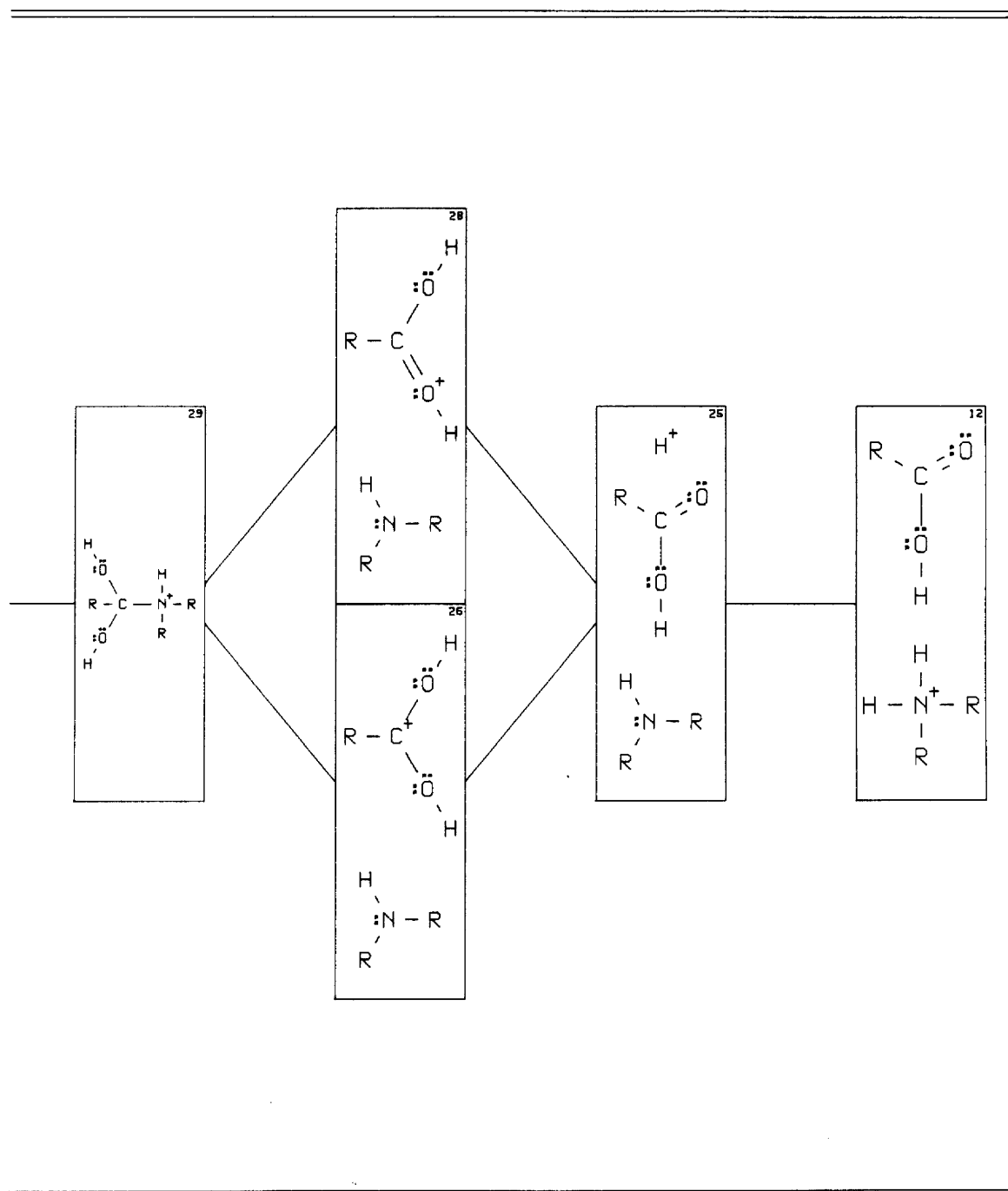
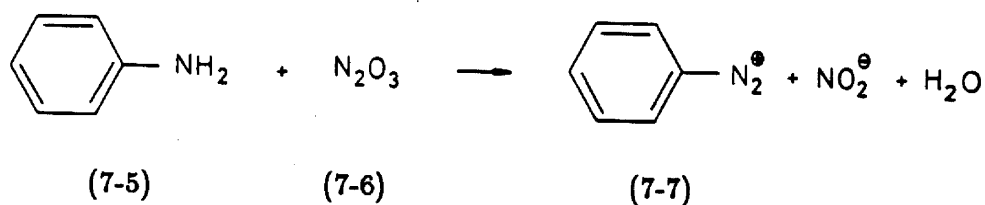


Abb. (7-4b) Reaktionsnetzwerk der sauren Amidspaltung (Teil b).

Das in Abbildung (7-4a und b) gezeigte Reaktionsnetzwerk repräsentiert die unter den gegebenen Randbedingungen möglichen kürzesten Reaktionspfade. Beide Alternativen für den ersten Reaktionsschritt entsprechen einer durch das Proton ausgelöste Aktivierung der Carbonylgruppe. Das in beiden Fällen entstehende Oxoniumion (Molekül 18) überträgt dann ein Proton an das *N*-Atom (Molekül 29). An diesem Molekül findet nun die Spaltung der *N*-*C*-Bindung statt, wobei die Ladung entweder am *O*-Atom (Molekül 28) oder am *C*-Atom (Molekül 26) verbleibt. Diese Ladung wird dann als Proton am Ende auf das Amin übertragen (Molekül 12).

7.2 Bildung von Diazoniumsalzen

Eine wichtige Reaktion zur Funktionalisierung von Aromaten ist die Bildung der Aryldiazoniumsalze (7-7). Diese entstehen aus den Arylaminen (7-5) mit salpetriger Säure, bzw. dem aus ihr entstehenden Distickstofftrioxid (7-6)⁵⁹.



Für die Analyse dieser "Lehrbuch-Reaktion" mit *RAIN* werden die folgenden Randbedingungen gesetzt :

1) Reaktionskern :

Alle Atome außer dem Arylrest *R* dürfen sich an den Reaktionsschritten beteiligen.

2) Valenzschemata :

N	$\Rightarrow N =$	$\begin{array}{c} \diagup \\ :N- \\ \diagdown \end{array}$	$\begin{array}{c} \diagup \\ N- \\ \diagdown \end{array}$	$\begin{array}{c} \\ -N^+- \\ \end{array}$	$\Rightarrow N -$	$\equiv N^+-$	$:N \equiv$
$\Rightarrow N =$	+	-	+	-	+	+	-
$\begin{array}{c} \diagup \\ :N- \\ \diagdown \end{array}$	-	+	+	+	+	-	-
$\begin{array}{c} \diagup \\ N- \\ \diagdown \end{array}$	+	+	+	-	+	-	-
$\begin{array}{c} \\ -N^+- \\ \end{array}$	-	+	-	+	-	-	-
$\Rightarrow N -$	+	+	+	-	+	+	+
$\equiv N^+-$	+	-	-	-	+	+	+
$:N \equiv$	-	-	-	-	+	+	+

H	H-	H ⁺
H-	+	+
H ⁺	+	+

O	$:O =$	$\rightarrow O -$	$:O^-$
$:O =$	+	+	+
$\rightarrow O -$	+	+	+
$:O^-$	+	+	+

Die Elemente O und N sind als Heteroatome markiert.

3) Grenzwerte für *R*- und *T*-Matrizen :

Maximaler Rang einer irreduziblen <i>R</i> -Matrix	4
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>R</i> -Dreieck	4
Maximaler Rang einer irreduziblen <i>T</i> -Matrix	4
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>T</i> -Dreieck	2
Maximaler Betrag eines außerdiagonalen Eintrags in einer <i>R</i> -Matrix .	1
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>R</i> -Zeile ...	2
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>T</i> -Zeile ...	2
Maximaler Betrag einer <i>T</i> -Zeilensumme	∞
Maximaler Betrag einer <i>T</i> -Matrixgesamtsumme	∞

4) Konstitutionelle Randbedingungen :

Maximale Ringspannung	264
Bredt'sche Regel beachten	ja
Kumulierte Mehrfachbindungen in Ringen erlaubt ab Ringgröße	10
Dreifachbindungen in Ringen erlaubt ab Ringgröße	10
Maximale Länge von Heteroatomketten	3
Maximale Anzahl geladener Atome	3
Maximale chemische Distanz zum Edukt	∞
Maximale chemische Distanz zum Produkt	∞
Anzahl weiter zu entwickelnder MCD-Gruppen	∞

5) Sonstiges :

- Es werden keine minimalen Bindungsordnungen gefordert.
- Automerisationsreaktionen sind als Reaktionsschritte nicht erlaubt.
- Unmittelbare Rückreaktionen sind nicht erlaubt.

Abbildung (7-8) zeigt das von *RA/W* unter diesen Randbedingungen für die Diazotierungsreaktion erzeugte Reaktionsnetzwerk.

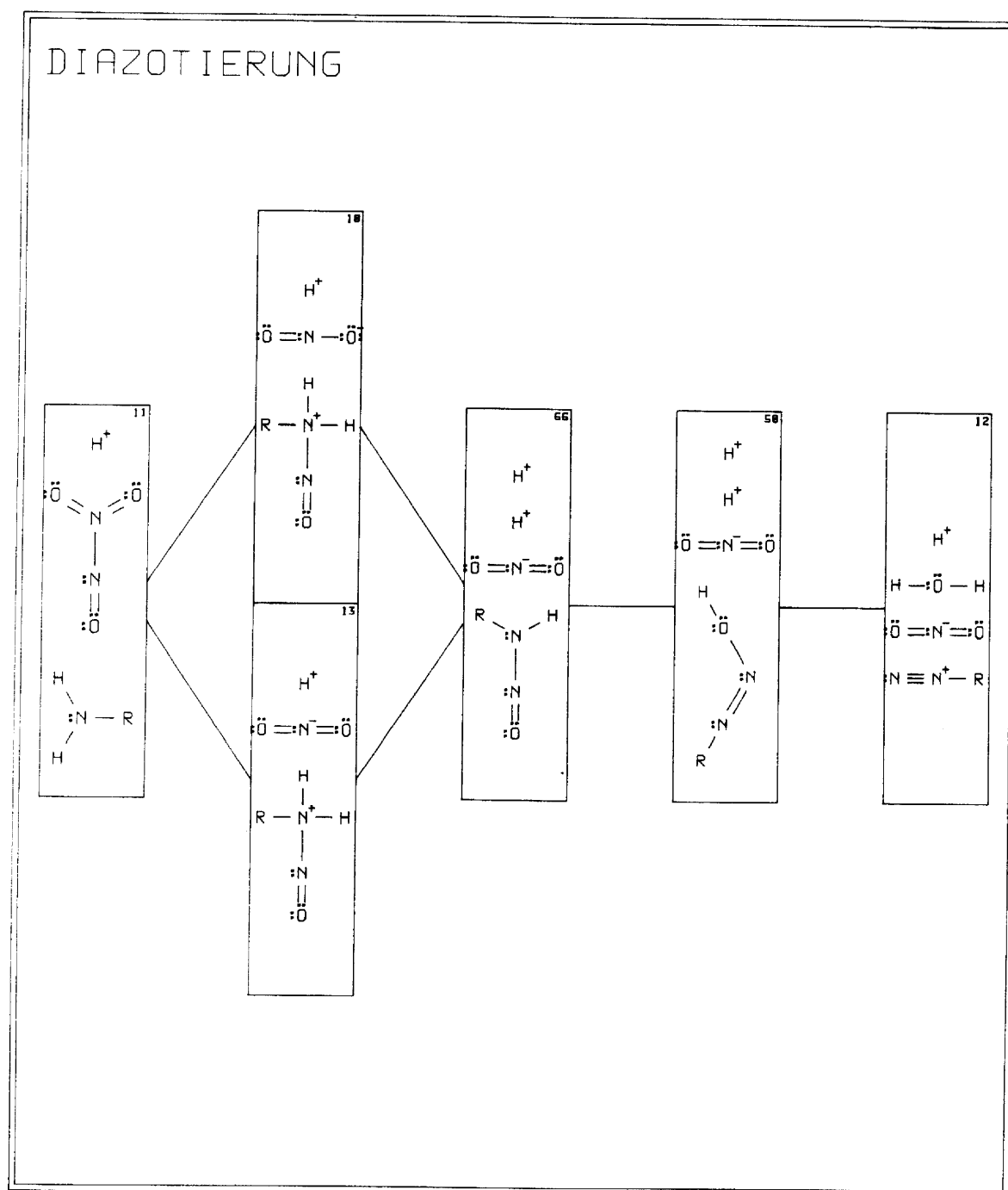


Abb. (7-8) Reaktionsnetzwerk der Diazotierungsreaktion.

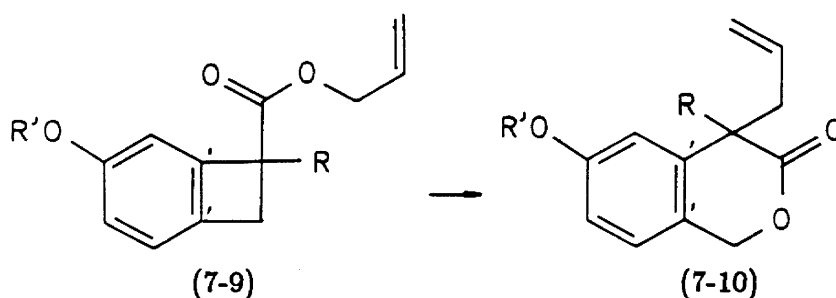
Dieses Reaktionsnetzwerk stellt unter den gegebenen Bedingungen die kürzest möglichen Reaktionspfade der Diazotierung dar. Der "Lehrbuch-Mechanismus" wird exakt wiedergegeben. Die beiden Alternativen mit den Molekülen 13 und 18 stellen lediglich verschiedene Resonanzstrukturen ein und desselben Moleküls dar (siehe Kap. 8).

Da die Reaktion unter mineralsauren Bedingungen abläuft, wurde dem Ensemble ein Proton beigefügt, um zu sehen, ob ein Eingriff dieses Protons in den Reaktionsverlauf stattfindet. Wie zu sehen ist, bleibt das Proton jedoch über alle Reaktionsschritte hinweg unverändert, ist also zur Formulierung des Reaktionsmechanismus nicht nötig.

7.3 Elektrocyclische Tandemreaktion von Benzocyclobutenen

Komplexe Gerüstumwandlungen eignen sich besonders als Studienobjekt für *RAIN*. Die strukturellen Veränderungen, die durch eine Abfolge von elektrocyclischen Reaktionen erzeugt werden, übersteigen ziemlich schnell die Möglichkeiten der menschlichen Kombinationsfähigkeit. Hier ist es besonders hilfreich, eine maschinelle Unterstützung zur Hand zu haben.

Betrachtet wird eine elektrocyclisch-sigmatrope Tandemreaktion eines Bicyclobutens (7-9) zu einem 4,4-disubstituierten Isochromanon⁶⁰ (7-10).



Für die Möglichkeiten der Atom-auf-Atom-Abbildung nach dem PMCD mit zugehörigen Reaktionskernen siehe (4-8). Um eine ausreichende Vielfalt der Reaktionsmöglichkeiten zu erhalten, müssen auch die Atome des Benzolrings, die am Vierring teilhaben, als reaktiv markiert werden. Um jedoch die Integrität des Benzolrings nicht aufzugeben, werden diese Atome mit einer Mindestbindungsordnung "zusammengehalten".

Für die Analyse mit *RAIN* werden die folgenden Randbedingungen gesetzt :

1) Reaktionskern :

Die Wasserstoffatome sollen nicht an den Reaktionsschritten beteiligt sein dürfen. Die Benzolringatome (außer den C'-Atomen) und der OR-Rest gehören ebenfalls nicht zum Reaktionskern (siehe (4-9b)).

2) Valenzschemata :

C'	$\begin{array}{c} \diagup \\ \text{C}' - \\ \diagdown \end{array}$
$\begin{array}{c} \diagup \\ \text{C}' - \\ \diagdown \end{array}$	+

O	$\rightarrow \ddot{\text{O}} -$	$:\ddot{\text{O}} =$
$\rightarrow \ddot{\text{O}} -$	+	+
$:\ddot{\text{O}} =$	+	+

C	$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	$\begin{array}{c} \diagup \\ \text{C}- \\ \diagdown \end{array}$	$\equiv \text{C}-$
$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	+	+	-
$\begin{array}{c} \diagup \\ \text{C}- \\ \diagdown \end{array}$	+	+	+
$\equiv \text{C}-$	-	+	+

Das Element O ist als Heteroatome markiert. Das Element C' steht für ein Pseudo-C-Atom, das nur das sp^2 -Valenzschema annehmen kann.

3) Grenzwerte für *R*- und *T*-Matrizen :

Maximaler Rang einer irreduziblen <i>R</i> -Matrix	6
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>R</i> -Dreieck	6
Maximaler Rang einer irreduziblen <i>T</i> -Matrix	4
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>T</i> -Dreieck	2
Maximaler Betrag eines außerdiagonalen Eintrags in einer <i>R</i> -Matrix .	1
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>R</i> -Zeile ...	2
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>T</i> -Zeile ...	2
Maximaler Betrag einer <i>T</i> -Zeilensumme	∞
Maximaler Betrag einer <i>T</i> -Matrixgesamtsumme	∞

4) Konstitutionelle Randbedingungen :

Maximale Ringspannung	264
Bredt'sche Regel beachten	ja
Kumulierte Mehrfachbindungen in Ringen erlaubt ab Ringgröße	10
Dreifachbindungen in Ringen erlaubt ab Ringgröße	10
Maximale Länge von Heteroatomketten	2
Maximale Anzahl geladener Atome	2
Maximale chemische Distanz zum Edukt	∞
Maximale chemische Distanz zum Produkt	∞
Anzahl weiter zu entwickelnder MCD-Gruppen	∞

5) Sonstiges :

- Zwischen den *C'*-Atomen wird eine minimale Bindungsordnung vom Wert 1 gefordert.
- Automerisationsreaktionen sind als Reaktionsschritte nicht erlaubt.
- Unmittelbare Rückreaktionen sind nicht erlaubt.

Abbildung (7-11) zeigt das von *RAIN* unter diesen Randbedingungen für diese Umlagerungsreaktion erzeugte Reaktionsnetzwerk.

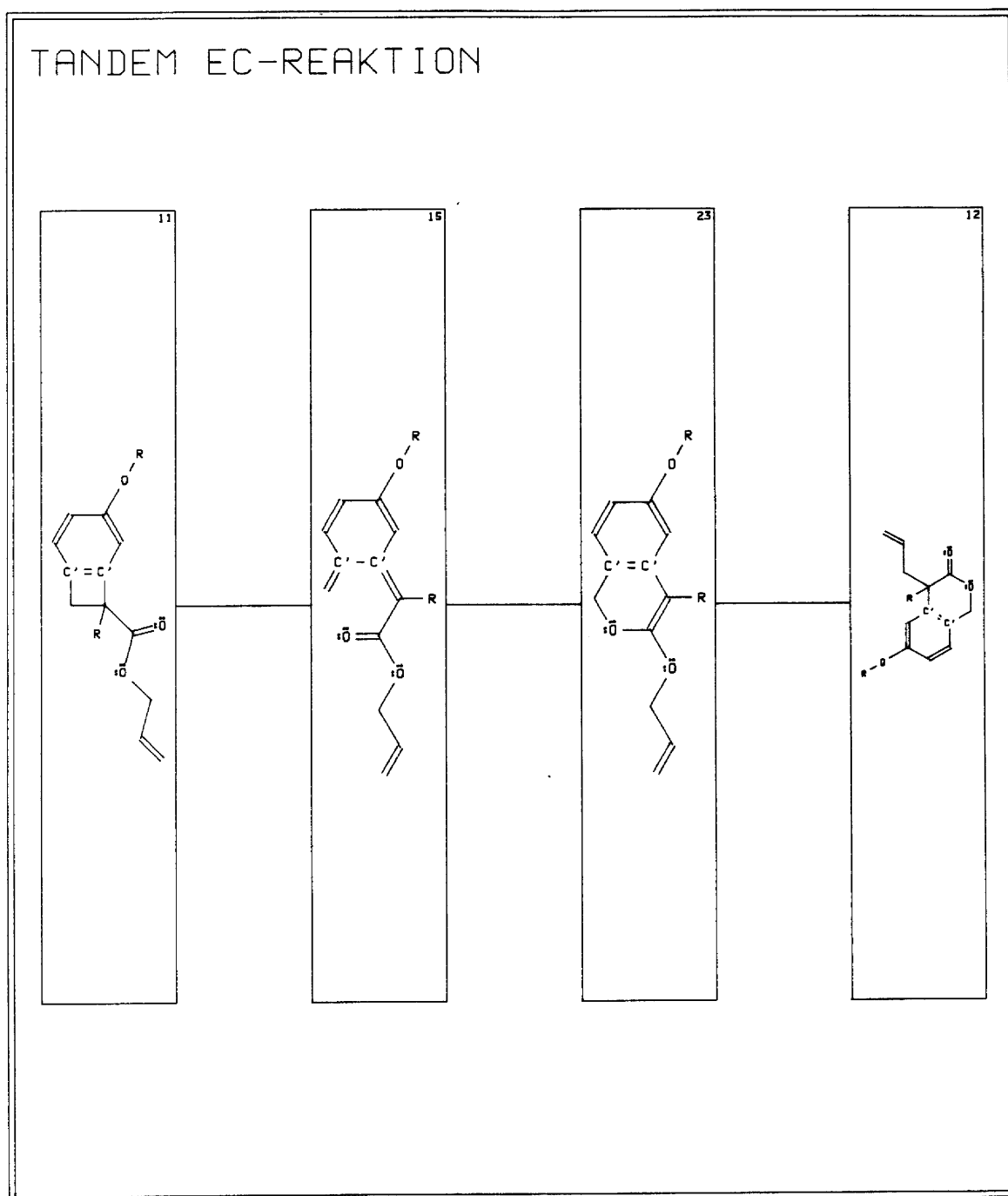
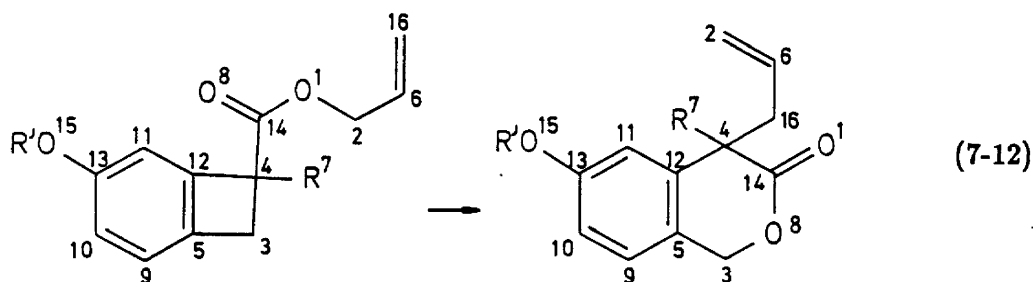


Abb. (7-11) Reaktionsnetzwerk einer electrocyclischen Tandemreaktion.

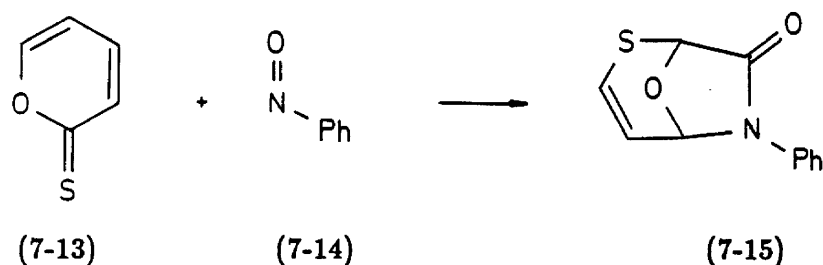
Dieses Reaktionsnetzwerk mit nur einem möglichen Reaktionspfad über zwei Zwischenstufen ist nicht das kürzest mögliche. Unter den gegebenen Randbedingungen existiert auch eine unmittelbare Verbindung der Moleküle 11 und 23, unter Auslassung von Molekül 15. Das gezeigte Reaktionsnetzwerk (7-11) entspricht jedoch vollkommen dem von den Autoren durch Identifizierung von Zwischenstufen gefundenen Reaktionsmechanismus.

Ein Vergleich dieser Lösung mit der in (4-8) durch Anwendung des PMCD erhaltenen Abbildungsvorschriften zeigt, daß keine dieser dort gezeigten Vorschläge dem wirklichen Reaktionsverlauf entspricht. Die Anwendung des PMCD hätte hier also das Auffinden dieser doch recht einfachen Lösung unmöglich gemacht. *RAIN* erzeugt die in (7-12) gezeigte Atomzuordnung mit einer chemischen Distanz von 16 (die minimale chemische Distanz beträgt 8).



7.4 Streith-Reaktion

Die 1982 entdeckte quantitative Umsetzung von Pyran-2-thion (7-13) mit Nitrosobenzol (7-14) zu einem bizyklischen Additionsprodukt (7-15) (*Streith-Reaktion*^{61,62}) war 1983 Gegenstand eines von *Jung* geleiteten "Summer Mechanism Contest" an der "University of California" in Los Angeles⁶³.



Eine Untersuchung dieser Reaktion mit dem Programm PEMCD³⁹ liefert 4 Zuordnungsmöglichkeiten mit einer minimalen chemischen Distanz von 28. Für die Analyse mit *RA/N* werden die folgenden Randbedingungen gesetzt :

1) Reaktionskern :

Alle Atome außer den Wasserstoffatomen sollen an den Reaktionsschritten beteiligt sein dürfen.

2) Valenzschemata :

C	$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	$\begin{array}{c} // \\ \text{C}- \\ / \end{array}$
$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	+	+
$\begin{array}{c} // \\ \text{C}- \\ / \end{array}$	+	+

N	$\begin{array}{c} \backslash \\ \text{:N}- \\ / \end{array}$	$\text{=N}-$	$\begin{array}{c} // \\ \text{N}^+- \\ / \end{array}$
$\begin{array}{c} \backslash \\ \text{:N}- \\ / \end{array}$	+	+	+
$\text{=N}-$	+	+	+
$\begin{array}{c} // \\ \text{N}^+- \\ / \end{array}$	+	+	+

O	$\text{--}\ddot{\text{O}}-$	$\text{:}\ddot{\text{O}}=$	$\text{:}\ddot{\text{O}}\text{--}$
$\text{--}\ddot{\text{O}}-$	+	+	+
$\text{:}\ddot{\text{O}}=$	+	+	+
$\text{:}\ddot{\text{O}}\text{--}$	+	+	+

S	$\text{--}\ddot{\text{S}}-$	$\text{:}\ddot{\text{S}}=$	$\text{:}\ddot{\text{S}}\text{--}$
$\text{--}\ddot{\text{S}}-$	+	+	+
$\text{:}\ddot{\text{S}}=$	+	+	+
$\text{:}\ddot{\text{S}}\text{--}$	+	+	+

Die Elemente *N*, *O* und *S* sind als Heteroatome markiert.

3) Grenzwerte für *R*- und *T*-Matrizen :

Maximaler Rang einer irreduziblen <i>R</i> -Matrix	6
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>R</i> -Dreieck	6
Maximaler Rang einer irreduziblen <i>T</i> -Matrix	4
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>T</i> -Dreieck	2
Maximaler Betrag eines außerdiagonalen Eintrags in einer <i>R</i> -Matrix .	1
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>R</i> -Zeile ...	2
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>T</i> -Zeile ...	2
Maximaler Betrag einer <i>T</i> -Zeilensumme	∞
Maximaler Betrag einer <i>T</i> -Matrixgesamtsumme	∞

4) Konstitutionelle Randbedingungen :

Maximale Ringspannung	264
Bredt'sche Regel beachten	ja
Kumulierte Mehrfachbindungen in Ringen erlaubt ab Ringgröße	10
Dreifachbindungen in Ringen erlaubt ab Ringgröße	10
Maximale Länge von Heteroatomketten	2
Maximale Anzahl geladener Atome	2
Maximale chemische Distanz zum Edukt	∞
Maximale chemische Distanz zum Produkt	∞
Anzahl weiter zu entwickelnder MCD-Gruppen	∞

5) Sonstiges :

- Es werden keine minimalen Bindungsordnungen gefordert.
- Automerisationsreaktionen sind als Reaktionsschritte nicht erlaubt.
- Unmittelbare Rückreaktionen sind nicht erlaubt.

Abbildung (7-16a mit 7-16b) zeigt das von *RAIN* unter diesen Randbedingungen für die *Streith*-Reaktion erzeugte Reaktionsnetzwerk.

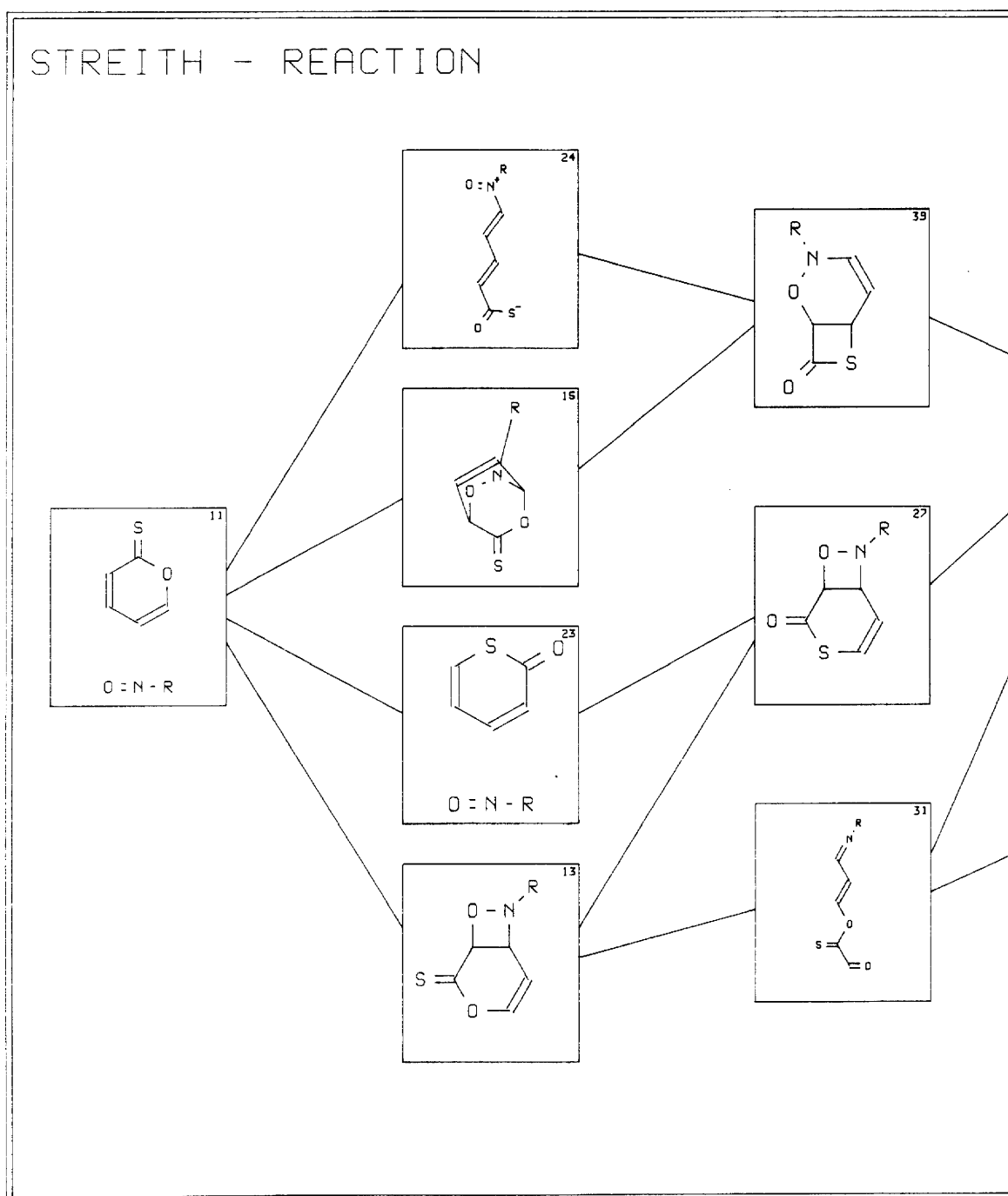


Abb. (7-16a) Reaktionsnetzwerk der *Streith*-Reaktion (Teil a).

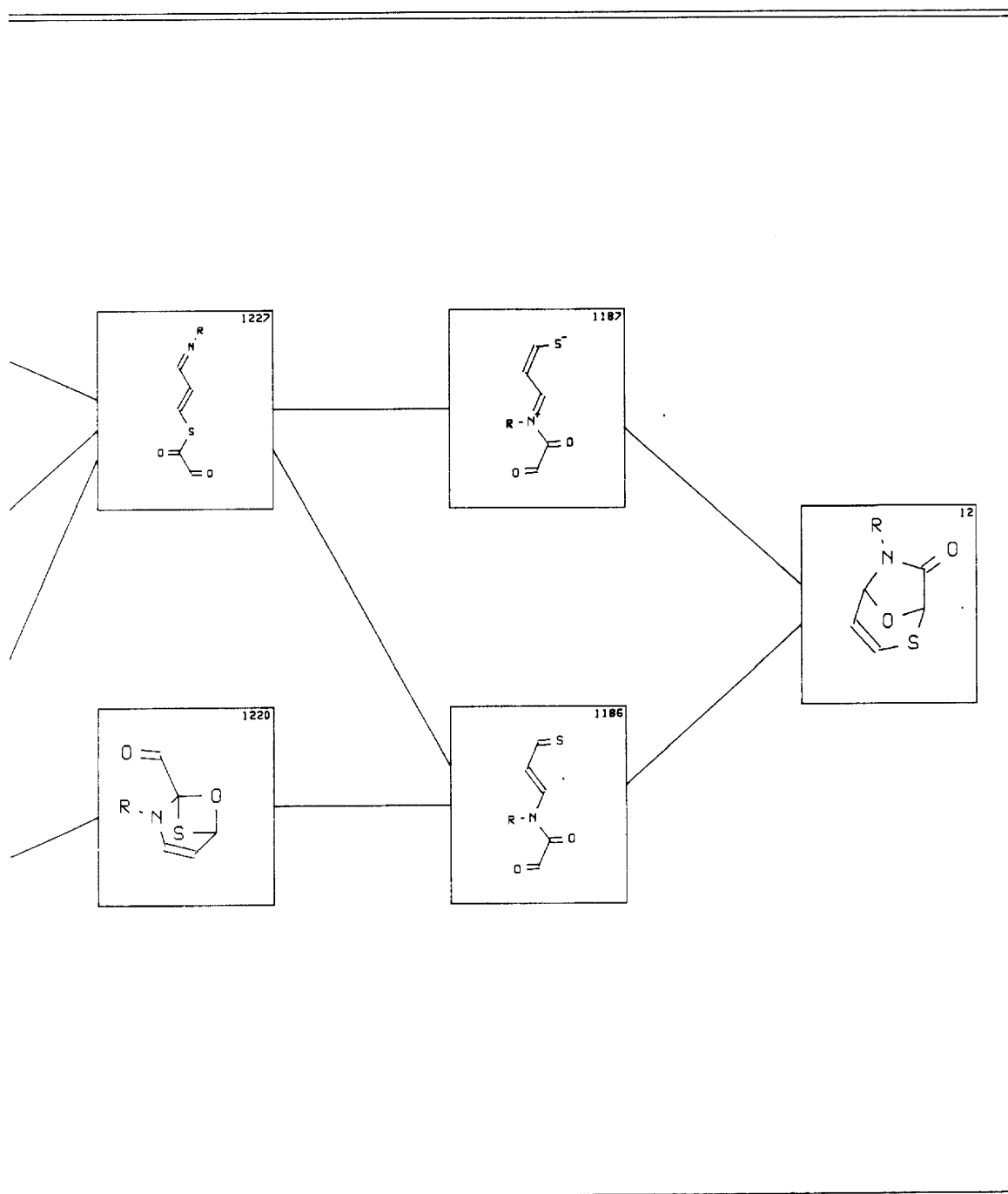


Abb. (7-16b) Reaktionsnetzwerk der *Streith*-Reaktion (Teil b).

Der Mechanismus der *Streith*-Reaktion wurde 1985 weitgehend aufgeklärt⁶². Der von den Autoren vorgeschlagene Reaktionspfad verläuft über die Zwischenstufen 15, 39, 1227 und 1186 und entspricht einer Verschiebung von netto 8 Elektronenpaaren ($CD = 32$). Er beginnt mit einer Hetero-*Diels-Alder*-Reaktion und führt über eine — für *RAIN* unter den gegebenen Randbedingungen nicht nötige — geladene Zwischenstufe (7-17)



zum Thiolacton (Molekül 39). Dieses ist bei einer Reaktionstemperatur von $-10\text{ }^{\circ}\text{C}$ isolierbar⁶², reagiert jedoch bereits bei Raumtemperatur unter einer electrocyclischen Ringöffnung weiter zum Formylthioester (Molekül 1227). Von hier führt der Weg über zwei electrocyclische 6-Zentrenumlagerungen zum bicyclischen Produkt (Molekül 12). Dieser Reaktionsmechanismus wurde bereits 1983 von *Jung* vorgeschlagen.

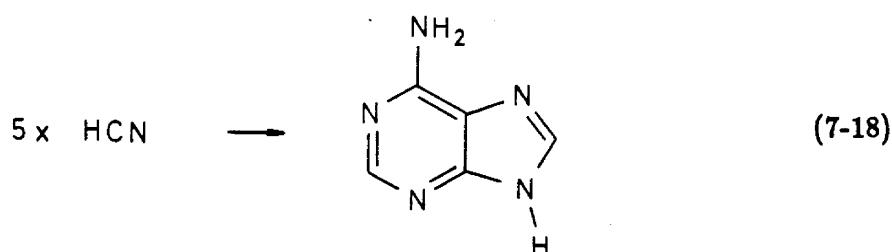
Ein Angriff des *N*-Atoms von Molekül 1227 am *C*(1)-Carbonyl- statt am *C*(2)-Carbonyl-*C*-Atom führt unter gleichzeitiger (1,2)-Hydridverschiebung zu einer Verbindung 1186' mit umgedrehter Orientierung des Dicarbonylrestes. Dieser Reaktionspfad, der unter den gesetzten Randbedingungen nicht erlaubt ist, beinhaltet die Verschiebung von netto 7 Elektronenpaaren und entspricht somit einer Reaktionsfolge, die dem Prinzip der minimalen chemischen Distanz gehorcht. Er wurde im Mechanismuswettbewerb neben einem Reaktionspfad über die Moleküle 13, 31, 1220 und 1186 gleichfalls favorisiert.

Eine Weiterentwicklung des Netzwerkes würde durch eine Erzeugung weiterer Zwischenstufen auch die in den publizierten Mechanismen auftretenden geladenen Intermediate (z.B. (7-17) zwischen 15 und 39) im Netzwerk integrieren.

7.5 Praebiotische Adeninsynthese

Seit den grundlegenden Experimenten von *Miller*^{64–67}, *Oró* und *Kimball*^{68–70} sind auf dem Gebiet der "chemischen Evolution der Biomoleküle" große Fortschritte gemacht worden. In simulierten Uratmosphären und unter den extremen Bedingungen, wie sie auf der jungen Erde geherrscht haben, sind nahezu alle biologisch wichtigen Moleküle wie Aminosäuren, Nucleinbasen und Zucker gefunden worden. Bei der praebiotischen Bildung dieser Biomoleküle spielen vor allen Dingen elektrische Entladungen (Blitze), Stoßwellen von Meteoriteinschlägen, starke Sonneneinstrahlung, die Abwesenheit von Sauerstoff in der Atmosphäre und nicht zuletzt radioaktive Strahlung eine wichtige Rolle.

Ein wesentlicher Ausgangsstoff für stickstoffhaltige Verbindungen ist der Cyanwasserstoff *HCN*, von dem angenommen wird, daß er einen nicht unbeträchtlichen Anteil an der Uratmosphäre gehabt hat⁷¹. Eine der am intensivsten studierte Reaktion des *HCN* ist die Oligomerisierung zum Adenin (7-18).



Diese für die Entstehung des Lebens entscheidende Reaktion bildet die Grundlage einer *RAIN*-Studie, um mögliche Mechanismen bzw. Zwischenstufen aufzuklären. Für die Analyse mit *RAIN* werden die folgenden Randbedingungen gesetzt :

1) Reaktionskern :

Alle 15 Atome des Ensembles ($5HCN$) sollen an den Reaktionsschritten beteiligt sein dürfen.

2) Valenzschemata :

C	$\begin{array}{c} \diagup \\ \text{C} - \\ \diagdown \end{array}$	$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	$\equiv \text{C} -$
$\begin{array}{c} \diagup \\ \text{C} - \\ \diagdown \end{array}$	+	+	-
$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	+	+	-
$\equiv \text{C} -$	+	-	+

N	$\begin{array}{c} \diagup \\ \text{N} - \\ \diagdown \end{array}$	$\Rightarrow \text{N} -$	$\text{N} \equiv$
$\begin{array}{c} \diagup \\ \text{N} - \\ \diagdown \end{array}$	+	+	-
$\Rightarrow \text{N} -$	+	+	-
$\text{N} \equiv$	-	+	+

H	H-
H-	+

Das Element N ist als Heteroatom markiert. Für die dreifachbindigen Valenzschemata der C - und N -Atome wird gefordert, daß diese zwar stabil sind, aber nicht aus anderen Valenzschemata hervorgehen können. Dies hat zur Folge, daß, sobald ein HCN -Molekül unter Aufgabe der Dreifachbindung in den Molekülverband eingetreten ist, das HCN -Molekül nicht wieder abgespalten werden kann. Der globalen Reaktionsrichtung ist also eine monotone Abnahme der $C \equiv N$ -Bindungen überlagert.

3) Grenzwerte für *R*- und *T*-Matrizen :

Maximaler Rang einer irreduziblen <i>R</i> -Matrix	6
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>R</i> -Dreieck	6
Maximaler Rang einer irreduziblen <i>T</i> -Matrix	4
Maximale Anzahl von außerdiagonalen Einträgen in einem <i>T</i> -Dreieck	3
Maximaler Betrag eines außerdiagonalen Eintrags in einer <i>R</i> -Matrix .	1
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>R</i> -Zeile ...	2
Maximale Anzahl von außerdiagonalen Einträgen in einer <i>T</i> -Zeile ...	2
Maximaler Betrag einer <i>T</i> -Zeilensumme	∞
Maximaler Betrag einer <i>T</i> -Matrixgesamtsumme	∞

4) Konstitutionelle Randbedingungen :

Maximale Ringspannung	200
Bredt'sche Regel beachten	ja
Kumulierte Mehrfachbindungen in Ringen erlaubt ab Ringgröße	10
Dreifachbindungen in Ringen erlaubt ab Ringgröße	10
Maximale Länge von Heteroatomketten	1
Maximale Anzahl geladener Atome	0
Maximale chemische Distanz zum Edukt	∞
Maximale chemische Distanz zum Produkt	∞
Anzahl weiter zu entwickelnder MCD-Gruppen	∞

5) Sonstiges :

- Es werden keine minimalen Bindungsordnungen gefordert.
- Automerisationsreaktionen sind als Reaktionsschritte nicht erlaubt.
- Unmittelbare Rückreaktionen sind nicht erlaubt.

Abbildung (7-18a mit 7-18b) zeigt das von *RAIN* unter diesen Randbedingungen für die Adeninsynthese erzeugte Reaktionsnetzwerk.

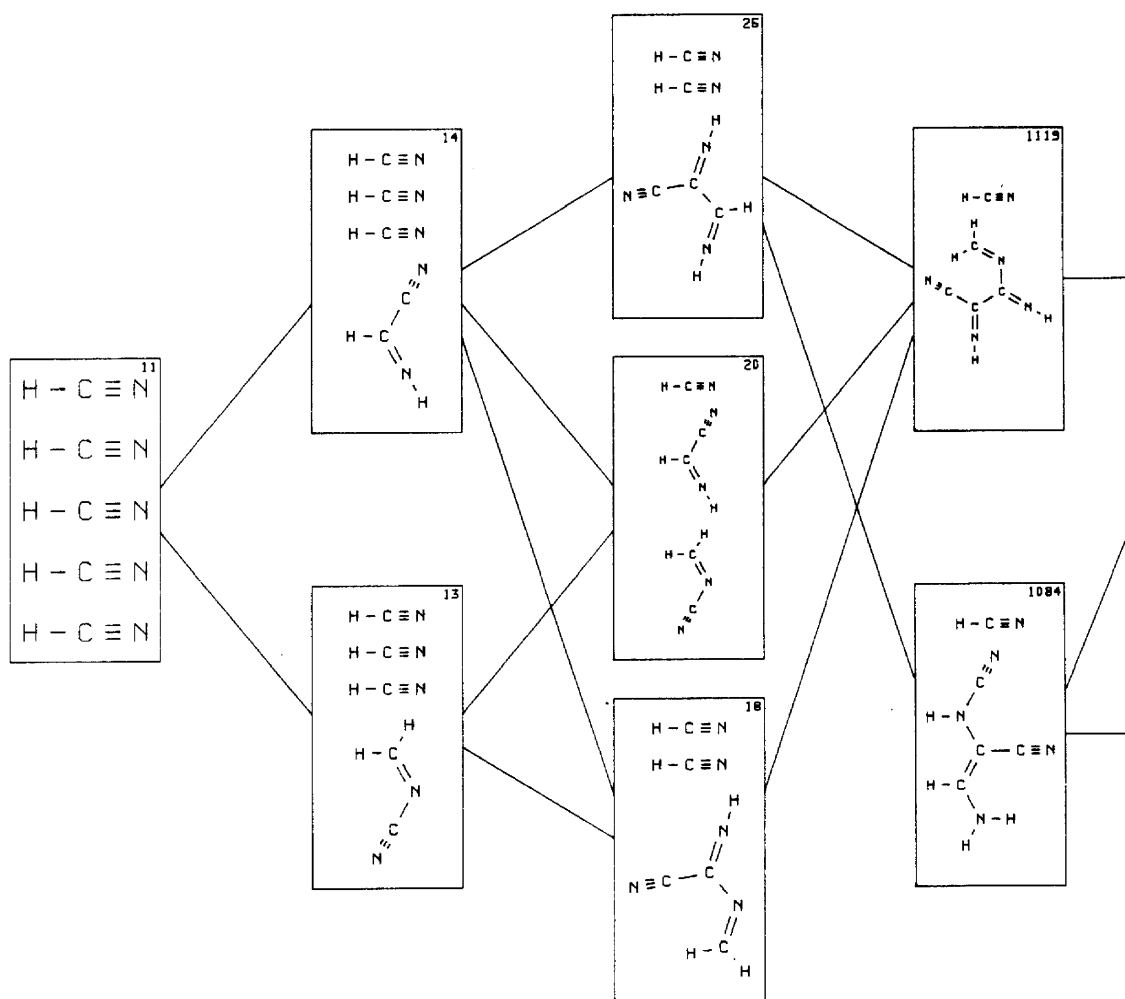
5 HCN $\rightarrow$ ADENIN

Abb. (7-18a) Reaktionsnetzwerk der prebiotischen Adeninsynthese (Teil a).

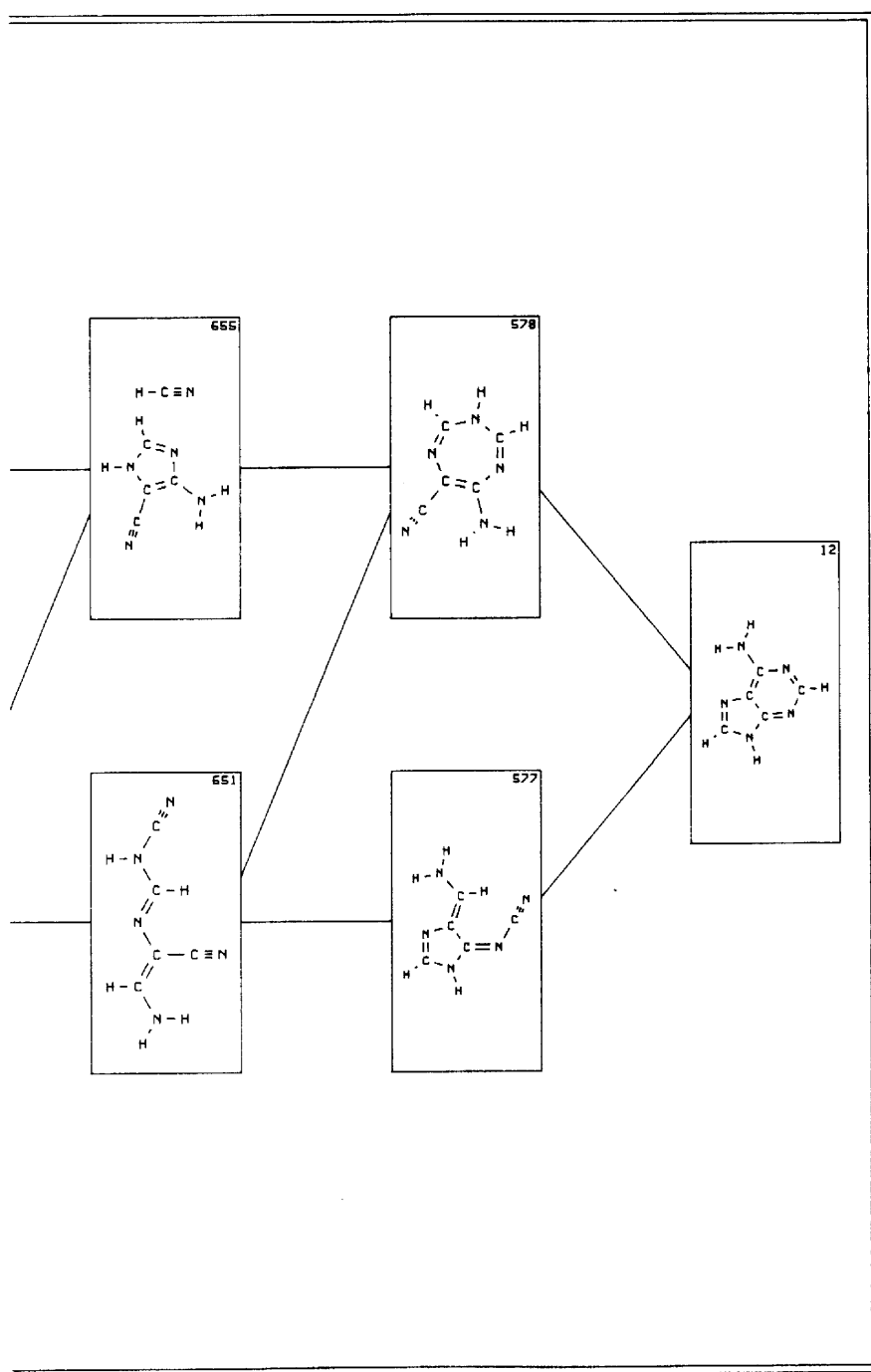
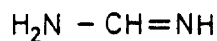


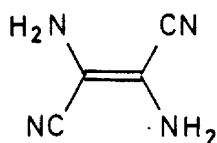
Abb. (7-18b) Reaktionsnetzwerk der praebiotischen Adeninsynthese (Teil b).

Das gezeigte Reaktionsnetzwerk repräsentiert die unter den gegebenen Randbedingungen kürzest möglichen Reaktionspfade. Ein Vergleich mit den von mehreren Autoren vorgeschlagenen Reaktionssequenzen^{72,73} zeigt folgendes :

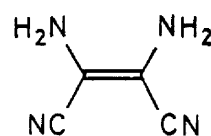
In allen vorgeschlagenen Mechanismen stellt das 4-Amino-imidazol-5-carbonitril (Molekül 655) oder das tautomere 5-Amino-imidazol-4-carbonitril zusammen mit einem Molekül *HCN* die unmittelbare Vorstufe des Adenin dar. Im Reaktionsnetzwerk (7-18) kann der Ringschluß des Sechsrings nicht in einem Schritt geschehen, sondern verläuft über eine Triazepin-Zwischenstufe (Molekül 578). In den publizierten Vorschlägen zum Reaktionsmechanismus tritt das *HCN*-Molekül alternativ auch als Formamidin (7-19) in den Molekülverband ein, was jedoch die Anwesenheit von *NH₃* erfordert, das in dem dieser Studie zugrunde liegenden Ensemble nicht vorkommt.



(7-19)



(7-20)



(7-21)

Als unmittelbare Vorstufe des 4-Amino-imidazol-5-carbonitril wird das Tetramere des HCN , das Diaminofumaronitril (7-20) angesehen, das durch *cis-trans*-Isomerisierung aus Diaminomaleonitril (7-21) entstehen soll. Diese Zwischenstufe und der ganze zugehörige davorliegende Reaktionspfad (siehe (7-22)) fehlt im Netzwerk (7-18). Der Grund liegt darin, daß für den Schlüsselreaktionsschritt eine komplizierte photochemische Umlagerung des Diaminofumaronitril zum 4-Amino-imidazol-5-carbonitril angenommen wird^{74,75}. Die Autoren ziehen für diesen Schritt, der die Orientierung einer CN -Funktion umkehren muß, eine Isomerisierung zum Isonitril mit anschließendem Ringschluß, oder einen über eine Vierring- oder Dreiringzwischentufe verlaufenden Weg in Erwägung. Alle diese Alternativen sind unter den gegebenen Randbedingungen für die Netzwerkgenerierung nicht zulässig, da weder ein geladenes oder carbenoides Valenzschema für die C -Atome definiert ist, noch die maximale Ringspannung einen Vierring oder Dreiring erlaubt. *RAIN* benötigt für diesen Reaktionsschritt zwei Zwischenstufen (siehe (7-23), Molekül 11 nach Molekül 562), sodaß bei seiner Integration das entstehende Reaktionsnetzwerk wesentlich länger und sehr viel komplizierter wäre, als das dargestellte. Die von *RAIN* vorgeschlagenen Reaktionswege beginnen zwar — wie die publizierten Reaktionsmechanismen auch — mit der Bildung von Iminocarbonitril (Molekül 14), verlaufen dann jedoch, ohne die Zwischenstufe (7-20) zu benötigen, auf kürzerem Weg zum 4-Amino-imidazol-5-carbonitril.

Um zu prüfen, ob *RAIN* auch längere Reaktionspfade findet, die über das Diaminomaleonitril führen, werden in zwei weiteren Generierungsläufen sowohl ein Netzwerk von $4HCN$ zum Diaminomaleonitril, als auch ein Netzwerk von Diaminomaleonitril mit HCN zum Adenin erzeugt. Die Randbedingungen entsprechen in beiden Fällen den für das Netzwerk (7-18) angegebenen.

Abbildung (7-22) zeigt das von *RAIN* für die HCN -Oligomerisierungsreaktion zum HCN -Tetramer Diaminomaleonitril erzeugte Reaktionsnetzwerk.

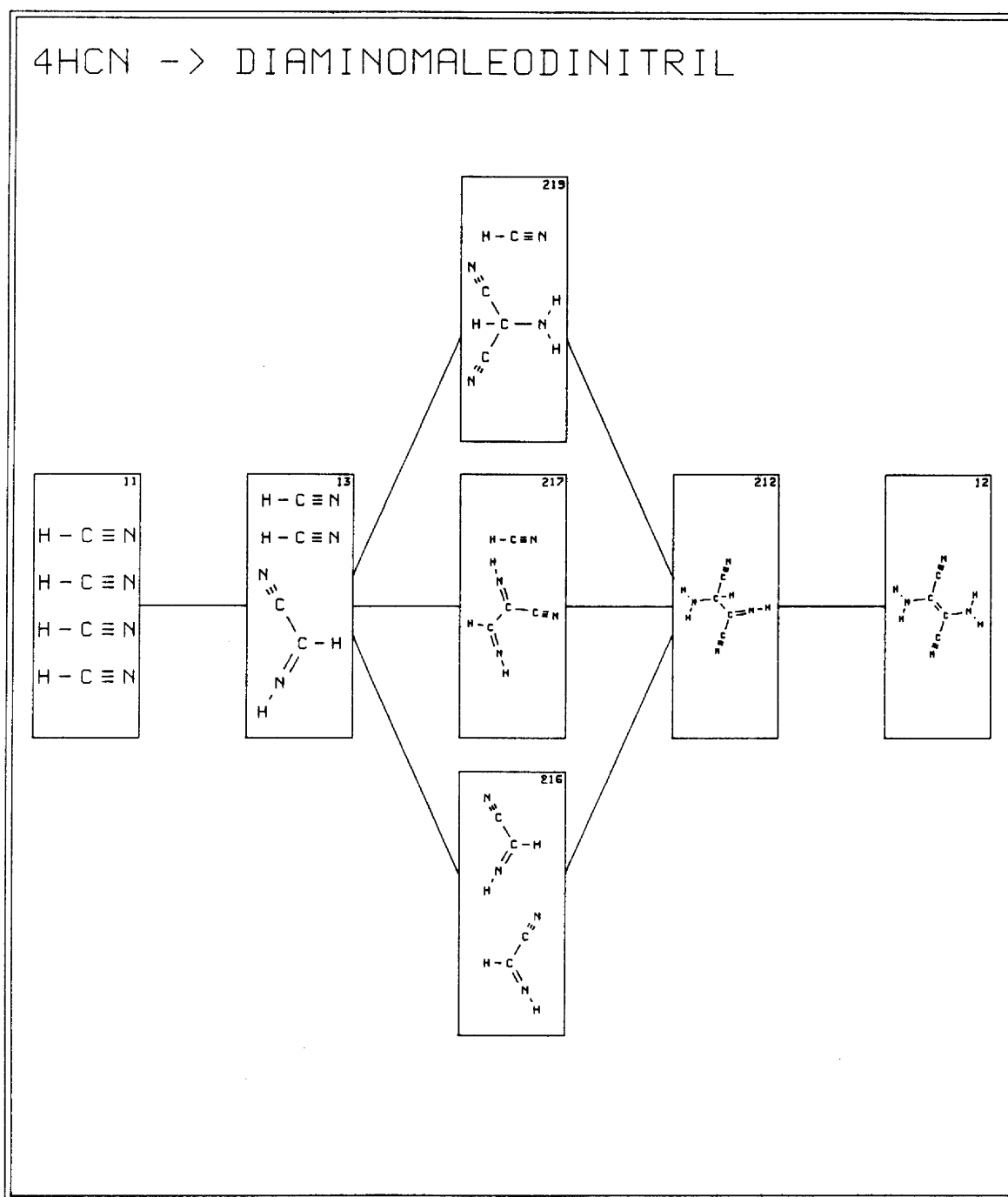


Abb. (7-22) Reaktionsnetzwerk der Bildung von Diaminomaleonitril.

Der Reaktionspfad über Aminomalonitril (Molekül 219) und Molekül 212 ist identisch mit den publizierten Vorschlägen. Da *RAIN* lediglich die konstitutionelle Molekülinformation bearbeitet, wird zwischen Diaminomaleonitril und Diaminofumaronitril nicht unterschieden.

Abbildung (7-23a mit 7-23b) zeigt das von *RAIN* für die Adeninsynthese aus Diaminomaleonitril erzeugte Reaktionsnetzwerk.

Dieses Netzwerk enthält — aus den bereits angegebenen Gründen — nicht die direkte Isomerisierung von Diaminomaleonitril (Molekül 11) in 4-Amino-imidazol-5-carbonitril (Molekül 562). Es werden jedoch Wege mit zwei Zwischenstufen vorgeschlagen, von denen einer (über die Moleküle 15 und 41) in einer publizierten Reaktionssequenz in Erwägung gezogen wird⁷⁶. Die Autoren geben jedoch, vom Molekül 15 ausgehend, der Abspaltung von NH_3 zum 4,5-Dicyano-imidazol den Vorzug gegenüber der Reaktion zum 4-Amino-imidazol-5-carbonitril. Da vom 4,5-Dicyano-imidazol innerhalb von drei Reaktionsschritten kein Weg zum Adenin führt, taucht diese Verbindung nicht im dargestellten Reaktionsnetzwerk (7-23) auf.

Vom 4-Amino-imidazol-5-carbonitril (Molekül 562), bzw. vom 5-Amino-imidazol-4-carbonitril (Molekül 558) führen die aus dem Netzwerk (7-18) bekannten Reaktionspfade zum Adenin.

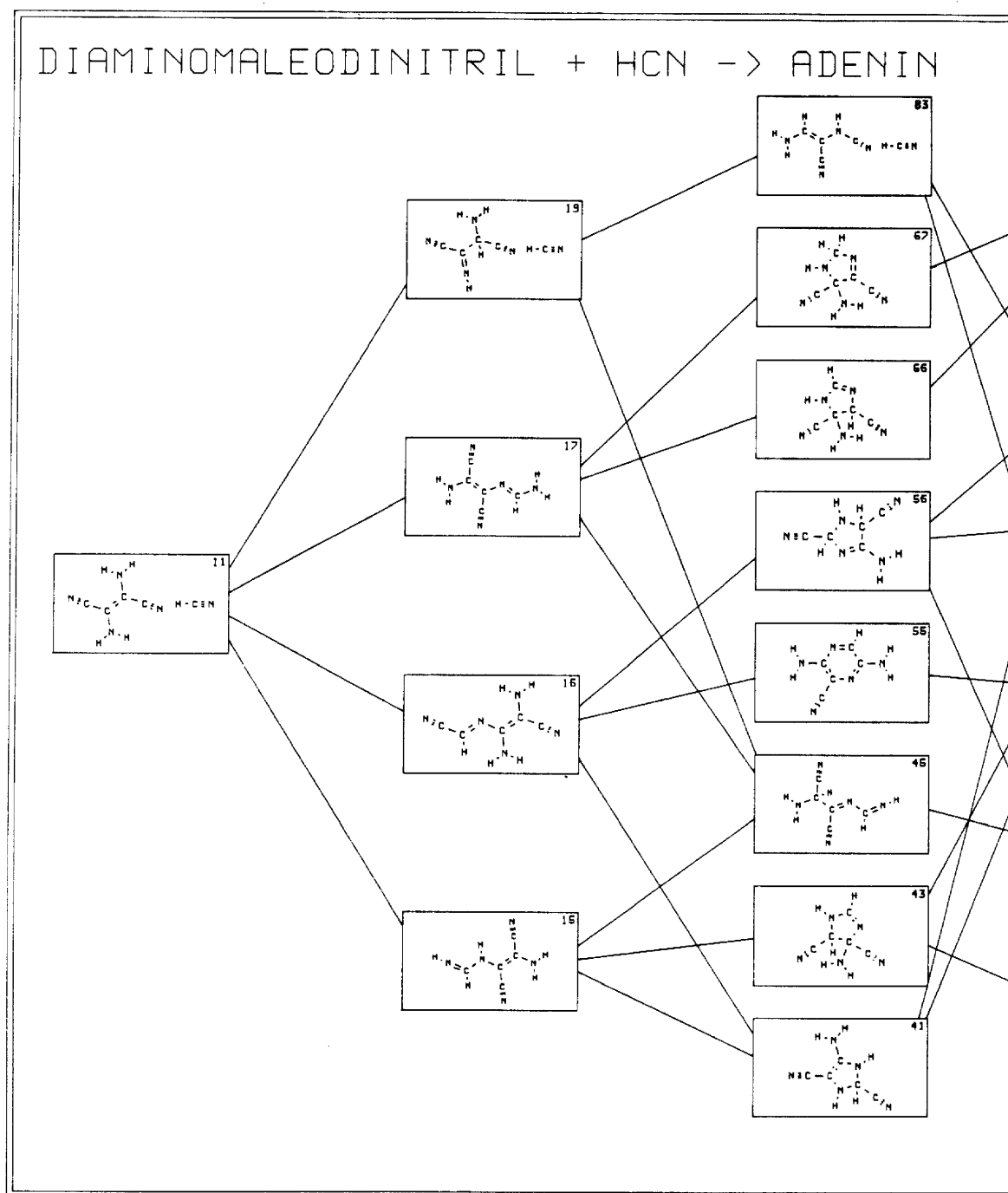


Abb. (7-23a) Reaktionsnetzwerk der Adeninsynthese aus Diaminomaleonitril(Teil a).

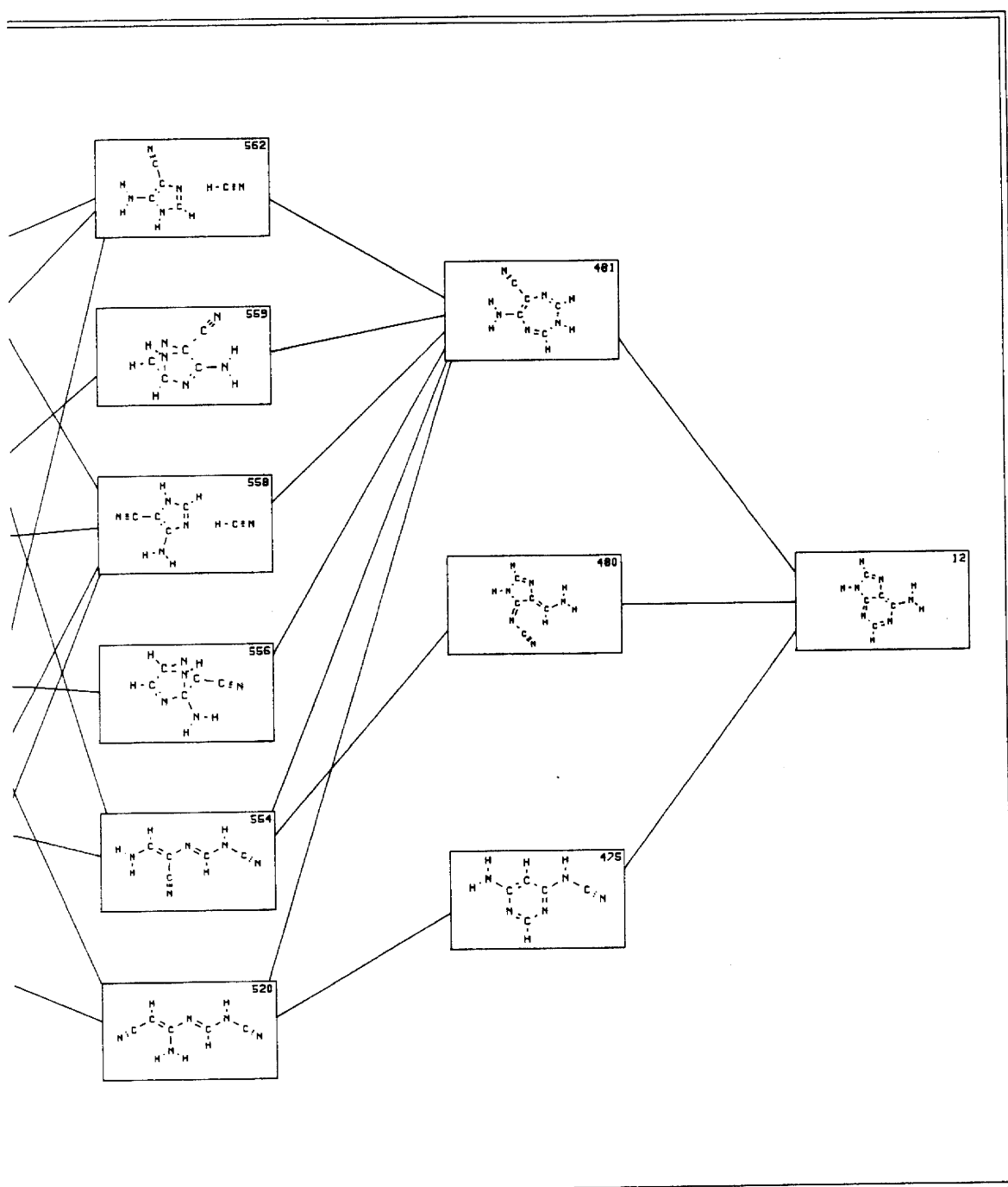


Abb. (7-23b) Reaktionsnetzwerk der Adeninsynthese aus Diaminomaleonitril(Teil b).

8 Zusammenfassung und Ausblick

Die vorliegende Arbeit beschreibt die Grundlagen eines Systems zur produkt- und eduktorientierten Erzeugung von Reaktionsnetzwerken. Diese Netzwerke enthalten die möglichen Reaktionspfade zwischen dem Edukt und dem Produkt einer chemischen Reaktion oder Reaktionsfolge. Sie können je nach Definition der valenzchemischen Elementeigenschaften von mechanistischer Natur sein oder Wege über stabile Intermediate beschreiben. Reaktionsnetzwerke dieser Art dienen zu mechanistischen Studien, zur Aufklärung biochemischer Auf- und Abbauege oder zur Planung von chemischen Reaktionssequenzen.

Der Aufbau der Netzwerke geschieht durch sukzessive Entwicklung von Reaktionssequenzen, die ihren Ursprung jeweils im Edukt bzw. Produkt der Gesamtreaktion haben. Moleküle, die in beiden Bäumen der Reaktionssequenzen vorkommen, stellen Brücken dar, über die Reaktionspfade vom Edukt zum Produkt führen. Die Reaktionssequenzen werden durch Anwendung eines neuartigen formalen Reaktionsgenerators erzeugt, für den der Anwender Regeln aufstellen kann. Hierzu wird ein Satz von Randbedingungen für die maximale Komplexität der erlaubten Reaktionsschritte und für die chemische Konstitution der Zwischenstufen definiert. Diese Randbedingungen machen es möglich, auf rein formalem Wege und unter der völligen Kontrolle des Anwenders, kombinatorisch vollständige Reaktionsnetzwerke zu erhalten.

Die zur Anwendung kommenden mathematischen Grundlagen der konstitutionellen Chemie auf der Basis der *BE*- und *R*-Matrizen werden erweitert um die Beschreibung der Topologieveränderungen chemischer Reaktionen mit Hilfe der *T*-Matrizen. Diese neue Art von Matrizen protokolliert die tatsächlichen Gerüstveränderungen in einer chemischen Reaktion. Die *T*-Matrizen ermöglichen eine effiziente Kontrolle über die maximale Komplexität von Reaktionsschritten. Dies ist mit den *R*-Matrizen alleine unmöglich, da diese lediglich die Änderungen in den formalen Bindungsordnungen, aber nicht die Gerüstveränderungen beschreiben.

Um die Leistungsfähigkeit dieses Systems zu demonstrieren, werden mit einer Implementation des Programmsystems auf einem Personalcomputer fünf Anwendungsbei-

spiele mit steigender Komplexität aus verschiedenen Gebieten der organischen Chemie durchgeführt. Diese Beispiele werden mit allen Randbedingungen und den erhaltenen Ergebnissen dargelegt.

Für die Zukunft sind folgende Erweiterungen und Neuorientierungen des Systems geplant :

Die Integration eines Substruktursuchalgorithmus, der Moleküle mit verbotenen Teilstrukturen eliminiert, sowie für bestimmte Substrukturen nur definierte Transformationen zuläßt, wird es möglich machen, die chemische Konstitution und ihre Veränderungen, noch effizienter zu beschreiben.

Auf der anderen Seite soll die Schaffung einer benutzerfreundlichen Anwenderoberfläche mit grafischen Ein- und Ausgabemöglichkeiten dem Chemiker einen noch größeren Einblick in das Geschehen geben und seine Möglichkeiten zur Einflußnahme sehr stark erweitern. So ist ein System denkbar, daß dem Anwender die volle Kontrolle über die Weiterentwicklung einzelner Moleküle im Netzwerk gibt, indem es die erzeugten Strukturen am Bildschirm präsentiert, und die Integration dieser Moleküle in das Netzwerk der Entscheidung des Chemikers überläßt. Dies wäre die ideale Kombination von maschineller Konsequenz und menschlicher Erfahrung, was zu einer sehr großen Effizienz in der Ausnutzung von Computern führen kann. Die Stärke der Computer liegt eben nicht in der Assoziationsfähigkeit, sondern vielmehr in der "sturen" — aber fehlerfreien — Anwendung formaler Prinzipien. Auch wenn der Begriff der "artificial intelligence" zur Zeit en vogue ist, wird es auf lange Sicht immer noch sehr nützlich sein, "dumme" Maschinen zu besitzen, die keine — ja doch nur vorgetäuschte — eigene Intelligenz besitzen, dafür aber in der Hand von intelligenten Anwendern ein äußerst hilfreiches Instrument zur Potenzierung menschlicher Kreativität werden können.

9 *Literatur*

- 1) P.G. Dittmar, N.A. Farmer, W. Fisanick, *J. Chem. Inf. Comput. Sci.* **23** (1983) 93.
- 2) R. Attis, *J. Chem. Inf. Comput. Sci.* **23** (1983) 102.
- 3) R. Brenner, *IBM Nachrichten* **35** (1985) 23.
- 4) J.H. Winter, "*Chemische Syntheseplanung in Forschung und Industrie*", Springer-Verlag, Berlin Heidelberg 1982.
- 5) J. Brandt, *Habilitationsschrift*, Technische Universität München (1981).
- 6) J. Bauer, I. Ugi, *J. Chem. Res. (S)* **11** (1982) 298; *J. Chem. Res. (M)* (1982) 3101, 3201.
- 7) J. Bauer, *Dissertation*, Technische Universität München (1981).
- 8) J. Bauer, R. Herges, E. Fontain, *Chimia* **39** (1985) 43.
- 9) E. Fontain, J. Bauer, I. Ugi, *Chem. Lett.* (1987) 37.
- 10) E. Fontain, J. Bauer, I. Ugi, *Z. Naturforsch.* (im Druck).
- 11) I. Ugi, P.D. Gillespie, *Angew. Chem.* **83** (1971) 980, 982; *Angew. Chem., Int. Ed. Engl.* **10** (1971) 914, 915.
- 12) J. Dugundji, I. Ugi, *Top. Curr. Chem.* **39** (1973) 19.

- 13) J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie in W.T. Wipke, S.R. Heller, R.J. Feldmann, E. Hyde, "*Computer Representation and Manipulation of Chemical Information*", John Wiley & Sons, New York London 1974.
- 14) I. Ugi, J. Brandt, J. Friedrich, C. Jochum, P. Lemmen, W. Schubert, *Pure Appl. Chem.* **50** (1978) 1303.
- 15) I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, *Angew. Chem.* **91** (1979) 99; *Angew. Chem., Int. Ed. Engl.* **18** (1979) 111.
- 16) J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, I. Ugi in W.T. Wipke, W.J. Howe, "*Computer-Assisted Organic Synthesis*", *Am. Chem. Soc. Symposium Series* **61** (1977) 33.
- 17) J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie, I. Ugi, *Tetrahedron* **30** (1974) 1845.
- 18) I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, *Informal Commun. Math. Chem.* **6** (1979) 159.
- 19) J.C.J. Bart, E. Garagnani, *Z. Naturforsch.* **31b** (1976) 1646.
- 20) J.C.J. Bart, E. Garagnani, *Z. Naturforsch.* **32b** (1977) 455.
- 21) J.C.J. Bart, E. Garagnani, *Z. Naturforsch.* **32b** (1977) 465.
- 22) J.C.J. Bart, E. Garagnani, *Z. Naturforsch.* **32b** (1977) 678.
- 23) J. Gasteiger, C. Jochum, *Top. Curr. Chem.* **74** (1978) 93.
- 24) J. Gasteiger, H.G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki, *Top. Curr. Chem.* **137** (1987) 21.

- 25) Z. Hippe, *Plenary Lecture VIIIth ICCCRE*, Garmisch Partenkirchen 1985.
- 26) W. Schubert, *Dissertation*, Technische Universität München (1978).
- 27) W. Schubert, *Informal Commun. Math. Chem.* **6** (1979) 213.
- 28) W.T. Wipke in W.T. Wipke, S.R. Heller, R.J. Feldmann, E. Hyde, "*Computer Representation and Manipulation of Chemical Information*", John Wiley & Sons, New York London 1974.
- 29) W.T. Wipke, W.J. Howe, "*Computer-Assisted Organic Synthesis*", *Am. Chem. Soc. Symposium Series* **61** (1977) 33.
- 30) E.J. Corey, *Quart. Rev. (Chem. Soc. London)* **25** (1971) 455.
- 31) E.J. Corey, R.D. Cramer, W.J. Howe, *J. Am. Chem. Soc.* **94/2** (1972) 440.
- 32) E.J. Corey, W.T. Wipke, R.D. Cramer, W.J. Howe, *J. Am. Chem. Soc.* **94/2** (1972) 421.
- 33) D.A. Pensak, E.J. Corey in W.T. Wipke, W.J. Howe, "*Computer-Assisted Organic Synthesis*", *Am. Chem. Soc. Symposium Series* **61** (1977) 33.
- 34) A. Weise, *Z. Chem.* **12** (1972) 81.
- 35) A. Weise, *Z. Chem.* **15** (1975) 33.
- 36) A. Weise, *Z. Chem.* **19** (1979) 49.
- 37) M. Bender, *J. Am. Chem. Soc.* **73** (1951) 1626.

- 38) C. Jochum, J. Gasteiger, I. Ugi, *Angew. Chem.* **92** (1980) 503; *Angew. Chem., Int. Ed. Engl.* **19** (1980) 495.
- 39) M. Wochner, J. Brandt, A. v. Scholley, *Chimia* (im Druck).
- 40) J. Gasteiger, O. Dammer, *Tetrahedron* **34** (1978) 2939.
- 41) J. Gasteiger, C. Jochum, *J. Chem. Inf. Comput. Sci.* **19** (1979) 43.
- 42) A.T. Balaban, "Chemical Applications of Graph Theory", Academic Press, London 1976.
- 43) S.W. Benson, F.R. Cruickshank, D.M. Golden, G.R. Haugen, H.E. O'Neal, A.S. Rodgers, R. Shaw, R. Walsh, *Chem. Rev.* **69** (1969) 279.
- 44) G. Köbrich, *Angew. Chem.* **85** (1973) 494; *Angew. Chem., Int. Ed. Engl.* **12** (1973) 464.
- 45) R. Keese, *Angew. Chem.* **87** (1975) 568; *Angew. Chem., Int. Ed. Engl.* **14** (1975) 528.
- 46) C.T. Fike, *The Computer Journal* **18** (1975) 21.
- 47) J.S. Rohl, *The Computer Journal* **19** (1976) 156.
- 48) M.K. Roy, *The Computer Journal* **21** (1978) 296.
- 49) M. Randić, *J. Chem. Inf. Comput. Sci.* **15** (1975) 105.
- 50) D.G. Corneil, C.G. Gotlieb, *J. Assoc. Comput. Mach.* **17** (1970) 51.
- 51) W.C. Herndon, *Stud. Phys. Theor. Chem.* **28** (1983) 231.

- 52) W. Schubert, I. Ugi, *J. Am. Chem. Soc.* **100** (1978) 37.
- 53) W. Schubert, I. Ugi, *Chimia* **33** (1979) 183.
- 54) I. Ugi, J. Dugundji, R. Kopp, D. Marquading, "*Perspektives in Theoretical Stereochemistry*", Springer-Verlag, Berlin Heidelberg 1984.
- 55) C. Jochum, J. Gasteiger, *J. Chem. Inf. Comput. Sci.* **17** (1977) 113.
- 56) H.L. Morgan, *J. Chem. Doc.* **5** (1965) 107.
- 57) J.B. Hendrickson, A.G. Toczko, *J. Chem. Inf. Comput. Sci.* **23** (1983) 171.
- 58) M. Gardner, *Spektrum der Wissenschaft* **1** (1981) 6.
- 59) J. March, "*Advanced Organic Chemistry*", John Wiley & Sons, New York 1985.
- 60) K. Shishido, E. Shitara, K. Fukumoto, T. Kametami, *J. Am. Chem. Soc.* **107** (1985) 5810.
- 61) G. Augelmann, H. Fritz, J. Streith, *J. Chem. Soc. Chem. Comm.* (1982) 1112.
- 62) A. Defoin, G. Augelmann, H. Fritz, G. Geoffroy, J. Streith, *Helvet. Chim. Acta* **68** (1985) 1998.
- 63) M.E. Jung, *priv. Mitteil.* 1983.
- 64) S.L. Miller, *J. Am. Chem. Soc.* **77** (1955) 2351.
- 65) S.L. Miller, *Biochem. et Biophys. Acta* **23** (1957) 480.

- 66) S.L. Miller, *Ann. N. Y. Acad. Sci.* **69** (1957) 260.
- 67) S.L. Miller, H. Urey, *Science* **130** (1959) 245.
- 68) J. Oró, A.P. Kimball, R. Fritz, F. Master, *Arch. Biochem. Biophys.* **85** (1959) 115.
- 69) J. Oró, S.S. Kamat, *Nature* **190** (1961) 442.
- 70) J. Oró, A.P. Kimball, *Arch. Biochem. Biophys.* **94** (1961) 217.
- 71) R.E. Dickerson, *Spektrum der Wissenschaft* **9** (1979) 99.
- 72) A.B. Voet, A.W. Schwartz, *Bioorg. Chem.* **12** (1983) 8.
- 73) J.P. Ferris, P.C. Joshi, E.H. Edelson, J.G. Lawless, *J. Mol. Evol.* **11** (1978) 293.
- 74) J.P. Ferris, L.E. Orgel, *J. Am. Chem. Soc.* **88** (1966) 1074.
- 75) J.P. Ferris, J.E. Kuder, *J. Am. Chem. Soc.* **92** (1970) 2527.
- 76) Y. Yamada, I. Kumashiro, T. Takenishi, *J. Org. Chem.* **33** (1968) 643.