

Kombinatorik und chemische Metrik formaler Reaktions- und Strukturgenerierung

**Der Fakultät für
Chemie, Biologie und Geowissenschaften
der Technischen Universität München
als Habilitationsschrift vorgelegt von**

Dr. rer. nat. Eric Fontain

aus Hindelang (Oberallgäu)

Garching 1995

© Copyright 1995
Dr. Eric Fontain
Technische Universität München
Institut für Organische Chemie und
Biochemie
Lichtenbergstraße 4
D-85748 GARCHING
Tel.: (089) 3209 3378
Fax: (089) 3209 3315

Danksagung

Diese Arbeit wäre ohne die Unterstützung zahlreicher Personen nicht zustande gekommen. Aus diesem Grunde möchte ich hier allen, die in irgendeiner Weise zum Gelingen beigetragen haben, meinen Dank aussprechen.

Herrn Prof. Ivar Ugi danke ich herzlich für die gewährte Förderung, für die Bereitstellung eines sicheren Arbeitsplatzes und für die Schaffung eines kreativen Arbeitsklimas.

Ich danke ebenfalls den folgenden Damen und Herren, die meine Arbeit im Rahmen von Praktikumsarbeiten, Diplomarbeiten und Dissertationen mit ihrem großen Engagement unterstützt haben:

Frau Dipl.-Inf. (FH) Heidi Axtmann, Herrn Dipl.-Chem. Michael Freymann, Herrn Dipl.-Chem. Claus-Rüdiger Heikenwälder, Herrn Dipl.-Chem. Martin Heilingbrunner, Herrn Dipl.-Chem. Udo Ottmann, Herrn Dipl.-Chem. Klaus Reitsam, Herrn Dipl.-Chem. Peter Schenk, Herrn Dipl.-Chem. Wolfgang Setzer und Frau Dipl.-Inf. (FH) Mechthild Smolinski.

Herrn Dr. Johannes Bauer danke ich für das ausgezeichnete Büroklima und die erfolgreiche Zusammenarbeit.

Meiner Frau Dr. Marina Fontain, sowie meinen Kindern Fabian, Pascal und Oliver danke ich für die Geduld und die Bereitschaft, alle Höhen und Tiefen im Leben eines Hochschulwissenschaftlers mitzumachen.

Inhalt

1	Einführung	1
2	Reaktionsgenerierung.....	7
2.1	Reaktionsgeneratoren, ein Überblick	9
2.1.1	Empirische Reaktionsgeneratoren	12
2.1.2	Mechanistische Reaktionsgeneratoren	15
2.1.3	Nicht-empirische Reaktionsgeneratoren.....	17
2.1.3.1	Logisch-formale Reaktionsgeneratoren.....	17
2.1.3.2	Matrizenbasierende Reaktionsgeneratoren	22
2.2	Der Reaktionsgenerator von RAIN	30
2.2.1	Valenzschemata und Übergangsmatrizen	30
2.2.2	Bindungsschemata.....	32
2.2.3	BR-Transformatoren.....	34
2.2.4	Numerische Randbedingungen.....	36
2.2.5	Generierungsalgorithmus.....	38
2.2.6	Kanonische Darstellung.....	42
2.2.7	Substruktursuche	44
2.2.8	Der heuristische Reaktivitätsindex (HRI)	48
2.2.8.1	Grundlagen des HRI.....	49
2.2.8.2	Definition des HRI	50
2.2.8.3	Anwendungsbeispiel für den HRI	52
2.2.9	Der formale Übergangszustand	56
2.2.10	Resonanzstrukturen und Tautomerie.....	60
2.2.10.1	Problematik von Resonanzstrukturen und Tautomeren in Reaktionsnetzwerken	62
2.2.10.2	Die Generierung von Resonanzstrukturen	64
2.2.10.3	Die Generierung von Tautomerstrukturen.....	68
3	Strukturgenerierung	71
3.1	Strukturgenerierung, ein Überblick.....	73
3.2	Der Reaktionsgenerator als Strukturgenerator	76
3.2.1	Warburganal : ein illustratives Beispiel.....	78
3.2.2	Das B ₆ H ₁₄ -Problem	84

4	Chemische Metrik	89
4.1	Der BE-Raum.....	91
4.2	Methoden zur Bestimmung der chemischen Distanz.....	101
4.2.1	Deterministische Methoden	101
4.2.2	Heuristische Methoden.....	104
4.2.3	Stochastische Methoden	105
4.3	Kontrollierte Reaktionsgenerierung im BE-Raum.....	108
4.3.1	Methode der monoton fallenden Chemischen Distanz	110
4.3.1.1	Beispiel: Siebenkomponentenreaktion.....	112
4.3.2	Methode des ellipsoidalen Einschlusses	116
4.3.2.1	Beispiel: Oxy-Cope-Umlagerung.....	118
4.3.2.2	Beispiel: Abiogenese von Adenin.....	121
5	Genetische Algorithmen	127
5.1	Einführung	129
5.2	Optimierungsstrategien.....	133
5.3	Techniken der Genetischen Algorithmen.....	137
5.3.1	Parameterrepräsentation	138
5.3.2	Selektive Reproduktion	140
5.3.3	Crossing over.....	143
5.3.4	Mutation.....	146
5.3.5	Wissensbasierte Operatoren.....	148
5.3.6	Vorzeitige Konvergenz	153
5.3.7	Abbruchkriterien	155
5.4	Ergebnisse.....	156
6	Schlußbemerkung	159
	Literaturverzeichnis.....	163

Kapitel 1

Einführung

1 Einführung

Der Einsatz moderner Computertechnik hat das Arbeitsumfeld und die Arbeitsinhalte der grundlagenforschenden und der angewandten Chemie genauso grundlegend verändert, wie dies in nahezu jedem anderen Bereich unseres Lebens auch geschehen ist und in zunehmenden Maße weiter geschehen wird. Dies gilt umso mehr, als durch eine stetige Senkung der Marktpreise und durch fortschreitende Standardisierung immer leistungsfähigere Hardware, Software und Vernetzungstechniken auch für Nicht-Computerexperten vielfältig nutzbar werden. So bedeutete zum Beispiel vor 30 Jahren eine umfassende Literaturrecherche zu einem komplexeren Themengebiet noch mehrerer Wochen an mühsamer Sucharbeit in Bibliotheken, wobei die Vollständigkeit des Ergebnisses selten garantiert war. Heutzutage leistet dies der Verbund aus Datenbanken, moderner Netz- und Kommunikationstechnik im Verein mit benutzerfreundlichen Programmen nahezu auf Knopfdruck. Doch nicht nur auf dem Informationssektor hat sich das Bild der Chemie gewandelt. Keine der modernen chemischen Analysemethoden, von der Massenspektroskopie über alle chromatographischen Verfahren bis hin zur mehrdimensionalen Kernresonanzspektroskopie, kommt auch nur ansatzweise ohne den Einsatz von Computern aus. Vielfach sind hierbei die Softwaretechniken nicht mehr nur ein Hilfsmittel zur Automatisierung und Beschleunigung der Arbeitsprozesse, sie sind oft ein integraler und essentieller Bestandteil der Methoden geworden. Auch im Bereich der theoretischen Chemie werden heute die schnellsten Rechner unter Ausnutzung moderner Parallelarchitektur zur *ab initio* Berechnung von immer größeren Molekülen eingesetzt.

Neben dem eher traditionellen Einsatzgebiet der Computer in der Chemie begann sich bereits Ende der 60er Jahre ein neues Arbeitsfeld, die sogenannte "Computerchemie" zu etablieren. Die Computerchemie umfaßt den Einsatz von Softwaremethoden auf den Gebieten der Syntheseplanung, Reaktions- und Strukturgenerierung, Retrievalsystemen, formaler Stereochemie, molekularer Topologie usw. Die chemometrischen Methoden im Bereich der chemischen Analytik, sowie die Komplexe des Molecular Modeling und des Computational Chemistry haben große Gemeinsamkeiten mit der Computerchemie, verstehen sich heutzutage jedoch weitgehend als eigenständige Forschungsgebiete.

Das neue Arbeitsfeld der Computerchemie hatte seinen Anfang zu einer Zeit, in der ein großer Teil der Chemiker die Strukturaufklärung und die Planung aufwendiger Synthesestrategien von komplizierten Naturstoffen als intellektuelle Herausforde-

rung betrachtete. Aus diesem Grund ist es nicht verwunderlich, daß bereits zu Beginn der Computerchemie die Entwicklung der computerunterstützten Strukturaufklärung (DENDRAL-Projekt) und von Syntheseplanungsprogrammen im Vordergrund standen. Dem damaligen allgemeinen Verständnis nach sollte es möglich sein, diese zwar intellektuellen, aber doch formalisierbaren Tätigkeiten den "künstlichen Denkmaschinen" beizubringen. Diese recht naive Vorstellung von den Fähigkeiten der Computer beginnt erst heute allmählich einer mehr realistischen Einschätzung zu weichen. Ähnlich wie es bei vergleichbaren Bemühungen auf anderen Gebieten, zum Beispiel bei den Schachcomputern war (und auch noch heute ist), stand der traditionsgebundene Teil der synthetisch arbeitenden Chemiker den anfänglichen Bestrebungen mit äußerster Skepsis und großen Vorbehalten gegenüber. Dennoch war es gerade einer der "Großmeister" der Synthesechemie, E.J. COREY, unter dessen Leitung das erste Syntheseplanungsprogramm entstand.

In den Anfängen der Computerchemie mußten naturgemäß die grundlegenden Arbeiten an den Algorithmen für die Struktur- und Reaktionsbehandlung erfolgen. Hier zeigte es sich, daß die rekognitiven Fähigkeiten des menschlichen Intellekts zur Struktur- und Substrukturerkennung, einer neben einem umfangreichen Detailwissen unabdingbaren Voraussetzung für die Planung chemischer Synthesen, nur mit sehr aufwendigen Verfahren und mit einem hohen Anteil an graphentheoretischen Algorithmen nachvollziehbar waren.

Schon bald begannen sich auch verschiedene Hauptarbeitsrichtungen der Computerchemie abzuzeichnen. In einem relativ naheliegenden Ansatz wurde zunächst versucht, die ungeheure Menge an Detailinformation über chemische Strukturen, und vor allem über chemische Reaktionen, formal zu erfassen und zur Vorhersage möglicher Reaktionspfade und Synthesewege einzusetzen. Der Vorteil solcher Verfahren liegt darin, daß zumindest die chemischen Einzelschritte, die von den Programmen vorgeschlagen werden, nahezu alle plausibel und über ihre Präzedenz aus der Literatur heraus nachvollziehbar sind. Die Inhomogenität und die große Diversifizierung des Expertenwissens setzt diesem empirischen Ansatz jedoch auch seine Grenzen.

Parallel hierzu bildete sich seit Anfang der 70er Jahre, hauptsächlich durch Arbeiten von I. UGI, J. DUGUNDJI und J.B. HENDRICKSON eine alternative, zunächst nichtempirische Vorgehensweise heraus. Diese basiert auf einer vollkommenen Formalisierung der Chemie mittels eines mathematischen Modells für Strukturen und Reaktionen und erzeugt Reaktionsschritte, zum Beispiel für die Syntheseplanung, durch kombinatorisch vollständige Anwendung von einigen wenigen mathematischen Grund-

operationen. Der Anspruch hierbei war, von einer meist voreingenommenen Wissensbasis unabhängig zu sein und somit auch präzedenzlose Vorschläge für die Lösung eines Problems präsentieren zu können. Obwohl zunächst davon ausgegangen wurde, daß ein erfahrener Chemiker mittels geeigneter Selektionsmethoden innerhalb der großen Anzahl von Reaktionsverzweigungen, die solche Programme generieren, "die Spreu vom Weizen trennen" kann, zeigte es sich doch, daß ohne eine Vorbewertung und eine Vorauswahl aufgrund empirischer und heuristischer Regeln diese kombinatorischen Programme keinen praktischen Nutzen haben. So lag das Schwergewicht der Arbeit in den letzten Jahren auf der Entwicklung von transparenten und wirkungsvollen Filterverfahren, die in Zusammenarbeit mit hierarchischen Reaktionsmodellen helfen, kombinatorische Reaktionsvorhersage- und Syntheseplanungsprogramme einsetzbar zu machen.

Die folgenden Kapitel geben einen Überblick über die Arbeiten des Verfassers auf dem Gebiet der formalen Computerchemie im Bereich der Reaktionsgenerierung, Strukturgenerierung, der chemischen Metrik und dem Einsatz der Genetischen Algorithmen. Hauptziel der Bemühungen war es, Methoden zur Verfügung zu stellen, die einen *zielorientierten* Aufbau von Reaktionspfaden ermöglichen. Die Berechnungen der hierzu erforderlichen Metrik im Raum der chemischen Konstitution sind durch die Anwendung von evolutionären Softwaretechniken erstmals so effizient geworden, daß sie in Reaktionsgenerierungsprogramme integriert werden können.

Alle vorgestellten Verfahren, Regeln, Konzepte und Algorithmen sind im Computerprogramm RAIN (Reactions And Intermediates Networks) implementiert, mit dessen Hilfe auch die Berechnungen der zur Illustration dienenden Anwendungsbeispiele durchgeführt wurden.

Kapitel 2

Reaktionsgenerierung

Zusammenfassung:

Nach einer Darstellung verschiedener Aspekte der automatischen Erzeugung chemischer Reaktionen mit einem Überblick über existierende Systeme zur Reaktionsgenerierung wird ein auf den sogenannten BR-Transformatoren beruhender formaler Reaktionsgenerator vorgestellt. Verschiedene Methoden wie Substruktursuche, heuristischer Reaktivitätsindex (HRI) und eine Analyse des formalen Übergangszustandes dienen zur Einschränkung der potentiellen Vielfalt erzeugter Reaktionen bzw. Reaktionsschritte. Mit Hilfe besonderer Einstellungen der Kontrollparameter kann der Reaktionsgenerator auch zur ausschließlichen Erzeugung von Mesomerstrukturen bzw. Tautomerstrukturen herangezogen werden.

2.1 Reaktionsgeneratoren, ein Überblick

Die Chemie befaßt sich als Wissenschaft mit den stofflichen Eigenschaften der Materie. Unter anderem sind es die möglichen *Umwandlungen* der chemischer Strukturen, die das Forschungs- und Anwendungsgebiet der Chemie darstellen. Aus diesem Grund wurden bereits in der Zeit der Anfänge des Computereinsatzes in der Chemie, neben Verfahren zur Beschreibung von *Strukturen* auch Algorithmen zur Erzeugung und Beschreibung von chemischen *Reaktionen* entwickelt. Gerade diese reaktionsgenerierenden Verfahren sind es nun, die durch ihre kreativen Möglichkeiten der Strukturumwandlung den Problemhorizont erweitern und zu phantasievollen Lösungsvorschlägen anregen. Dies gilt nun sowohl für die Anwender solcher Programme, die hierbei lernen, ihr Problem aus den verschiedensten Richtungen zu betrachten, als auch für die Systementwickler, die gezwungen werden, unscharf definierte Begriffe, wie zum Beispiel den der chemischen Reaktivität zu formalisieren, und somit viel zu einem tieferen Verständnis mancher Sachverhalte beitragen.

Ein Verfahren, das zu einem gegebenen Molekül oder Ensemble von Molekülen diejenigen Strukturen erzeugt, die in einer Reaktionssequenz unmittelbare *Vorläufer*, oder auch *Nachfolger* dieses gegebenen Moleküls sein können, nennt man *Reaktionsgenerator*.

Es existieren viele verschiedene Ansätze und Modelle für Reaktionsgeneratoren, die je nach der Zielsetzung ihrer Anwendung spezifische Methoden einsetzen, um die Menge der erzeugten Transformationen auf eine realistische, dem Kenntnisstand entsprechende, aber auch innovative Möglichkeiten beinhaltende Menge einzugrenzen. Reaktionsgeneratoren setzen hierzu entweder eine möglichst *umfassende* Menge empirischer Regeln ein, die dem chemischen Erfahrungsschatz entnommen sind, oder sie versuchen, mit wenigen möglichst *allgemein gültigen* Generierungsregeln zu arbeiten, die unter formalen Einschränkungen kombinatorisch vollständige Lösungsmengen erzeugen.

Die Generierungskapazitäten der Reaktionsgeneratoren reichen von mechanistischen Elementarschritten bis hin zu Folgen von Synthesereaktionen. Manche Verfahren eignen sich für beide Reaktionsrichtungen (siehe Abbildung 2.1), während andere (z.B. die meisten in Syntheseplanungsprogrammen integrierten Generatoren) nur retrosynthetisch arbeiten können. Einige wenige spezialisierte Systeme erzeugen in einer Art Reaktionsvorhersage die möglichen Folgeprodukte eines Ensembles von Molekülen (Prosyntheserichtung).

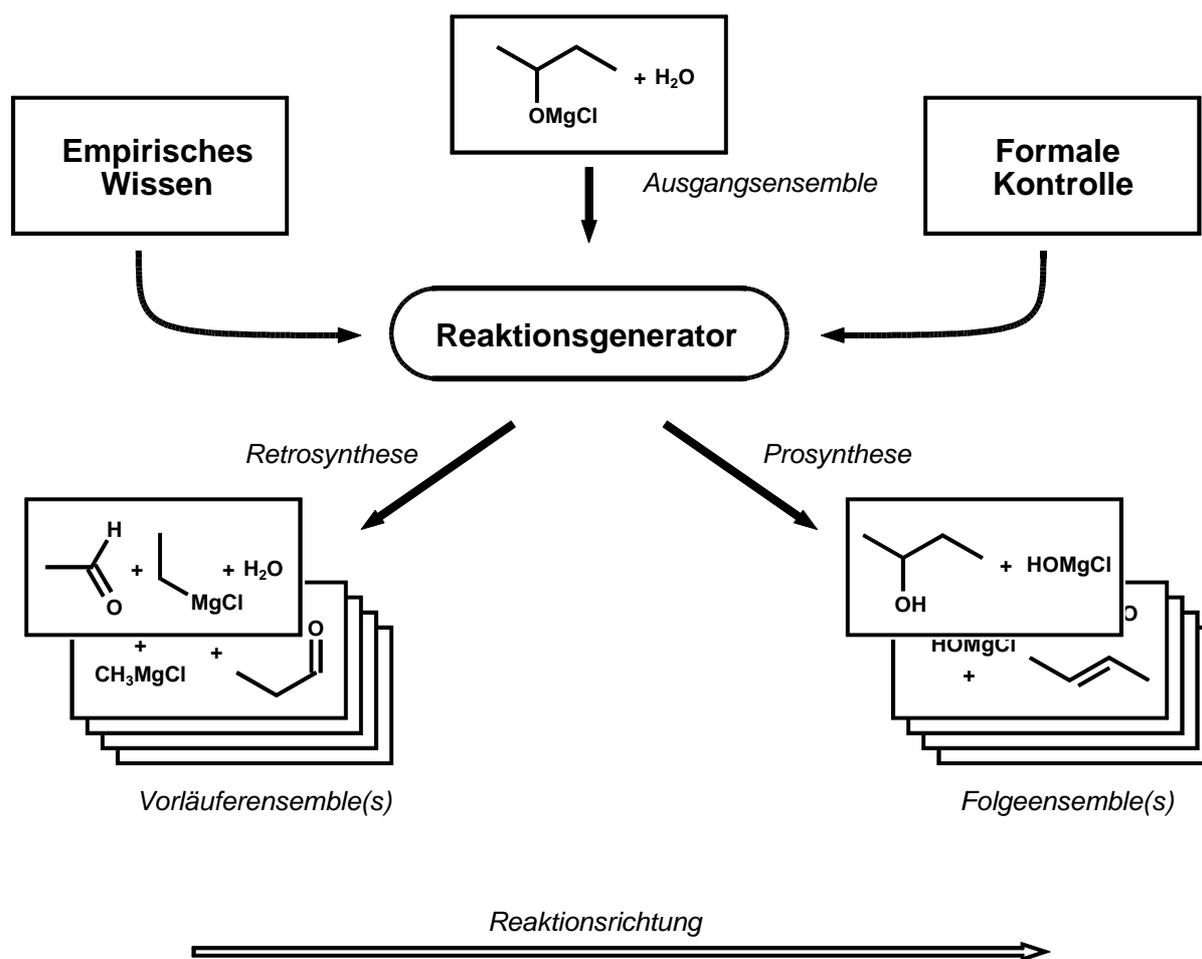


Abb. 2.1 Reaktionsgeneratoren erzeugen zu gegebenen (Substrat)-Molekülen die unter formaler Kontrolle und unter Berücksichtigung von empirischem Wissen möglichen Folgeprodukte bzw. Vorstufen.

Sogenannte materialerhaltende Reaktionsgeneratoren, hierzu zählen die meisten matrizenbasierenden Systeme, erzeugen Reaktionen lediglich durch Umverteilung der vorhandenen Valenzelektronen zwischen den Atomen der gegebenen Moleküle. Im Zuge der Reaktionserzeugung werden hierbei keine Atome oder Moleküle hinzugefügt bzw. entfernt. Da dieses Vorgehen in abgeschlossenen Ensembles einer eigentlichen Synthesepaltung nicht gerecht werden kann, operieren die hauptsächlich retrosynthetischen Reaktionsgeneratoren der Synthesepaltungsprogramme in offenen Systemen, die eine nahezu beliebige Erhöhung oder Verminderung der Atomzahlen im Ensemble zulassen. Für die Vorhersage biochemischer Reaktionspfade ist wegen der Vielzahl relevanter Auf- und Abbaureaktionen ein materialerhaltendes System, das alle zu betrachtenden Atome und Elektronen bereits im Eingabeensemble enthalten muß, ebenfalls nicht adäquat einsetzbar.

Die Klassifizierung reaktionsgenerierender Systeme erfolgt meist nach der Art des generierenden Prinzips. Transform-orientierte Reaktionsgeneratoren arbeiten auf der Grundlage einer Wissensbasis in der Form einer verallgemeinerten Reaktionsbibliothek (Transformbibliothek). Diese stellen eine Sammlung des chemischen Erkenntnisstandes dar und begrenzen damit gleichzeitig den Ereignishorizont der zugehörigen Programme. Es können hiermit jedoch immer nur Reaktionsschritte erzeugt werden, deren zugehörige Transformbeschreibung in der Bibliothek enthalten ist und die nach Maßgabe der heuristischen Zusatzregeln auf die gegebene Struktur anwendbar sind. Trotzdem kann man diesen Verfahren eine mehr oder weniger große Kreativität nicht absprechen, da die verallgemeinerte Darstellung der Reaktionen in der Transformbibliothek auch die Erzeugung präzedenzloser *Anwendungen* dieser Umwandlungen zuläßt.

Weitaus weniger voreingenommen, und damit sehr viel universeller, sowohl im Einsatz, als auch in der Generierungskapazität, sind jedoch Reaktionsgeneratoren, die auf der Grundlage von allgemein gültigen Regeln arbeiten. Trotz des auch hier nötigen Einsatzes heuristischer Auswahlregeln zur Vermeidung von unermeßlichen Lösungsmengen, sind diese Systeme *a priori* kreativer als transformbasierende Systeme. Das Problem hierbei ist, daß es nicht leicht ist, allgemein gültige Regeln, die zudem noch einfach und transparent anwendbar sein sollen, aufzustellen und zu implementieren. Erfolgreich eingesetzt werden können diese Verfahren nur, wenn mächtige und vor allem transparente Möglichkeiten existieren, die erzeugte Lösungsmengen formal, oder heuristisch zu klassifizieren und einzuschränken.

2.1.1 Empirische Reaktionsgeneratoren

OCSS

Der Vorläufer aller prominenten Syntheseplanungsprogramme und das erste wirklich funktionsfähige reaktionsgenerierende System war das Programm OCSS (Organic Chemical Simulation of Synthesis), das in den späten 60er Jahren von E.J. COREY und W.T. WIPKE in Harvard entwickelt wurde ¹. Neben einem, auf COREYS richtungsweisenden Konzept der retrosynthetischen Analyse ^{2, 3, 4} beruhenden Verfahren zur Erkennung synthetisch relevanter, sogenannter strategischer Bindungen, enthielt OCSS einen in mehreren Programmmodulen fest codierten Reaktionsgenerator, der Strukturmanipulationen in Form von "symbolic mechanisms" und "symbolic functional group modifications" ermöglichte. Ein einfaches Evaluierungsmodul verhinderte hierbei unwahrscheinliche formale Ladungskombinationen und geometrisch inplausible Topologien.

Die Unbeweglichkeit und schlechte Erweiterbarkeit dieser "festverdrahteten" Reaktionsregeln führte Anfang der 70er Jahre zum Aufbau von externen Wissensbasen, den Transformbibliotheken. Transforms sind verallgemeinerte, retrosynthetische Reaktionsschritte, die in einer, der natürlichen Sprache der Chemiker angepaßten Form in einer Wissensbasis abgelegt sind. Sie enthalten die strukturverändernden Manipulationen als Vorschriften in Form von lesbarem Text, wie z.B.

"BREAK BOND2*1"

oder

"JOIN ATOM*2 AND ATOM*3".

Der weitaus wichtigere Teil eines Transforms besteht jedoch aus einer Sammlung von Deskriptoren und Anweisungen, die die Bedingungen für die Anwendbarkeit des Transforms definieren. Über die Angabe von strukturellen Minimalanforderungen, von Substrukturen, die als Promotoren oder Inhibitoren wirken, wie z.B.

"GROUP*1 MUST BE KETONE OR CYANO OR ESTER"

oder

"KILL* IF MULTIPLE BOND ON ATOM*2"

und über Auswahlregeln sowie spezielle Vorbedingungen findet möglichst das gesamte empirische Wissen über eine synthetische Reaktion Eingang in die Transform-

bibliothek. Ein numerisches Wertungssystem mit Inkrementen und Dekrementen für die aktuell vorgefundenen Strukturparameter, wie z.B.

"ADD 15 IF ANOTHER WITHDRAWING BOND ON ATOM*3"

oder

"SUBTRACT 10 IF ATOM*3 IS A TERTIARY*CENTER"

soll die Anwendung eines Transforms auf die chemisch sinnvollen Fälle einschränken.

Um die Transformbibliothek aufbauen zu können, wurden spezielle Sprachen, wie CHMTRN⁵ und ALCHEM⁶ entwickelt, die wie Programmiersprachen von einem Compiler in programminterne Codetabellen umgesetzt werden. Bei CHMTRN ist die Möglichkeit gegeben, Funktionen und Routinen zu definieren, die dann von vielen Transforms als gemeinsame Bausteine genutzt werden können.

LHASA, SECS, PASCOP, CASP

Die Programmsysteme mit transformbasierenden Reaktionsgeneratoren unterscheiden sich hauptsächlich in ihren synthetisch-strategischen Fähigkeiten, ein Aspekt, der an dieser Stelle nicht relevant ist. Prominente Vertreter dieser Systeme sind COREYS LHASA^{5, 7, 8, 9} (Logic And Heuristics Applied to Synthetic Analysis) und WIPKES SECS⁶ (Simulation and Evaluation of Chemical Synthesis), beide unmittelbare Nachfolgerprogramme von OCSS. Das System SECS war weiterhin Grundlage für das auf die Bedürfnisse der phosphor-organischen Chemie zugeschnittene Programm PASCOP (Programme d'Aide à la Synthèse des Composés Organo Phosphores)^{10, 11, 12} und für das Gemeinschaftsprojekt CASP¹³ (Computer Aided Synthesis Planning) der größten Deutschen und Schweizer Pharmazieunternehmen.

SYNCHEM

Das Syntheseplanungsprogramm SYNCHEM^{14, 15} (SYNthetic CHEMistry) von H.L. GELERNTER enthält spezielle, nach Reaktionstypen gegliederte Reaktionsbibliotheken, die infolge einer Suche nach sogenannten "synthemes", d.h. synthesesrelevanten funktionellen Gruppen und Struktureigenschaften, geladen werden. In einer neueren Version SYNCHEM2^{16, 17} können die Transforms erstmals nicht nur einfach, sondern gleichzeitig mehrfach, und zudem auch in Vorwärtsrichtung, d.h. zur Reaktionsvorhersage eingesetzt werden.

Zu einem zentralen Problem der transformorientierten Reaktionsgeneratoren wurde im Laufe der Zeit die Erstellung, Erweiterung und Pflege der Transformbibliotheken, da diese möglichst das gesamte empirische Wissen über synthetisch relevante Reaktionen enthalten sollte. Die Transformbibliothek von LHASA umfaßt heute ca. 2.000 Transforms. Die Einträge müssen mit großer Sorgfalt gemacht werden, da Fehler unmittelbar die erreichbare Lösungsmenge beeinflussen, die Effekte und die Ursachen solcher Unzulänglichkeiten aber nur schwer aus den Ergebnissen herausgelesen werden können. Aus diesem Grunde wurden in den letzten Jahren zunehmend automatisierte Verfahren¹⁸ geschaffen, die in einer standardisierten, und vor allem nachvollziehbaren Weise, Transforms aus Sammlungen von Reaktionen, die in vollständiger Form notiert sind, generieren.

Die große Vielfalt bekannter Reaktionen stellt hierbei ein weiteres Problem dar. Für das Syntheseprogramm SYNCHM2 wurde hierzu ein ganzer Komplex von Programmen zum automatisierten Wissenserwerb zusammengestellt¹⁹. Nach einer Vorklassifizierung der Ausgangsreaktionen durch das Programm BRANGÄNE wird vom System ISOLDE versucht, induktiv auf verallgemeinerte Reaktionstypen zu schließen. Der Reaktionsbibliotheksaufbau wird weiterhin unterstützt durch das Programm TRISTAN, das in deduktiver Weise Reaktionen ableitet. TRISTAN basiert auf Regeln für die Durchführbarkeit von Reaktionen, die denen des Reaktionssimulationsprogramms CAMEO²⁰ entsprechen.

RDSS

Einen weiteren Ansatz zur automatisierten Exzerption von sogenannten Synthongleichungen aus umfangreichen Reaktionsdatenbanken verfolgt das System RDSS^{21, 22} (Reaction Design by Synthon Substitution). Der Aufbau seines ReaktionsSpeichers SPRESI²³ (SPeicherung und REcherche Strukturchemischer Informationen) geschieht hierbei durch automatische Extraktion von Synthons (hier: an Reaktionszentren beteiligte Substrukturen) und Kontextdeskriptoren aus der Reaktionsdatenbank des Allunionsinstituts für wissenschaftliche und technische Information der ehemaligen UdSSR (VINITI).

2.1.2 Mechanistische Reaktionsgeneratoren

Die Aufgabe, diejenigen Reaktionssequenzen zu generieren, die aus einem gegebenen Molekül oder Ensemble von Molekülen in vorwärtsgerichteter (synthetischer) Richtung hervorgehen können, erfordert eine eingehende Analyse der gegebenen Struktur. Dieser Evaluierungsphase folgt dann die Erzeugung potentieller Strukturveränderungen in der Form mechanistischer Einzelschritte. Zuverlässige Programme dieser Art wären insbesondere wertvoll, um zum Beispiel mögliche Reaktions- und Abbauprodukte von in die Umwelt gelangenden Chemikalien vorherzusagen.

AHMOS

Das Reaktionssimulationssystem AHMOS²⁴ (Automatisierte Heuristische Modellierung Organisch-chemischer Synthesen), von A. WEISE bereits Mitte der 70er Jahre vorgestellt, beschränkt sich in seiner Generierungskapazität auf die Elementarschritte der heterolytischen (polaren) Einzentrenreaktionen. Ein Elementarschritt involviert maximal zwei Reaktionszentren und maximal drei sich verändernde Bindungen. Der Reaktionsgenerator von AHMOS enthält Routinen für die Elementarschritte Addition, Substitution, Dissoziation, Protonenaustausch, sowie anionotrope 1,2-Umlagerung, Sextettpolarisierung, aromatische elektrophile Substitution und Eliminierung. Die Anwendbarkeit der Elementarschritte wird gesteuert durch Reaktivitätskennwerte funktioneller Gruppen unter Einbeziehung des HSAB-Konzepts²⁵ von R.G. PEARSON. AHMOS arbeitet mit abgeschlossenen Ensembles, d.h. es werden nur die Atome betrachtet, die in der Eingabestruktur vorhanden sind. Ein Hinzufügen von weiteren Reaktionspartnern geschieht nicht.

CAMEO

Einen sehr direkten, wenn auch wenig transparenten Ansatz zur mechanistischen Evaluierung organischer Reaktionen verfolgt W.L. JORGENSEN mit seinem Programm CAMEO^{20, 26} (Computer-Assisted Mechanistic Evaluation of Organic reactions). In einer Serie mit über 20 Einzelpublikationen, die seit Anfang der 80er Jahre erschienen, werden einzelne Programmmodule beschrieben, die jeweils einen ganz spezifischen Aspekt der Reaktivität organischer Verbindungen behandeln. JORGENSEN geht in den Generierungskapazitäten damit weit über das AHMOS-System von WEISE hinaus. So wurden z.B. Routinen zur Behandlung von electrocyclischen, bzw. pericyclischen Reaktionen^{27, 28}, Oxidationsreaktionen²⁹, organometallischen Reak-

tionen³⁰ etc. geschaffen. Das empirische Wissen findet sich hier also weder in einigen wenigen, möglichst allgemein gültigen Regeln, noch in individuell zugeschnittenen und programmunabhängigen Transforms, sondern ist zum Teil tabelliert, zum Teil in algorithmischer Form in einzelnen Programmteilen integriert, die nach einer Perzeptionsphase selektiv angesprochen werden. Für manche Entscheidungen des Programms werden auch direkt quantenmechanische Berechnungen herangezogen. Der Vorteil eines solchen Konzeptes ist, daß hier auch schwer zu verallgemeinernde Sonderfälle durch eine einfache Erweiterung des Programmcodes berücksichtigbar sind. CAMEO ist auch das einzige Programm, das Reaktionsbedingungen, wie Temperatur, Lösungsmittel oder pH-Wert in seine Betrachtungen mit einbeziehen kann. Ein Nachteil dieses Konzepts ist es jedoch zweifellos, daß durch die Vielzahl von berücksichtigten Spezialfällen keine klare Linie mehr erkennbar ist, und die zugehörigen komplizierten Programme wenig Transparenz aufweisen.

2.1.3 Nicht-empirische Reaktionsgeneratoren

2.1.3.1 Logisch-formale Reaktionsgeneratoren

DENDRAL-REACT

Eine Zwischenstellung zwischen den empirischen Reaktionsgeneratoren auf Transformbasis (siehe oben) und den rein formalen Reaktionsgeneratoren nimmt der im Rahmen des DENDRAL-REACT-Projekts^{31, 32, 33, 34} entwickelte Reaktionsgenerator ein. DENDRAL-REACT dient hauptsächlich zur mechanistischen Strukturaufklärung. Hiermit sollen Strukturkandidaten, die von dem Programm CONGEN (ebenfalls Mitglied der DENDRAL Familie) für eine unbekannte Verbindung vorgeschlagen werden, gezielten Abbau-, Oxidations- oder anderen Reaktionen unterworfen werden, um über die entstehenden Produkte im Vergleich zu unter Laborbedingungen erhaltenen Ergebnissen die Vorschlagslisten weiter einzuschränken. Hierzu wird eine einzelne Reaktion aus einer Bibliothek ausgewählt oder vom Benutzer ad hoc definiert und auf die Liste von Strukturkandidaten in synthetischer (Vorwärts)-Richtung angewandt. Die Definition der Reaktion erfolgt in einer Transform-ähnlichen Notation, unter zusätzlicher Beschreibung von strukturellen Mindestvoraussetzungen. Überdies kann der gesamte Einschränkungsmechanismus von CONGEN hinsichtlich verbotener oder geforderter Substrukturen herangezogen werden. Die Erwähnung von DENDRAL-REACT hier an dieser Stelle und nicht bei den empirischen Transform-Reaktionsgeneratoren erfolgt deshalb, weil bei diesem Programm keine Möglichkeit gegeben ist, detailliertes empirisches Wissen mit Inkrementen und Dekrementen in einem Bewertungssystem von Applikationsmöglichkeiten eines Transforms zu spezifizieren.

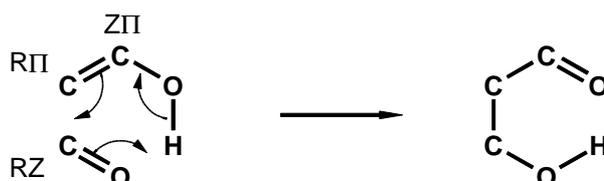
SYNGEN

Die Arbeiten von J.B. HENDRICKSON stellen einen der wichtigsten Beiträge bei den Bemühungen dar, die Chemie des Kohlenstoffs, dem zentralen Punkt des Interesses in der organischen Chemie, weitestgehend zu formalisieren. Durch eine Generalisierung der an C-Atomen gebundenen Nachbarn werden die Variationsmöglichkeiten auf die abstrahierten Reaktionskategorien am C-Atom vereinfacht. Die gebun-

denen Heteroatome werden nach ihrer Elektronegativität in Bezug auf Kohlenstoff als elektropositiv oder elektronegativ klassifiziert. Eine Unterscheidung gebundener Atome erfolgt in σ -Bindungen (betrachtet als Skelettbindungen) und π -Bindungen (betrachtet als funktionelle Gruppen). Die Funktionalität eines C-Atoms läßt sich demgemäß durch ein Quadrupel von Zahlen beschreiben:

Typ	Anz.	Bedeutung
H	h	Bindungen zu H, oder elektropos. Atomen (B, Al, Si, Met., etc.)
R	σ	σ -Bindungen zu Kohlenstoff (Skelett)
Π	π	π -Bindungen zu Kohlenstoff (funktionell)
Z	z	Bindungen zu elektroneg. Heteroatomen (N, O, S, Hal., etc.)
	$\Sigma=4$	

Eine Reaktion wird nunmehr dargestellt durch eine Folge von 2-Buchstabencodes, wobei jeweils der erste Buchstabe für die Art der geknüpften Bindung und der zweite Buchstabe für die Art der gebrochenen Bindung steht. Die Refunktionalisierung der C-Atome einer Aldolreaktion läßt sich demnach linear notieren als: RZ·R Π ·Z Π .



Betrachtet man die Funktionalität der α -, β - und γ -Atome ($z+\pi$), so gelangt man zu insgesamt 25 sogenannten "half reactions" (16 nucleophil und 9 elektrophil), die die Änderungen an jeweils einer Seite einer reagierenden Bindung protokollieren. Aus den möglichen Kombinationen der nucleophilen mit den elektrophilen half reactions ergeben sich 114 verschiedene "construction reactions", die zusammen mit den erforderlichen $z\pi$ -Listen der α -, β - und γ -Atome die Basis des Reaktionsgenerators von SYNGEN^{35, 36} bilden. Findet der Generator im Substratmolekül ein passendes $z\pi$ -Muster, so kann die entsprechende construction reaction in retrosynthetischer Richtung angewandt werden. Weitere mechanistische Tests betrachten die nötige Akti-

vierung der reagierenden Atome und eventuell störende bzw. inkompatible funktionelle Gruppen im nicht reagierenden Teil des Moleküls. Für diese mechanistische Evaluierung werden die z-Listen weiter diversifiziert in E (Acceptor), O (Donator), L (Abgangsgruppe) und W (Acceptor mit Mehrfachbindung). Für alle construction reactions existieren Anforderungs- und Kompatibilitätslisten der entsprechenden E-, O-, L- und W-Substituenten³⁶. Als Syntheseprogramm zeichnet sich SYNGEN hauptsächlich durch seine Eigenschaft aus, strategische Bindungen zu finden, die in konvergenter Retrosynthese mit wenigen Reaktionsschritten Targetmoleküle in einfachere Bruchstücke ähnlicher Größe zerlegen³⁷.

HENDRICKSONS Konzept der "half reactions" bildete bereits einige Jahre vor der Veröffentlichung von SYNGEN die Grundlage des Syntheseprogramms MASSO³⁸.

TOSCA

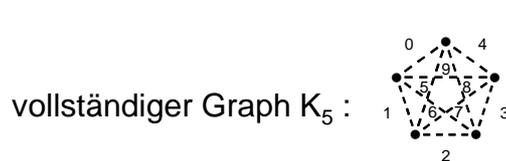
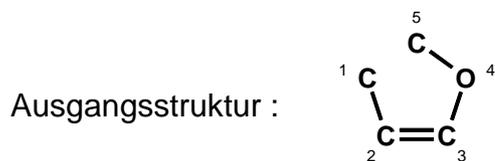
Das Syntheseprogramm TOSCA³⁹ (TOPological Synthesis design by Computer Application), von J. SANDER bei der Hoechst AG in Frankfurt entwickelt, benutzt ebenfalls eine deduktive Methode zum Design von Reaktionen, sowohl in pro-synthetischer, als auch in retrosynthetischer Richtung. Der Reaktionsgenerator von TOSCA läßt bis zu 11 brechbare und bis zu 11 knüpfbare Bindungen zu. Die Anzahl der Bindungen an jedem Atom des Reaktionszentrums bleibt im Verlauf der Reaktion konstant, d.h., es werden keine formal geladenen Atom und Radikale erzeugt, noch werden freie Elektronenpaare erzeugt oder verschoben (keine echten Redoxreaktionen). Die Reaktionserzeugung erfolgt unter Kontrolle heuristischer Prinzipien nach dem Konsonanz/Dissonanz-Konzept von D.A. EVANS^{40, 41}. Eine weitere Ausfilterung unwahrscheinlicher Reaktionen erfolgt auf der Basis von instabilen Strukturelementen, die in einer entsprechenden Liste dem Programm zur Verfügung stehen. TOSCA arbeitet mit abgeschlossenen Ensembles von Molekülen, Reaktionen werden also lediglich als Valenzisomerisierungen betrachtet, wie dies auch bei der Mehrzahl der matrizenbasierenden Systeme (siehe unten) der Fall ist. Leider wurden die Arbeiten an TOSCA (von 1976 bis 1980) nach der Veröffentlichung im Jahre 1985 (!) nicht mehr weiter verfolgt.

RENE

Einen sehr einfachen mathematischen Ansatz zur Reaktionsgenerierung verfolgt R. BARONE mit dem Programm RENE⁴² (REaction geNERator). Er basiert auf der Erzeugung von sogenannten Homotopomeren. Ein Homotopomer ist ein Konstitutionsisomer zu einer gegebenen Struktur, welches, im Gegensatz zur geläufigen Verwendung des Begriffs "isomer", aus mehreren unverbundenen Teilen bestehen kann. Sowohl die Anzahl, als auch die Art der Bindungen können variieren, lediglich die Anzahl der Atome bleibt konstant. I. UGI und J. DUGUNDJI definierten schon sehr viel früher (bereits Anfang der 70er Jahre) eine nahezu identische Form der erweiterten Konstitutionsisomerie. Sie benutzten den Begriff "Ensemble von Molekülen" (EM) und definierten als "Familie der Isomeren Ensembles von Molekülen" (FIEM) alle valenzchemisch möglichen Verteilungen einer gegebenen Anzahl von Valenzelektronen *auf* (als freie Elektronen) und *zwischen* (als Bindungen) einer gegebenen Menge von Atomen. Der Reaktionsgenerator von RENE betrachtet die n Atome einer Ausgangsstruktur als Knoten eines vollständigen Graphen K_n . Die $n(n-1)/2$ möglichen Kanten dieses Graphen bilden nun ein Gerüst für eine einfache Zählmaschine, die die Zahlen der Basis 4 mit n Stellen erschöpfend erzeugt. Die Basis 4 ergibt sich hierbei aus den möglichen Bindungsordnungen 0 bis 3. Die entsprechenden Ziffern an den Stellen 1 bis n repräsentieren die Bindungsordnungen zwischen den Atomen in den generierten Strukturen (siehe Abbildung 2.2).

Das in Abbildung 2.2 angeführte Beispiel zeigt auch die schwerwiegenden Nachteile dieses Verfahrens. Die Anzahl möglicher Bindungskonstellationen steigt mit $4^{(n(n-1)/2)}$ und begrenzt damit den Generator auf maximal 4 oder 5 Atome im Reaktionszentrum. Gleichzeitig werden enorme Mengen valenzchemisch unrealistischer Strukturen erzeugt, die zum Teil durch eine Beachtung maximaler Bindungsordnungen während der Generierung, sowie zum Teil mit einer nachgeschalteten Valenzüberprüfung eliminiert werden müssen. Die zwangsläufig mitgenerierten Redundanzen werden durch ein Kanonisierungsverfahren nach R. MOREAU⁴³ erkannt und bei der Generierung ausgesondert. Das Fehlen von geeigneten Einschränkungsmöglichkeiten zur Begrenzung des Generators, das algorithmisch naive Zählverfahren, sowie die hieraus resultierende Begrenzung auf eine nur kleine Anzahl von reaktiven Zentren läßt die Anwendbarkeit von RENE auf wenige Ausnahmefälle begrenzt erscheinen. Der im Rahmen dieser Arbeit beschriebene Reaktionsgenerator von RAIN hat einige Wesenszüge mit RENE gemeinsam, nutzt jedoch bereits während der Erzeugung von Elektronenumverteilungen alle valenzchemischen Randbedingungen zum frühestmöglichen Zeitpunkt, was im Normalfall bis zu 32 reaktive

Atome zuläßt. Gleichzeitig ist das RAIN-Verfahren durch eine große Auswahl von Grenzwerten steuerbar und kann somit in seinen Generierungskapazitäten an das zu bearbeitende Problem angepaßt werden.



10-stellige Zahl zur Basis 4 :

generierte Struktur (Homotopomer) :

9	8	7	6	5	4	3	2	1	0
0	0	0	0	0	0	0	0	0	0
9	8	7	6	5	4	3	2	1	0
0	0	0	0	0	0	0	0	0	1
...									
9	8	7	6	5	4	3	2	1	0
0	0	0	0	0	0	2	1	1	2
9	8	7	6	5	4	3	2	1	0
0	0	0	0	0	0	2	1	1	3
...									
9	8	7	6	5	4	3	2	1	0
0	0	1	0	2	2	1	1	1	2
...									
9	8	7	6	5	4	3	2	1	0
3	3	3	3	3	3	3	3	3	3

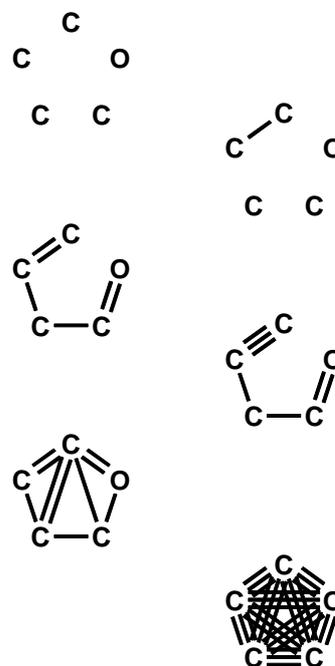
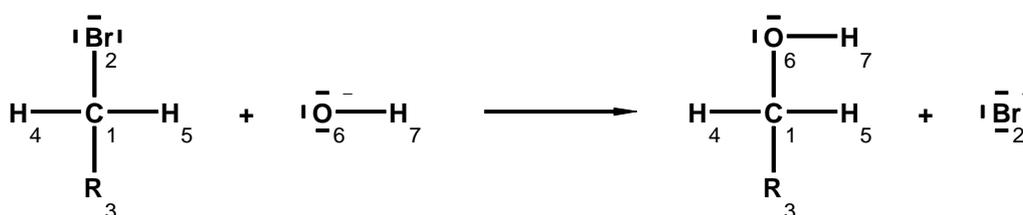


Abb. 2.2 Das Reaktionsgenerierungsprinzip von RENE auf der Basis einer einfachen Zählmaschine für Zahlen der Basis 4.

2.1.3.2 Matrizenbasierende Reaktionsgeneratoren

Den wohl höchsten Grad an formaler Abstraktion erfuhr die organische Chemie durch die Einführung eines streng mathematischen Konzepts der konstitutionellen Chemie^{44, 45, 46} durch I. UGI und J. DUGUNDJI zu Anfang der 70er Jahre. Die Betrachtung einer chemischen Reaktion als additive Wirkung einer die Elektronenverschiebungen repräsentierende R-Matrix auf eine das Ausgangsensemble beschreibende BE-Matrix (siehe Abbildung 2.3) ebnete den Weg für eine ganze Reihe von formalen Reaktionsgeneratoren, zum Teil mit vollkommen neuen Eigenschaften.

Reaktionsgleichung:



BE-Matrizengleichung:

		1	2	3	4	5	6	7			1	2	3	4	5	6	7			1	2	3	4	5	6	7		
		C	Br	R	H	H	O	H	+		C	Br	R	H	H	O	H	=		C	Br	R	H	H	O	H		
1	C	0	1	1	1	1	0	0		1	C	0	-1	0	0	0	+1	0		1	C	0	0	1	1	1	1	0
2	Br	1	6	0	0	0	0	0		2	Br	-1	+2	0	0	0	0	0		2	Br	0	8	0	0	0	0	0
3	R	1	0	0	0	0	0	0		3	R	0	0	0	0	0	0	0		3	R	1	0	0	0	0	0	0
4	H	1	0	0	0	0	0	0		4	H	0	0	0	0	0	0	0		4	H	1	0	0	0	0	0	0
5	H	1	0	0	0	0	0	0		5	H	0	0	0	0	0	0	0		5	H	1	0	0	0	0	0	0
6	O	0	0	0	0	0	6	1		6	O	+1	0	0	0	0	-2	0		6	O	1	0	0	0	0	4	1
7	H	0	0	0	0	0	1	0		7	H	0	0	0	0	0	0	0		7	H	0	0	0	0	0	1	0
		B							+									=									E	

Abb. 2.3 Reaktionsgleichung und zugehörige BE- bzw. R-Matrizen nach dem Modell von UGI und DUGUNDJI.

Je nachdem, welchen Teil der Fundamentalgleichung $\mathbf{B} + \mathbf{R} = \mathbf{E}$ man als variabel und welchen Teil man als gegeben ansieht, eröffnen sich die folgenden operativen Vorschriften für potentielle Reaktionsgeneratoren:

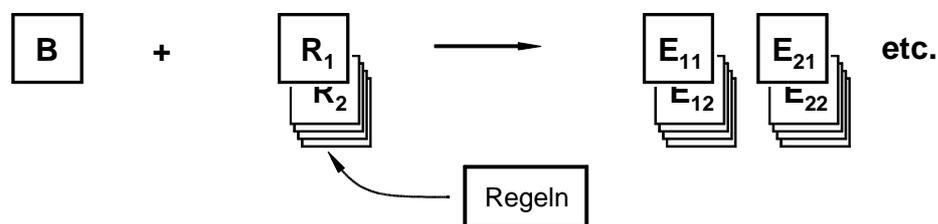
\mathbf{R}_{fix}

Nehme einen *festen Satz* von R-Matrizen (R_1, R_2, \dots) und wende diesen auf alle formal möglichen Weisen und unter Einhaltung valenzchemischer Randbedingungen auf das Ausgangsensemble B an.



\mathbf{R}_{gen}

Erzeuge den, einer gegebenen Menge von *Auswahlregeln* gehorchenden Satz von R-Matrizen (R_1, R_2, \dots) und wende diesen wie bei \mathbf{R}_{fix} an.

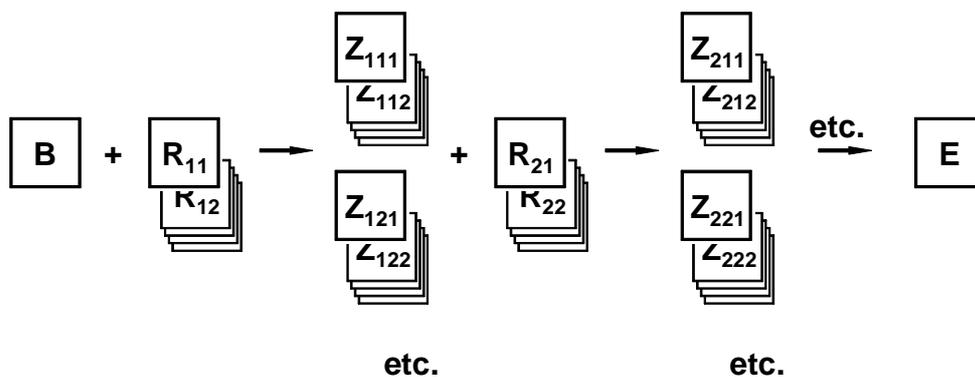


BE_{gen}

Erzeuge zu einer gegebenen R-Matrix R alle *Paare* von Ensembles B und E ($B_1/E_1, B_2/E_2, \dots$), die unter den gegebenen Randbedingungen möglich sind.

**R**_{seq}

Erzeuge zu einem gegebenen Paar von Ensembles E und B alle *Sequenzen* von R-Matrizen ($R_{11}, R_{12}, \dots, R_{21}, R_{22}, \dots$), die nach Addition jeweils Zwischenstufen (Z_{111}, Z_{112}, \dots) des Pfades von B nach E ergeben. Diese Zwischenstufen müssen dabei gegebenen Randbedingungen gehorchen.



In Umkehrung der Fundamentalgleichung als $B = E - R$ lassen sich die oben angeführten Reaktionsgeneratoren natürlich auch in retrosynthetischer Richtung (Austausch von B und E) formulieren.

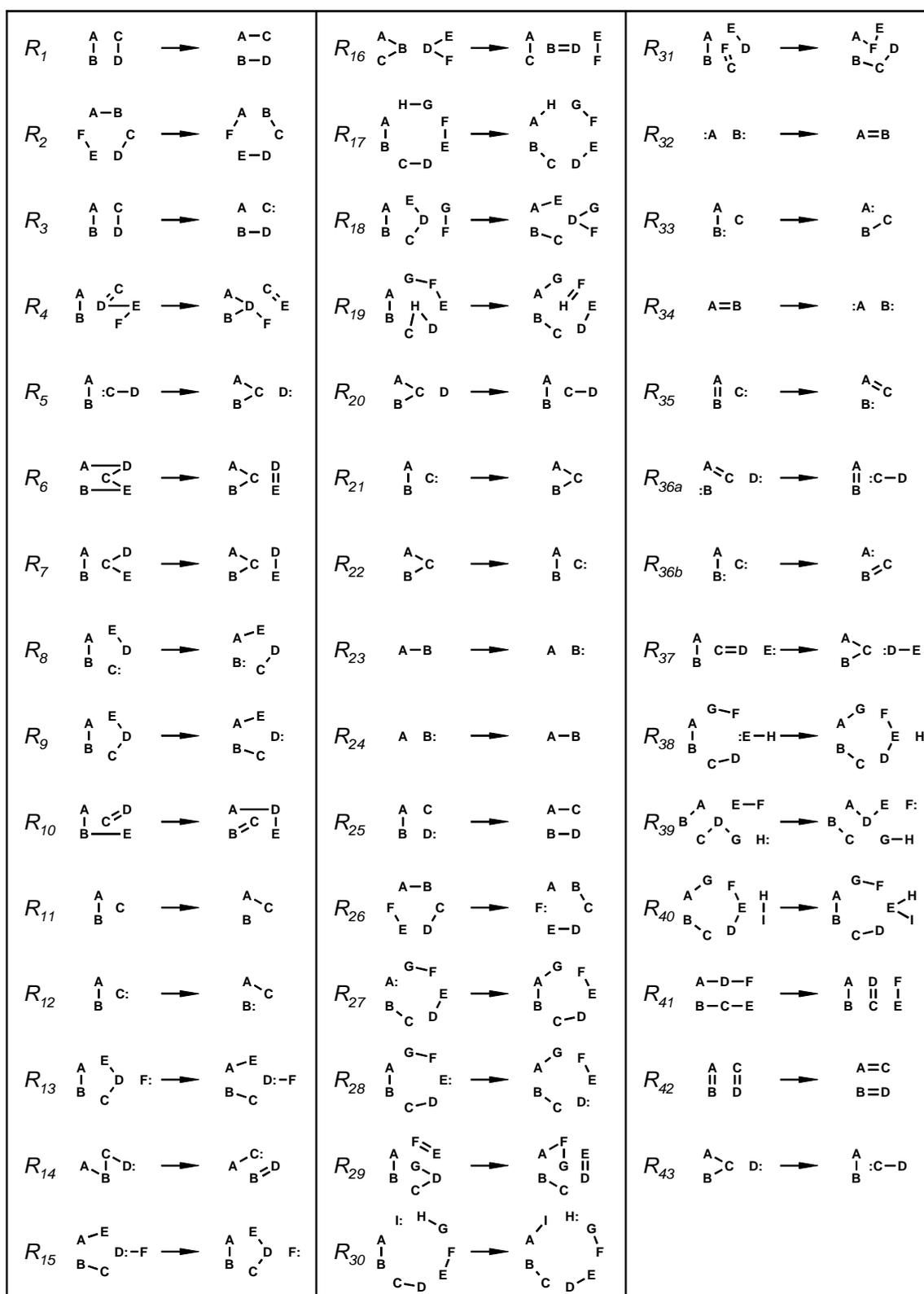
Im allgemeinen wird durch ein Kanonisierungsverfahren (siehe Abschnitt 2.2.6) dafür gesorgt, daß Redundanzen, die sich durch die unterschiedlichen Anwendungsmöglichkeiten einer R-Matrix auf konstitutionell äquivalente Atome,

oder durch Symmetrien innerhalb der formalen Übergangszustände ergeben können, mit einer einheitlichen Numerierung bei der Speicherung der Ensembles eliminiert werden. In speziellen Fällen, z.B. beider Simulation von Markierungsexperimenten, kann es jedoch unumgänglich sein, die Individualität der einzelnen Atome zu bewahren. Die Reaktionsgeneratoren müssen dann die Numerierung der Atome innerhalb der Ensembles beibehalten.

CICLOPS

Bereits kurz nach UGIS grundlegenden Publikationen über R-Matrizen entstand in seinem Arbeitskreis das Programm CICLOPS^{47, 48} (Computers In Chemistry, Logic-Oriented Planning of Syntheses). CICLOPS war mit einem Reaktionsgenerator des Typs \mathbf{R}_{fix} (siehe oben) mit 38 fest vordefinierten Reaktionsschemata ausgerüstet. Ein Reaktionsschema wird hier beschrieben durch eine irreduzible R-Matrix⁴⁹ R^i , beschränkt sich also auf das Reaktionszentrum nach Streichung aller Spalten und Zeilen einer R-Matrix, die nur Nullen enthalten. Auswahlkriterium für die Reaktionsschemata waren: maximal 1, 2 oder 3 außerdiagonale Paare von Einträgen mit ± 1 , maximal 1 Paar mit einer Spalten/Zeilensumme von ± 1 , die anderen Spalten/Zeilensummen müssen gleich 0 sein. Nach einem Verfahren zur vollständigen Permutation der relevanten Matrixelemente wurden die nach der Anzahl der gebrochenen Bindungen in drei Klassen eingeteilten Reaktionsschemata auf die Atome des Ausgangsensembles angewandt. Eine valenzchemische Überprüfung und eine grobe Vorauswahl brechbarer Bindungen (Mehrfachbindungen, Bindungen zu Heteroatomen und α -, bzw. β -ständige Bindungen hierzu) sollten in dieser Programmstudie die Anzahl der Ergebnisse sinnvoll einschränken.

Die Auswahl der angewandten irreduziblen R-Matrizen in CICLOPS war rein formal begründet und stützte sich nicht auf eine Analyse von tatsächlich vorkommenden Reaktionen. E. GARAGNANI und J.C.J. BART erarbeiteten eine eingehende statistische Untersuchung^{50, 51, 52, 53} von 1900 synthetisch relevanten C-C-Verknüpfungsreaktionen⁵⁴ mit Hinblick auf das Auftreten verschiedener Reaktionsschemata. Das Ergebnis dieser Klassifikation war, daß von den insgesamt 43 auftretenden Schemata R_1 bis R_{43} lediglich 10 verschiedene Reaktionsschemata ausreichen ($R_1, R_2, R_{21}, R_{31}, R_{10}, R_{15}, R_8, R_{12}, R_9$ und R_3 , in fallender Häufigkeit), um 92,2 % der Synthesereaktionen zu beschreiben (siehe Tabelle 2.1). Hierbei wurden allerdings nur Reaktionsschemata betrachtet, die Elektronenpaarverschiebungen enthalten, d.h. Radikalreaktionen sind nicht berücksichtigt.



Tab. 2.1 Die 42 wichtigsten Reaktionsschemata für C-C-Verknüpfungsreaktionen nach BART und GARAGNANI.

EROS

Die ersten Erfahrung mit CICLOPS und die Erkenntnis, daß ein Großteil relevanter Reaktionen durch nur wenige Reaktionsschemata beschrieben werden kann, führten zu einer starken Reduktion der eingesetzten Reaktionsschemata und zu einer Verschiebung des Betrachtungsschwerpunktes hin zu Problemen der chemischen Reaktivität. Das Programm EROS⁵⁵ (Elaboration of Reactions for Organic Synthesis) von J. GASTEIGER enthielt in den ersten Versionen 1 bis 4 einen Reaktionsgenerator vom Typ R_{fix} (siehe oben) mit einem stark verkleinerten Satz von Reaktionsschemata (R_1 , R_2 , R_8 , R_{21} und R_{22} , siehe Tabelle 2.1). Die Anwendung von R_8 wurde auf zwei verschiedene Doppelbindungssysteme eingeschränkt.

Die Erkennung von brechbaren Bindungen erfolgte durch eine heuristische Analyse von Bindungspolarität, Bindungsdissoziationsenergie und Resonanzeffekten. Die Werte für diese Größen wurden durch schnelle Parameterverfahren abgeschätzt^{56, 57}. Eine weitere Reduzierung in der Anzahl verwendeter Reaktionsschemata auf R_1 und R_2 erfuhr EROS in der Version 5.0⁵⁸. Der Reaktionsgenerator von EROS 5.2⁵⁸ erzeugte lediglich heteropolare Bindungsbrüche und -neubildungen nach den Schemata R_{23} und R_{24} (siehe Tabelle 2.1). Die Methoden zum automatisierten Wissenserwerb von SYNCHEM2 (siehe oben) standen Pate bei der Entwicklung von EROS 6.0^{59, 60}. Hier wurde die Wissensbasis und der Reaktionsgenerator vom Programm abgekoppelt und somit der Übergang von einem inhärent begrenzten zu einem offenen Erzeugungssystem vollzogen.

ASSOR

Jede Auswahl von Reaktionsschemata, die einem reaktionsgenerierenden System vorgegeben werden, begrenzt dieses in seiner Generierungskapazität. Wählt man jedoch einen kleinsten vollständigen Satz von sogenannten Basistransformationen, dann lassen sich damit alle erdenklichen chemischen Reaktionen beschreiben. W. SCHUBERT konnte zeigen⁶¹, daß durch Linearkombination der Basistransformationen $\pm\theta_{AB}$ ($1e^-$ -Redoxreaktion), $\pm\tau_{AB}$ (Homoapsis, Homolyse), $\pm(\theta_{AB},\theta_{AB})$ ($2e^-$ -Redoxreaktion) und $\pm(\theta_{AB},\tau_{AB})$ (Heteroapsis, Heterolyse) jedes beliebige Reaktionsschema erzeugt werden kann (siehe Abbildung 2.4).

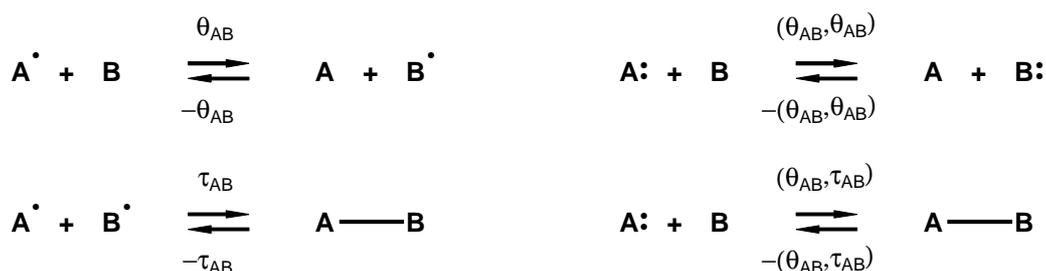


Abb. 2.4 Mit diesen elementaren Reaktionsschemata (Basistransformationen) lassen sich alle formal denkbaren Reaktionsschritte erzeugen.

Das hierauf beruhende Programm ASSOR⁶² (Allgemeines SimulationsSystem Organischer Reaktionen) zeigte jedoch auch deutlich den Nachteil dieser Methode. Die Aufeinanderfolge vieler kleiner Elementarschritte, die z.B. für einen synthetisch relevanten Reaktionsschritt nötig sind, führt zu weit auseinandergefächerten Reaktionsbäumen, die ohne sinnvolle heuristische Einschränkung sehr schnell unüberschaubar werden.

R-STRINGS

Die spezielle Notation von R-Matrizen in der Form von sogenannten R-Strings^{63, 64}, zu Zwecken der Reaktionsdokumentation eingeführt von J. BRANDT, führte zu einem Reaktionsgenerator, der alle formal möglichen Anwendungen einer vom Anwender gegebenen R-Matrix (in R-Stringnotation) auf ein Ausgangsensemble erzeugt. Die graphentheoretische Grundlage der Problemstellung, alle erlaubten Zuordnungen eines gegebenen Pfades (R-String) auf einen Molekülgraphen zu finden, legt die Anwendung eines rekursiven Algorithmus nahe, der in dieser Form von K. STADLER implementiert wurde⁶⁵.

IGOR

Einen Reaktionsgenerator des bis dahin unbekanntens Typs **BE**_{gen} (siehe oben) enthält das Programm IGOR^{66, 67, 68, 69} (Interactive Generation of Organic Reactions) von J. BAUER. IGOR dient nicht zur Behandlung der eher klassischen Probleme wie Reaktionssimulation bzw. Syntheseplanung, es stellt vielmehr Methoden zur Verfügung, zu beliebigen Elektronenverschiebungsmustern (R-Matrizen) alle, nach gegebenen Randbedingungen möglichen Edukt-Produktpaare zu generieren, die durch das gegebene Reaktionsschema ineinander umgewandelt werden. Hiermit wird die gezielte Suche nach präzedenzlosen Reaktionen mit beliebigem Neuheitsgrad möglich⁷⁰. Der Reaktionsgenerator von IGOR arbeitet auf der Basis eines kombi-

natorischen Verfahrens, das unter möglichst frühzeitiger Ausnutzung der gegebenen Grenzwerte in einem Backtrackingalgorithmus die Zeilen der gesuchten BE-Matrizen B und E unter Berücksichtigung valenzchemischer Randbedingungen schrittweise auffüllt.

RAIN

Das zur Zeit aktuellste Programm zur Reaktionsgenerierung auf der Grundlage der R-Matrizen ist das System RAIN^{71, 72, 73} (Reactions And Intermediates Networks). Der im folgenden eingehend beschriebene Reaktionsgenerator von RAIN hat die Eigenschaft, jede beliebige R-Matrix "*in situ*" zu erzeugen und auf ein Ensemble von Molekülen anzuwenden. Die R-Matrizen sind hierbei nicht, wie in CICLOPS, EROS und ASSOR von vorneherein festgelegt, sondern werden über ihre numerischen Eigenschaften in Form von beliebig einstellbaren Randbedingungen definiert. Je nachdem, wie diese Grenzwerte gesetzt sind, und wie die valenzchemischen Randbedingungen definiert sind, erzeugt RAIN Elementarschritte (entsprechend Basistransformationen), ganze Reaktionsschritte oder komplexere Bindungsumlagerungen, die Folgen von Reaktionsschritten darstellen können. Das Programm RAIN dient in der Hauptsache dazu, mit einem Reaktionsgenerator des Typs \mathbf{R}_{gen} Probleme des Typs \mathbf{R}_{seq} (siehe oben) zu lösen, d.h. es hilft bei der Aufklärung von Reaktionsmechanismen und beim Aufbau von Synthesepfaden zwischen gegebenen Edukten und Produkten einer chemischen Reaktion.

Im Rahmen eines Projekts zur sogenannten "bilateralen Syntheseplanung" werden Methoden erarbeitet, die bei der Lösungsfindung gleichzeitig das Syntheseziel sowie einen Katalog verfügbarer Ausgangsstoffe betrachten. Nach einer optimierten Abbildung geeigneter Ausgangsmaterialien auf das Zielmolekül⁷⁴ werden dann in einer stöchiometrischen Ergänzung Coprodukte und kleinere Hilfsmoleküle zum Ensemble hinzugefügt^{75, 76}. In einem letzten Schritt stellt dann RAIN Verfahren zur Verfügung, um die Ausgangsmaterialien mit den Endprodukten über mechanistische oder syntheseorientierte Reaktionspfade zu verbinden.

2.2 Der Reaktionsgenerator von RAIN

2.2.1 Valenzschemata und Übergangsmatrizen

Im Verlauf eines chemischen Reaktionsschrittes wird das Verhalten eines Atoms in starkem Maße durch seine chemische Natur (Ordnungszahl), sowie durch seine Einbettung in den Molekülverband (Bindungen zu Nachbaratomen) beeinflusst. Eine formale Beschreibung dieser Einbettung erfolgt durch sogenannte Valenzschemata⁷⁷. Das Valenzschema eines Atoms beschreibt die Verteilung seiner Valenzelektronen auf Bindungen verschiedener Ordnung und freie Valenzelektronen. Die chemische Natur der Nachbaratome wird hierbei unberücksichtigt gelassen. Die algebraischen Eigenschaften der Valenzschemata und deren wechselseitige Umwandlungen durch Elementarschritte wird durch J. KOCA et al.^{78, 79} beschrieben.

Die formale Darstellung eines Valenzschemas erfolgt durch einen Valenzschemavektor VS mit fünf Elementen. Dieses Pentupel beschreibt die Anzahl der Einfach-, Doppel-, und Dreifachbindungen, sowie die freien Elektronen und die formale Ladung eines Atoms (siehe Abbildung 2.5).

Valenzschema VS = (E,D,T,F,L) mit

- E : Zahl der Einfachbindungen
- D : Zahl der Doppelbindungen
- T : Zahl der Dreifachbindung
- F : Zahl der freien Elektronen (Radikale und lone-pairs)
- L : Formale Ladung

Beispiele :

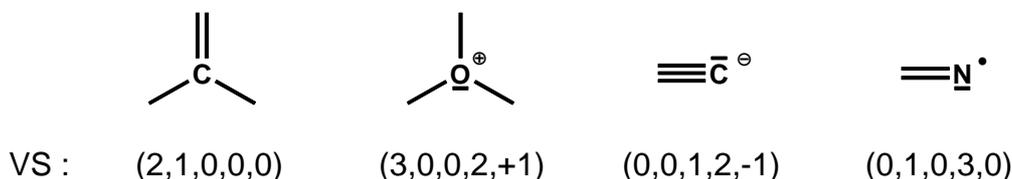


Abb. 2.5 Ein Valenzschema beschreibt die Verteilung der Valenzelektronen auf Bindungen und freie Elektronen. Um auch Pseudoatome behandeln zu können, wird die formale Ladung mit angegeben.

Die valenzchemischen Eigenschaften eines Atoms werden durch die ihm zugeordneten erlaubten Valenzschemata festgelegt. In den meisten Fällen erfolgt die Zuordnung dadurch, daß für das betreffende chemische Element eine Liste von Valenzschemata festgelegt wird. Für Atome mit besonderen Eigenschaften (z.B. Radikale oder Atome in aromatischen Systemen) können spezielle, nur für diese ausgewählten Atome gültige Valenzschemalisten definiert werden.

Nimmt ein Atom an einem chemischen Reaktionsschritt teil, so kann sich sein Valenzschema ändern, oder es kann unverändert bleiben. Die Konversionsmöglichkeiten der einem Atom zugewiesenen Valenzschemata werden in sogenannten Übergangsmatrizen⁸⁰ festgelegt. Eine Übergangsmatrix umfaßt alle erlaubten Valenzschemata und deren innerhalb eines Reaktionsschrittes zugelassenen Übergangsmöglichkeiten. Die zugelassenen Valenzschemata werden den Reihen i und den Spalten j einer quadratischen Matrix TT zugeordnet (siehe Abbildung 2.6). Ein Matrixeintrag tt_{ij} bestimmt, ob der Reaktionsgenerator das Valenzschema i innerhalb eines Reaktionsschrittes in das Valenzschema j umwandeln darf. Dies gilt für die vorwärtsgerichtete Reaktionsgenerierung (Prosynthese). Um rückwärts gerichtete Reaktionsschritte zu erzeugen (Retrosynthese), werden die Übergangsmatrizen von den Spalten zu den Reihen hin interpretiert.

C	$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	$\begin{array}{c} \\ \text{C} \\ \diagup \quad \diagdown \end{array}$	$\begin{array}{c} \\ \text{C}^{\oplus} \\ \diagup \quad \diagdown \end{array}$	$-\text{C}\equiv$
$\begin{array}{c} \\ -\text{C}- \\ \end{array}$	●	●	●	○
$\begin{array}{c} \\ \text{C} \\ \diagup \quad \diagdown \end{array}$	●	●	●	●
$\begin{array}{c} \\ \text{C}^{\oplus} \\ \diagup \quad \diagdown \end{array}$	●	●	○	○
$-\text{C}\equiv$	○	●	○	●

O	$\begin{array}{c} \bar{\text{O}} \\ - \\ \bar{\text{O}} \end{array}$	$=\bar{\text{O}}$	$\begin{array}{c} \bar{\text{O}} \\ \\ \bar{\text{O}}^{\ominus} \end{array}$
$\begin{array}{c} \bar{\text{O}} \\ - \\ \bar{\text{O}} \end{array}$	●	●	●
$=\bar{\text{O}}$	●	●	●
$\begin{array}{c} \bar{\text{O}} \\ \\ \bar{\text{O}}^{\ominus} \end{array}$	●	●	●

● : $tt_{ij} = 1$ Übergang erlaubt
○ : $tt_{ij} = 0$ Übergang verboten

Abb. 2.6 Valenzschema-Übergangsmatrizen. Diese Matrizen können beliebig den Anforderungen der Problemstellung angepaßt werden.

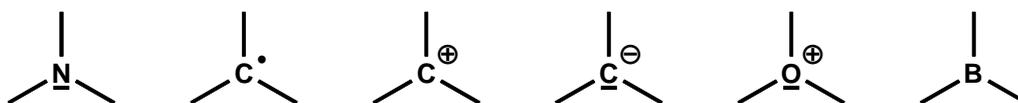
Valenzschemata und Übergangsmatrizen legen somit den valenzchemischen Ereignishorizont der Atome fest und stellen ein zentrales Steuermedium für den Reaktionsgenerator dar. So können durch einen entsprechenden Hauptdiagonaleintrag ($tt_{ii} = 0$) instabile Spezies wie Radikale oder Carbene, Nitrene etc. beschrieben werden. Solche Valenzschemata dürfen nicht innerhalb eines Reaktionsschritts in sich selbst übergehen, müssen also abreagieren oder sie müssen über Umlagerungsreaktionen an einem anderen Atom auftreten.

2.2.2 Bindungsschemata

Ein Bindungsschema beschreibt die Einbettung eines Atoms in den Molekülverband, ohne die Information über freie Elektronen, formale Ladung und den Elementtyp zu berücksichtigen. Es repräsentiert somit einen höheren Grad der Abstraktion als die Valenzschemata, da ein Bindungsschema die Bindungsverhältnisse in einer Reihe von Valenzschemata für verschiedene Elemente darstellen kann. So findet sich das Bindungsschema

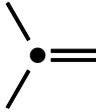
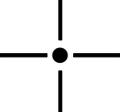
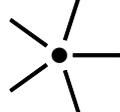
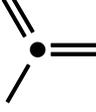
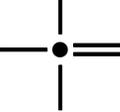
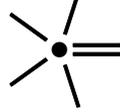
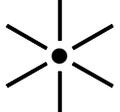


z.B. in folgenden Valenzschemata :



Die Gesamtzahl von verschiedenen Bindungsschemata, die in der organischen Chemie vorkommen, ist begrenzt. Sie hängt lediglich ab von der maximalen Koordinationszahl und Wertigkeit der betrachteten chemischen Elemente.

Tabelle 2.2 zeigt die vollständige Liste der Bindungsschemata, die vom Reaktionsgenerator behandelt werden können. Es sind dies alle Kombinationen mit Einfach-, Doppel- und Dreifachbindungen bis zu einer maximalen Koordinationszahl von sechs. Kombinationen mit mehr als einer Dreifachbindung, sowie Kombinationen mit einer Dreifachbindung und einer Doppelbindung, oder einer Dreifachbindung mit mehr als einer Einfachbindung werden nicht berücksichtigt.

					
(0,0,0)	(0,0,1)	(1,0,1)	(2,1,0)	(4,0,0)	(5,0,0)
					
(1,0,0)	(2,0,0)	(0,2,0)	(1,2,0)	(3,1,0)	(4,1,0)
					
(0,1,0)	(1,1,0)	(3,0,0)	(0,3,0)	(2,2,0)	(6,0,0)

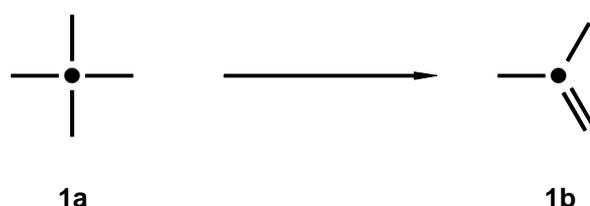
Tab. 2.2 Liste aller Bindungsschemata, die der Reaktionsgenerator behandelt. Die eingeklammerten Zahlentripel geben die jeweiligen numerischen Beschreibungen der Bindungsschemata an (Anzahl der Einfach-, Doppel-, Dreifachbindungen).

2.2.3 BR-Transformatoren

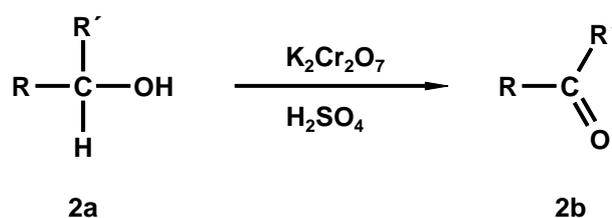
Die lokalen Bindungsverhältnisse von Atomen können durch ihre Valenzschemata bzw. Bindungsschemata eindeutig beschrieben werden. Anders verhält es sich bei der Betrachtung der Veränderung der unmittelbaren Umgebungen der Atome im Verlauf von chemischen Reaktionen.

Der Übergang zwischen zwei Bindungsschemata eines Atoms kann innerhalb eines chemischen Reaktionsschrittes im allgemeinen auf mehrere verschiedene Weisen erreicht werden. So überführt z.B. folgende Transformationsregel das Bindungsschema **1a** in das Bindungsschema **1b** :

- *Breche* eine Einfachbindung
- *Transformiere* eine Einfachbindung in eine Doppelbindung



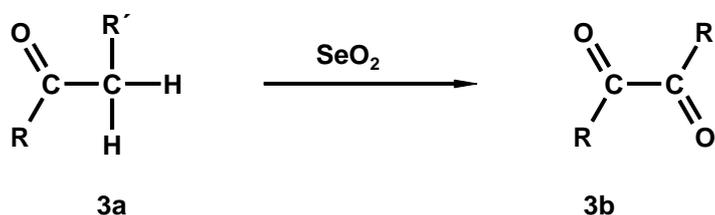
Die Oxidation eines sekundären Alkohols **2a** zum Keton **2b** ist eine Realisierung dieser Transformation⁸¹. Zu beachten ist hierbei, daß lediglich die Änderungen am Valenzschema des C-Atoms berücksichtigt werden.



Eine weitere Möglichkeit zur Umwandlung von **1a** in **1b** besteht in der Anwendung der Transformationsregel

- *Breche* zwei Einfachbindungen
- *Erzeuge* eine Doppelbindung

Die Oxidation des α -C-Atoms eines Ketons durch SeO_2 ist ein Beispiel hierfür ⁸².



Die verschiedenen Möglichkeiten zur Umwandlung eines Bindungsschemas in ein anderes werden formal durch ihre Transformationsregeln beschrieben. Diese legen fest, wieviele Bindungen welcher Ordnung auf welche Weise verändert werden. Wir nennen solch eine Transformationsregel BR-Transformator ⁷³. Er besteht aus zwei Komponenten :

- einer Liste von Bindungsordnungen B
- einer Liste von Veränderungen R , die an B vorgenommen werden.

Der Bindungsschemaübergang von **1a** nach **1b** kann z.B. durch folgende BR-Transformatoren erreicht werden :

$$\begin{array}{ll}
 \text{BR}_1 : & \begin{array}{l} B_1 = 1 \quad 1 \\ R_1 = -1 \quad +1 \end{array} & \begin{array}{l} \text{Bindungsordnungen} \\ \text{Änderungen} \end{array} \\
 \text{BR}_2 : & \begin{array}{l} B_2 = 0 \quad 1 \quad 1 \\ R_2 = +2 \quad -1 \quad -1 \end{array} & \begin{array}{l} \text{Bindungsordnungen} \\ \text{Änderungen} \end{array}
 \end{array}$$

Die Tabelle 2.3 zeigt eine Liste aller formal möglicher BR-Transformatoren, die den Bindungsschemaübergang von **1a** nach **1b** bewirken können. Die komplette Liste aller 324 Bindungsschemaübergänge der 18 Bindungsschemata in Tabelle 1 mit insgesamt 1624 BR-Transformatoren dient als Basissatz für den Reaktionsgenerator. Diese Liste der BR-Transformatoren kann mit Hilfe eines speziellen Programms ⁸³ unabhängig von der eigentlichen Reaktionsgenerierung erzeugt werden. Sie hängt lediglich ab von der grundsätzlichen Auswahl der zu betrachtenden Bindungsschemata.

BR ₁ :	$B_1 = 1 \quad 1$	BR ₄ :	$B_4 = 0 \quad 0 \quad 1 \quad 1 \quad 1$
	$R_1 = -1 \quad +1$		$R_4 = +1 \quad +2 \quad -1 \quad -1 \quad -1$
BR ₂ :	$B_2 = 0 \quad 1 \quad 1$	BR ₅ :	$B_5 = 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1$
	$R_2 = +2 \quad -1 \quad -1$		$R_5 = +1 \quad +1 \quad -1 \quad -1 \quad -1 \quad +1$
BR ₃ :	$B_3 = 0 \quad 1 \quad 1 \quad 1$	BR ₆ :	$B_6 = 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1$
	$R_3 = +1 \quad -1 \quad -1 \quad +1$		$R_6 = +1 \quad +1 \quad +2 \quad -1 \quad -1 \quad -1 \quad -1$

Tab. 2.3 Liste der formal möglichen BR-Transformatoren für den Übergang von 1a nach 1b. Die obere Zeile eines BR-Transformators gibt die erforderlichen Bindungsordnungen, die untere Zeile die Änderungen der Bindungsordnungen, die der Transformator ausführt an.

2.2.4 Numerische Randbedingungen

Der im folgenden beschriebene Reaktionsgenerator ist in der Lage, für einen einzelnen Reaktionsschritt jede Umlagerung von verfügbaren Valenzelektronen zu erzeugen, die durch die definierten Übergangsmatrizen erlaubt ist. Diese erschöpfende Erzeugung aller denkbaren Valenzisomerisierungen kommt lediglich in bestimmten Anwendungsgebieten, so z.B. bei der Strukturgenerierung, zum Einsatz.

Soll jedoch ausschließlich eine chemisch einigermaßen realistische Untermenge der möglichen Elektronenumverteilungen erzeugt werden, so muß das generierende Verfahren durch die Anwendung von Randbedingungen eingeschränkt werden. Diese Randbedingungen limitieren die maximale Komplexität des Elektronenflusses, die Änderung in der molekularen Topologie, sie berücksichtigen geometrische Faktoren des formalen Übergangszustandes und grenzen die strukturelle Vielfalt der erzeugten Intermediate ein. Je nach Anwendungsgebiet kann der Reaktionsgenerator durch die Wahl geeigneter Werte für die Grenzbedingungen mit gewünschten Eigenschaften versehen werden. Mechanistische Studien mit kleinen Elementarreaktionsschritten sind genauso möglich, wie die Erzeugung komplexerer synthetischer Reaktionsschritte. In besonderen Fällen können die Randbedingungen so gewählt werden, daß nur sehr verwandte Strukturen der Ausgangsmoleküle, wie Mesomere oder Tautomere entstehen (siehe Abschnitt 2.2.10).

Die Randbedingungen können anhand ihres Wirkungsbereichs in zwei Gruppen eingeteilt werden. Interne Grenzwerte wirken direkt auf die Eigenschaften des Reaktionsgenerators. Sie steuern unmittelbar den Ablauf im Algorithmus des generierenden Verfahrens. Diese Randbedingungen werden in diesem Abschnitt behandelt.

Externe Randbedingungen bewerten die erzeugten Intermediate, bzw. den gesamten Reaktionsschritt, nachdem der Reaktionsgenerator diesen komplett erzeugt hat. Diese Grenzbedingungen werden in den Abschnitten 2.2.7, 2.2.8 und 2.2.9 behandelt.

Die auf den Generator direkt einwirkenden Randbedingungen limitieren die Komplexität der Elektronenumverteilung, sowie die Änderungen der Topologie. Sie begrenzen für einen Reaktionsschritt u.a. die :

- maximale Anzahl der beteiligten Atome und Bindungen
- maximale Anzahl der gebildeten bzw. gebrochenen Bindungen
- maximale Anzahl der an einem Atom veränderten Bindungen

Eine vollständige Liste aller Randbedingungen wird in Tabelle 2.4 gegeben.

Randbedingung	Typischer Bereich	Satz R01	Satz R02	Satz R03	Satz R04	Satz R05	Satz R06
Max. Anz. reagierender Atome	2...6	∞	∞	∞	6	4	6
Max. Anz. adjazenzändernder Atome	1...6	∞	4	4	4	4	4
Max. Anz. reduz./oxid. Atome	0...2	∞	∞	∞	∞	2	∞
Max. Anz. wechselnder freier Elektronen	0...2	∞	∞	∞	∞	2	∞
Max. Anz. reagierender Bindungen	1...6	∞	∞	∞	6	4	6
Max. Anz. reagierender Bindungen/Atom	1...2	∞	2	2	2	2	2
Max. Anz. wechselnder Adjazenzen	1...4	∞	∞	∞	4	4	4
Max. Anz. wechselnder Adjazenzen/Atom	1...2	∞	∞	∞	2	2	2
Max. Änderung der Bindungsordnung	1...2	∞	1	1	1	1	1
Max. HRI-Wert (siehe unten)	0...10	∞	∞	0	6	4	0,1
Max. Anz. geladener Atome/Ensemble	0...4	∞	∞	∞	∞	2	∞
Max. Anz. geladener Atome/Molekül	0...2	∞	∞	∞	∞	2	∞
Max. Nettoladung/Ensemble (Absolutwert)	0...1	∞	∞	∞	∞	1	∞
Max. Nettoladung/Molekül (Absolutwert)	0...1	∞	∞	∞	∞	1	∞
Max. Ringzahl	0...6	∞	∞	∞	3	2	∞
Max. Anz. der Ringe der Größe n (n = 3...9)	0...6	∞	∞	∞	RG>3	RG>3	∞
Min. Ringgröße für Dreifachbindungen	8	0	0	0	∞	∞	0
Min. Ringgröße für Allenbindungen	8	0	0	0	∞	∞	0
Max. Moleküle/Ensemble	1...4	∞	∞	∞	∞	1	∞
Max. Übergangskomplexität (siehe unten)	0...1	∞	∞	∞	1	2	1
Max. MCD zum Edukt (siehe Absch. 4.3)	N.A.	∞	∞	∞	∞	∞	∞
Max. MCD zum Produkt	N.A.	∞	∞	∞	∞	∞	∞
Max. MCD-Summe zu Edukt und Produkt	N.A.	∞	∞	∞	$\infty(20)$	$\infty(20)$	∞

Tab. 2.4 Numerische Randbedingungen des Reaktionsgenerators von RAIN. Die Sätze von Randbedingungen R01 bis R06 beziehen sich auf verschiedene Anwendungsbeispiele im Text.

2.2.5 Generierungsalgorithmus

Das reaktionsgenerierende Verfahren basiert auf einer Vorauswahl der in Betracht kommenden BR-Transformatoren. Sowohl bei dieser Vorauswahl, als auch bei der anschließenden kombinatorisch vollständigen Erzeugung aller Kombinationen von Anwendungen dieser BR-Transformatoren werden die numerischen Randbedingungen immer zum frühest möglichen Zeitpunkt eingesetzt. Erst dieser optimierte Einsatz von Einschränkungen macht die erschöpfende Enumeration zu einem praktikablen Verfahren mit akzeptablen Ausführungszeiten. Abbildung 2.7 gibt eine schematische Übersicht über den Ablauf des Enumerationsverfahrens. Eine detaillierte Beschreibung des Algorithmus, sowie des individuellen Einsatzes der diversen Randbedingungen wird in Lit. ⁷³ gegeben.

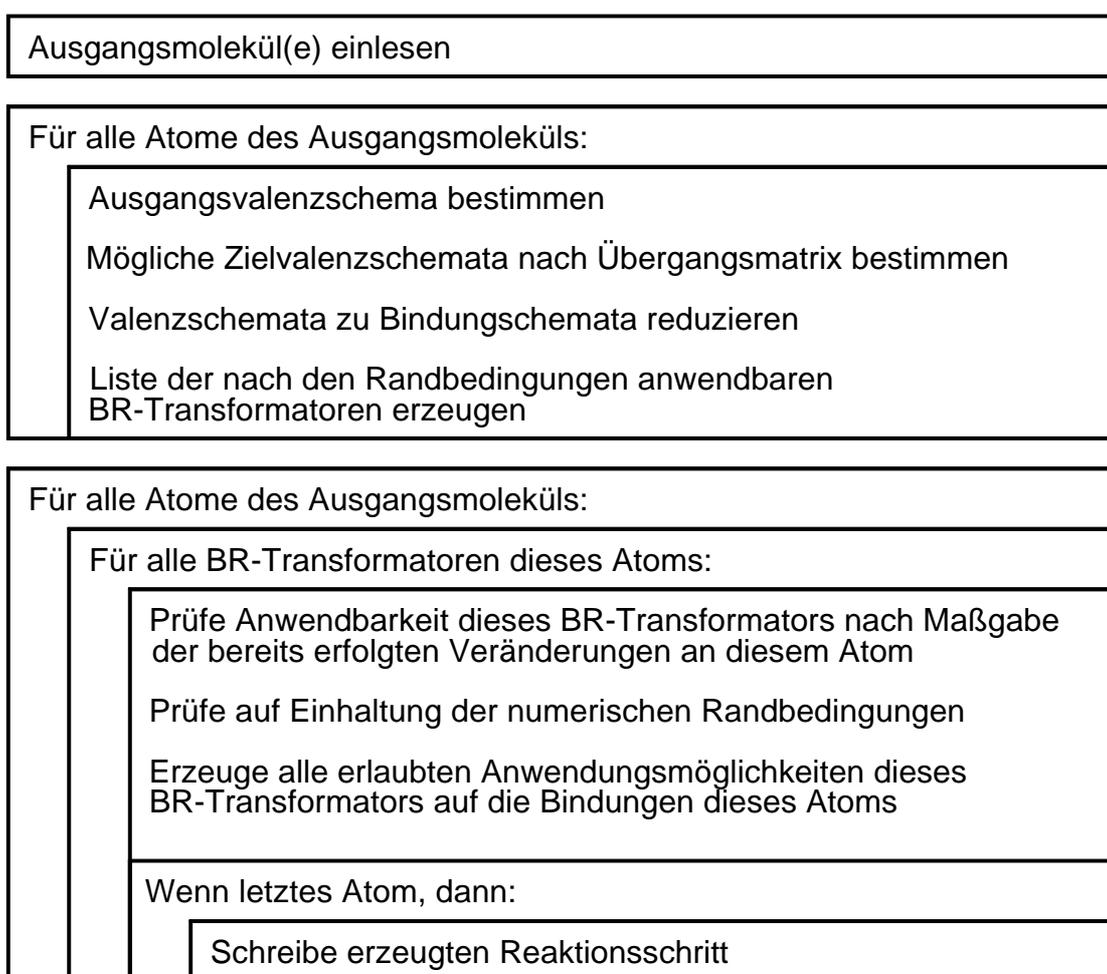


Abb. 2.7 Algorithmus des BR-Transformator-basierenden Reaktionsgenerators (schematisch).

Beispiel

Als illustratives Beispiel für die Auswirkung der Randbedingungen auf das Generierungsergebnis diene die folgende Anwendung des Reaktionsgenerators auf das Ensemble Acetaldehyd/Wasser mit der Ensemblesummenformel $C_2H_6O_2$. Abbildung 2.8 zeigt das Ausgangsensemble zusammen mit den für dieses Beispiel gewählten Valenzschema-Übergangsmatrizen.

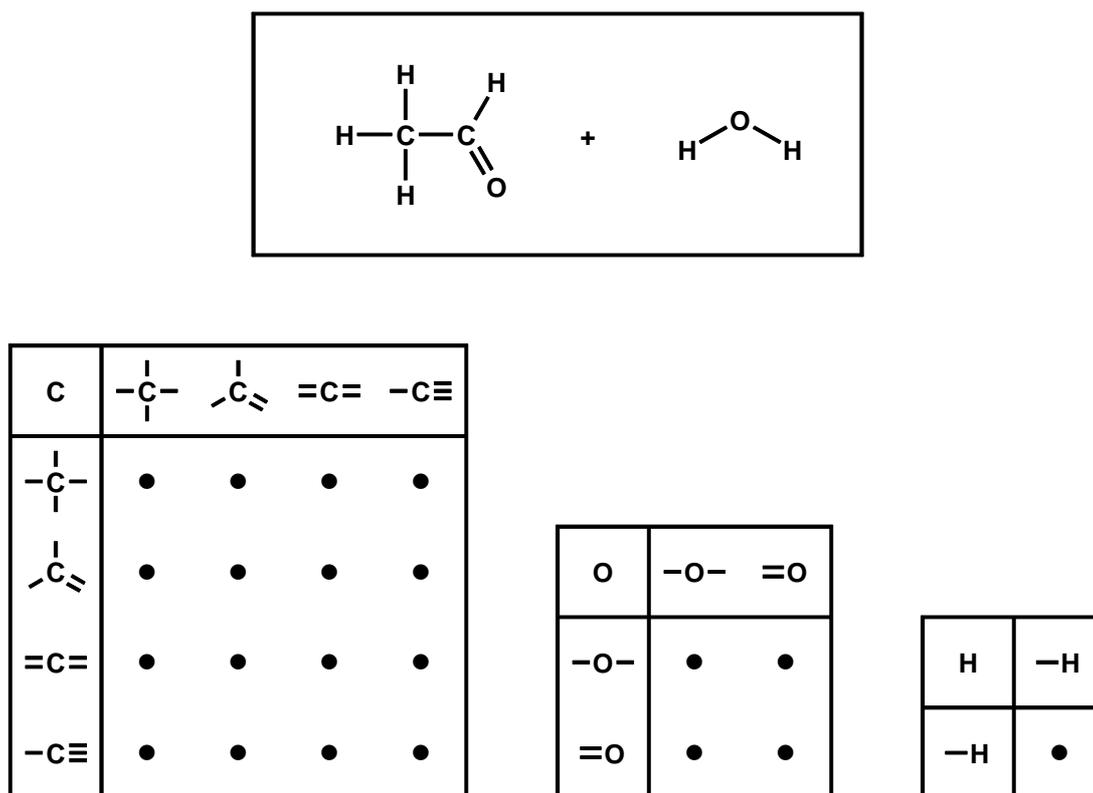


Abb. 2.8 Ausgangsensemble Acetaldehyd/Wasser mit Valenzschema-Übergangsmatrizen. Alle Valenzschemaübergänge seien erlaubt.

Wird das Ausgangsensemble Acetaldehyd/Wasser dem Reaktionsgenerator übergeben und werden zugleich alle Grenzwerte der Randbedingungen auf "unendlich" gesetzt (Randbedingungssatz R01 in Tabelle 2.4), so erzeugt der Generator alle formal möglichen Zusammenstellungen der Ensembleatome unter Beibehaltung der Elektronenzahl (siehe Abbildung 2.9). Im Sinne des erweiterten Isomeriebegriffs von UGI und DUGUNDJI entspricht dies der vollständigen Familie isomerer Ensembles⁴⁶ (FIEM).

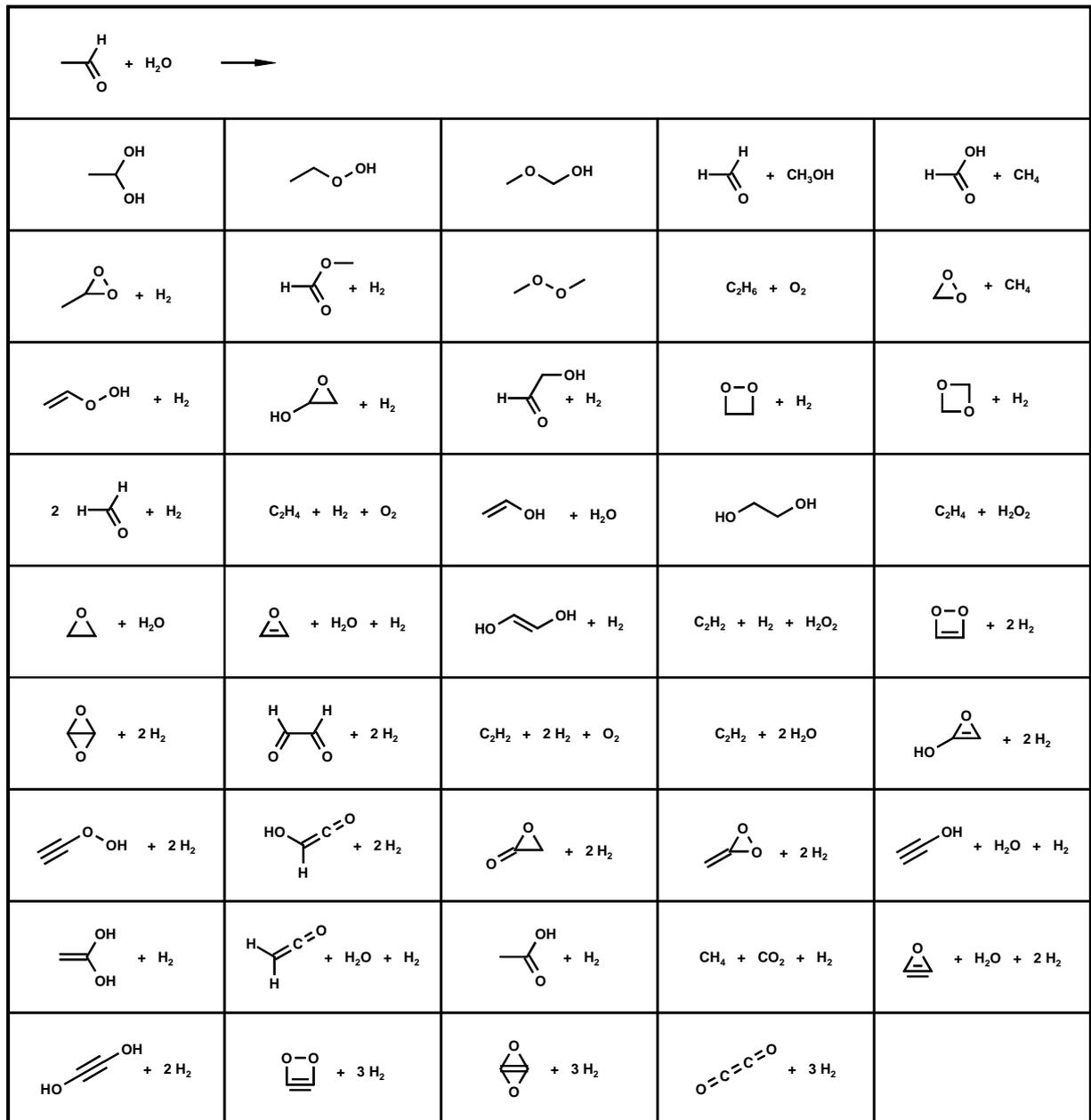


Abb. 2.9 Reaktionsprodukte aus Acetaldehyd/Wasser, generiert ohne Randbedingungen (Satz R01). Dies entspricht der Familie isomerer Ensembles (FIEM) der Summenformel $\text{C}_2\text{H}_6\text{O}_2$.

Eine Begrenzung der Anzahl reagierender Atome und Bindungen, sowie die Beschränkung der Änderungsmöglichkeiten der Bindungsordnungen (Satz R02 in Tabelle 2.4) führt zu einer stark reduzierten Anzahl von erzeugten Reaktionsschritten (siehe Abbildung 2.10). Da die gewählten Randbedingungen jedoch noch keine heuristische Bewertung der Reaktionen einschließen, finden sich neben chemisch sinnvollen Vorschlägen auch solche, die weniger realistisch zu sein scheinen.

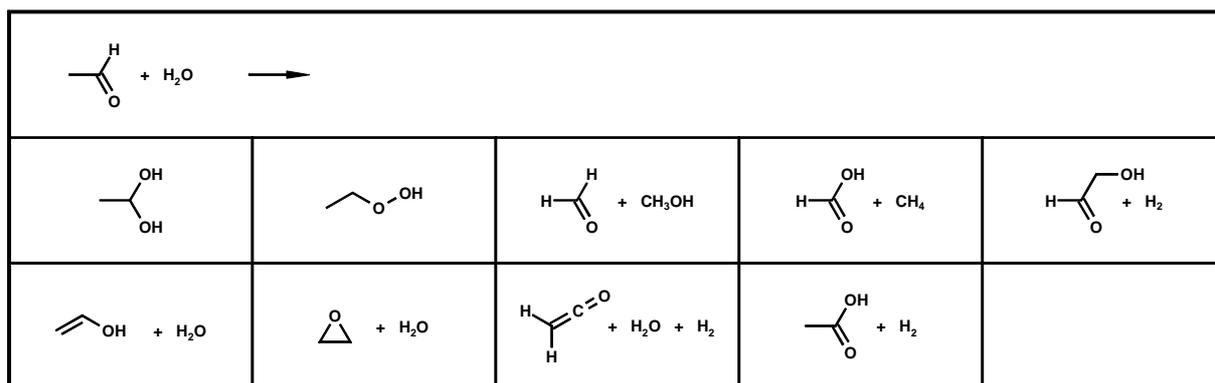


Abb. 2.10 Reaktionsprodukte aus Acetaldehyd/Wasser, generiert mit Randbedingungen (Satz R02, keine Heuristik).

Erst die Berücksichtigung lokaler Atomeigenschaften, wie Ladungsaffinität u.a. (HRI Heuristischer Reaktivitätsindex, siehe Abschnitt 2.2.8) führt zu einer Begrenzung der generierten Reaktionsschritte auf nur mehr zwei chemisch plausible Alternativen (Acetalbildung und Enolisierung, siehe Abbildung 2.11).

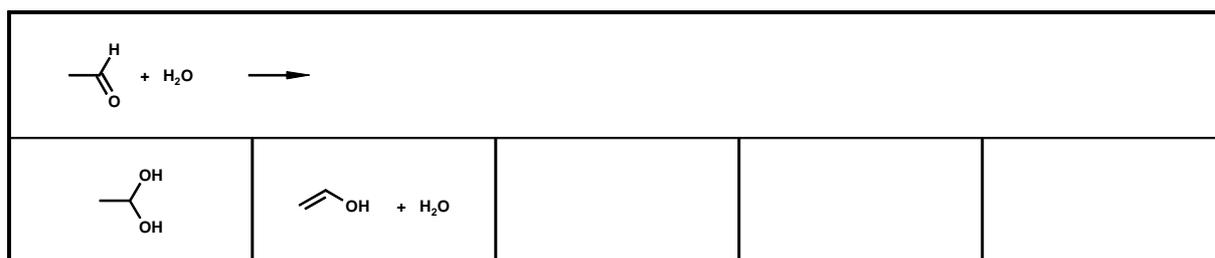


Abb. 2.11 Reaktionsprodukte aus Acetaldehyd/Wasser, generiert mit Randbedingungen (Satz R03, unter Anwendung des Heuristischen Reaktivitätsindex HRI, siehe Abschnitt 2.2.8).

2.2.6 Kanonische Darstellung

Das im vorgehenden beschriebene Verfahren zur Reaktionsgenerierung arbeitet naturgemäß nicht redundanzfrei. Konstitutionell äquivalente Atome im Ausgangsensemble, sowie Symmetrien in den zur Anwendung kommenden BR-Transformatoren führen zu Situationen, in denen bereits generierte Strukturen mehrfach, und dies aus ein und demselben Ausgangsensemble, erzeugt werden.

Eine Möglichkeit, diese Redundanzen zu vermeiden, besteht darin, jedes erzeugte Ensemble nach seiner Komplettierung einem Verfahren zur eindeutigen Numerierung (Kanonische Bezifferung) zu unterwerfen und über diese, lediglich von der Konstitution abhängige Reihenfolge der Atome, eine eindeutige Darstellung zu gewinnen, die letztendlich zur Abspeicherung kommt. Die eindeutige Numerierung molekularer Graphen stellt einen Sonderfall des klassischen Graphisomorphieproblems dar, das als np-vollständiges Problem innerhalb der Graphentheorie nur für wenige spezielle Anwendungen, aber niemals allgemein mit effizienten Algorithmen gelöst werden kann. Da bei Datenbankanwendungen und anderen Retrievalsystemen die wesentliche Funktionalität unmittelbar von einer Eindeutigkeit der Molekülrepräsentation abhängt, gab es bereits zu den Anfängen der computer-gestützten chemischen Dokumentation erhebliche Anstrengungen, dieses Problem in den Griff zu bekommen.

Das bekannteste und auch weithin am häufigsten eingesetzte Kanonisierungsverfahren ist der MORGAN-Algorithmus^{84, 85}. In diesem Verfahren wird die lokale Information über die Knoten des Molekülgraphen (hier: die Konnektivität, d.h. die Anzahl der Bindungsnachbarn) iterativ über die Nachbarn der Atome hinaus solange propagiert, bis sich sie lokalen Atomdeskriptoren nicht mehr ändern.

Abwandlungen des ursprünglichen Verfahrens durch G. MOREAU⁴³ und W. SCHUBERT^{86, 87} benutzen verschiedenartige Atomdeskriptoren, wie Ordnungszahl, Ringinformation, π -Bindungen, Formalladungen, Stereodeskriptoren etc. Sowohl der Algorithmus von MOREAU, als auch CANON von SCHUBERT erzeugen Atomdeskriptoren aus den lokalen Atomeigenschaften und leiten hieraus eine Partition in Äquivalenzklassen ab. Die Zugehörigkeit eines Atoms zu einer Äquivalenzklasse und die Äquivalenzklassen der Bindungsnachbarn führen zu neuen lokalen Atomdeskriptoren usw. Liefert das Verfahren keine weitere Verfeinerung der Äquivalenzklassen und sind dennoch noch nicht alle Atome singular partitioniert (konstitutionell äquivalente Atome), so wird ohne Verlust der Eindeutigkeit des Verfahrens ein Vertreter solch einer Äquivalenzklasse willkürlich ausgezeichnet und das

Verfahren fortgesetzt. Die Implementation des oben beschriebenen Reaktionsgenerators umfaßt ein Relaxationsverfahren nach CANON zur eindeutigen Bezifferung der Molekülensembles, sowie zur Erkennung konstitutionell äquivalenter Atome.

Die Kanonisierung erzeugter Moleküle ist jedoch nicht nur zur Vermeidung von Redundanzen von Bedeutung, vielmehr muß ein reaktionsgenerierendes System, das zur Erzeugung von Reaktionspfaden dient, auch in der Lage sein, zu erkennen, daß Reaktionspfade konvergieren, d.h. über verschiedene Zwischenstufen ein und dasselbe Ensemble erreicht werden kann. Da die Identität einzelner Atome infolge der Kanonisierung verloren geht, kann es bei mechanistischen Untersuchungen vonnöten sein, durch eine gezielte Markierung (z.B. durch Anhängen spezieller Symbolerweiterungen) einzelne Atome individuell auszuzeichnen. Solche Atome bewahren dann über die gesamte Netzwerkgenerierung ihre individuelle Identität und können somit in ihren Veränderungen einzeln verfolgt werden. Eine Konsequenz hiervon ist jedoch die Mehrfachabspeicherung von chemisch eigentlich identischen und somit ununterscheidbaren Molekülen. Diese Technik ist in vielerlei Hinsicht ähnlich den gängigen Markierungsexperimenten mit radioaktiv oder durch ihren Kernspin ausgezeichneten Isotopen in der mechanistischen Aufklärung von chemischen Reaktionen. Ein prominentes Beispiel hierfür sind die Arbeiten zur Klärung des Mechanismus der hydrolytischen Esterspaltung⁸⁸.

2.2.7 Substruktursuche

Die einfache Überprüfung der vom Reaktionsgenerator erzeugten Strukturen hinsichtlich konstitutioneller Aspekte geschieht über lokale Eigenschaften (Valenzzustände, Ladungen etc.) und durch Analyse der Molekültopologie (Ringsysteme). Darüberhinausgehende Einschränkungen erfordern ein System zur Substruktursuche. Dieses erlaubt es in sehr flexibler Weise, bestimmte Teilstrukturen, vom Anwender definiert, in den generierten Strukturen zu suchen, und die fraglichen Moleküle gegebenenfalls zuzulassen oder abzulehnen.

Die hierbei auftretende Problemstellung lautet: Suche eine große Menge von (relativ kleinen) Substrukturen in einem einzelnen (relativ großen) Molekül. Innerhalb der Generierungsphase wechselt die Liste der relevanten Substrukturen nicht, während ständig neue Moleküle der Substrukturanalyse unterworfen werden müssen.

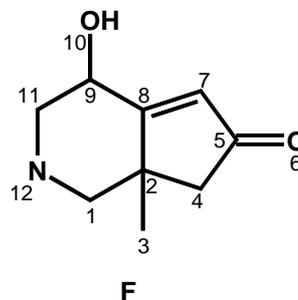
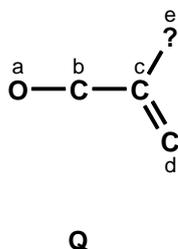
Die Implementation des Reaktionsgenerators enthält ein Substruktursuchverfahren, das Teilstrukturen mit den folgenden Prädikaten verwendet:

- (-) : verbotene Substruktur
- (+) : geforderte Substruktur
- (++) : absolut geforderte Substruktur

Teilstrukturen mit dem Prädikat (-) dürfen in den erzeugten Molekülen nicht auftreten. Die Substrukturen mit dem Prädikat (+) bilden eine Menge, deren einzelne Elemente mit einer "oder"-Verknüpfung auf die erzeugten Moleküle angewandt werden. Es genügt also, wenn die generierte Struktur mindestens eine dieser geforderten Teilstrukturen enthält. Das Prädikat (++) beschreibt Substrukturen, die auf jeden Fall in den erzeugten Molekülen enthalten sein müssen. Eine Anwendung der Substrukturanalyse gibt das Beispiel der Warburganal-Strukturen im Kapitel über Strukturgenerierung (Abschnitt 3.2.1).

Der eingesetzte Substruktursuchalgorithmus beruht im wesentlichen auf der Technik der Partitionierung und "set reduction" von E.H. SUSSENGUTH⁸⁹. Screening-Methoden und fragmentorientierte Verfahren, wie sie in modernen Datenbanken und Strukturretrievalsystemen zur Anwendung kommen^{90, 91}, können hier nicht eingesetzt werden. Eine hierzu nötige Präprozessierung der Moleküle, in denen Substrukturen gesucht werden, kommt nicht in Frage, da die jeweils neu erzeugten Strukturen unmittelbar der Substrukturanalyse mit vielen Teilstrukturen unterworfen werden müssen (siehe oben).

Das Verfahren zur Substruktursuche löst die Frage: Ist die Substruktur **Q** (query structure) in der Struktur **F** (file structure) enthalten oder nicht.



Das implementierte Verfahren liefert gegebenenfalls nur eine und nicht alle der möglichen Abbildungen der Atome von **Q** auf die Atome von **F**. Für die Atome und Bindungen sind Platzhalter (z.B. "?" für ein beliebiges Atom) möglich.

Die im Laufe des Algorithmus immer weiter reduzierten Möglichkeiten, die Atome von **Q** auf die Atome von **F** abzubilden, werden in Partitionen zusammengefaßt. Für ein Atom i der Substruktur **Q** wird die primäre Partition P_i der Atome f aus Struktur **F** nach folgender Vorschrift gebildet:

$$P_i = \{ \text{Atome } f \in \mathbf{F} \mid (\text{Symbol}(f) = \text{Symbol}(i) \vee \text{Symbol}(i) = "?") \wedge \\ (\text{Bindungsschema}(f) \supseteq \text{Bindungsschema}(i)) \wedge \\ (\text{Formalladung}(f) = \text{Formalladung}(i)) \wedge \\ (\text{Anz. freie Elektronen}(f) = \text{Anz. freie Elektronen}(i)) \}$$

Dies ergibt für das obige Beispiel folgende Partitionen:

- a: 10
- b: 1, 2, 3, 4, 5, 7, 8, 9, 11
- c: 5, 7, 8
- d: 5, 7, 8
- e: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12

Aufgrund geeigneter Schnittmengenbildung können nun die Partitionen erheblich reduziert werden: sei nP_i die Häufigkeit, mit der eine Partition P_i auftritt und gelte weiterhin:

$$|P_i| = nP_i$$

so bilden diese nP_i Mengen eine sogenannte "abgeschlossene Partition" aP_i , d.h. die Abbildungen aller nP_i Atome der Substruktur müssen erschöpfend innerhalb der Partition P_i vollzogen werden. Die Elemente von P_i können demzufolge aus anderen

Partitionen $P_j \notin aP_i$ entfernt werden. Wird hierbei eine Partition auf die Nullmenge reduziert, kann die Substruktur nicht in der Struktur enthalten sein. Das folgende Beispiel möge die Partitionsreduktion aufgrund der Schnittmengenbildung erläutern:

Es seien folgende Partitionen gegeben:

m:	3, 4	q:	4, 5, 7, 9, 10
n:	3, 5, 7, 8, 9	r:	10
o:	3, 4	s:	1, 2, 5, 7, 9
p:	3, 5, 7, 9	t:	5, 6, 7, 9

Die Partitionen P_m und $P_o = \{3, 4\}$ bilden eine abgeschlossene Partition, da:

$$P_m = P_o \wedge |P_m| = 2 = nP_m$$

Gleiches gilt für die Partition $P_r = \{10\}$. Entfernt man die Elemente von P_m und P_r aus den anderen Partitionen, so entsteht:

m:	3, 4	q:	5, 7, 9
n:	5, 7, 9	r:	10
o:	3, 4	s:	1, 2, 5, 7, 9
p:	5, 7, 9	t:	5, 6, 7, 9

Nun ergibt sich wiederum die Möglichkeit, die abgeschlossene Partition $P_n = \{5, 7, 9\}$ aus anderen Partitionen zu entfernen und man gelangt letztendlich zu:

m:	3, 4	q:	5, 7, 9
n:	5, 7, 9	r:	10
o:	3, 4	s:	1, 2
p:	5, 7, 9	t:	6

Das Atom q_r der Substruktur **Q** mit der kleinsten Partition (Atom a im obigen Beispiel) wird nun auf die entsprechenden Atome f_r der Struktur **F** (Atom 10 im obigen Beispiel) abgebildet. Für jede dieser möglichen (Root)-Abbildung wird eine weitere Reduktion der Partitionen aufgrund der graphentheoretischen Distanz D zum entsprechenden Root-Atom durchgeführt. Hierfür werden sowohl für q_r , als auch für f_r die kürzestmöglichen Bindungspfade zu den Atomen in **Q** und **F** ermittelt.

$$D(q_r): \quad a:0, b:1, c:2, d:3, e:3$$

$$D(f_r) : \quad 1:4, 2:3, 3:4, 4:4, 5:4, 6:5, 7:3, 8:2, 9:1, 10:0, 11:2, 12:3$$

Für die Reduktion der Partitionen muß gelten:

$$D(f_r) \leq D(q_r)$$

Hieraus ergeben sich für das Beispiel (**Q**, **F**) folgende (Root)-Partitionen:

a: 10
b: 9
c: 8
d: 7, 8
e: 2, 7, 8, 9, 11

Werden diese Partitionen der Reduktion aufgrund der Schnittmengen (siehe oben) unterzogen, verbleibt:

a: 10
b: 9
c: 8
d: 7
e: 2, 11

Die verbleibenden Abbildungsmöglichkeiten werden nun in einem gewöhnlichen Backtracking-Verfahren (Atom-auf-Atom-Abbildung) vollständig kombiniert, wobei auf die Einhaltung der geforderten Bindungsordnungen geachtet wird. Die endgültige Abbildung der Atome von **Q** auf die Atome von **F** ergibt sich damit zu:

a: 10
b: 9
c: 8
d: 7
e: 2

Die für die (Root)-Partitionen verwendeten graphentheoretischen Distanzen sind ebenfalls ein wesentliches Kriterium für die Auswahl von Zentralatomen des Substruktursuchalgorithmus CABASS⁹² (Central Atom Based Algorithm for Substructure Search) von A. DENGLER.

Das im vorigen beschriebene Verfahren zur Substruktursuche löst das Subgraphisomorphieproblem für den eingeschränkten Problemkreis der Zielanwendung RAIN sehr effizient und schnell. Hierbei sei jedoch noch einmal angemerkt daß die Anwendung dieses Verfahrens auf die Suche von relative kleinen Substrukturen (bis ca. 16 Atome) in größeren Molekülen (bis ca. 100-150 Atomen) begrenzt ist.

2.2.8 Der heuristische Reaktivitätsindex (HRI)

Die numerischen Randbedingungen des in diesem Kapitel beschriebenen Reaktionsgenerators erlauben eine rein formale Begrenzung der erzeugten Reaktionsschritte. Dies geschieht ohne direkte Berücksichtigung chemischer Gesichtspunkte. Die von sehr kleinen Reaktionsschritten bis hin zu komplexen Umordnungen einstellbaren Generierungskapazitäten geben Zugang zu einer großen Vielfalt sowohl möglicher als auch chemisch unsinniger Reaktionen.

Eine Alternative zur Einschränkung der erzeugten Reaktionssequenzen auf rein formaler Basis besteht in der Anwendung von chemisch-heuristischen Regeln. Diese Reaktivitätsmodelle auf heuristischer Basis wurden von mehreren Autoren in breiter Form zur Grundlage ihrer Reaktionsvorhersageprogramme gemacht.

Bereits Anfang der 80er Jahre begann W.L. JORGENSEN mit der Entwicklung des Reaktionssimulationsprogrammes CAMEO²⁰ (Computer-Assisted Mechanistic Evaluation of Organic Reactions). Als Gegenstück zu den damals aufkommenden Transformbibliothek-basierenden Systemen entwarf er ein rein mechanistisch orientiertes Programm, das detaillierte Informationen über chemische Reaktivität in algorithmischer Form implementierte. Im Laufe der Zeit wurden in seinem Arbeitskreis viele Routinen zur Erkennung von reaktiven Substrukturen und zur Vorhersage ihrer möglichen Reaktionen entwickelt. Dieses modulare System enthält mechanistische Segmente für eine ganze Reihe von Teilgebieten der organischen Chemie 27, 30, 93, 94, 95, 96, 97, 98, 99. Dieses detaillierte chemische Hintergrundwissen macht es CAMEO als einzigem System seiner Art möglich, bei der Reaktionssimulation auch äußere Einflüsse, wie Druck, Temperatur und Lösungsmittel zu berücksichtigen. Anders als die bibliotheksorientierten Programme hat CAMEO kein Wissen über spezifische Reaktionen, wie Aldol- oder Dieckmannkondensation, Michael-Reaktion oder α -Alkylierungen. Vielmehr können diese, und auch viele präzedenzlose Reaktionen, in einer kreativen Weise durch die Anwendung von mechanistischen Einzelschritten generiert werden. Ein Nachteil der algorithmischen Implementierung ist die unmittelbare Verquickung von Programmcode und Wissensbasis. Diese unflexible Konzeption macht Änderungen und Anpassungen in den Basisregeln sehr schwer, wenn nicht gar unmöglich.

Im Versuch, auf die explizite Erkennung und Bearbeitung von reaktiven Substrukturen zu verzichten, und eine allgemein gültige, berechenbare Größe für das Maß der chemischen Reaktivität zu gewinnen, wendet J. GASTEIGER in seinem Programm EROS (Elaboration of Reactions for Organic Synthesis) ein Multiparameter-

Diese "latenten Polaritäten" bestimmen sowohl die Reaktivität der Reaktionspartner, als auch deren relative Anordnung während der Reaktion. Die + und - Zeichen in Abbildung 2.12 (entnommen aus LAPWORTHS Originalpublikation) stellen hier keine formalen Ladungen dar. A. LAPWORTH ¹⁰⁴ :

" ... It must be emphasized, however, that in attaching the + and - signs to the oxygen and the carbon atoms no hypothesis is invoked, nor is it necessary or even desirable to assume that electrical charges are developed on these two atoms (except perhaps at the actual instant of chemical change). The signs are applied, in the first instance, merely as expressing the relative polar characters which the two atoms seem to display at the instant of the chemical change in question ... "

D.A. EVANS griff Anfang der 70er Jahre dieses stark vereinfachende Modell auf und erarbeitete hieraus Grundlagen für ein allgemeines Synthesekonzept polyfunktioneller Verbindungen ^{40, 41}. Die in seinem unveröffentlichten Manuskript eingeführten Begriffe der dissonanten und konsonanten Strukturen sind seitdem ein wichtiges Werkzeug zur automatisierten Planung komplexer chemischer Synthesen ^{39, 108}.

2.2.8.2 Definition des HRI

Das von LAPWORTH eingeführte Prinzip der alternierenden Polaritäten ist Grundlage eines Teils des Konzepts des heuristischen Reaktivitätsindex. Ausgehend von als formal electrophil, nucleophil oder geladen erkannten Atomen wird dem Molekül ein alternierendes Muster von positiven bzw. negativen Ladungsaffinitäten aufgeprägt. Im Gegensatz zu den Betrachtungen von LAPWORTH und EVANS beschränkt sich hier der Einfluß der funktionellen Substituenten jedoch lediglich auf ihre unmittelbare und mittelbare Umgebung.

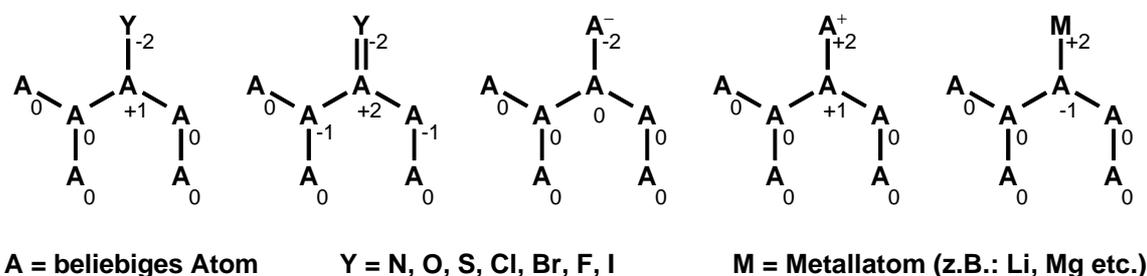


Abb. 2.13 Ladungsaffinitäten aufgrund von Heterosubstitution. Der Einflußbereich eines Heteroatoms erstreckt hierbei sich maximal auf eine Sphäre mit einem Radius von zwei Bindungen.

Abbildung 2.13 zeigt, wie sich das Ladungsaffinitätsmuster in Abhängigkeit der Heterosubstitution und der Bindungsordnungen auf die Nachbaratome ausbreitet. Positive Zahlen der Ladungsaffinität stehen für formal electrophile Zentren, negative Zahlen für formal nucleophile Zentren.

Die nach obiger Abbildung definierten formalen Ladungsaffinitäten dienen nun zur Abschätzung der möglichen Reaktionen, die das Ausgangsmolekül eingehen kann. Hierzu wird für einen Reaktionsschritt der HRI-Wert auf folgende Weise berechnet :

- 1.) $HRI = 0$
- 2.) Für jede geknüpfte Bindung zwischen Atom i und Atom j :
 $HRI = HRI + inc \ HRI(i,j)$
 wobei $inc \ HRI(i,j)$ folgender Matrix entnommen wird :

LA(j) LA(i)	-2	-1	0	+1	+2
-2	+4	+3	0	-1	-1
-1	+3	+3	+1	0	-1
0	0	+1	+2	+1	0
+1	-1	0	+1	+3	+3
+2	-1	-1	0	+3	+4

LA(i): Ladungsaffinität des Atoms i
 LA(j): Ladungsaffinität des Atoms j

- 3.) Für jede geknüpfte Bindung zwischen Atom i und Atom j :
 $HRI = HRI + 2$ wenn Atom i und Atom $j = C, Y, M$
- 4.) Für jede gebrochene Bindung zwischen Atom i und Atom j :
 $HRI = HRI + 2$ wenn Atom i und Atom $j = C, Y, M$

Die Berechnung des HRI berücksichtigt also die Tatsache, daß bei ladungskontrollierten Reaktionen vornehmlich diejenigen Atome miteinander reagieren, die Ladungsaffinitäten unterschiedlichen Vorzeichens tragen. Gleichzeitig wird die Bildung bzw. der Bruch von C-C und C-Hetero-Bindungen mit HRI-Inkrementen ausgezeichnet.

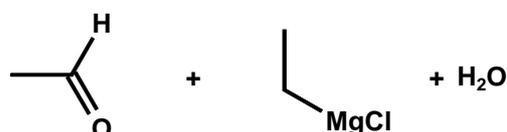
Dieses sehr einfache formale Mittel zur Diskriminierung zwischen möglichen und unmöglichen Reaktionen erlaubt den schrittweisen Übergang von reinen "Lehrbuchreaktionen" mit niedrigen HRI-Werten zu mehr und mehr unwahrscheinli-

chen, aber vielleicht doch interessanten Alternativen mit höheren Werten für den heuristischen Reaktivitätsindex.

2.2.8.3 Anwendungsbeispiel für den HRI

Um den Einsatz und die Wirkungsweise des heuristischen Reaktivitätsindex zu demonstrieren, wird folgendes Beispiel konstruiert :

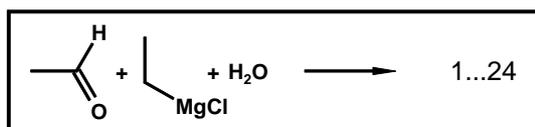
Das Ausgangsensemble einer chemischen Reaktionsfolge sei für dieses Beispiel eine Zusammenstellung der Moleküle einer typischen Grignard-Reaktion :

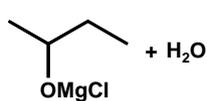
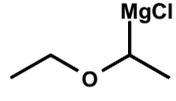
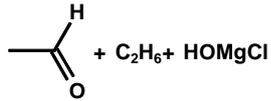
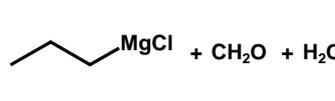
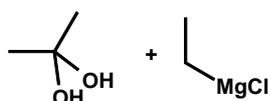
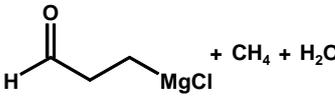
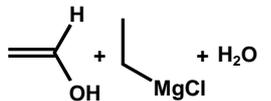
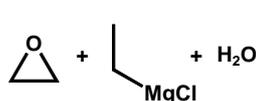
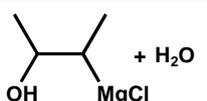
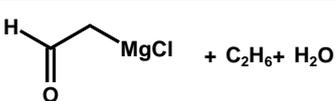
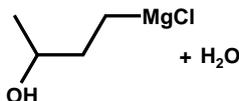
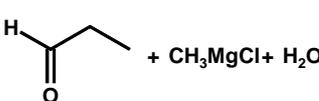
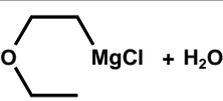
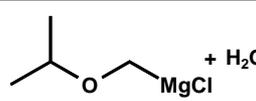
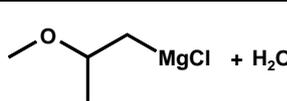
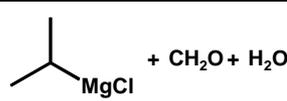
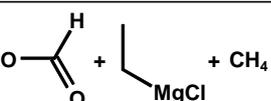
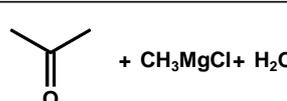
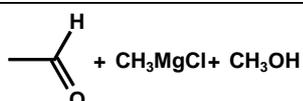
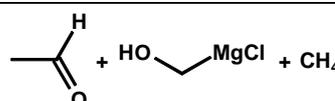
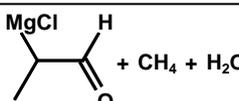
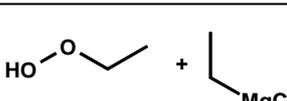
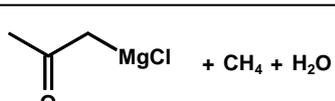
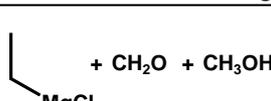


Das Wassermolekül wurde in das Ausgangsensemble übernommen, da es für den zweiten (Aufarbeitungs)-Schritt der Grignard-Reaktion vonnöten ist. Da der Reaktionsgenerator in seiner aktuellen Fassung nur mit abgeschlossenen Ensembles arbeiten kann, d.h. es können im Verlauf der Reaktionssequenzen keine Moleküle hinzukommen bzw. entfernt werden, müssen alle beteiligten Reaktionspartner in einem Ensemble vereinigt sein. Für dieses Ausgangsensemble erzeugt das reaktionsgenerierende Verfahren unter den Standardeinstellungen für die Randbedingungen (R06 in Tabelle 2.4) die 24 Reaktionen in Tabelle 2.5. Für jeden Reaktionsschritt ist dort der entsprechende Wert des HRI angegeben. Die zwei wahrscheinlichsten Reaktionen mit den niedrigsten HRI-Werten sind :

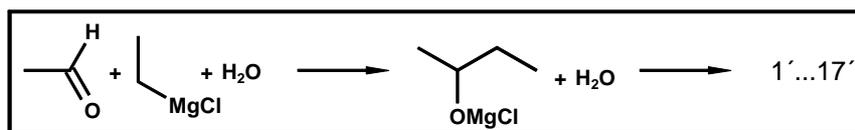
- 1) die Addition des Grignard-Reagenz an die Carbonylverbindung
- 2) die Hydrolyse des Grignard-Reagenz durch das beigefügte Wasser.

Werden für das Reaktionsprodukt der Reaktion 1 in einem weiteren Generierungscyclus alle hierauf folgenden Reaktionsmöglichkeiten erzeugt, so erhält man die 17 Reaktionen in Tabelle 2.6. Dort ist die Hydrolyse des Zwischenprodukts zum sekundären Alkohol (Reaktion 1') durch ihren niedrigen HRI-Wert von 0 als die bei weitem wahrscheinlichste Möglichkeit ausgezeichnet.



Nr.	Produkt 1. Reaktionsschritt	HRI	Nr.	Produkt 1. Reaktionsschritt	HRI
1	 + H ₂ O	0	13	 + H ₂ O	7
2		1	14		7
3		2	15		7
4		2	16		7
5		3	17		8
6		4	18		8
7		4	19		9
8		5	20		9
9		5	21		9
10		6	22		10
11		6	23		11
12		6	24		12

Tab. 2.5 Primärreaktionsschritte der Grignard-Reaktion, geordnet nach steigenden HRI-Wert.



Nr.	Produkt 2. Reaktionsschritt	HRI	Nr.	Produkt 2. Reaktionsschritt	HRI
1'	+ HOMgCl	0	10'	+ CH ₄ + H ₂ O	6
2'	+ H ₂ O + HOMgCl	4	11'	+ CH ₃ OH	7
3'	+ H ₂ O + HOMgCl	4	12'	+ CH ₄	7
4'	+ C ₂ H ₄ + H ₂ O	5	13'	+ HOMgCl + H ₂ O	7
5'	+ CH ₃ OMgCl + H ₂ O	6	14'	+	9
6'	+ C ₂ H ₆	6	15'	+ CH ₃ OH	9
7'	+ CH ₄	6	16'	+ CH ₄ + H ₂ O	9
8'	+ C ₂ H ₆ + H ₂ O	6	17'	+ HOOMgCl	13
9'	+ CH ₄ + H ₂ O	6			

Tab. 2.6 Auf Reaktion Nr. 1 (Tab. 2.5) folgende Reaktionsschritte, geordnet nach steigenden HRI-Wert.

Eine Begrenzung des heuristischen Reaktivitätsindex auf einen Wert von 1 würde den Reaktionsgenerator veranlassen, ausgehend vom Startensemble der Grignard-Reaktion ausschließlich die Reaktionssequenzen in Abbildung 2.14 zu produzieren :

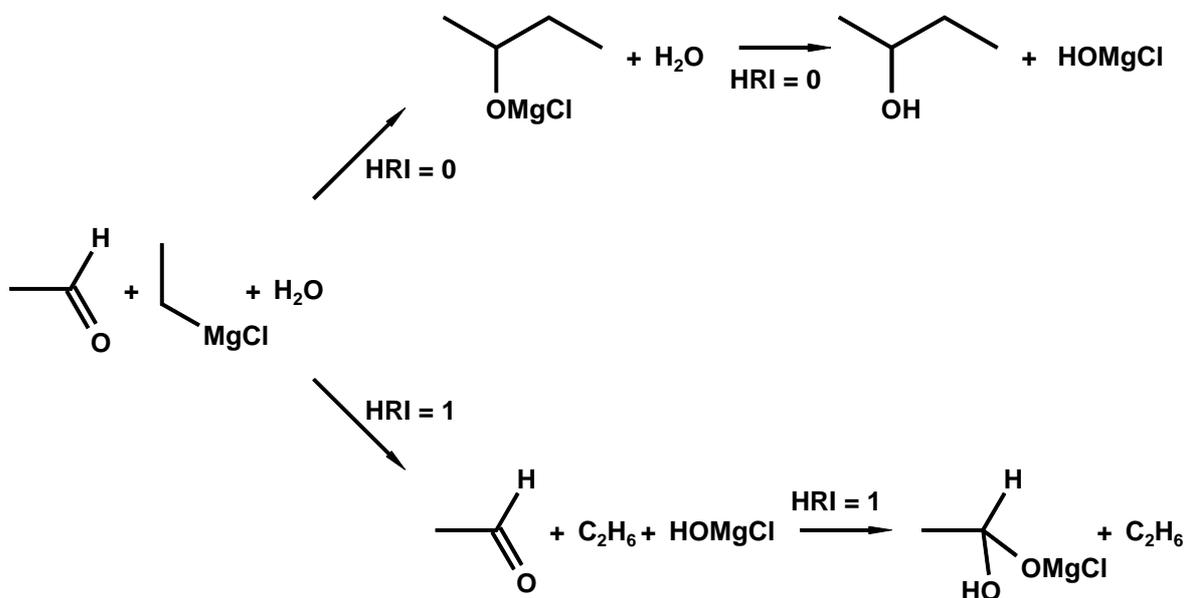


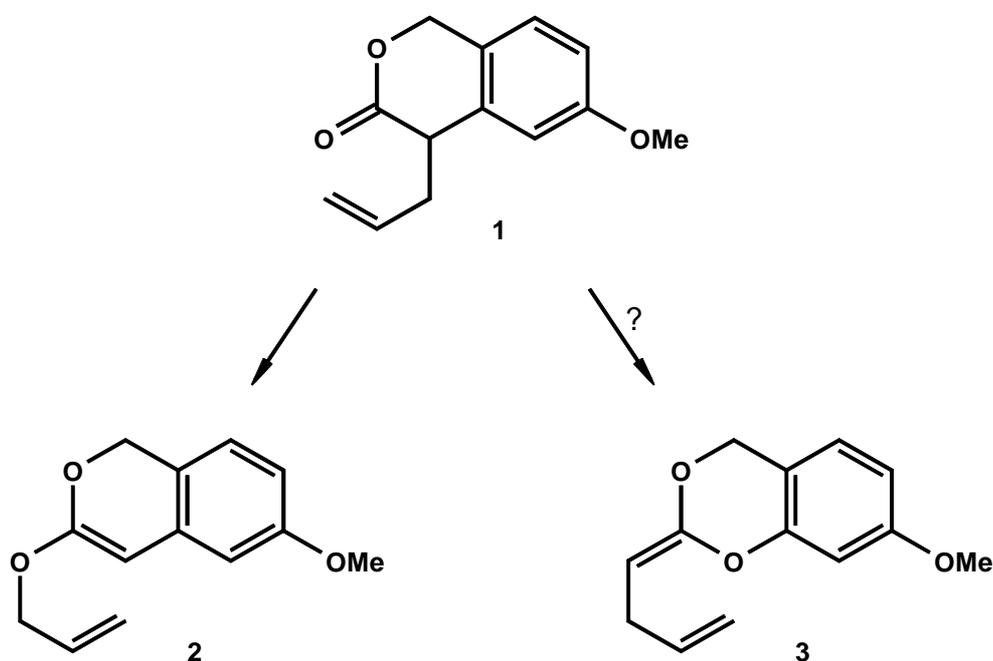
Abb. 2.14 Kompletter Reaktionsbaum eines Grignard-Ensembles. Generiert wurde der Baum unter Kontrolle eines Heuristischen Reaktivitätsindex vom Wert 1.

Von beiden Endprodukten gibt es keine weiteren Reaktionsschritte mehr, die einen $HRI \leq 1$ aufweisen.

Die Anwendung des heuristischen Reaktivitätsindex als Ausschlußverfahren für den Reaktionsgenerator ermöglicht es also, aus einer Vielzahl von Reaktionssequenzen diejenigen auszuwählen, die unter Maßgabe der Ladungskontrolle die wahrscheinlichsten sind. Probleme bei der Anwendung des HRI tauchen auf, wenn Reaktionen nicht ladungskontrolliert ablaufen, bzw. anderen Gesetzmäßigkeiten gehorchen (Cycloadditionen etc.).

2.2.9 Der formale Übergangszustand

Die Begrenzung der Generierungskapazität des Reaktionsgenerators durch die Festlegung der maximalen Komplexität der Elektronenumverteilung und durch den Heuristischen Reaktivitätsindex (HRI) kann im allgemeinen nicht verhindern, daß Reaktionsschritte erzeugt werden, die zu hohe Anforderungen an die topologischen Veränderungen im Molekül stellen. So wird unter gegebenen Randbedingungen z.B. für das Molekül **1** unter anderem sowohl die Hetero-Copeumlagerung zu **2**, als auch die komplexere Reaktion zu **3** generiert.

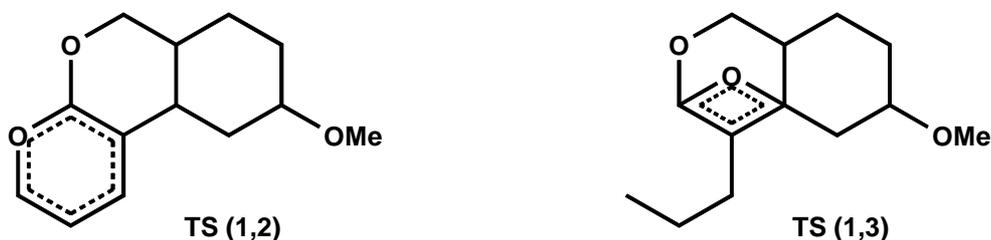


Um nun zwischen diesen beiden Möglichkeiten zu differenzieren, ist es erforderlich, eine topologische Analyse des formalen Übergangszustandes der Reaktion vorzunehmen¹⁰⁹. Der formale Übergangszustand $TS(\mathbf{B}, \mathbf{E})$ zweier Ensembles **B** und **E** ist definiert über :

$$TS(\mathbf{B}, \mathbf{E}) = \text{ad}(\text{BE}(\mathbf{B})) \vee \text{ad}(\text{BE}(\mathbf{E}))$$

Es handelt sich hierbei um die logische Kombination der Adjazenzmatrizen von **B** und **E**. $TS(\mathbf{B}, \mathbf{E})$ enthält demnach sowohl die Bindungen *vor*, als auch die Bindungen

nach dem Reaktionsschritt. Für die beiden obigen Reaktionen ergeben sich hierbei folgende formale Übergangszustände :



Die größere topologische Komplexität des formalen Übergangszustandes TS(**1,3**) gegenüber TS(**1,2**) drückt sich in einer komplexeren Ringstruktur aus. Die Komplexität einer Ringstruktur wird hierbei gemessen an der Zahl einfacher Ringe ¹¹⁰ (Definition siehe unten). Demzufolge wird eine überzählige Ringkomplexität $R^+TS(\mathbf{B},\mathbf{E})$ wie folgt definiert :

$$R^+TS(\mathbf{B},\mathbf{E}) = \text{Ringzahl}(\text{TS}(\mathbf{B},\mathbf{E})) - \max(\text{Ringzahl}(\mathbf{B}), \text{Ringzahl}(\mathbf{E}))$$

Für die beiden obigen Reaktionen ergeben sich demnach :

$$R^+TS(\mathbf{1},\mathbf{2}) = 3 - \max(2,2) = 1$$

$$R^+TS(\mathbf{1},\mathbf{3}) = 4 - \max(2,2) = 2$$

Die Reaktion von **1** nach **3** ist demnach topologisch aufwendiger, als die Reaktion von **1** nach **2**, bei der der formale Übergangszustand keine überaus komplexe Ringstruktur aufweist. Ein maximal gültiger Wert für R^+TS ist Bestandteil des Satzes von Randbedingungen für den Reaktionsgenerator. In der überwiegenden Zahl der Fälle ist hier ein Maximalwert von 1 anzunehmen. Dies erlaubt die Reaktionen mit formal cyclischen Übergangszuständen (nicht unbedingt electrocyclische Reaktionen) und verhindert Reaktionen, in denen, über Ringsysteme hinweg, topologisch komplexe Anordnungen der Atome erforderlich wären.

Für Reaktionsschritte, die mehrere kleinere Einzelschritte in einem zusammenfassen, ist diese Option jedoch außer Kraft zu setzen, da dann sehr wohl (über geeignete Zwischenstufen) eine komplexere Topologie des Übergangszustandes der Gesamtreaktion möglich ist.

Die Wahl des Satzes von Ringen, der für die Berechnung von R^+TS verwendet wird, ist von wesentlicher Bedeutung. Der sogenannte "Smallest Set of Smallest Rings"

SSSR kommt hierbei nicht in Frage, da seine Erzeugung nicht willkürfrei ist ¹¹⁰, und der SSSR nicht für *alle Atome* die Einbettung in kleinere Ringe wiedergibt. So würden zum Beispiel bei dem polycyclischen Kohlenwasserstoff Cuban (einem Pentacyclus gemäß SSSR und IUPAC-Nomenklatur) willkürlich vier C-Atome jeweils drei Vierlingen angehören, während die anderen vier C-Atome nur in zwei Vierringen enthalten sind. Dieser willkürliche Bruch der Molekülsymmetrie ist für eine topologische Analyse der Molekülstruktur (bzw. eines formalen Übergangszustandes) nicht tolerabel.

Die Implementation des Reaktionsgenerators umfaßt deshalb einen Algorithmus zur Ringerkennung ¹¹¹, der mit einem einfachen Backtrackingverfahren für ein Ensemble einen kompletten und eindeutigen Satz einfacher Ringe ¹¹⁰ berechnet. Die Definition dieser Ringmenge geschieht hierbei wie folgt :

Die Menge der einfachen Ringe umfaßt alle Ringe, die keinen kleineren oder gleichgroßen Ring bezüglich der Atome komplett enthalten.

Abbildung 2.15 zeigt zur Erläuterung eine Auswahl komplexerer Ringsysteme mit den zugehörigen Ringgrößen der einfachen Ringe, die das verwendete Verfahren berechnet.

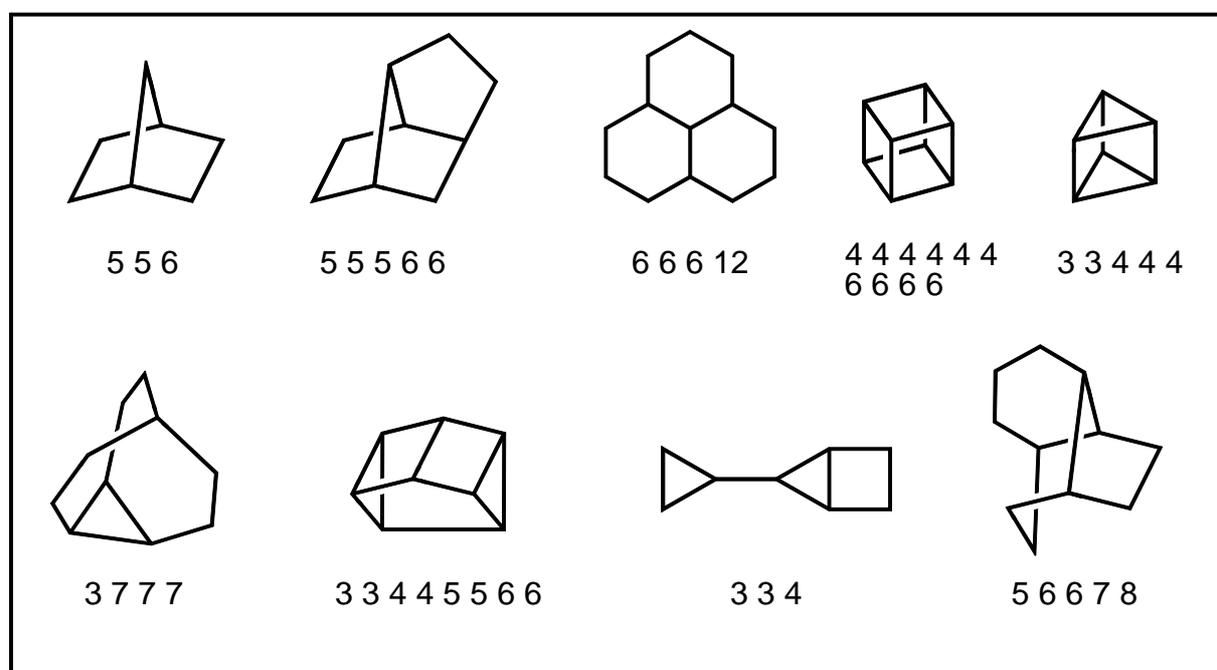
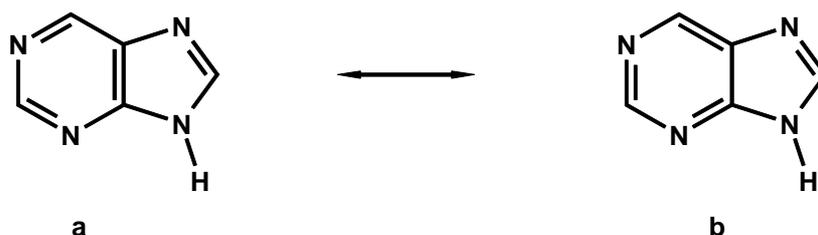


Abb. 2.15 Komplett Sätze sogenannter "einfacher Ringe" in einigen Ringsystemen. Der Satz einfacher Ringe im Übergangszustand bildet Grundlage zur Ermittlung des R^+TS -Wertes.

Die obige Definition des formalen Übergangszustandes entspricht im wesentlichen der des sogenannten "Imaginary Transition State" (ITS) von S. FUJITA¹¹². Mit Hilfe des ITS legt FUJITA ein Ordnungssystem fest, mit dem er in einer sehr umfangreichen Serie von Publikationen (z.B. Lit. 113, 114, 115, 116) eine große Menge organischer Reaktionen katalogisiert. Der ITS umfaßt aber nicht nur die Topologie des Übergangszustandes, sondern auch die formalen Änderungen der Bindungsordnungen während der Reaktion.

2.2.10 Resonanzstrukturen und Tautomerie

Wie es in Computerchemieprogrammen generell üblich ist, speichert der Reaktionsgenerator von RAIN Ensembles von Molekülen in der Form von kompakten Bytestrings, die aus kanonisch bezifferten (siehe oben) Bindungslisten oder BE-Matrizen gewonnen werden. Die konstitutionelle Identität zweier Strukturen kann so durch einen einfachen Stringvergleich verifiziert werden. Für die formalen Bindungsordnungen werden, wie bei dem Ansatz von UGI und DUGUNDJI üblich, ganzzahlige Werte von 1 bis 3 verwendet. Die Interpretation der chemischen Konstitution ähnelt demnach sehr dem Valence-Bond-Konzept. Folglich werden auch die beiden Resonanzstrukturen **a** und **b** als zwei verschiedene chemische Spezies erkannt. Sie erfordern zwei separate Einträge in einem generierten Reaktionsnetzwerk, obwohl sie eigentlich chemisch identische Moleküle repräsentieren.

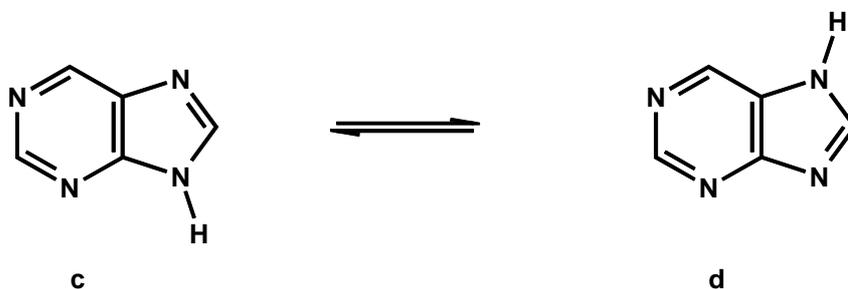


Die bekannten Datenbankapplikationen, wie CAS-Online und Beilstein-Online stehen hier demselben Problem gegenüber. Das Resultat einer Strukturanfrage sollte von der Art der Benutzereingabe unabhängig sein, wenn das gesuchte Molekül durch verschiedene Resonanzstrukturen beschrieben werden kann. Innerhalb einer statischen Beschreibung molekularer Strukturen (Datenbanken) wurde dieses Problem gelöst durch die Einführung von alternierenden (delokalisierten) Bindungen¹¹⁷, oder durch sogenannten PI-Listen von Elektronen als zusätzliche Deskriptoren für Atome, die π -Bindungen angehören, bzw. freie Elektronenpaare tragen¹¹⁸. Bei dem letztgenannten Verfahren, das bei der Erzeugung der sogenannten BRCT (Beilstein Registry Connection Table) zu Anwendung kommt, werden die formalen Bindungsordnungen überhaupt nicht mehr zur Molekülbeschreibung verwendet.

Die Erzeugung von Reaktionsnetzwerken erfordert jedoch einen dynamischen Ansatz. Über lange Zeit gab es keinen Erfolg bei dem Entwurf eines dynamischen Reaktionsgenerators auf der Basis der oben angeführten statischen Konzepte. Erst seit kurzem ist es durch eine Erweiterung des ursprünglichen Modells von UGI und DUGUNDJI, hauptsächlich durch das sogenannte sXBE/sXR-Modell^{119, 120} von N. STEIN gelungen, auch delokalisierte Elektronensysteme in einem mathematisch konsistenten Ansatz, sowohl statisch zu beschreiben, als auch Reaktionsgeneratoren hierfür zu konzipieren.

Dasselbe Problem wie bei den Resonanzstrukturen taucht auf, wenn Moleküle behandelt werden, die in verschiedenen tautomeren Strukturen auftreten können. Innerhalb dieser Arbeit sei der Begriff "Tautomerie" auf das Phänomen der Protonenwanderung oder "Prototropie" beschränkt. Alle beschriebenen Prozeduren können jedoch auf relativ einfache Weise auf andere Arten der Tautomerie, wie "Kationotropie", oder "Anionotropie" erweitert werden.

Für einen Chemiker ist es ein leichtes, die beiden Strukturformeln **c** und **d** als im Gleichgewicht zueinander stehende Tautomere von Purin zu erkennen.



Aufgrund ihrer verschiedenen Konstitutionsformeln, belegen die Strukturen **1c** und **1d** zwei separate Einträge in einem Reaktionsnetzwerk, wie es z.B. von RAIN erzeugt wird. In den oben angeführten Datenbankapplikationen wird das Problem dadurch gelöst, daß die Bindungen, die am Tautomerieprozeß beteiligt sind, durch sogenannte "normalisierte" Bindungen ersetzt werden¹¹⁷, oder indem ein spezielles Speicherkonzept für tautomere Strukturen angewandt wird¹²¹.

W.L. JORGENSEN hat in seinem Programm CAMEO einen Satz von Prozeduren implementiert⁹³, die Moleküle hinsichtlich Aromatizität und Tautomerie (beschränkt auf Ringsysteme) klassifizieren und entsprechend prozessieren können.

2.2.10.1 Problematik von Resonanzstrukturen und Tautomeren in Reaktionsnetzwerken

Das Programm RAIN dient unter anderem zur bilateralen Generierung von Reaktionsnetzwerken zwischen bekannten Edukten und bekannten Produkten einer chemischen Reaktion¹¹¹. Hierbei werden zwei unabhängige "Reaktionsbäume" erzeugt, einer enthält Reaktionspfade, die aus dem Edukt herausführen (prosynthetische Richtung), der andere solche, die zum Produkt hinführen (retrosynthetische Richtung). Diese Reaktionsbäume wachsen durch Erzeugung neuer Reaktionsschritte solange, bis Zwischenstufen gefunden werden, die in beiden Reaktionsbäumen vorkommen (siehe Abbildung 4.8). Sind nun die Strukturen **A** und **A'** (Abbildung 2.16) verschiedene Resonanz-, oder Tautomerstrukturen *ein und desselben* Moleküls, so kann ihre Äquivalenz aus den oben erwähnten Gründen nicht erkannt werden.

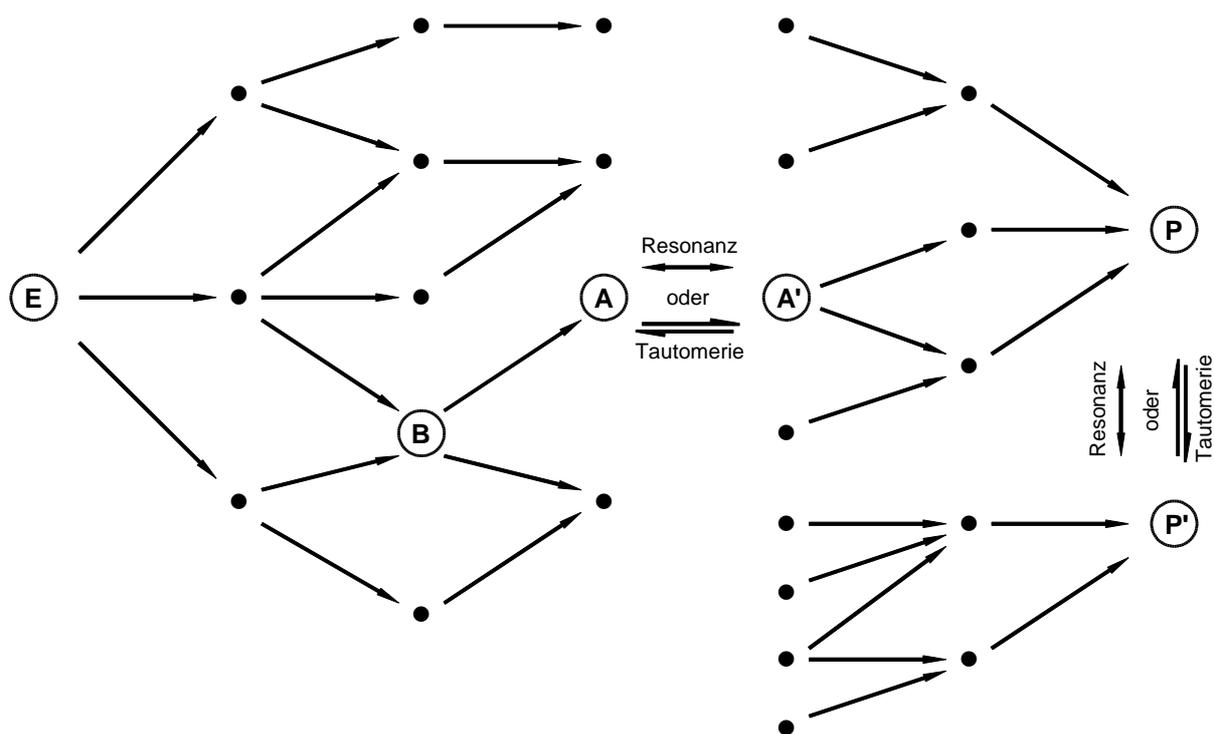


Abb. 2.16 Bei der Netzwerkgenerierung auftretende Probleme durch unterschiedliche Bindungsverhältnisse in Resonanz- bzw. Tautomerstrukturen

Um die Reaktionspfade zwischen **E** und **P** zu schließen, wäre eine weitere Propagation der Reaktionsbäume nötig, bis erkannt wird, daß **A** durch einen erlaubten Reaktionsschritt in **A'** (oder umgekehrt) umgewandelt werden kann. Da nun aber Reaktionsbäume in ihrer Mächtigkeit im allgemeinen mit jedem neuen Propagations-

schritt sehr schnell anwachsen, ist es wünschenswert, den Koaleszenzpunkt ($\mathbf{A} = \mathbf{A}'$) möglichst frühzeitig zu erkennen.

Eine Möglichkeit, dieses Problem zu lösen, besteht darin, die Prozedur zur eindeutigen Darstellung der Moleküle (Kanonisierung, siehe Abschnitt 2.2.6) dahin abzuändern, daß sie zwei beliebige Moleküle als identisch, resonant oder in einer tautomeren Beziehung zueinander klassifiziert. Eine solche Modifikation der Routine CANON wurde in unserem Arbeitskreis erstellt¹²². Dieser Ansatz bringt jedoch keine Lösung bei folgender Problematik :

Das Produkt \mathbf{P} der untersuchten Reaktion sei durch eine alternative (resonante oder tautomere) Form \mathbf{P}' beschreibbar (siehe Abbildung 2.16). Im schlimmsten Fall würde dies nun, da \mathbf{P}' und \mathbf{P} in ihren kovalenten Bindungen differieren, zu zwei vollkommen, oder zumindest wesentlich verschiedenen Reaktionsnetzwerken führen, in Abhängigkeit davon, welche der Strukturen \mathbf{P} oder \mathbf{P}' der Benutzer als Produkt der Reaktion eingegeben hat.

Da dies eine zu große Begrenzung der Allgemeingültigkeit wäre, und eine große Unsicherheit bezüglich der erschöpfenden Behandlung von Problemen darstellt, wurde RAIN mit einem Mechanismus ausgestattet, der jedes Ensemble von Molekülen, das neu in das Reaktionsnetzwerk eingetragen wird (entweder durch den Anwender, oder durch den Reaktionsgenerator) in alle seine Resonanz- bzw. Tautomerstrukturen expandiert, und diese sowohl untereinander, als auch mit ihrem Stammolekül verbindet.

Damit gelingt es dann z.B., daß, wenn \mathbf{A} als Folgeprodukt von \mathbf{B} generiert wird, automatisch eine Verbindung zu \mathbf{A}' hergestellt wird, und das Reaktionsnetzwerk damit geschlossen ist. Auf gleiche Weise würde eine vom Benutzer eingetragene Produktstruktur \mathbf{P} auch als \mathbf{P}' repräsentiert und damit könnten alle individuellen Reaktionsbäume generiert werden.

Es ist nun bemerkenswert, daß für diese Aufgaben keine weiteren Prozeduren und Algorithmen erarbeitet werden mußten. Vielmehr kann der originäre Mechanismus des Reaktionsgenerators selbst, wenn er durch einen speziellen Satz von Randbedingungen kontrolliert wird, entweder Resonanzstrukturen oder Tautomere generieren. Einen wesentlich komplizierteren und in seinen Ergebnissen nicht erschöpfenden Ansatz zur Behandlung dieser Problematik beschreibt W. ELLENRIEDER im Zusammenhang mit Arbeiten am Programm ASSOR^{123, 124}.

2.2.10.2 Die Generierung von Resonanzstrukturen

Eine ganze Reihe von Arbeiten ^{125, 126, 127, 128, 129, 130, 131, 132} beschäftigt sich mit den graphentheoretischen Aspekten und der Enumeration von konjugierten Kohlenwasserstoffen (in den meisten Fällen von benzenoiden Charakter). E.C. KIRBY ¹³³ beschreibt einen Algorithmus zur Generierung von Resonanzstrukturen für beliebige Klassen von konjugierten Kohlenwasserstoffen. Dieser basiert auf einer semirandomisierten Pfadsuche innerhalb einer Adjazenzmatrix. Es gibt keinen Hinweis, daß dieser Algorithmus erweitert werden kann, um organische Moleküle im allgemeinen zu behandeln. Desweiteren ist KIRBYs Programm nicht in der Lage, alle Redundanzen aus den erhaltenen Strukturmengen zu entfernen. So generiert es z.B. für Naphthalen vier "verschiedene" Kekulé-Strukturen. Jedoch sind zwei davon symmetrie-äquivalent (siehe Abbildung 2.17). Dieses Problem könnte jedoch durch Anwendung einer einfachen Kanonisierungsroutine bewältigt werden.

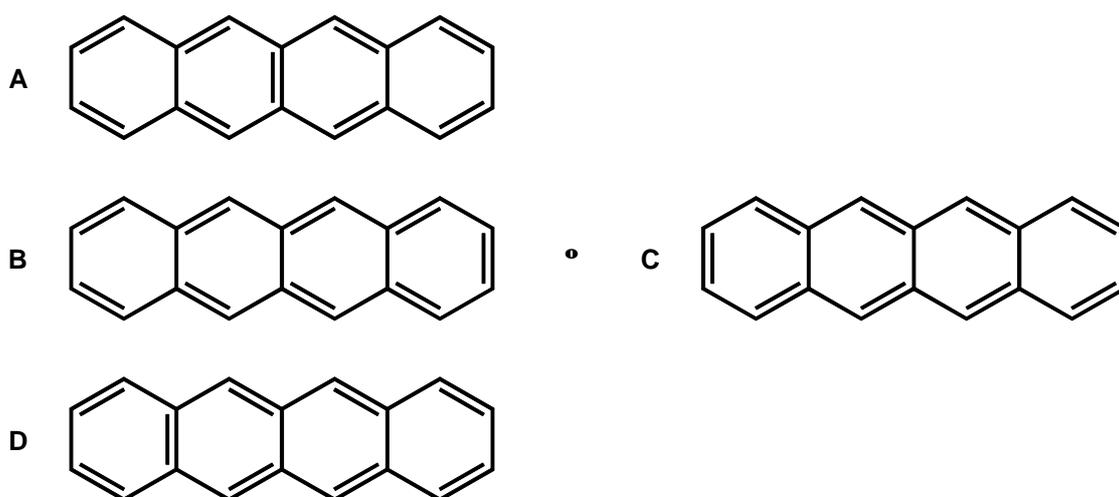


Abb. 2.17 Von KIRBYs Algorithmus erzeugten Resonanzstrukturen von 2,3 Benzanthracen (Naphthalen). Bei Struktur C handelt es sich um eine Symmetriedoublette von Struktur B. Das im Programm RAIN implementierte Verfahren zur Generierung von Resonanzstrukturen erzeugt ausschließlich die Kekulé-Strukturen A, B, und D.

Das Programm RAIN verfolgt einen sehr allgemeinen Ansatz und benutzt die inhärenten Eigenschaften des Reaktionsgenerators zur Erzeugung von Resonanzstrukturen. Dies geschieht mit den üblichen Mechanismen dieses Generators, also einfacher Elektronenumverteilung unter Berücksichtigung folgender Regeln :

- a) Erhalte das σ -Skelett des Moleküls, d.h. verbiete jegliche Änderung der Topologie (Adjazenzen).
- b) Erlaube jede mögliche Verschiebung von "beweglichen" Elektronen, wie π -Elektronen in Mehrfachbindungen und freie Valenzelektronen.

Der Reaktionsgenerator wird durch einen Satz von numerischen Randbedingungen kontrolliert (siehe Tabelle 2.4). Die Forderung, das molekulare σ -Skelett zu erhalten (a), kann erfüllt werden, wenn die entsprechenden Randbedingungen der Topologieänderungen auf den Wert Null gesetzt werden. Dies betrifft die folgenden Grenzwerte :

Max. Anz. adjazenzändernder Atome	→ 0
Max. Anz. wechselnder Adjazenzen/Atom	→ 0
Max. Anz. wechselnder Adjazenzen	→ 0

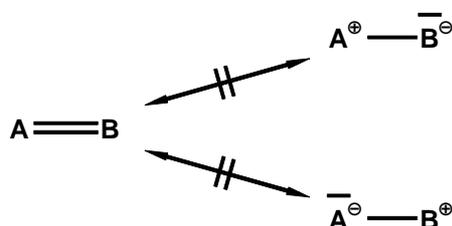
Die Möglichkeit π -Elektronen beliebig zu verschieben (b), wird erreicht, indem die Randbedingungen, die die Elektronenbewegungen kontrollieren, auf einen "unendlichen" Wert, d.h. auf den größtmöglichen ganzzahligen Wert des Computersystems) gesetzt werden. Dies betrifft die folgenden Grenzwerte :

Max. Anz. reagierender Atome	Ⓜ ∞
Max. Anz. Reduz./oxid. Atome	Ⓜ ∞
Max. Anz. wechselnder freier Elektronen	Ⓜ ∞
Max. Anz. reagierender Bindungen	Ⓜ ∞
Max. Anz. reagierender Bindungen/Atom	Ⓜ ∞
Max. Änderung der Bindungsordnung	Ⓜ ∞

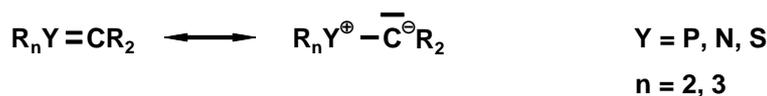
Alle anderen Randbedingungen, die formale Ladungen und die konstitutionellen Aspekte der generierten Moleküle (Ringe etc.) betreffen, werden nicht beeinflusst. Diese werden wie üblich angewandt. Im normalen Generierungsmodus, d.h. bei der

Erzeugung von Reaktionen, würde der Generator keine Strukturen erzeugen, die diesen Anforderungen nicht genügten. Es besteht somit auch keine Notwendigkeit, für die Resonanzstrukturgenerierung den chemischen Ereignishorizont zu erweitern.

Eine einzige weitere Regel muß hinzugefügt werden, um ein spezielles Problem zu behandeln, das unter gewissen Voraussetzungen auftreten kann. Wenn eine Doppelbindung zwei Atome verknüpft, die jeweils sowohl formal positive wie negative Valenzzustände einnehmen können, so ist damit die Umwandlung dieser Doppelbindung in eine dipolare Bindung erlaubt. Da dies zu einer sehr großen Anzahl von evtl. unerwünschten "Resonanzstrukturen" führen würde, kann dem Generator die Trennung von Ladungen nach dem folgenden Schema verboten werden :



In Fällen, in denen eine Ylid-Ylen-Resonanz (siehe unten), bzw. die Polarisierung einer Carbonylbindung relevant ist, sollte diese Regel jedoch nicht angewandt werden.



Die Prozedur zur Resonanzstrukturgenerierung ist in den Reaktionsgenerierungsprozeß integriert. Jedesmal, wenn ein neue Struktur dem Netzwerk hinzuzufügen ist, wird diese dem Reaktionsgenerator übergeben, der dann im speziellen Resonanzstrukturmodus (siehe oben) alle formal möglichen Resonanzstrukturen erzeugt und diese mit dem Stammolekül verbindet. So ist jedes Molekül im Netzwerk durch seinen vollständigen Satz von Resonanzstrukturen repräsentiert. Abbildung 2.18 zeigt einige Beispiele für diese Anwendung.

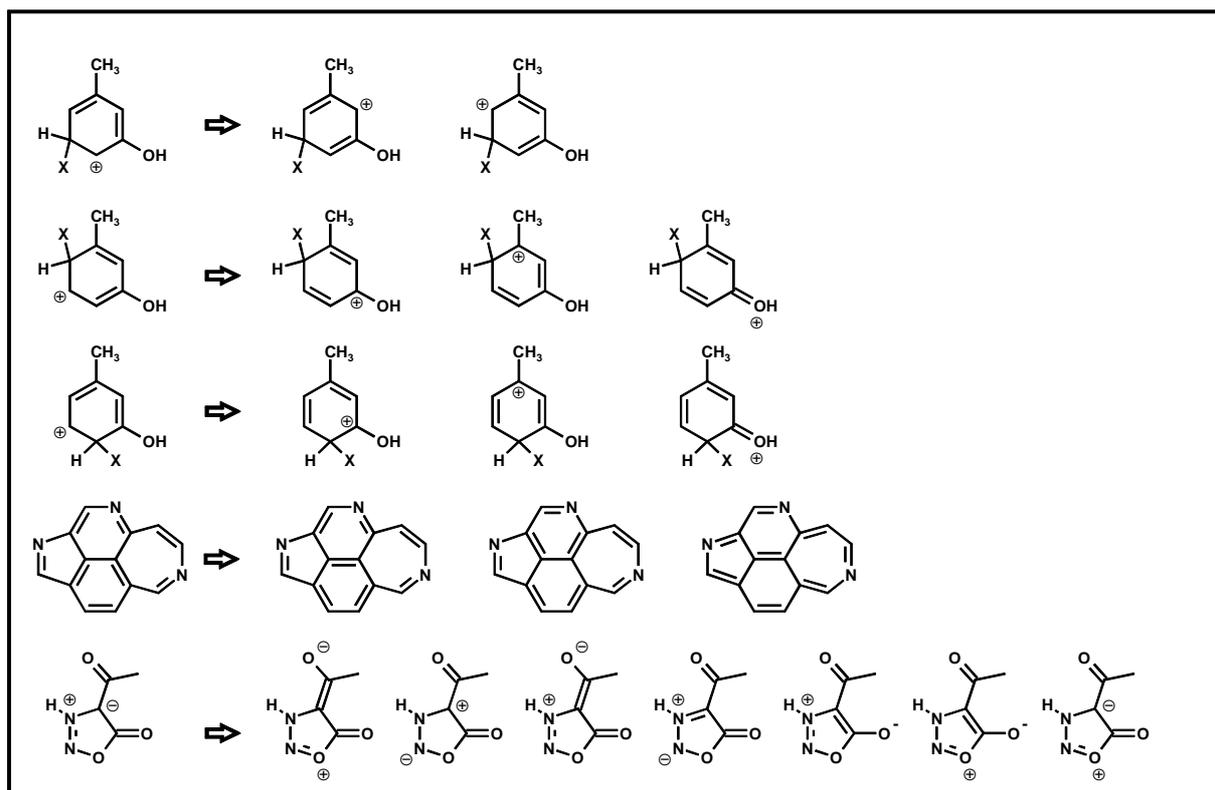


Abb. 2.18 Automatisch generierte Resonanzstrukturen. Die erste Struktur jeder Reihe wurde dem Generator als Stammmolekül übergeben.

2.2.10.3 Die Generierung von Tautomerstrukturen

Von der Technik zur Resonanzstrukturgenerierung hin zur Erzeugung von tautomeren Molekülen ist lediglich ein kleiner Schritt. Wiederum wird der eigentliche Prozeß zur Reaktionsgenerierung mit speziellen Randbedingungen verwendet. Es gelten hierbei folgenden Regeln :

- a) Erhalte das σ -Skelett des Moleküls bis auf die Bindung eines H-Atoms (Prototropie !), das seine Adjazenz innerhalb desselben Moleküls verändern darf.

- b) Erlaube jede mögliche Verschiebung von "beweglichen" Elektronen, wie π -Elektronen in Mehrfachbindungen und freie Valenzelektronen (analog Resonanz).

Die Beschränkung auf ein H-Atom, das eine prototrope Verschiebung erfahren darf (a), geschieht durch Definition folgender Werte für die Adjazenzänderungen beschreibende Randbedingungen :

Max. Anz. adjazenzändernder Atome	→ 3
Max. Anz. wechselnder Adjazenzen/Atom	→ 2
Max. Anz. wechselnder Adjazenzen	→ 2

Zusätzlich muß geprüft werden, ob das Atom, das seine Adjazenz wechselt, ein H-Atom ist. Die Umverteilung der Elektronen (b) geschieht analog zu den Randbedingungen bei der Generierung von Resonanzstrukturen (siehe oben).

Innerhalb des Reaktionsgenerators prüft der Algorithmus zur Elektronenumverteilung lediglich die *oberen* Grenzen der gegebenen Beschränkungen. Da eine Überprüfung *unterer* Grenzwerte nicht vorgesehen ist, kommt es dazu, daß der Generator auch Strukturen mit genau dem σ -Skelett des Stammoleküls produziert. Die Erzeugung von Tautomeren umfaßt also implizit auch die Erzeugung von Resonanzstrukturen. Dies ist insofern von Interesse, als unterschiedliche Tautomerstrukturen auch unterschiedliche Familien von Resonanzstrukturen aufbauen können.

In Fällen, in denen mehr als nur ein H-Atom einer prototropen Verschiebung unterliegen kann, wird das Verfahren iterativ auf den Ergebnissatz von Tautomeren angewandt, bis dieser komplett ist, d.h. keine neuen Tautomer- bzw. Resonanzstrukturen mehr hinzukommen. Abbildung 2.19 zeigt einige illustrative Ergebnisse von Tautomererzeugungen. Für diese Beispiele wurden nur formal ungeladene Valenzschemata zugelassen, um eine kleinere Anzahl resultierender Strukturen zu erhalten. So würde z.B. das Acetylsydnon (siehe Abbildung 2.18) mit zwei formal geladenen Atomen einen Satz von 175 verschiedenen Tautomer- bzw. Resonanzstrukturen erzeugen.

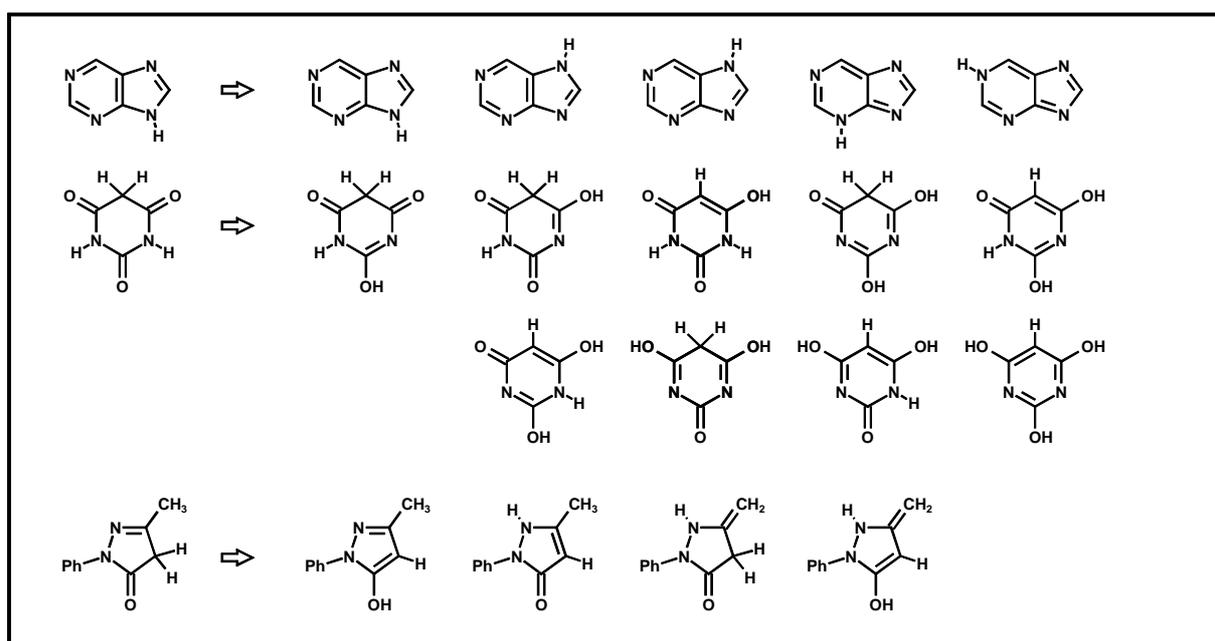


Abb. 2.19 Automatisch generierte Tautomerstrukturen. Die erste Struktur jeder Reihe wurde dem Generator als Stammmoleküle übergeben.

Kapitel 3

Strukturgenerierung

Zusammenfassung:

Nach einer kurzen Einführung in das Problemfeld und den Kenntnisstand der redundanzfreien und erschöpfenden Erzeugung von Konstitutionsisomeren wird beschrieben, wie der in Abschnitt 2.2 vorgestellte Reaktionsgenerator als Strukturgenerator verwendet werden kann. Die Erzeugung von 42 nach spektroskopischen Daten möglichen Konstitutionsisomeren des Warburganals dient als illustratives Beispiel für die Leistungsfähigkeit des Systems. Als weiteres aktuelles Beispiel ist die Generierung eines kompletten Strukturkatalogs für die bislang noch ungeklärte Struktur von B_6H_{14} beschrieben.

3.1 Strukturgenerierung, ein Überblick

Die Aufklärung unbekannter Molekülstrukturen ist eine der Hauptaufgaben moderner chemischer Forschung. Aus diesem Grund ist es nicht verwunderlich, daß sich seit Beginn der Computeranwendung in der Chemie viele Arbeitsgruppen mit der Entwicklung von Algorithmen und Programmen für dieses Problemfeld auseinandergesetzt haben. Die computerunterstützte Strukturaufklärung läuft typischerweise in drei Phasen ab (siehe Abbildung 3.1).

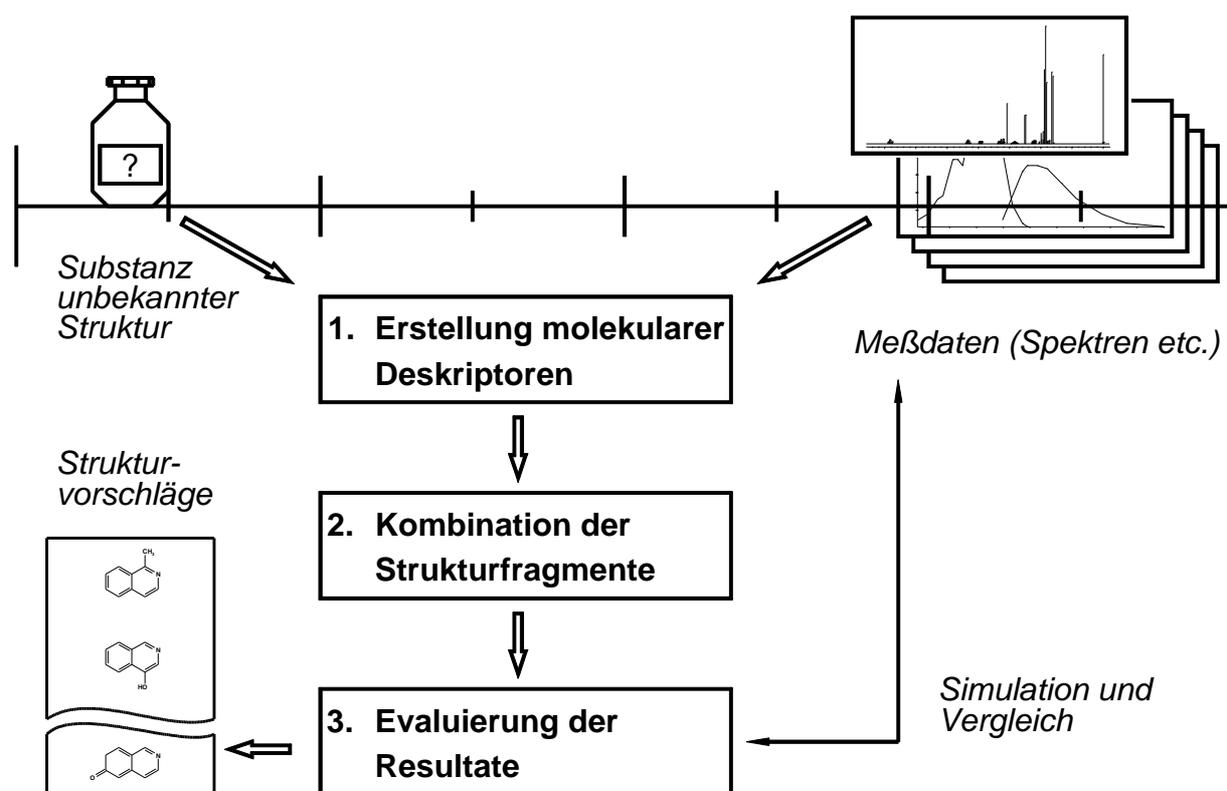


Abb. 3.1 Die drei Phasen der computerunterstützten Strukturaufklärung.

In der ersten Phase werden alle verfügbaren spektroskopischen, chemischen und physikalischen Daten der unbekannt Substanz, soweit möglich, in strukturellen Anforderungen ausgedrückt. Dies geschieht meist durch mehr oder weniger genaue Definition von geforderten oder ausgeschlossenen Substrukturen, einer angenommenen Summenformel, oder anderen Deskriptoren, wie z.B. teilweise definierte Umgebungen einzelner Atome. Bis zu einem gewissen Grad ist dieser Prozeß automatisierbar.

Im zweiten Schritt werden die in der ersten Phase gewonnenen Strukturfragmente unter Berücksichtigung aller weiterer Nebenbedingungen zu vollständigen Strukturen zusammengesetzt. Hierbei ist es unumgänglich, daß diese Strukturgenerierung *erschöpfend* und *redundanzfrei* durchgeführt wird. In diesem Abschnitt der Strukturaufklärung liegt deswegen auch der Schwerpunkt der Computerunterstützung.

In der dritten Phase werden alle gewonnenen Strukturvorschläge einer weiteren Evaluierung unterzogen. Hierbei können die unterschiedlichsten Kriterien, wie die Vorgeschichte einer möglichen *in-vitro*-Synthese, die potentielle Biogenese, ein Vergleich der realen mit simulierten Spektren, und vieles mehr herangezogen werden. Im Idealfall verbleiben am Ende einige wenige Strukturkandidaten für die unbekannte Substanz.

Im Prinzip umfaßt die Strukturgenerierung natürlich die Erzeugung von allen formal möglichen Konstitutionsisomeren, Stereoisomeren und Konformeren. Sind jedoch die denkbaren Stereoisomeren^{134, 135, 136} zu einer gegebenen Molekülkonstitution meist noch handhabbar, so scheidet eine Erzeugung aller Konformationen¹³⁷ eines Moleküls oft an der unübersehbar großen Zahl der Möglichkeiten, die meist nur durch sehr flache Minima in der Potentialhyperfläche^{138, 139} charakterisiert sind. Aus diesem Grund beschränkt sich die Mehrzahl der publizierten Strukturgenerierungsprogramme auf die Erzeugung aller Konstitutionsisomere aus gegebenen Fragmenten und einzelnen Atomen unter definierten Randbedingungen.

Die Entwicklung von Algorithmen zur redundanzfreien Generierung von Konstitutionsisomeren begann bereits in den 30er Jahren mit der Publikation eines speziellen Verfahrens für acyclische Alkane¹⁴⁰. Naturgemäß brachte jedoch erst die Verfügbarkeit von wissenschaftlich nutzbaren Computern einen weiteren wesentlichen Fortschritt auf diesem Gebiet. Das von J. LEDERBERG initiierte DENDRAL Projekt^{141, 142} stellt hierbei einen wichtigen Beitrag dar. DENDRAL wurde weit über die Grenzen der Computerchemie hinaus als einer der Wegbereiter für Systeme der *artificial intelligence* bekannt, ist in seiner Bedeutung allerdings auch umstritten^{143, 144, 145}. LEDERBERG erkannte, daß sich die Topologien aller Moleküle auf vorteilhafte Weise mit Hilfe von Baumgraphen und sogenannten "vertex graphs" darstellen lassen¹⁴⁶. Letztere entstehen durch Reduktion (poly)-cyclischer Systeme auf die Knoten, die mehr als zwei Nachbarknoten besitzen. Ein erschöpfendes Generierungsverfahren für "vertex graphs" bei vorgegebenen Knotenvalenzen ermöglichte die Konstruktion des Algorithmus STRGEN¹⁴⁷, der die Basis des Strukturaufklärungsprogramms CONGEN¹⁴⁸ bildet.

Das von Y. KUDO und S. SASAKI beschriebene System CHEMICS^{149, 150, 151, 152} erzeugt Strukturen aus ungefähr 200 vordefinierten Fragmenten. Ein automatisiertes Spektreninterpretationssystem selektiert dabei zuvor aus dieser Substrukturbasis für die aktuelle Fragestellung relevante Untermengen von Teilstrukturen. Die Kombination dieser Fragmente zu kompletten Strukturen erfolgt auf der Basis des sogenannten "connectivity stacks"^{153, 154}, einem Verfahren, das bereits während der Generierung eine nahezu vollständige Redundanzfreiheit garantieren soll. Allerdings ist diese Methode auf Fragmente beschränkt, die nur eine Art von konstitutionell nicht unterscheidbaren Anknüpfungspunkten besitzen.

Der von M.E. MUNK beschriebene Algorithmus ASSEMBLE¹⁵⁵ des Strukturaufklärungsprogramms CASE^{155, 156} besitzt diese Beschränkung nicht. Hier können für beliebige Atome in den Fragmenten spezifische Umgebungen, wie weitere gebundene Atomarten oder Ringsysteme, definiert werden. Jedoch besitzt dieses Verfahren, wie L.A. GRIBOV'S MASS¹⁵⁷ den Nachteil, daß *a priori* keine zusammenhängenden Strukturen erzeugt werden. Die molekulare Verbundenheit, sowie die Redundanzfreiheit muß durch ein nachgeschaltetes Überprüfungssystem und Kanonisierungsverfahren gewährleistet werden.

Das modernste, aus dem DENDRAL Projekt hervorgegangene Strukturgenerierungsprogramm GENOA¹⁵⁸ besitzt gegenüber vergleichbaren Systemen den Vorteil, daß es *überlappende* Substrukturen als Bausteine zuläßt. Oftmals läßt sich aus spektroskopischen Daten kein eindeutiger Satz von nichtüberlappenden Fragmenten gewinnen, sodaß eine Beschränkung auf relativ kleine, eindeutig platzierbare Substrukturen nötig wird. R.E. CARHARTS "constructive substructure search" erlaubt eine mehrdeutige und redundante Definition von Substrukturanforderungen, die von GENOA auch überlappend zu kompletten Molekülen zusammengesetzt werden können.

Einen Übergang von den eher kombinatorischen Verfahren zu einer algebraischen Vorgehensweise beschreibt A. KERBER mit seinem Strukturgenerator MOLGRAPH bzw. MOLGEN^{159, 160, 161}. MOLGEN wurde im Jahr 1993 mit dem Deutsch-Österreichischen Hochschul-Softwarepreis ausgezeichnet¹⁶². Die spezifischen Eigenheiten der Programmiersprache PROLOG nutzt H.J. LUNGE im Algorithmus AEGIS^{163, 164}, das Kernstück des Expertensystems EXSPEC¹⁶⁵ für die wissensbasierte Interpretation von IR-Spektren bildet.

Obwohl für alle strukturgenerierenden Verfahren Vollständigkeit und Redundanzfreiheit gefordert ist, zeigt ein Vergleich der von verschiedenen Programmen erzeugten Isomerenzahlen schon bei kleineren Summenformeln deutliche Divergenzen. Tab. 3.1 zeigt eine Gegenüberstellung der Resultate mehrerer Programme¹⁶⁶ (auch die des im folgenden beschriebenen, im Programm RAIN implementierten Verfahrens) für einige einfachere Summenformeln. In allen Fällen wurden keine weiteren Substrukturanforderungen gestellt, sodaß diese Zahlen jeweils die Gesamtheit aller formal möglichen Konstitutionsisomeren ausdrücken sollten. Abweichungen in der erzeugten Isomerenzahl deuten demnach auf Probleme bzw. Implementierungsfehler der zugrundeliegenden Algorithmen hin.

	MOLGRAPH ¹⁵⁹ AEGIS ^{163, 164} RAIN ¹⁶⁷ GENMAS ¹⁶⁶	ASSEMBLE ¹⁵⁵	CHEMICS ^{151, 152}	DENDRAL ¹⁴²	MASS ¹⁵⁷
C₂H₅NO₂	84	84	-	84	86
C₄H₇NO	764	764	802	764	762
C₃H₄BrCl	10	-	8	-	-
C₅H₈BrCl	140	-	108	-	-
C₆H₁₀BrCl	477	-	367	-	-
C₆H₁₀O	747	-	745	747	-

Tab. 3.1 Anzahl der generierten, redundanzfreien Konstitutionsisomere einiger Strukturgenerierungsalgorithmen an ausgewählten Beispielen.

3.2 Der Reaktionsgenerator als Strukturgenerator

Der im Abschnitt 2.2 beschriebene Reaktionsgenerator arbeitet nach dem Prinzip der vollständigen Enumeration aller formal möglichen Valenzelektronenverschiebungen in einem Ausgangsensemble. Für den Fall der Reaktionsgenerierung werden die erlaubten Valenzelektronenverschiebungen durch eine Reihe von formalen und heuristischen Randbedingungen eingeschränkt, was sich sowohl auf die Effizienz des generierenden Verfahrens, als auch auf die Menge und "chemische Relevanz" der erhaltenen Ergebnisse auswirkt.

Werden nun alle Kontrollmechanismen zur Begrenzung der erlaubten Valenzelektronenverschiebungen außer Kraft gesetzt, so produziert der Reaktionsgenerator zu einem gegebenen Ausgangsensemble die Menge aller hierzu isomerer Ensembles, d.h. die Familie isomerer Ensembles FIEM⁴⁶. Der selektive Einsatz von Randbedingungen, die nicht die Valenzelektronenverschiebungen betreffen (z.B. Ringsysteme, Substrukturlisten etc.) erlaubt die Generierung von Isomeren spezieller Verbindungsklassen als Teilmengen der FIEM¹⁶⁷.

Als Ausgangsensemblen für die Strukturgenerierung kommen komplette Moleküle (d.h. ein ausgewählter Vertreter der Isomerenfamilie), oder Fragmente mit ungepaarten Elektronen, die im Zuge der (Reaktions-)Generierung zusammengesetzt werden, in Frage. Insbesondere die Möglichkeit der Fragmentassemblierung erlaubt es, eine Strukturgenerierung unter Maßgabe von Vorausinformationen, z.B. aus spektroskopischen Daten vorzunehmen. Zu diesem Zweck müssen in den Valenzschemaübergangsmatrizen spezielle Valenzschemata von Pseudoradikalen aufgenommen werden, die ausschließlich in abgesättigte Valenzschemata übergehen dürfen. So kann zu diesem Zweck für ein spezielles C-Atom z.B. folgende Übergangsmatrix definiert werden :

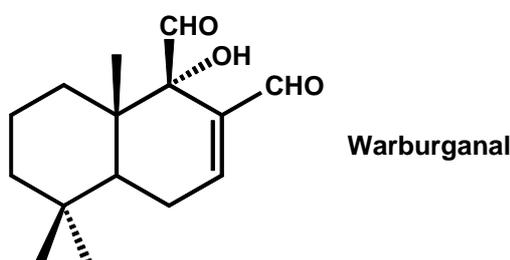
C	$-\dot{\text{C}}\cdot$	$-\overset{ }{\underset{ }{\text{C}}}-$	$-\overset{ }{\underset{ }{\text{C}}}=\overset{ }{\text{C}}$	$-\text{C}\equiv$
$-\dot{\text{C}}\cdot$	○	●	●	●
$-\overset{ }{\underset{ }{\text{C}}}-$	○	●	●	●
$-\overset{ }{\underset{ }{\text{C}}}=\overset{ }{\text{C}}$	○	●	●	●
$-\text{C}\equiv$	○	●	●	●

Abb. 3.2 Spezielle Valenzschema-Übergangsmatrix zur Strukturgenerierung. Die pseudoradikalischen Valenzzustände erlauben die variable Fragmentassemblierung.

3.2.1 Warburganal : ein illustratives Beispiel

Zur Illustration des Vorgehens bei der Strukturgenerierung dient ein Beispiel, das von N.A.B GRAY ¹⁶⁸ zur Demonstration der Leistungsfähigkeit des Strukturgenerierungsprogramms GENOA ¹⁵⁸ benutzt wurde. Da die Beweisbarkeit, und damit die uneingeschränkte Korrektheit von komplexen Algorithmen, wie der des beschriebenen Reaktionsgenerators, noch nicht gezeigt werden konnte, existiert als einzige Möglichkeit zur Überprüfung des Verfahrens der Vergleich mit Ergebnissen anderer Programme. Stimmen die Resultate überein, und sind die Verfahren hinreichend verschieden, so kann mit großer Wahrscheinlichkeit von der Korrektheit der Algorithmen ausgegangen werden.

Das gewählte Beispiel beschreibt die Generierung einer Liste von Strukturkandidaten für Warburganal, das als fraßhemmender Wirkstoff aus der Baumrinde von *Warburgia Ugandensis* isoliert werden kann ¹⁶⁹.



Aus einem hochaufgelösten Massenspektrum wurde die Summenformel von Warburganal zu $C_{15}H_{22}O_3$ ermittelt. Als weitere Informationsquellen für die Strukturaufklärung stehen das 1H - und das ^{13}C -NMR Spektrum zur Verfügung. Durch Analyse der Spektren ¹⁶⁸ konnte eine Liste von Substrukturanforderungen erstellt werden, die in Tabelle 3.2 gezeigt ist.

Substruktur	Anzahl	Substruktur	Anzahl
SUB1 $\begin{array}{c} \\ -\text{C}-\text{OH} \\ \end{array}$	mind. 1	SUB7 $-\text{CH}_3$	3
SUB2 $\begin{array}{c} \\ -\text{C}-\text{CH}=\text{O} \\ \end{array}$	mind. 1	SUB8 $\begin{array}{c} \\ \text{HO}-\text{C}-\text{CH}_3 \\ \end{array}$	kein
SUB3 $-\text{CH}=\overset{ }{\text{C}}-\text{CH}=\text{O}$	mind. 1	SUB9 $\text{C}=\text{CH}-\text{CH}_2-\overset{\text{C}}{\underset{\text{C}}{\text{CH}}}$	mind. 1
SUB4 $\text{C} (sp^2)$	4	SUB10 $-\text{CH}_2-$	4
SUB5 $\begin{array}{c} \\ -\text{C}-\text{CH}_3 \\ \end{array}$	mind. 3	SUB11 $-\text{CH}_2\text{CH}_2\text{CH}_2-$	mind. 1
SUB6 $\begin{array}{c} \text{CH}_3 \\ \\ -\text{C}-\text{CH}_3 \\ \\ \text{CH}_3 \end{array}$	kein	SUB12 $\begin{array}{c} \text{C} \\ \diagup \quad \diagdown \\ \text{C}-\text{C} \end{array}$	kein

Tab. 3.2 Substrukturanforderungen an Warburganal. Grundlage hierfür ist eine Analyse der ^1H - und ^{13}C -NMR-Spektren.

Die Umsetzung dieser Substrukturtafel in nichtüberlappende Bausteine und weitere Randbedingungen für den beschriebenen Strukturgenerator geschieht nach Tabelle 3.3.

Ergebnis :

Trotz der wenigen Strukturelemente und der stark eingeschränkten Kombinationsmöglichkeiten in Tab. 3.3 werden 42 gültige Strukturen vom Generator erzeugt. (siehe Abbildung 3.3). Diese decken sich vollkommen mit den von GENOA erzeugten Molekülen. Zu beachten ist hierbei, daß sowohl GENOA, als auch das hier beschriebene Verfahren von RAIN keine stereochemische Informationen berücksichtigen. Die in Abbildung 3.3 angezeigten Strukturen verstehen sich also lediglich als Vorschläge für die *Konstitution* von Warburganal.

Fragment bzw. Randbedingung	Erläuterung	Valenzschemaübergänge der "reaktiven" Atome
1x $\begin{array}{c} H & H & H & \cdot\dot{C}_q \\ & & & \\ O=C & -C_d & =C & -C & -\dot{C}_q \\ & & & & \\ & & & H & H \end{array}$	einzige Kombinationsmöglichkeit von SUB3 und SUB9 bei 4 sp ² -C-Atomen	$-\dot{C}_d= \rightarrow -\overset{ }{C}_d= \quad \quad -\dot{C}_q \rightarrow -\overset{ }{C}_q-$
3x $\begin{array}{c} H \\ \\ H-C_m \\ \\ H \end{array}$	SUB7	$-\overset{ }{C}_m \rightarrow -\overset{ }{C}_m-$
1x $\begin{array}{c} H \\ \\ O=C_a \end{array}$	SUB2	$-\dot{C}_a= \rightarrow -\overset{ }{C}_a=$
1x $\begin{array}{c} H & H & H \\ & & \\ \cdot\dot{C}_c & -C & -C_c \\ & & \\ H & H & H \end{array}$	kommt nur einmal vor, da sonst Anzahl C-Atome > 15	
1x $\cdot O-H$	kommt nur einmal vor, da sonst Anzahl O-Atome > 3	$-\dot{O} \rightarrow -O-$
1x $\cdot\dot{C}$	Anzahl C-Atome = 15. Ist sp ³ -C-Atom, da bereits 4 sp ² -C-Atome definiert sind.	$\cdot\dot{C} \rightarrow -\overset{ }{C}-$
keine Bindungen zwischen: $C_m-C_m \quad C_m-C_d \quad C_m-C_c$	Me (SUB7) ist gebunden an quartären C-Atomen (SUB5)	
keine Bindungen zwischen: $C_a-C_m \quad C_a-C_d \quad C_a-C_c$	CHO (SUB2 und SUB3) ist gebunden an quartären C-Atomen (SUB5)	
keine Bindungen zwischen: $O-C_m \quad O-C_d \quad O-C_c \quad O-C_a$	OH ist gebunden an tertiärem C-Atom (SUB1)	
minimale Ringgröße = 4	SUB12	
verbotene Substrukturen: $H_3C-C-OH \quad H_3C-\overset{ }{C}-CH_3$ $\quad \quad \quad $ $\quad \quad \quad CH_3$	SUB6 und SUB8	

Tab. 3.3 Nichtüberlappende Fragmente und weitere Randbedingungen für den Strukturgenerator zur Ausführung der Anforderungen gemäß Tab. 3.2. Kursiv gesetzte Atomsymbole bezeichnen Atome mit unveränderbaren Bindungen.

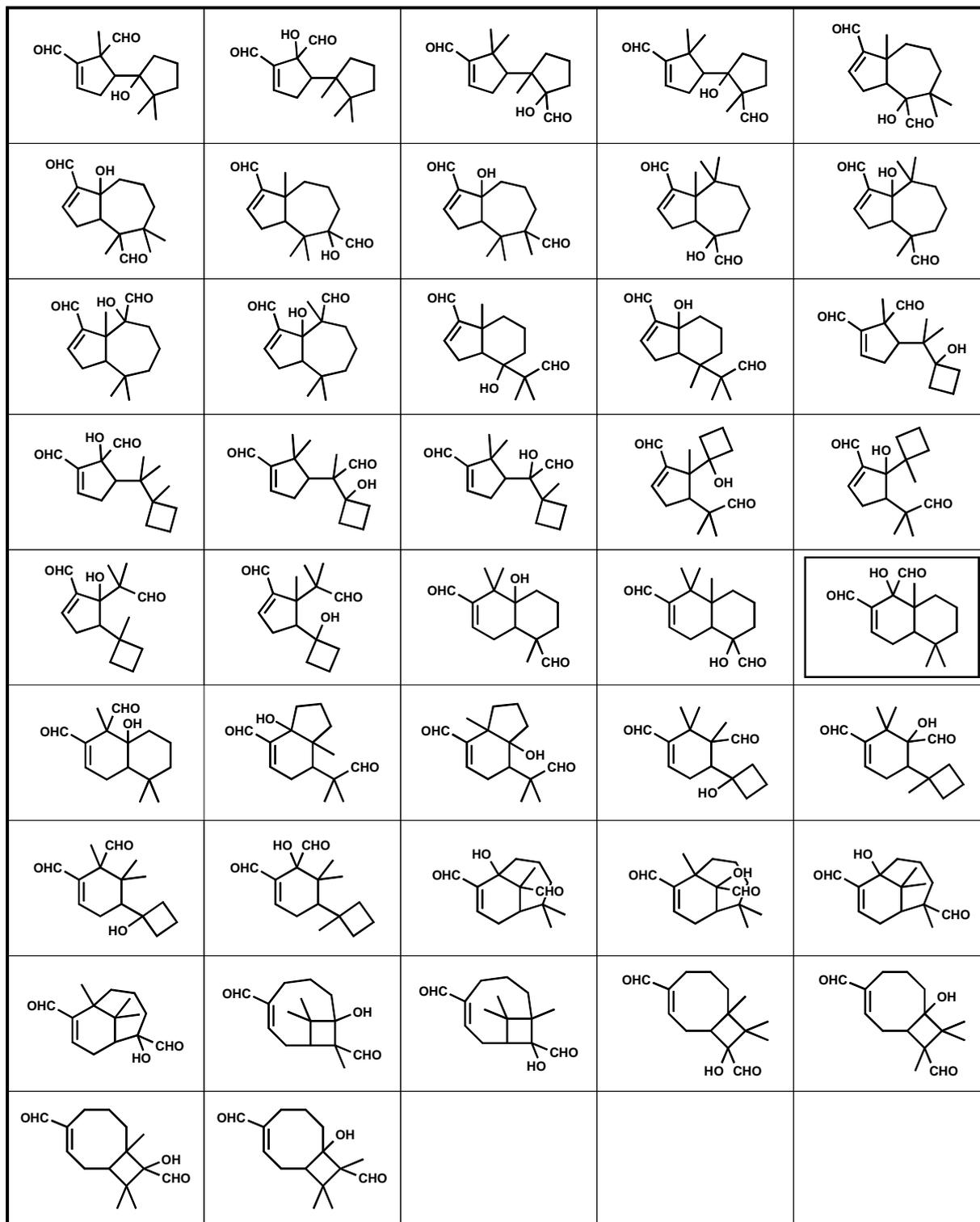


Abb. 3.3 Vom Strukturgenerator erzeugte Vorschläge für die Konstitution von Warburganal ($C_{15}H_{22}O_3$). Alle Strukturen entsprechen den Anforderungen von Tabelle 3.2 und Tabelle 3.3.

Die Summenformel von Warburganal und die Herkunft der Substanz legen nahe, daß es sich hierbei um eine Verbindung aus der Naturstoffklasse der Sesquiterpene handelt. Es besteht daher die Möglichkeit, die erzeugten Strukturen weiter einzuschränken. Hierzu wird das Substruktursuchverfahren verwendet, wie es in Abschnitt 2.2.7 beschrieben wurde. Dieses Verfahren erlaubt es, eine Liste von Substrukturen als geforderte Strukturmerkmale zu definieren. Die Teilstrukturen erhalten das Prädikat (+), d.h. es genügt, wenn eine erzeugte Struktur mindestens eine der geforderten Substrukturen enthält. Abbildung 3.5 zeigt die verwendeten Substrukturen. Es handelt sich hierbei um 49 Grundgerüste bzw. Gerüstfragmente von natürlich vorkommenden Sesquiterpenen¹⁷⁰.

Wird die Strukturgenerierung unter Kontrolle dieser Fragmentliste durchgeführt, so verbleiben nurmehr 10 erzeugte Strukturvorschläge für Warburganal (Abbildung 3.4), davon 8 Strukturen mit einem Azulengrundkörper und 2 (darunter das Warburganal) mit einer Drimanstruktur als Basisgerüst.

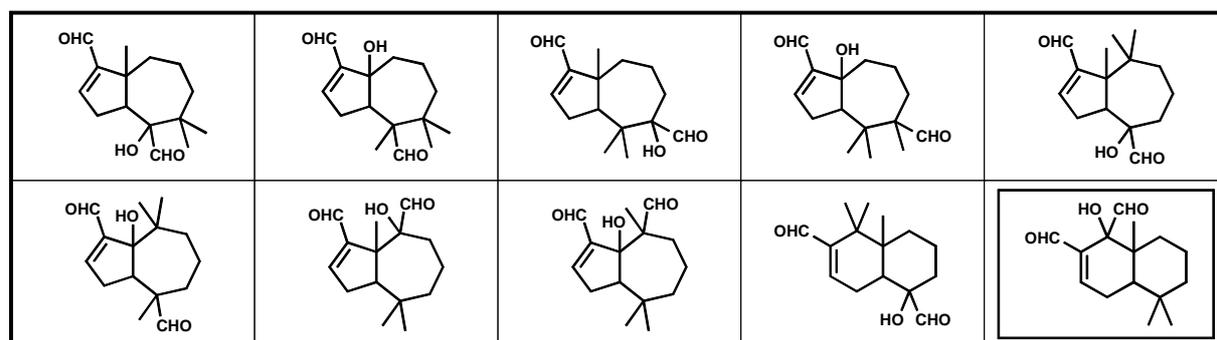


Abb. 3.4 Strukturkandidaten für Warburganal mit natürliche vorkommenden Sesquiterpengerüsten.

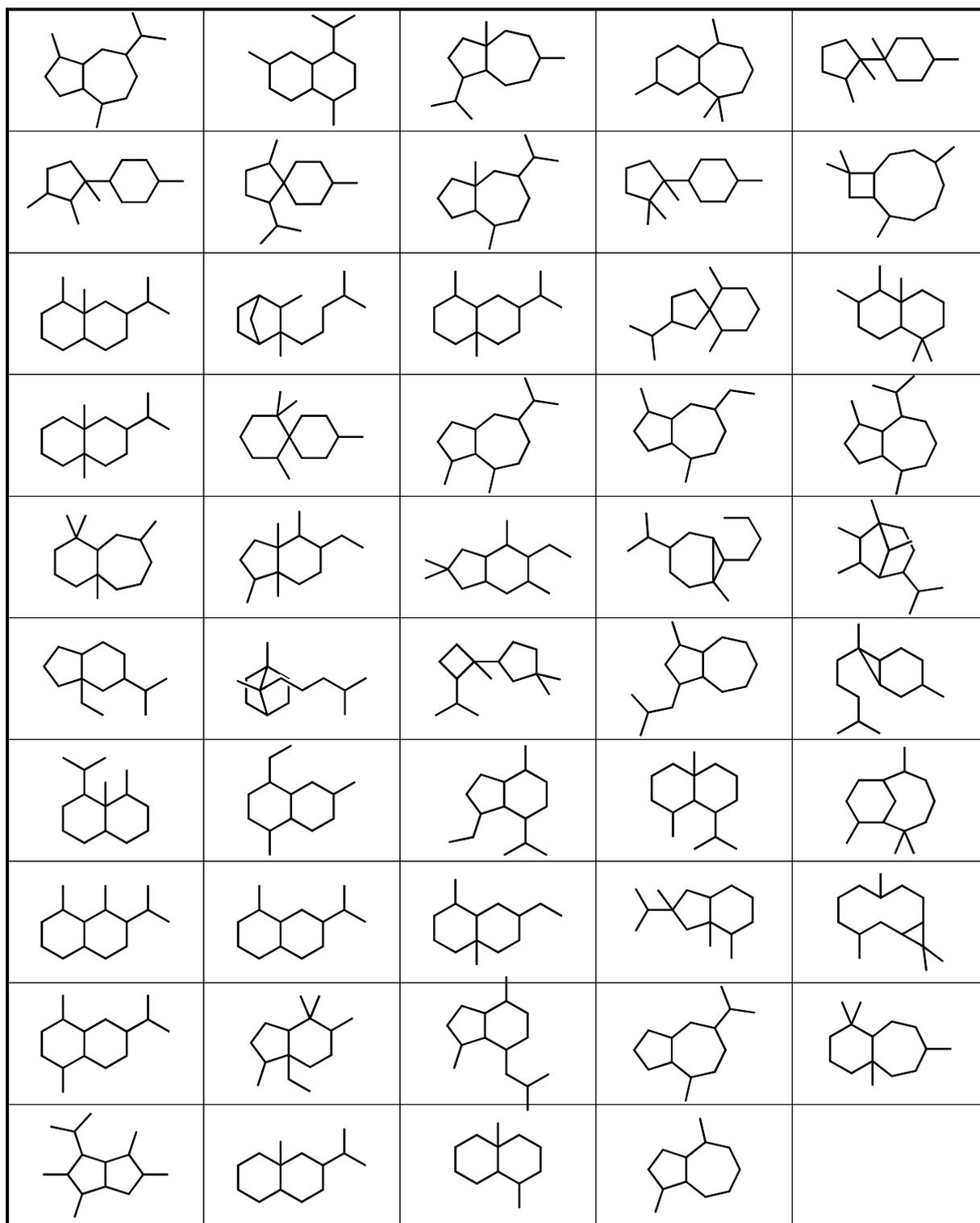


Abb. 3.5 Liste der natürlich vorkommenden Sesquiterpengerüste und -gerüstfragmente, die zur weiteren Selektion von relevanten Strukturkandidaten herangezogen wurde.

3.2.2 Das B_6H_{14} -Problem

Sechzig Jahre nach den grundlegenden Arbeiten über die Chemie der Borane von A. STOCK¹⁷¹ fanden die theoretischen Studien an dieser interessanten Verbindungs-klasse in den 70er Jahren einen ersten Höhepunkt, der hauptsächlich mit dem Namen W.N. LIPSCOMB¹⁷² verbunden wird. Die völlig unerwarteten Strukturprinzipien der Borane erforderten eine beträchtliche Erweiterung der MO-Theorien und werfen auch heute noch viele ungelöste Fragen auf.

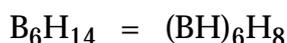
Die neutralen Borane B_nH_m und die Borananionen $B_nH_m^{x-}$ lassen sich heute in fünf Hauptklassen unterteilen. Den durch die WADE'schen Regeln^{173, 174} definierten *closo*-, *nido*- und *arachno*-Strukturen wurden die *hypho*- und *conjuncto*-Strukturen hinzugefügt. Die *hypho*-Borane^{175, 176} besitzen weit geöffnete Clusterstrukturen mit n Boratomen in den Ecken eines $(n+3)$ eckigen Polyeders.

Ein Vertreter dieser Klasse ist B_6H_{14} , das 1988 durch Dimerisierung von *in situ* erzeugtem B_3H_7 synthetisiert und durch sein ^{11}B -NMR-Spektrum charakterisiert wurde¹⁷⁷. Die wahrscheinlich außergewöhnliche Struktur von B_6H_{14} ist Gegenstand zahlreicher theoretischer und experimenteller Studien^{178, 179, 180, 181}. P.V.R. SCHLEYER kommt in einem Bericht über *ab initio* Rechnungen und NMR-Simulationen zu dem Schluß, daß, trotz aller Bemühungen, die Natur dieser Spezies noch ungeklärt ist¹⁸¹:

" ... Since none of the isomers considered reproduces the experimental chemical shifts, the actual nature of this compound is still in doubt ... "

Für weitere theoretische Betrachtungen ist es sicher sinnvoll, einen kompletten Katalog aller formal möglichen B_6H_{14} -Isomeren zur Verfügung zu haben. Der in Abschnitt 2.2 beschriebene Strukturgenerator ist in der Lage diesen Katalog zu erarbeiten¹⁸². Da jedoch die Borane im Hinblick auf ihre elektronische Struktur eine Besonderheit darstellen, und der Generator nur ganzzahlige Bindungsordnungen bearbeiten kann, muß ein besonderer Weg eingeschlagen werden:

Es wird angenommen, daß die Konstitutionsisomere von B_6H_{14} ausnahmslos aus BH- bzw. BH_2 -Einheiten konstruiert werden können, entsprechend:



Die in der Literatur bisher in Betracht gezogenen Konstitutionsisomere enthalten nahezu alle an jedem Boratom mindestens eine "gewöhnliche" Zweizentren-zweielek-

tronen-Bindung zu einem Wasserstoffatom. Die einzige Ausnahme hierzu ist die Struktur 2f in Lit. 181.

Die B_6H_{14} -Konstitutionsisomere werden in dieser Studie aus folgenden Bausteinen zusammengesetzt (Abbildung 3.6) :

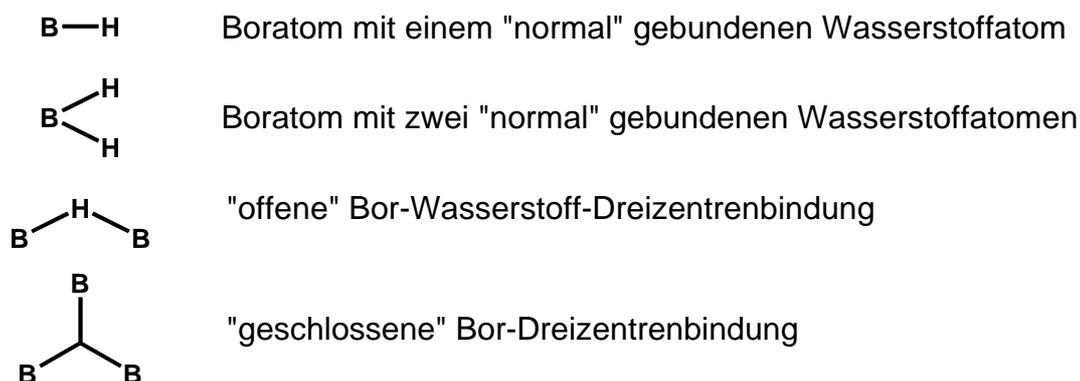


Abb. 3.6 Grundbausteine, die für die Strukturgenerierung der B_6H_{14} -Konstitutionsisomeren verwendet wurden.

Jede BH-Einheit ist in der Lage, zwei Elektronen für Gerüstbindungen beizutragen. Ein Wasserstoffatom liefert naturgemäß ein Elektron, das entweder in einem Zweizentren-zweielektronen-MO oder in einem Dreizentren-zweielektronen-MO eingesetzt wird. Um der Ganzzahligkeit der Bindungsordnungen im Strukturgenerator Rechnung zu tragen, werden alle Valenzelektronenzahlen und die Bindungsordnungen der BH- bzw. H-(Pseudo)-Atome formal mit dem Faktor zwei multipliziert. Zusätzlich werden "Zentralatome" definiert, die die "Mittelpunkte" von geschlossenen Bor-Dreizentrenbindungen darstellen. Die Bausteine und ihre gültigen "Valenzzustände" sind also die folgenden (Abbildung 3.7) :

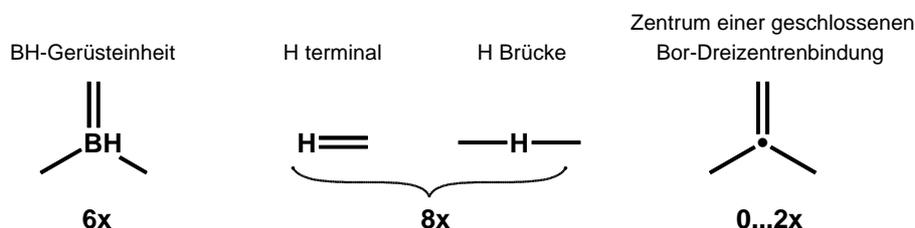


Abb. 3.7 Valenzzustände und Multiplizität der Bor-, bzw. H-Atome, sowie der Pseudozentren der geschlossenen Dreizentrenbindungen. Alle Bindungsordnungen wurden formal mit dem Faktor zwei multipliziert.

Eine auf diese Weise generierte B_6H_{14} -Struktur würde folgendermaßen aussehen (Abbildung 3.8) :

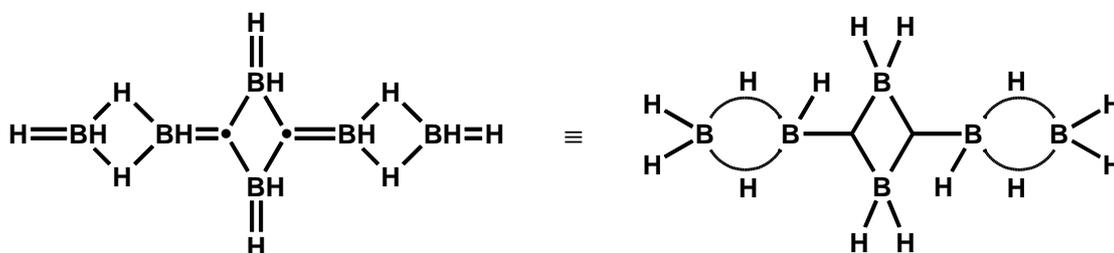


Abb. 3.8 Aus den Bausteinen in Abb. 3.7 erzeugte B_6H_{14} -Struktur. Im rechten Teil ist die Struktur in der üblichen halbtologischen Darstellung gezeigt.

Unter den gegebenen Voraussetzungen erzeugt der Generator insgesamt 130 Konstitutionsisomere von B_6H_{14} . Es sind keine Strukturen möglich, die mehr als zwei geschlossene Bor-Dreizentrenbindungen haben. Die Boranstrukturen besitzen jedoch im allgemeinen mindestens ein höheres Symmetrieelement. In seinem vielbeachteten Buch über Borane legt LIPSCOMB sieben Regeln für die Stabilität von Boranstrukturen fest. Die wichtigste Regel lautet:

" ... all known boron hydrides have at least a twofold element of symmetry ... "

Da Symmetrieeigenschaften *a priori* nicht zu den Einschränkungsmöglichkeiten des Strukturgenerators gehören, wird ein zusätzliches Filter definiert, das es erlaubt, Konstitutionsisomere ohne zweizählige Symmetrie auszuschließen. Hierzu dienen die Atom-Äquivalenzklassen, die bei der eindeutigen Numerierung durch CANON⁸⁷ erzeugt werden. Gehören zwei Atome in einem Molekül derselben Äquivalenzklasse an, so sind diese beiden Atome konstitutionell äquivalent. Für die Erzeugung gültiger Boran-Strukturen mit zweizähliger Konstitutionssymmetrie gilt nun folgende Regel :

Gibt es für eine Äquivalenzklasse in einem Molekül nur *ein* einzelnes Atom, so dürfen die Nachbarn dieses Atoms maximal *zwei* verschiedenen Äquivalenzklassen angehören.

Nach Anwendung dieser zusätzlichen Regel verbleiben von den 130 Isomeren noch 48 Strukturen, die in Abbildung 3.9 aufgelistet sind.

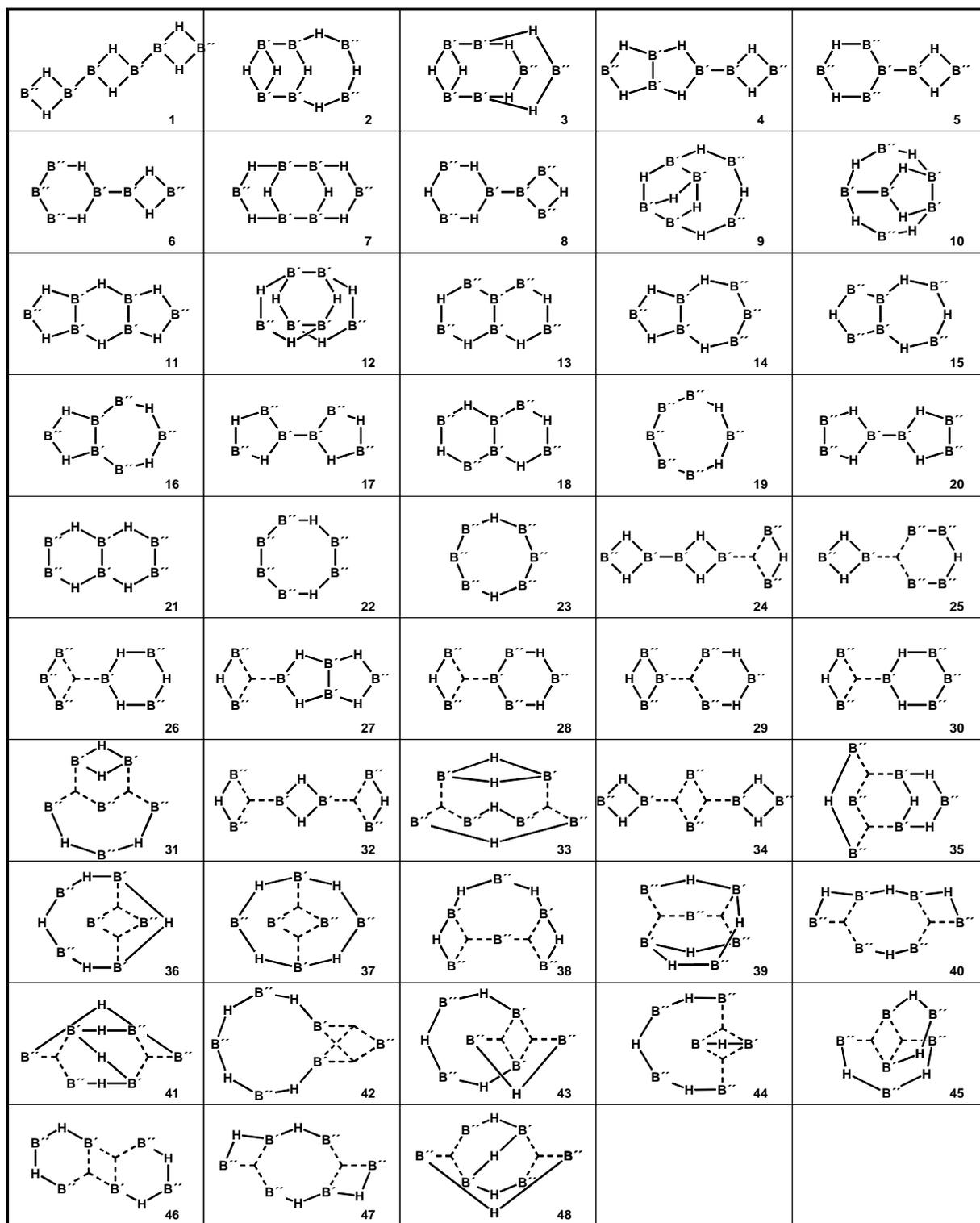


Abb. 3.9 Liste der vom Strukturgenerator erzeugten 48 B_6H_{14} -Konstitutionsisomeren mit mindestens zweizähliger Konstitutionssymmetrie ($B' = BH$, $B'' = BH_2$).

Insgesamt acht Strukturvorschläge wurden bisher in der einschlägigen Literatur in Betracht gezogen. Der älteste Vorschlag stammt von W.N. LIPSCOMB¹⁷². Aufgrund der kombinatorischen Vollständigkeit der erzeugten Lösung sind alle diese Strukturisomere in der Menge der generierten Konstitutionsisomeren enthalten. Tabelle 3.4 gibt einen Überblick über diese Strukturen mit ihrer entsprechenden Referenz in den Artikeln, die sich mit der B_6H_{14} -Problematik auseinandersetzen.

Nr.	Nr. bzw. styx-Notation der entsprechenden Struktur in Lit.:					
	Ref. 172	Ref. 177	Ref. 178	Ref. 179	Ref. 180	Ref. 181
1	-	-	2a	1	1a	2a
11	-	6022	2b	2	1b	2b/2d
24	-	-	2c	-	1c	-
32	-	-	2d	3	1d	2c
34	-	-	2e	-	1e	-
38	-	-	-	-	-	2e
46	4204	-	-	-	-	2g
47	-	4204	-	-	-	-

Tab. 3.4 Bereits in der Literatur behandelte B_6H_{14} -Isomere (siehe Abbildung 3.9). Die angegebenen Nummern bzw. Bezeichnungen entsprechen den in den zugehörigen Originalpublikationen verwendeten Strukturidentifikatoren.

Es verbleiben demnach noch 40 Konstitutionsisomere von B_6H_{14} , die für weitere theoretische oder experimentelle Studien in Betracht gezogen werden sollten. Die Situation wird allerdings dadurch kompliziert, daß es sich bei B_6H_{14} möglicherweise um eine stark fluktuierende Struktur handelt, da viele Isomere von B_6H_{14} auf der Potentialhyperfläche Minima mit sehr ähnlicher Energie besetzen¹⁸¹.

Kapitel 4

Chemische Metrik

Zusammenfassung:

Jedes Molekül einer Reaktionssequenz bildet ein Punkteraster in einem vieldimensionalen Hyperraum (BE-Raum). Die jeweils kürzesten Abstände zwischen diesen Punkterastern verschiedener Moleküle definieren ein Maß für die konstitutionelle Ähnlichkeit der zugehörigen Strukturen. Nach einer Besprechung verschiedener Methoden zur Bestimmung dieser sogenannten "chemischen Distanz" werden die Einsatzmöglichkeiten dieser Metrik bei der zielgerichteten Generierung von Reaktionsbäumen besprochen. Für die Methode der monoton fallenden chemischen Distanz, sowie für die Methode des ellipsoidalen Einschlusses werden illustrative Anwendungsbeispiele gegeben.

4.1 Der BE-Raum

Eine der wichtigsten Konsequenzen des mathematischen Modells der konstitutionellen Chemie von I. UGI und J. DUGUNDJI ist der direkte Zugang zu einer geometrischen Interpretation von chemischen Strukturen *und Reaktionen*.

Die geometrische Interpretation molekularer Konstitution und die hierin definierte chemische Metrik bilden die grundlegende Basis von Klassifizierungssystemen und Netzwerkgenerierungsprogrammen. Diese Metrik erlaubt die Definition von konstitutioneller Ähnlichkeit, sowie die zielgesteuerte Propagation von Reaktionsbäumen bei der computerunterstützten Aufklärung von Reaktionsmechanismen.

Grundlage der geometrischen Interpretation der konstitutionellen Chemie ist die Einbettung von BE-Matrizen in einen hochdimensionalen Hyperraum. Eine $n \times n$ Matrix M kann als Raumvektor mit n^2 Komponenten angesehen werden :

$$m = (m_{11}, m_{12}, \dots, m_{1n}, m_{21}, \dots, m_{n1}, \dots, m_{nn})$$

Die Symmetrie der BE-Matrizen bezüglich ihrer Hauptdiagonalen ($BE_{ij} = BE_{ji}$) reduziert die Anzahl der nötigen Raumdimensionen von n^2 auf $n(n+1)/2$. Jede BE-Matrix läßt sich demgemäß als ein Punkt (BE-Punkt) in einem $n(n+1)/2$ -dimensionalen Hyperraum (BE-Raum) darstellen :

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix} \Rightarrow b_{BE} = (b_{11}, b_{12}, \dots, b_{1n}, b_{22}, \dots, b_{2n}, \dots, b_{nn})$$

In besonderen Fällen kann der für die synthetische Chemie wichtigere Bindungsaspekt stärker betont werden, indem unter Verzicht auf die freien Valenzelektronen der Hauptdiagonalen eine weitere Reduktion auf $n(n-1)/2$ Dimensionen im sogenannten Bindungsraum (B-Raum) erfolgt :

$$B \Rightarrow b_B = (b_{12}, b_{13}, \dots, b_{1n}, b_{23}, \dots, b_{2n}, \dots, b_{(n-1)n})$$

Die BE-Punkte sind in einen euklidischen Hyperraum eingebettet und bilden darin ein Punktraster. Der Abstand zweier Punkte BE_1 und BE_2 ist festgelegt durch die L_1 -Distanz (entsprechend der sogenannten "Taxigeometrie"), die sich ergibt nach :

$$d (BE_1 , BE_2) = \sum_{i,j} | be_{1ij} - be_{2ij} |$$

Diese L_1 -Distanz unterscheidet sich (außer für $BE_1 = BE_2$) von der L_2 -Distanz (euklidische Distanz) :

$$D (BE_1 , BE_2) = \sqrt{ \sum_{i,j} (be_{1ij} - be_{2ij})^2 }$$

Hierbei gilt $d (BE_1, BE_2) > D (BE_1, BE_2)$. Abbildung 4.1 erläutert diese Distanzmetrik für den einfachen zweidimensionalen Fall.

Die L_1 -Distanz zweier BE-Punkte $d (BE_1, BE_2)$ nennt man "chemische Distanz" ⁴⁶. Der durch die BE-Matrizen definierte Hyperraum bildet einen metrischen Raum, da die chemische Distanz alle Bedingungen einer Metrik erfüllt ^{183, 184}, das heißt, es gilt für alle BE_n :

- | | | |
|----|------------------------|---|
| a) | Nichtnegativität : | $d (BE_1, BE_2) \geq 0$ |
| b) | Symmetrie : | $d (BE_1, BE_2) = d (BE_2, BE_1)$ |
| c) | Dreiecksungleichheit : | $d (BE_1, BE_2) \leq d (BE_1, BE_3) + d (BE_3, BE_2)$ |

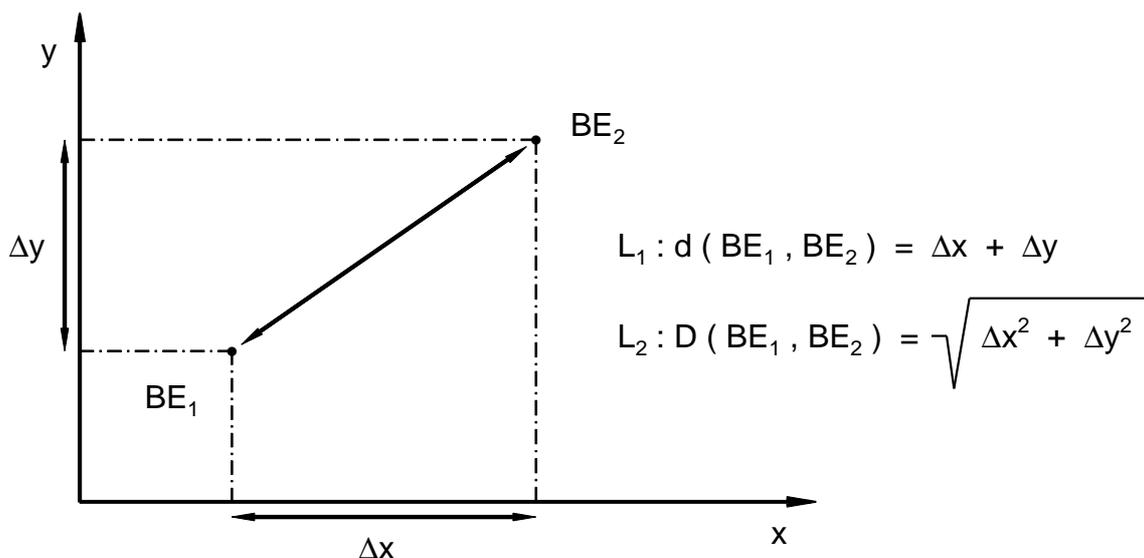


Abb. 4.1 Abstand zweier Punkte nach L_1 -Metrik (Taxigeometrie) und nach L_2 -Metrik (Euklidische Distanz).

Die BE-Punkte einer Familie isomerer Ensembles (FIEM) mit n Atomen und k Valenzelektronen liegen dabei alle auf der Oberfläche einer L_1 -Kugel mit dem Radius k im positiven "Hyperoktanten" des BE-Raums. Dies folgt aus den grundlegenden Eigenschaften aller Ensembles innerhalb einer FIEM :

$$k = \sum_{ij} be_{ij} = \text{const}$$

$$\text{und } b_{ij} > 0$$

Die n Atome eines Ensembles von Molekülen können nun beliebig numeriert sein, d.h. sie können den Spalten und Zeilen einer BE-Matrix auf beliebige Weise zugeordnet werden. Dies führt dazu, daß ein und dasselbe Ensemble im BE-Raum durch bis zu $n!$ verschiedene BE-Punkte repräsentiert ist. Bedingt durch interne Symmetrie, können allerdings nun mehrere dieser Punkte zusammenfallen (Entartung), sodaß die wirkliche Punktmenge weniger mächtig ist.

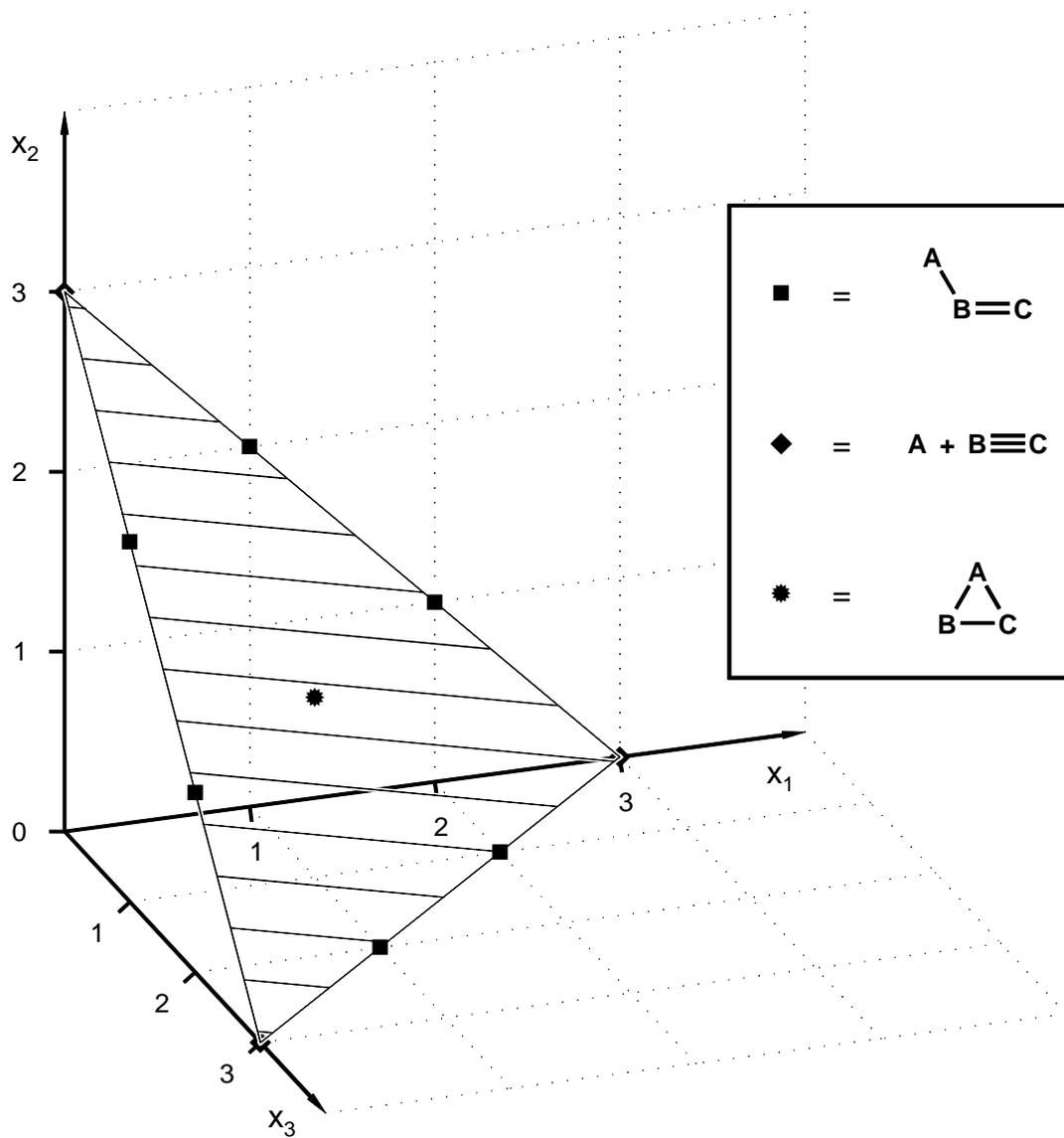


Abb. 4.2 B-Raumdarstellung für die drei möglichen Bindungen eines dreiatomigen Ensembles (Atome A, B und C). Die zum Teil multiplen Abbildungspunkte der drei Spezies resultieren aus den unterschiedlichen Numerierungsmöglichkeiten der einzelnen Atome, wobei jedesmal andere Bindungen auf die Koordinatenachsen x_1 , x_2 und x_3 zugeordnet werden. Die Punkte bilden ein reguläres Gitter auf einem Ausschnitt der L_1 -Kugeloberfläche. Die Form einer dreidimensionalen L_1 -Kugel entspricht der eines Oktaeders. Die Ebene wird festgelegt durch die Anzahl der Bindungselektronen (2×3).

Die BE-Punkte eines Ensembles bilden auf der Hyperkugeloberfläche ein regelmäßiges Gitter. Abbildung 4.2 zeigt dies für ein dreiatomiges Ensemble. Um für die Anzahl der darzustellenden Dimensionen den Wert von drei nicht zu überschreiten, aber trotzdem noch eine aussagekräftige Darstellung zu erhalten, wurde für dieses Beispiel nicht der BE-Raum, sondern der B-Raum gewählt, in dem die Raumachsen für die freien Elektronen wegfallen.

Die Gesamtheit aller BE-Punkte p (BE) einer BE-Matrix kann durch Anwendung aller formal möglichen Permutationsvektoren p der Länge n auf die Spalten und Zeilen der BE-Matrix erzeugt werden.

$$p(\text{BE}) = p(\text{BE})$$

Eine alternative Möglichkeit besteht in der Anwendung einer Permutationsmatrix P .

$$p(\text{BE}) = P^t \cdot \text{BE} \cdot P \quad \text{mit} \quad P_{ij} \in \{0,1\}$$

$$\text{und} \quad \sum_i P_{ij} = 1, \quad \sum_j P_{ij} = 1$$

Wenn nun die *Ensembles* einer FIEM durch *Punktecluster* im BE-Raum beschrieben werden können, so entsprechen die *Reaktionen* zwischen einzelnen Ensembles L_1 -Vektoren, die deren BE-Punkte miteinander verbinden. Die heuristische Regel des "Prinzips der kleinsten Strukturveränderung" ¹⁸⁵ besagt, daß chemische Reaktionen in der Weise ablaufen, daß eine Minimalanzahl von Elektronen umverteilt werden müssen.

Die geometrische Interpretation dieser Regel lautet nun : die Reaktion zwischen zwei Ensembles von Molekülen **A** und **B** wird beschrieben durch einen L_1 -Vektor eines BE-Punktes von **A** zum *nächstgelegenen* (L_1 -Distanz) BE-Punkt des Punkteclusters von **B**. Die L_1 -Distanz zwischen diesen beiden BE-Punkten nennt man die Minimale Chemische Distanz (MCD) ¹⁸⁶. In Abbildung 4.3 wird deutlich gemacht, daß der nächstliegende BE-Punkt von **B** gemäß der L_1 -Metrik nicht unbedingt der nächstliegende BE-Punkt gemäß der euklidischen Distanz sein muß.

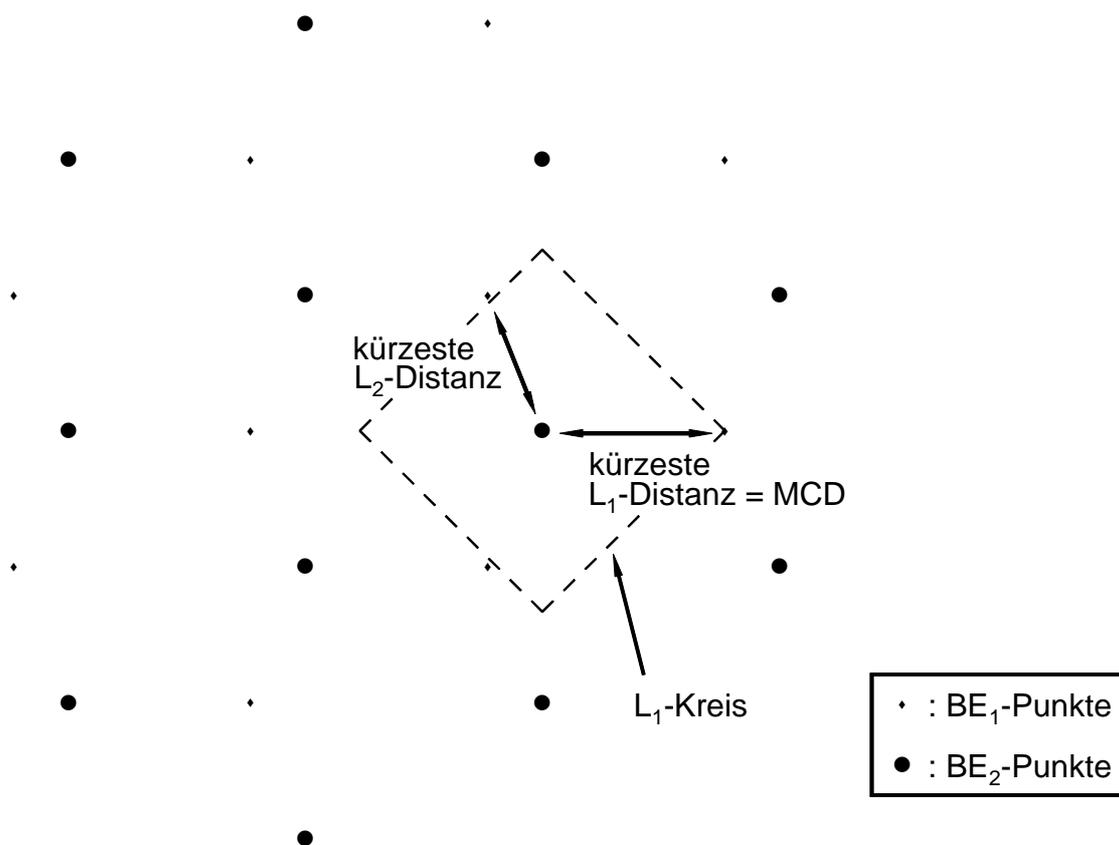


Abb. 4.3 Minimale Distanzen der Punktkluster zweier Ensembles BE₁ und BE₂ nach L₁-Metrik und nach L₂-Metrik.

Es konnte nun bewiesen werden¹⁸⁶, daß für *jeden* BE-Punkt $p \in BE_1$ mindestens ein BE-Punkt $p \in BE_2$ existiert, dessen Distanz zum Punkt von BE₁ der minimalen chemischen Distanz $MCD(BE_1, BE_2)$ entspricht. Zur Ermittlung der MCD genügt es also, von einem beliebigen BE-Punkt von BE₁ auszugehen, alle Permutationen von BE₂ zu erzeugen und die minimale L₁-Distanz der Punkte BE₁ und $p \in BE_2$ zu ermitteln.

$$MCD(BE_1, BE_2) = \min \left(\sum_{ij} |be_{1ij} - p(be_{2ij})| \right)$$

Die Berechnung der minimalen chemischen Distanz stellt somit ein quadratisches Zuordnungsproblem dar, das zur Klasse der np-vollständigen Probleme gehört. Dies bedeutet, daß keine allgemein gültigen Algorithmen zur Verfügung gestellt werden können, die das Zuordnungsproblem für beliebige Ensemblegrößen effizient lösen. Die Abschnitte 4.2 und 5.3 dieser Arbeit beschäftigen sich mit Lösungsansätzen für die Berechnung der MCD, die sich entweder der chemischen Beschränkungen des Problems bedienen oder neuartige Verfahren zur näherungsweise Lösung einführen (Genetische Algorithmen).

Wenn man auf die Berücksichtigung des Wechsels freier Elektronen verzichtet, so gibt der Wert $MCD(BE_1, BE_2)$ die doppelte Anzahl der Bindungsordnungen wider, die gebrochen bzw. neu geknüpft werden müssen, um die beiden Ensembles ineinander überzuführen. Für besondere Anwendungen in der Syntheseplanung und der zielgerichteten Entwicklung von Reaktionsbäumen werden nun spezielle Abwandlungen der MCD definiert, die sich mehr auf die topologische Aspekte, bzw. auf die Veränderungen der C-C und C-Hetero-Gerüste beziehen.

Die minimale topologische Distanz $MTD(BE_1, BE_2)$ vernachlässigt die unterschiedlichen Bindungsordnungen, indem die minimale L_1 -Distanz der Abbildungspunkte der Adjazenzmatrizen $ad(BE_1)$ und $ad(BE_2)$ ermittelt wird :

$$MTD(BE_1, BE_2) = \min \left(\sum_{i,j} |ad(be_{1ij}) - \rho(ad(be_{2ij}))| \right)$$

Der Schwerpunkt der Betrachtungen liegt hier also bei den eigentlichen Veränderungen der Bindungsgerüste hinsichtlich neu geknüpfter Bindungen bzw. bei Bindungen, die komplett gebrochen werden.

Oftmals liegt das Augenmerk der synthetischen und der mechanistischen Untersuchungen auf den Veränderungen im Skelett der Schweratome (nicht-H-Atome). Viele Programme in der Computerchemie vereinfachen die Strukturbehandlungen sogar so weit, daß sie Wasserstoffatome nicht explizit, sondern implizit, als mehr oder weniger relevante Atomdeskriptoren von C-Atome und Heteroatomen protokollieren. Um diesem Aspekt ebenfalls gerecht werden zu können, werden zusätzlich die minimale chemische Distanz und die minimale topologische Distanz der Schweratome MCD_{nH} und MTD_{nH} eingeführt :

$$\left. \begin{aligned} MCD_{nH} (BE_1 , BE_2) &= \min \left(\sum_{ij} | be_{1ij} - \rho (be_{2ij}) | \right) \\ MTD_{nH} (BE_1 , BE_2) &= \min \left(\sum_{ij} | ad(be_{1ij}) - \rho (ad(be_{2ij})) | \right) \end{aligned} \right\} \begin{array}{l} \text{Atom } i \neq H \\ \text{Atom } j \neq H \end{array}$$

Abbildung 4.4 gibt eine Übersicht über die verschiedenen Distanzwerte für einige ausgewählte Beispiele. Hierbei wird deutlich, wie die einzelnen Werte für MCD, MTD, MCD_{nH} und MTD_{nH} die konstitutionelle Ähnlichkeit der Ensemblepaare, bzw. die mögliche Komplexität einer die Ensemble verbindenden Reaktionssequenz widerspiegeln. Kleine Werte sprechen für große konstitutionelle Ähnlichkeit, bzw. kurze Reaktionssequenzen, große Werte für konstitutionelle Unähnlichkeit bzw. lange und komplexere Reaktionssequenzen.

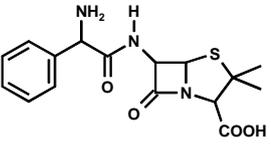
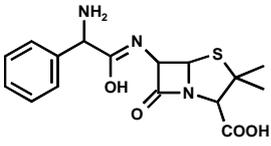
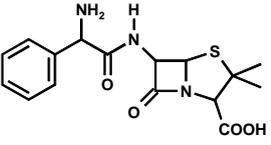
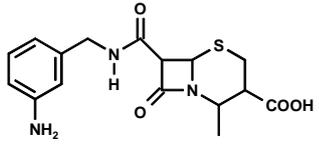
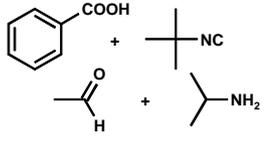
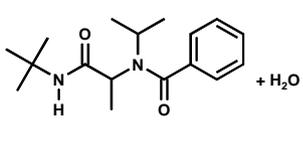
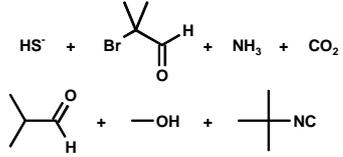
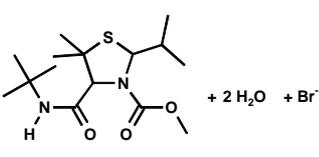
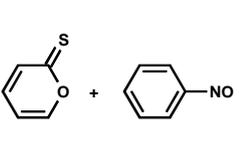
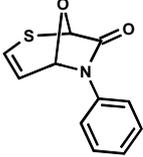
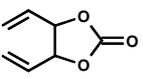
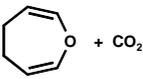
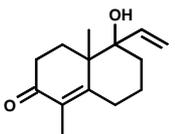
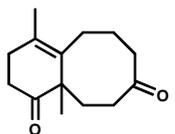
Ensemble 1	Ensemble 2	Distanzwerte			
		MCD	MTD	MCD _{nH}	MTD _{nH}
<p>Ampicillin</p> 	<p>Ampicillin Tautomer</p> 	8	4	4	0
<p>Ampicillin</p> 	<p>Ampicillin Isomer</p> 	32	32	24	24
<p>4CC Edukte</p> 	<p>4CC Produkte</p> 	30	20	18	12
<p>7CC Edukte</p> 	<p>7CC Produkte</p> 	60	40	44	24
<p>Streith-Reaktion Edukte</p> 	<p>Streith-Reaktion Produkte</p> 	28	16	24	12
<p>sym-Divinylethylencarbonat</p> 	<p>4,5-Dihydrooxepin</p> 	20	10	12	10
<p>Vinylcarbinol</p> 	<p>1,5-Dion</p> 	16	8	12	4

Abb. 4.4 Minimale chemische und topologischen Distanzen für ausgewählte Ensemblepaare. Die Werte wurden mit Hilfe eines Genetischen Algorithmus (siehe Abschnitt 5.3) ermittelt.

Die Behandlung von chemischen Reaktionen als formale Aneinanderreihung sogenannter "unit reactions" erlaubt auch im Rahmen des Konzepts von J.B. HENDRICKSON die Definition einer Reaktionsdistanz^{187, 188, 189}. Diese Reaktionsdistanz entspricht hierbei der Mindestanzahl N von unit reactions, die zwei Strukturen trennt.

$$N = \frac{1}{2} \sum_i \left(\left| \Delta h_i \right| + \left| \Delta z_i \right| \right)$$

Die Reaktionsdistanz wird C-Atom für C-Atom nach obiger Gleichung ermittelt, wobei hier Δh_i und Δz_i für die Unterschiede in den jeweils gebundenen H-Atomen und Heterofunktionalitäten stehen (siehe Abschnitt 2.1.3.1). Die Berechnung der Reaktionsdistanz erfordert jedoch Kenntnis über die durch die Reaktion festgelegte Abbildung der Eduktatome auf die Produktatome. Eine Betrachtung unterschiedlicher Abbildungsmöglichkeiten, die insbesondere bei der Beteiligung von komplexeren Umlagerungsschritten von Bedeutung ist, sowie die Ermittlung des wahren Minimums der Reaktionsdistanz findet hierbei nicht statt.

Eine weitere, mehr graphentheoretisch orientierte Betrachtungsweise der chemischen Distanz geben J. POSPICHAL und V. KVASNICKA¹⁹⁰. Werden zwei Moleküle durch die Graphen G_1 und G_2 beschrieben, so ist die chemische Distanz $CD(G_1, G_2)$ gleich der symmetrischen Differenz der Kantenmengen von G_1 und G_2 , unter Berücksichtigung aller formal möglichen Knotenindizierungen (Atom-auf-Atom-Abbildungen). Dies bedeutet, daß die chemische Distanz gegeben ist durch die Zahl der Kanten der Graphen G_1 und G_2 , die nicht Element der größten gemeinsamen Substruktur $G_1 \cap G_2$ sind.

$$CD(G_1, G_2) = \left| G_1 \right| + \left| G_2 \right| - 2 \left| G_1 \cap G_2 \right|$$

Diese Definition ist äquivalent zu der oben gegebenen Definition der minimalen chemischen Distanz MCD, wenn entweder auf die Berücksichtigung freier Elektronen verzichtet wird, oder freie Elektronen als Schleifen mit in die Graphen G_1 und G_2 aufgenommen werden.

4.2 Methoden zur Bestimmung der chemischen Distanz

Die exakte Berechnung der chemischen Distanz $CD(BE_1, BE_2)$ erfordert eine Berücksichtigung aller Abbildungsmöglichkeiten der Atome von BE_1 auf die Atome von BE_2 . Aufgrund der Einschränkungen, die aus der Zugehörigkeit der Atome zu chemischen Elementen resultieren (eine Abbildung eines Atoms ist selbstverständlich nur auf ein Atom gleichen Elementtyps möglich), ergeben sich für zwei n -atomiges Ensemble zwar weit weniger als $n!$ Abbildungsmöglichkeiten, die Anzahl der durchzuführenden Permutationen steigt jedoch immer noch äußerst stark mit der Größe der zu betrachtenden Ensembles an. Abbildung 4.5 gibt einen Eindruck über die kombinatorischen Probleme, die bereits bei Molekülen mittlerer Größe entstehen.

4.2.1 Deterministische Methoden

Als quadratisches Zuordnungsproblem gehört die Bestimmung der Minimalen Chemischen Distanz in die Klasse der sogenannten np -vollständigen Probleme, d.h. es sind keine Algorithmen konstruierbar, die das Problem allgemein und für alle Problemgrößenordnungen effizient lösen. Nun sind die Graphen chemischer Strukturen nicht beliebiger Natur, sondern haben einige ausgezeichnete Eigenschaften, die es spezialisierten Verfahren möglich machen, das np -Vollständigkeitsproblem zumindest teilweise zu umgehen. Zu diesen besonderen Eigenschaften gehört unter anderem die Beschränkung der Kantengrade (Koordinationszahlen) auf Werte ≤ 4 , was gesteuerte Backtracking-Verfahren aus dem Gebiet des Operations Research möglich macht.

Diese sogenannten Branch-and-Bound-Algorithmen¹⁹¹ nutzen innerhalb des Verfahrens den bislang kleinsten gefundenen Wert der chemischen Distanz dazu, ganze "Zweige und Äste" des aufzuspinnenden Baums von Permutationen dann abzuschneiden, wenn eindeutig erkennbar ist, daß alle diese Permutationen zu größeren Werten der chemischen Distanz führen würden.

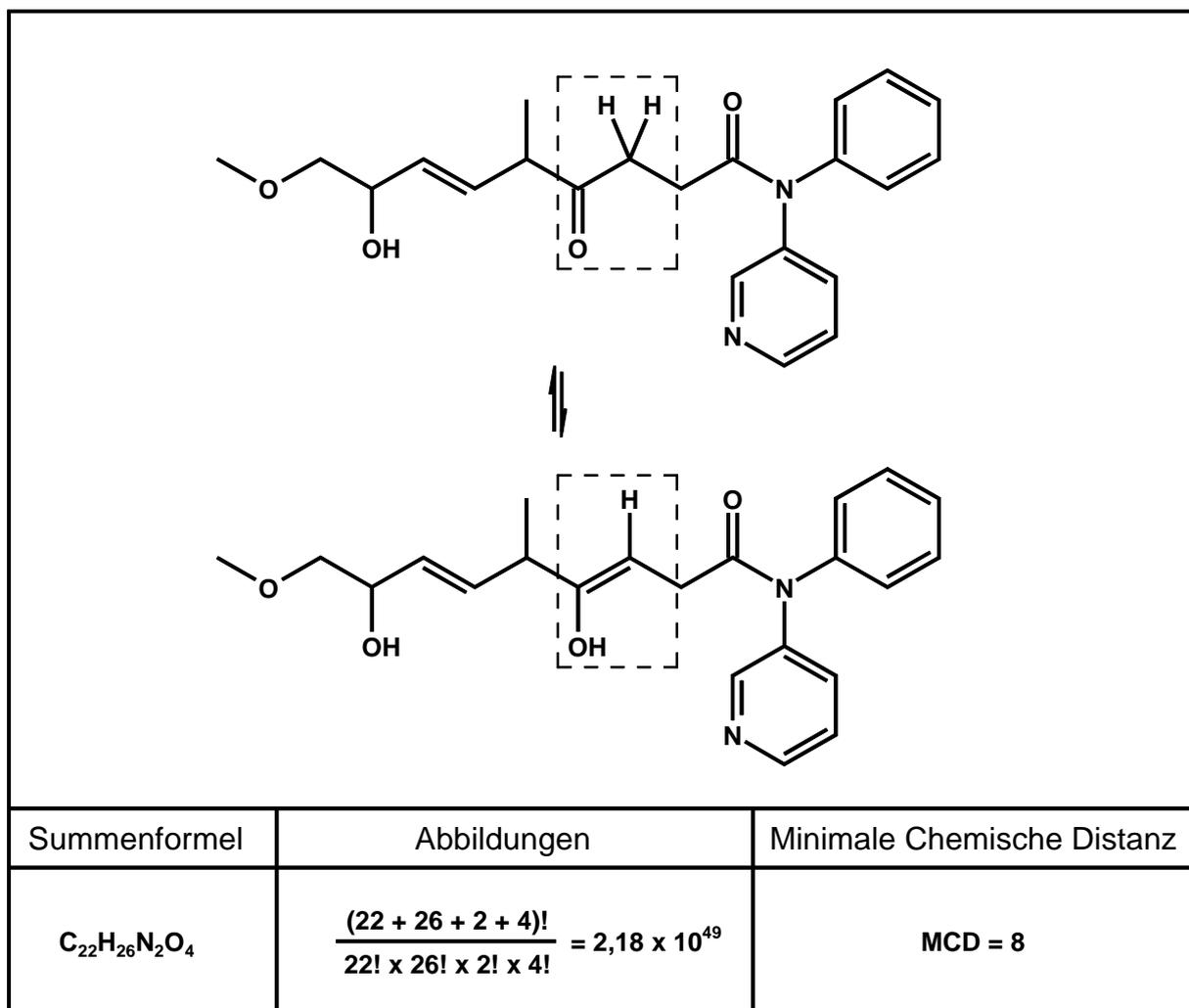


Abb. 4.5 Kombinatorische Probleme tauchen bereits bei der Atom-auf-Atom-Abbildung von Molekülen mittlerer Größe auf. Von den nahezu 10^{50} Abbildungsmöglichkeiten dieses konstruierten Beispiels entsprechen lediglich zwei Abbildungen der Minimalen Chemischen Distanz.

C. JOCHUM untersuchte eine ganze Reihe von Methoden des Operation Research, bis hin zu Methoden der Störungsrechnung, sowie vereinfachenden Linearisierungen des Problems, auf ihre Anwendbarkeit zur Berechnung der chemischen Distanz ¹⁸³. Alle diese Verfahren, auch eine verbesserte Methode von J.J. MCGREGOR, die Kanten(Bindungs)abbildungen statt Knoten(Atom)abbildungen verwendet ¹⁹², haben ihre obere Grenze der Anwendbarkeit bei 10-20 abzubildenden Atomen und sind daher für viele Probleme nicht einsetzbar.

Ein zunächst sehr vielversprechender Lösungsansatz aus der Graphentheorie erwies sich ebenfalls als ungeeignet für Moleküle mittlerer Größe: V. KVASNICKA et al. ¹⁹³ vervollständigten und verallgemeinerten ein zuvor von V. NICHOLSON et al. ¹⁹⁴

veröffentlichtes Theorem über maximale gemeinsame Substrukturen. Das Auffinden der maximalen gemeinsamen Substrukturen ist ja, wie erwähnt (siehe Abschnitt 4.1), äquivalent zur Berechnung der minimalen chemischen Distanz. Nach NICHOLSON resp. KVASNICKA wird die maximale gemeinsame Substruktur beschrieben durch die "maximale Clique" (siehe unten) im Kreuzprodukt $G^{\wedge}_1 \times G^{\wedge}_2$ der Kantengraphen der abzubildenden Ensemble. Kantengraphen G^{\wedge} entstehen aus den (eher geläufigen) Knotengraphen G durch Abbildung der Kanten von G auf Knoten von G^{\wedge} und durch Verknüpfung dieser Knoten in G^{\wedge} , wenn die ursprünglichen Kanten in G benachbart waren (siehe Abbildung 4.6).

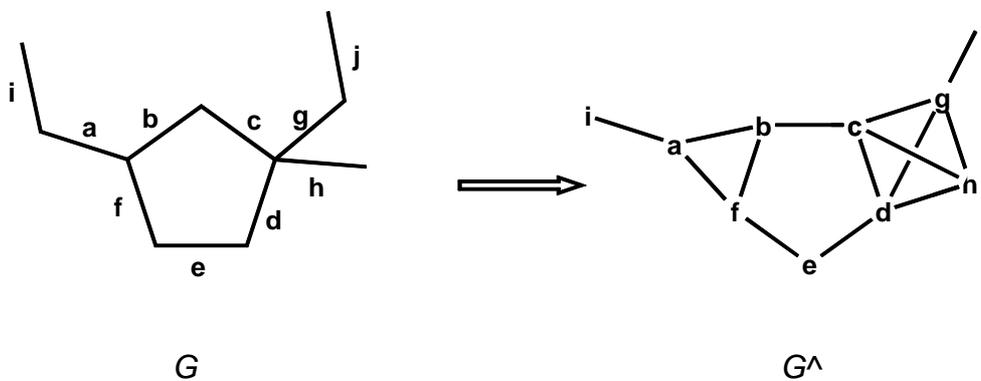


Abb. 4.6 Übergang eines Knotengraphen in den zugehörigen Kantengraphen.

Das Kreuzprodukt $G_1 \times G_2$ zweier Graphen enthält $n_1 \times n_2$ Knoten (n_1 und n_2 sind die jeweiligen Knotenzahlen der Graphen). Zwei Knoten in $G_1 \times G_2$ sind genau dann durch eine Kante verbunden, wenn die sie bildenden Knotenpaare keine übereinstimmenden Knoten enthalten und beide in G_1 und in G_2 verbunden, oder in beiden Graphen unverbunden waren (siehe Abbildung 4.7).

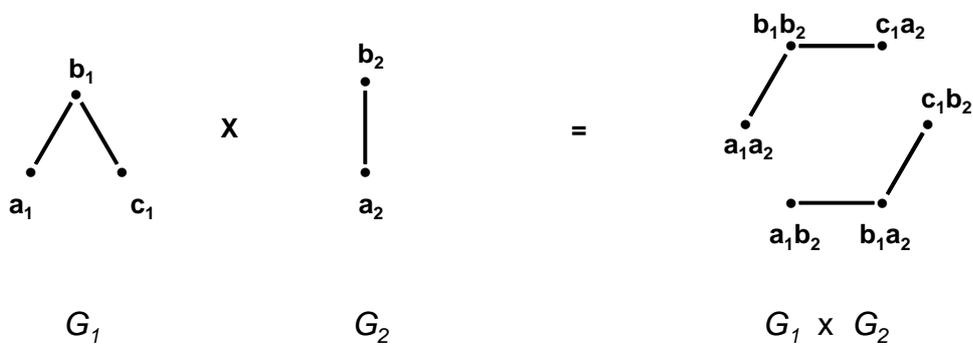


Abb. 4.7 Kreuzprodukt zweier Kantengraphen.

Die maximale Clique eines Graphen ist die größte Untermenge aller Knoten eines Graphen, die vollständig verbunden sind, d.h. jeder Knoten des entsprechenden Subgraphen ist mit jedem anderen Knoten des Subgraphen verknüpft. So enthält zum Beispiel der Graph G^{\wedge} in Abbildung 4.6 eine maximale Clique mit 4 Knoten (c, d, g, und h). Für die Ermittlung von maximalen Cliques in ungerichteten Graphen existieren gut eingeführte Methoden¹⁹⁵. Trotz allem scheitert dieses elegante graphentheoretische Verfahren wohl an der Tatsache, daß die Mächtigkeit des nach Cliques abzusuchenden Graphen (Kreuzprodukt $G^{\wedge}_1 \times G^{\wedge}_2$) quadratisch mit der Problemgröße wächst.

4.2.2 Heuristische Methoden

Wie in allen Fällen, in denen deterministische Verfahren nicht zur Lösung eines Problems führen, sind auch bei der Bestimmung der minimalen chemischen Distanz bzw. der maximalen gemeinsamen Substruktur auch Wege eingeschlagen worden, denen ein mehr oder minder großer Anteil an heuristischen Prinzipien zugrunde liegt. Bereits Ende der 70er Jahre beschrieben M.F. LYNCH und P. WILLETT ein Verfahren zur Bestimmung von Reaktionszentren¹⁹⁶, das sich eines Relaxationsverfahrens (des MORGAN-Algorithmus^{84, 85}) bedient.

In einem Relaxationsverfahren werden Atomdeskriptoren gebildet, die zunächst lokale Eigenschaften der Atome (Ordnungszahl, Koordinationszahl, freie Elektronen etc.) enthalten. Im Laufe des Verfahrens werden neue Atomdeskriptoren unter Einbeziehung der Atomdeskriptoren der Nachbaratome gebildet. Nach einigen Iterationsschritten hat sich die lokale Eigenschaft des Primärdeskriptors jedes Atoms über das ganze Molekül ausgebreitet (relaxiert). Die Atomdeskriptoren, die sich in weiteren Iterationen nicht mehr weiter diversifizieren, beschreiben nun für jedes Atom seine konstitutionelle Einbettung in das Molekül. Bei dem von LYNCH und WILLETT beschriebenen Verfahren werden nun die Atome der Ausgangsmaterialien auf ihnen entsprechende Atome der Produkte abgebildet, wenn sich deren Atomdeskriptoren nach komplettem Ablauf des MORGAN-Algorithmus möglichst ähnlich sind.

Einen hierzu verwandten Ansatz, bei dem die Reduktion der verschiedenen Abbildungsmöglichkeiten eine nachgeschaltete Enumeration möglich machen soll, verfolgten W. SCHUBERT und W. ELLENRIEDER^{123, 197}. Hierbei kam als Relaxationsverfahren der CANON-Algorithmus^{86, 87} zum Einsatz.

Die Verkleinerung der Problemdimension hat auch das Programm PEMCD (Programm zur Exakten Bestimmung der Minimalen Chemischen Distanz) von M. WOCHNER^{198, 199} zum Ziel. Mit Hilfe einer Relaxationsmethode werden hierbei Atome der abzubildenden Ensembles in Klassen äquivalenter Knoten eingeteilt, die in beiden Ensembles gleich mächtig sind. Hieraus werden nun durch Zusammenfassen dieser Atome sogenannte Hyperatome gebildet. Dies führt im allgemeinen zu einer starken Verkleinerung der Anzahl abzubildender Atome. Führt die iterative Anwendung von Relaxation und Bildung von Hyperatomen zu keiner weiteren Reduktion, so wird der verbleibende Rest von Atomabbildungen einer erschöpfenden Permutation unterzogen.

Die genannten heuristischen Verfahren sind in vielen Fällen in der Lage, eine korrekte Atom-auf-Atom-Abbildung zu ermitteln. Probleme treten dann auf, wenn durch eine Überbewertung der Ähnlichkeitskriterien falsche Primärzuordnungen erfolgen, oder wenn sich aufgrund einer relativ großen chemischen Distanz (große Unähnlichkeit der zu betrachtenden Ensemble) keine hinreichenden Atomkorrelationen ergeben, und somit der für die Enumeration verbleibende Rest zu groß wird.

4.2.3 Stochastische Methoden

Neben der Anwendung heuristischer Prinzipien werden für Probleme mit sehr großen Lösungsräumen im allgemeinen auch Verfahren entwickelt, die mit Hilfe von zufallgesteuerten Mechanismen zumindest suboptimale Lösungen auffinden können. Die hierbei eingesetzten Techniken reichen von rein zufälligen Auswahlverfahren (Monte-Carlo Techniken) über sogenanntes "Simulated Annealing" bis hin zu Evolutionsstrategien und Genetischen Algorithmen.

Die meist immens große Anzahl möglicher Atom-auf-Atom-Abbildungen verbietet den Einsatz ungesteuerter Monte-Carlo Verfahren²⁰⁰. Ein einfach zufälliges Auswählen von Atomzuordnungen führt so gut wie nie zu einem verlässlichen Wert für die chemische Distanz. Um eine einigermaßen sichere Antwort zu erhalten, müßten eine derart große Anzahl von einzelnen, zufällige Zuordnungsoperationen ausgeführt werden, daß der entsprechende Aufwand einer erschöpfenden Enumeration gleichkäme.

Ein sehr modernes Optimierungsverfahren, das insbesondere in chemischen Computeranwendungen zunehmend eingesetzt wird, ist das sogenannte "Simulated An-

nealing". Das Konzept des Simulated Annealing²⁰¹ wurde abgeleitet aus einem physikalischen Prozeß, bei dem durch Aufheizen und anschließendem langsamen Abkühlen Gitterdefekte in Kristallen entfernt werden. Die erhöhte Temperatur gibt den Atomen genügend Energie, über Potentialbarrieren hinweg andere Gitterpositionen zu erreichen. Während des langsamen Abkühlvorgangs besteht eine gewisse Wahrscheinlichkeit, daß die Atome in einem insgesamt (sub)optimalen Kristallgitter eingefroren werden. Es gibt selbstverständlich keine Gewißheit, daß der energetisch günstigste (global optimale) Zustand eingenommen wird, jedoch kann, insbesondere durch wiederholtes Anwenden des Verfahrens mit variierender (abnehmender) Aufheiztemperatur eine sehr gute Annäherung an das Optimum erreicht werden.

In einem Optimierungsverfahren auf der Basis des Simulated Annealing werden ein oder mehrere Strings aufgebaut, die die zu optimierenden Parameter und eventuell weitere Operatoren enthalten. Anschließend werden diese Strings zufälligen Veränderungen unterworfen, woraufhin mit Hilfe des METROPOLIS Algorithmus²⁰² Verbesserungen in Richtung auf das Optimum selektiert werden.

Das Prinzip besteht darin, auch neue Strings mit etwas schlechteren Eigenschaften zuzulassen, und zwar mit einer Wahrscheinlichkeit, die sich aus der Boltzmannverteilung bei einer gegebenen (Pseudo-)Temperatur ergibt. Diese Temperatur wird nun im Verlauf des Verfahrens langsam abgesenkt (Analogie zum Tempern), das Selektionskriterium also ständig verschärft, bis ein definierter Endzustand erreicht ist. Zwischen den einzelnen Abkühlphasen werden Perioden konstanter Temperatur eingeschoben, die das künstliche System in eine Art (Pseudo-)Gleichgewichtszustand überführt. Das gesamte Verfahren wird meist sehr oft wiederholt, um eine große Anzahl von Parametersets zu erhalten, die nahe beim eigentlichen Optimum liegen. Simulated Annealing wird seit kurzem sehr erfolgreich eingesetzt, um niederenergetische Konformationen von Makromolekülen aufzufinden²⁰³.

V. KVASNICKA et al. berichten über den Einsatz eines Simulated Annealing Algorithmus zur Berechnung der chemischen Distanz¹⁸⁴. Sie bedienen sich hierbei zusätzlich eines speziellen Umordnungsoperators²⁰⁴, der Nachbarschaftsverhältnisse der Atome berücksichtigt und zu einer sehr schnellen Konvergenz der Methode führt. Innerhalb dieses Simulated Annealing Verfahrens wird aus einem Permutationsvektor der Länge N ein String P gebildet, der die Atom-auf-Atom-Abbildung mit der chemischen Distanz $CD(P)$ definiert. Dieser String P wird anfangs zufällig initialisiert und in jedem Iterationsschritt zufälligen Veränderungen unterworfen. Ein daraus resultierender String P' wird für die Weiterverarbeitung mit einer Wahrschein-

lichkeit $p(P, P')$ akzeptiert, die sich aus den chemischen Distanzen $CD(P)$ und $CD(P')$ wie folgt ergibt :

$$p(P', P) = \begin{cases} 1 & \text{für } CD(P') \leq CD(P) \\ e^{-\frac{CD(P') - CD(P)}{T}} & \text{für } CD(P') > CD(P) \end{cases}$$

Der Temperaturparameter T (Anfangswert z.B. 3-5) wird für jeweils $100N$ Iterationen oder $10N$ erfolgreichen Schritten (d.h. $CD(P') < CD(P)$) auf einem konstanten Wert gehalten und dann um jeweils 10% des aktuellen Werts vermindert, bis ein gesetzter Endwert von T (z.B. 0,01) erreicht ist. KVASNICKAS Ergebnisse erster Experimente mit diesem Verfahren zeigen, daß bei Molekülgrößen von ca. 30-40 Atomen das Simulated Annealing Verfahren ausgezeichnete Ergebnisse liefert. Hiermit eröffnen moderne stochastische Verfahren Möglichkeiten der Berechnung der chemischen Distanz, die für deterministische Algorithmen aufgrund unüberwindbarer kombinatorischer Probleme unlösbar sind.

Ein weitere stochastische Methode zur optimierten Abbildung von Atomen ist die Anwendung Genetischer Algorithmen. Diese Vorgehensweise lehnt sich an die Evolution einer Population von Individuen nach dem DARWINistischen Prinzip an. Genetische Algorithmen eignen sich hervorragend für die Lösung dieses Abbildungsproblems. Für die Besprechung Genetischer Algorithmen^{204, 205, 206} und ihrer Anwendung allgemein in der Chemie und speziell zur Berechnung der chemischen Distanz bzw. von größten gemeinsamen Substrukturen, siehe Kapitel 5.

4.3 Kontrollierte Reaktionsgenerierung im BE-Raum

Die Konstruktion von mechanistischen oder synthetischen Reaktionspfaden, die gegebene Ausgangsmaterialien mit den Produkten chemischer Reaktionen oder Reaktionsfolgen verbinden, gleicht als Aufgabe sehr stark einer Navigation im komplexen multidimensionalen Raum der chemischen Konstitution. Die Lösung dieses Problems erfolgt im allgemeinen durch Erzeugung von sich immer weiter verzweigenden Reaktionsbäumen (siehe Abbildung 4.8). Die Generierung der Reaktionsbäume kann hierbei einseitig (unilateral), d.h. allein von der Eduktseite oder allein von der Produktseite, oder beidseitig (bilateral), d.h. von der Edukt- und der Produktseite geschehen. Die bilaterale Reaktionsgenerierung ist im allgemeinen die effizientere Vorgehensweise ¹¹¹.

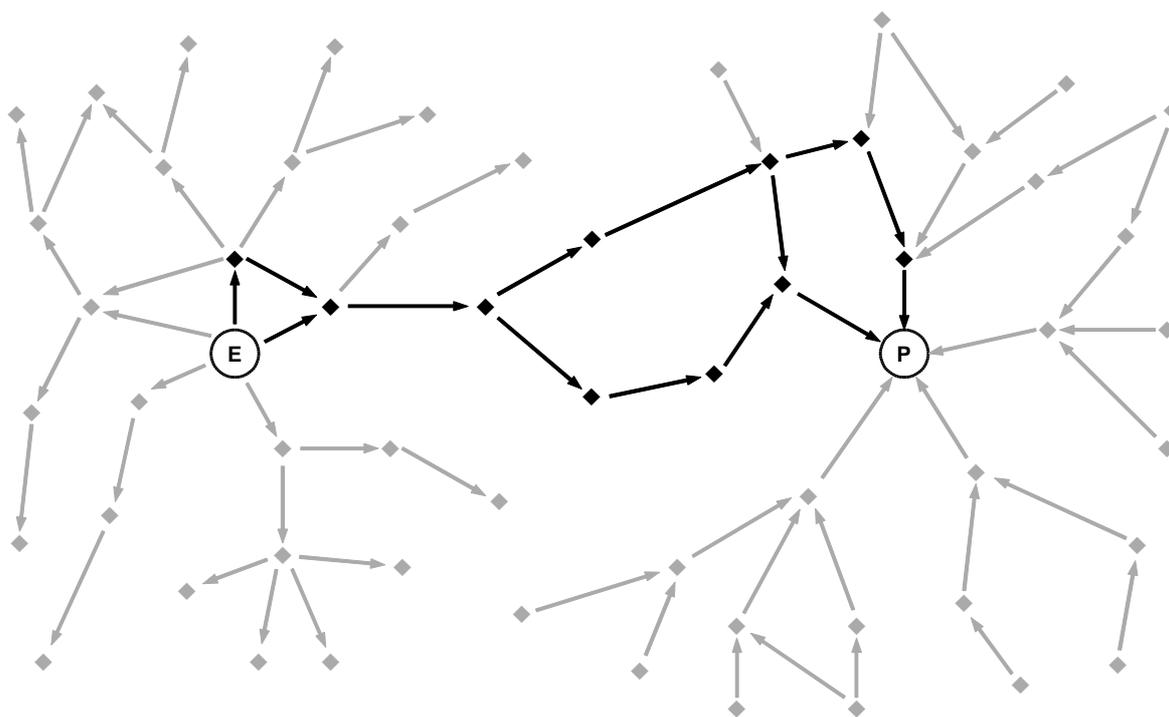


Abb. 4.8 Bilateral generiertes Reaktionsnetzwerk vom Edukt (E) zum Produkt (P) einer chemischen Reaktion (◆ sind die erzeugten Zwischenstufen). Die Edukt und Produkt verbindenden Reaktionspfade sind schwarz, abweichende Pfade sind grau dargestellt.

Eine formal mögliche Lösung des gestellten Problems, d.h. ein unter den gegebenen Randbedingungen gültiger Reaktionspfad von den Edukten zu den Produkten der Reaktion ist gefunden, wenn die weitere Verzweigung des gerade "wachsenden" Re-

aktionsbaums auf das angestrebte gegenüberliegende Ziel (Edukt oder Produkt) trifft oder ein Mitglied des von dort "entgegenwachsenden" Reaktionsbaums erreicht.

Das Erzeugen von Reaktionspfaden "in alle Richtungen" des BE-Raums, bis mehr oder weniger zufällig das Ziel erreicht wird, ähnelt jedoch einem orientierungslosen Herumtappen im Nebel. Sehr viel effizienter wäre ein zielorientiertes Generieren der Reaktionsbäume, das Abwege erkennt und rechtzeitig Verzweigungen abschneidet, die nicht, oder nur nach sehr vielen Reaktionsschritten zum Ziel führen können.

Dies erfordert jedoch, wie jede Art der Navigation, eine möglichst präzise Ortsbestimmung. Es muß eine Methode zur Verfügung gestellt werden, die es erlaubt festzustellen, an welchem Punkt im BE-Raum, in welcher Entfernung zu den Produkten bzw. Edukten sich ein generiertes Zwischenprodukt befindet. Ist die relative Positionierung eines Zwischenproduktes bezüglich des Ausgangspunktes und des Zieles bekannt, so kann für jeden einzelnen der vom aktuellen Punkt ausgehenden Reaktionsschritte festgestellt werden, ob er sich auf das angestrebte Ziel zubewegt oder in eine mehr oder weniger abweichende Richtung verläuft.

Damit ist ein zielorientierter Aufbau von Reaktionsbäumen möglich, der hauptsächlich kurze und direkte Reaktionsmechanismen in Betracht zieht.

Im folgenden werden mehrere Verfahren vorgestellt, mit denen diese kontrollierte Reaktionsgenerierung im BE-Raum durchgeführt werden kann.

4.3.1 Methode der monoton fallenden Chemischen Distanz

Eine chemische Reaktionssequenz von einem Ausgangsensemble zu einem Produktensemble verläuft im allgemeinen über Zwischenstufen, deren Strukturen den Produktverbindungen von Reaktionsschritt zu Reaktionsschritt immer ähnlicher werden. Dies gilt insbesondere bei mechanistischen Reaktionspfaden, während bei synthetisch orientierten Sequenzen zum Beispiel durch die Einführung und Abspaltung von Schutzgruppen in einem oft unvermeidbaren Umweg auch Zwischenstufen angenommen werden, die dem Endprodukt strukturell weniger verwandt sind, als ihre unmittelbaren Vorläuferstrukturen.

Die Ähnlichkeit zweier Molekülensembles im Sinne konstitutioneller Verwandtheit läßt sich messen, indem die schrittweisen Veränderungen gezählt werden, die ein Ensemble in das andere transformiert. Dies führt unmittelbar zu einer Beziehung zwischen konstitutioneller Ähnlichkeit und minimaler chemischer Distanz ²⁰⁴.

Eine zunehmende Ähnlichkeit der Zwischenstufen vom Produkt zum Edukt drückt sich darin aus, daß die chemischen Distanzen der Intermediate zum Produkt monoton abnehmend. Abbildung 4.9 zeigt dies anhand des Beispiels der Bildung von Cyansulfinen aus *N,N*-Bis(trimethylsilyl)inaminen ²⁰⁷.

Das Kriterium der monoton fallenden Chemischen Distanzen ist ein ausgezeichnetes Mittel zur zielgerichteten Erzeugung von Reaktionsbäumen. Um auch geringfügige Abweichungen von der (strengen) Monotonie zu erlauben, wird hierzu ein Toleranzwert CD_{tol} definiert. Wird nun im Zuge der monolateralen (zum Beispiel vom Edukt ausgehenden) Generierung eines Reaktionsnetzwerks für ein Molekül *M* eine Nachfolgestruktur *N* erzeugt, so ist *N* nur dann eine gültige Struktur des Reaktionsnetzwerkes, wenn gilt:

$$CD(N, \text{Produkt}) \leq CD(M, \text{Produkt}) + CD_{tol}$$

Wird die Toleranzschwelle $CD_{tol} = 0$ gesetzt, so ist garantiert, daß alle erzeugten Intermediate dem Produkt nicht unähnlicher sind als es das Edukt von vorneherein ist. Der Ereignishorizont im BE-Raum beschränkt sich somit auf eine L_1 -Kugel, in deren Mittelpunkt sich das Produkt befindet. Der Radius dieser Kugel ist dann gleich der Chemischen Distanz des Edukts vom Produkt.

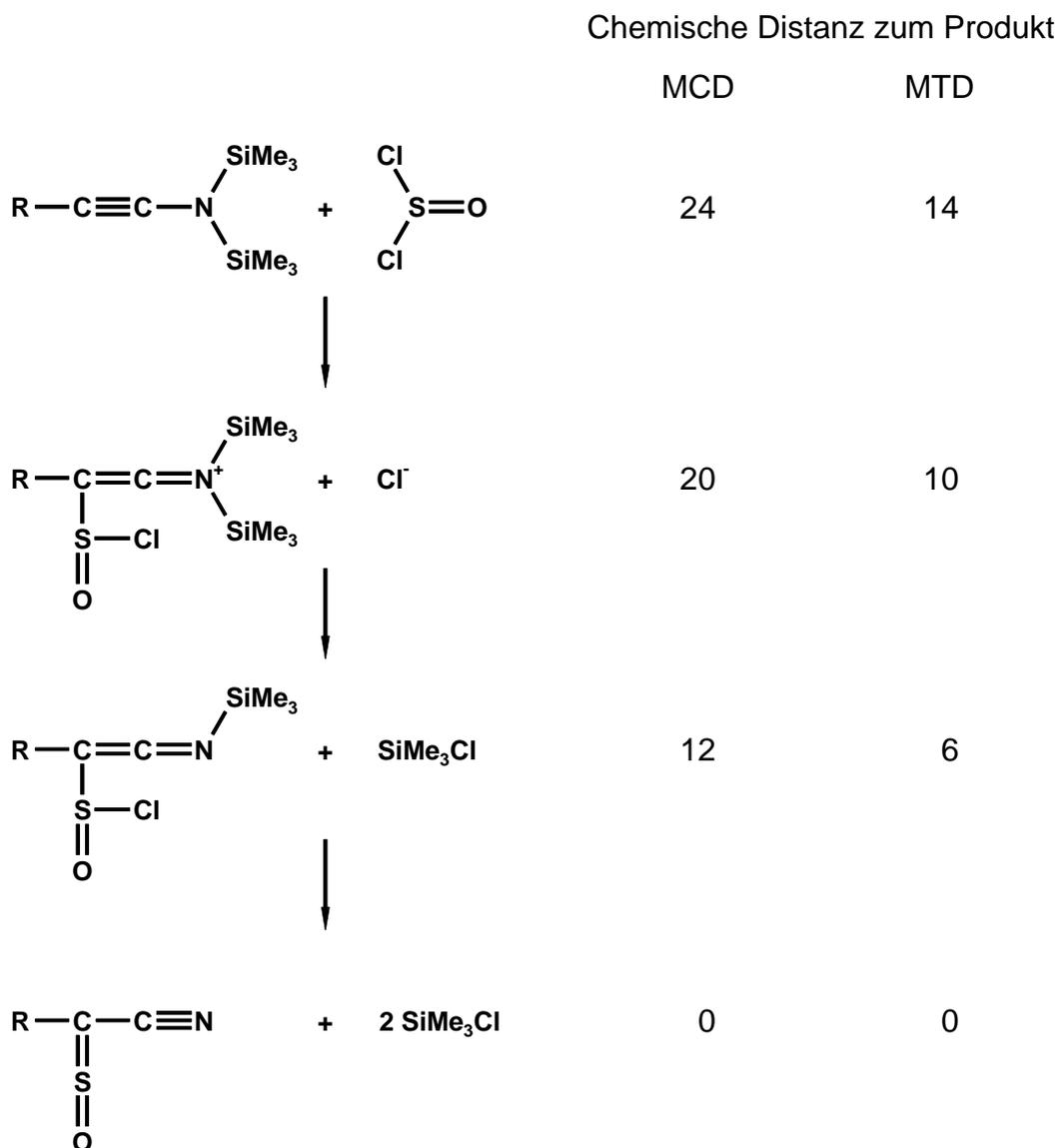


Abb. 4.9 Die Bildung von Cyansulfinen aus *N,N*-Bis(trimethylsilyl)inaminen zeigt als Beispiel, wie im Verlauf des Reaktionsmechanismus die Chemischen Distanzwerte (streng) monoton fallen, d.h. die konstitutionelle Ähnlichkeit der Intermediate zum Produkt hin immer größer wird.

Für $CD_{\text{tol}} < 0$ werden ausschließlich selektiv auf das Produkt hin orientierte Reaktionssequenzen generiert, die allerdings auch keine Umwege mehr erlauben (strenge Monotonie der Chemischen Distanz). Das folgende Beispiel der Siebenkomponentenreaktion zeigt eine Anwendungsmöglichkeit dieses Verfahrens.

4.3.1.1 Beispiel: Siebenkomponentenreaktion

Multikomponentenreaktionen sind Eintopfreaktionen, an denen sich mindestens drei unterschiedliche Reaktionspartner beteiligen, wobei Produkte einheitlicher Konstitution gebildet werden²⁰⁸. Meist handelt es sich hierbei um eine komplizierte Abfolge von Reaktionsschritten, die ein komplexes System von Gleichgewichten aufbauen, aus denen das Produkt in einem oder mehreren irreversiblen Reaktionsschritten entfernt wird.

Die von A. DÖMLING und I. UGI im Jahre 1993 gefundene Siebenkomponentenreaktion²⁰⁹ hält zur Zeit den "Weltrekord" bezüglich der Anzahl gleichzeitig eingesetzter unterschiedlicher Ausgangskomponenten. Bei der Siebenkomponentenreaktion (7CC-Reaktion) handelt es sich um eine Kombination zweier Multikomponentenreaktionen. Der seit Anfang der 60er Jahre bekannten Vierkomponentenkondensation (4CC-Reaktion, Ugi-Reaktion)²¹⁰, bei der sich ein Isocyanid, ein Amin, eine Carbonylverbindung und eine Säurekomponente in einer Kondensationsreaktion vereinigen, wird eine Asinger-Kondensation²¹¹ vorgeschaltet, die eine Mercaptocarbonylverbindung in einen Heterocyclus überführt (siehe Abbildung 4.10).

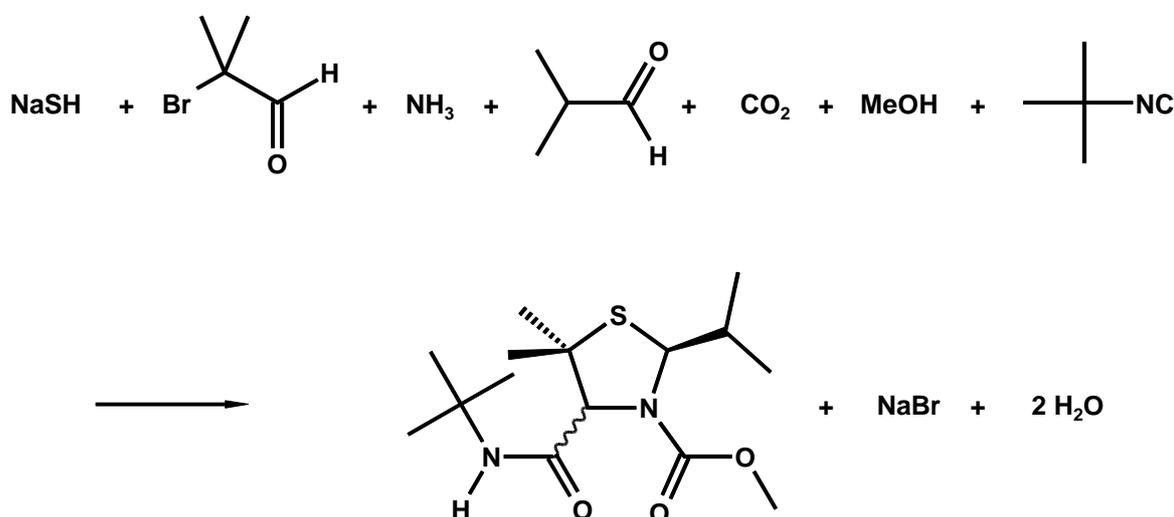


Abb. 4.10 Siebenkomponentenreaktion (7CC-Reaktion).

Die große Anzahl von Reaktionspartnern mit vielen potentiell reaktiven Zentren (hier sind es 21 aktive Atome) und eine zu erwartende lange Reaktionssequenz machen eine Erarbeitung möglicher Reaktionsmechanismen mittels formaler Computerprogramme außerordentlich schwer. Eine zielgerichtete Generierung eines

monolateralen Reaktionsnetzwerkes, ausgehend vom Edukt, ist jedoch möglich. Abbildung 4.11 und Tabelle 4.1 zeigen die Resultate einer Distanzanalyse von Reaktionsintermediaten der Siebenkomponentenreaktion.

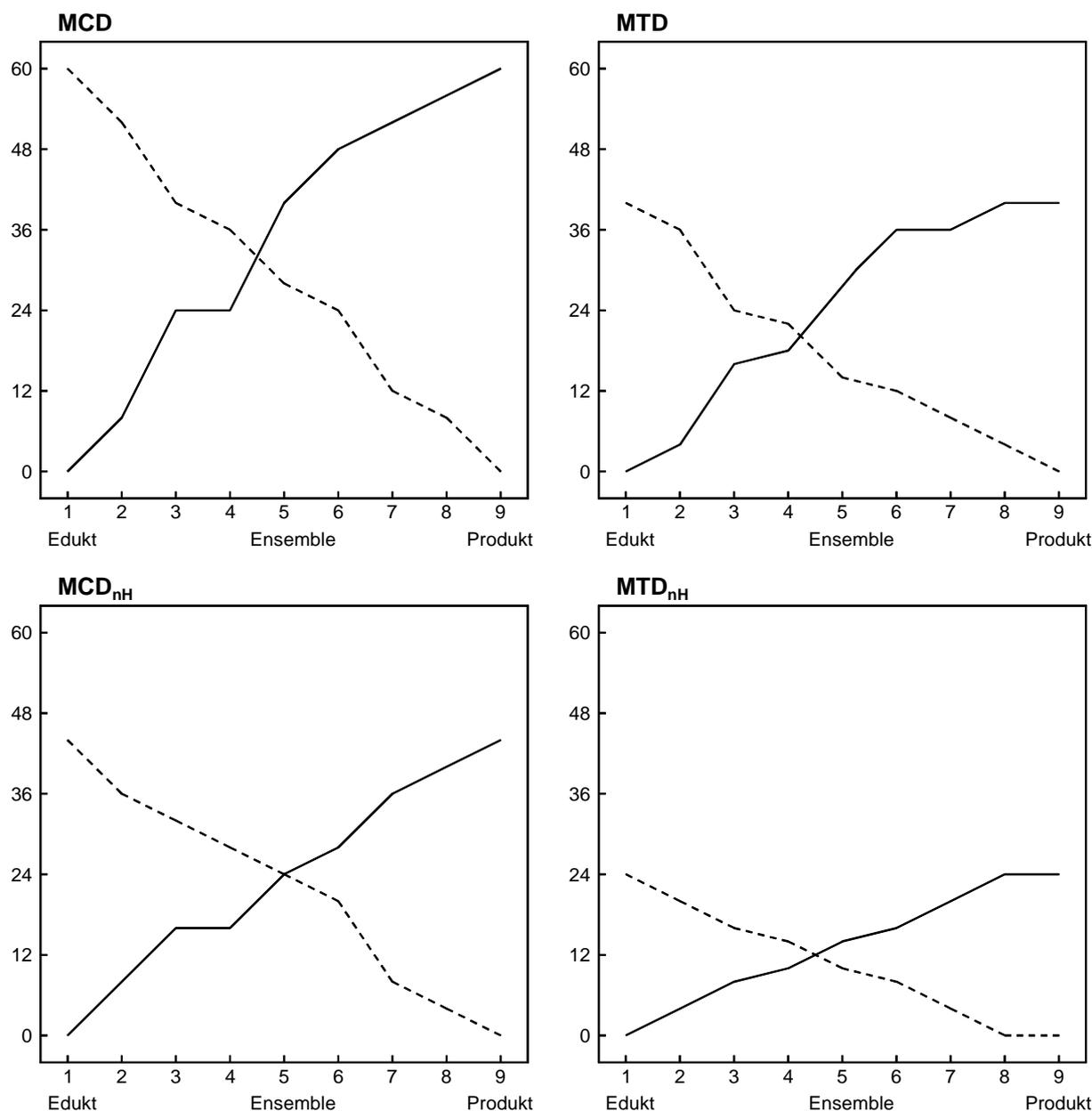
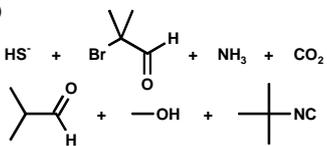
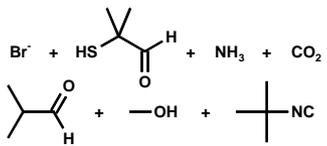
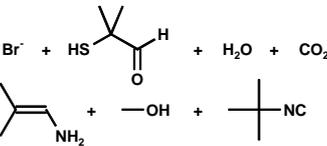
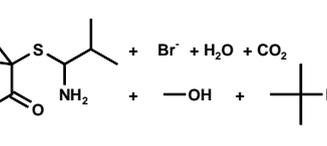
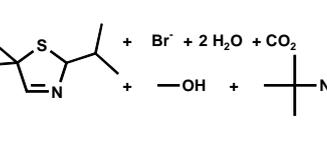
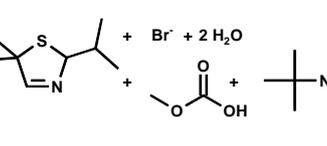
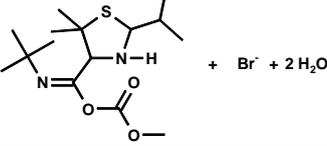
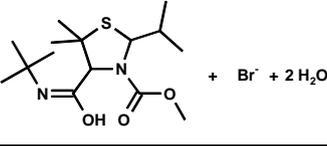
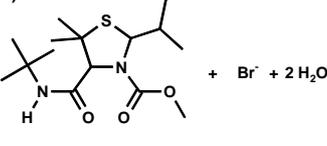


Abb. 4.11 Änderung der chemischen Distanzen der Zwischenstufen zum Edukt (—) bzw. Produkt (---) im Verlauf einer Siebenkomponenten-Kondensation (7CC). Für Einzelwerte und Strukturen der Ensembles, siehe Tab. 4.1.

7CC - Ensemble	Distanz zum Edukt				Distanz zum Produkt			
	MCD	MTD	MCD _{nH}	MTD _{nH}	MCD	MTD	MCD _{nH}	MTD _{nH}
1 (Edukt) 	0	0	0	0	60	40	44	24
2 	8	4	8	4	52	36	36	20
3 	24	16	16	8	40	24	32	16
4 	24	18	16	10	36	22	28	14
5 	40	30	24	14	28	14	24	10
6 	48	36	28	16	24	12	20	8
7 	52	36	36	20	12	8	8	4
8 	56	40	40	24	8	4	4	0
9 (Produkt) 	60	40	44	24	0	0	0	0

Tab. 4.1 Chemische Distanzen der Ensembles einer Reaktionsfolge für die 7CC-Reaktion. Die Distanzwerte wurden mit Hilfe eines Genetischen Algorithmus ermittelt.

Die 7CC-Reaktion folgt demnach einer streng monotonen Abnahme der chemischen Distanzwerte vom Edukt über die Intermediate bis hin zum Produkt. Dies gilt für alle vier unterschiedlichen Distanzwerte MCD, MTD, MCD_{nH} , MTD_{nH} . Lediglich der letzte Reaktionsschritt zeigt für den MTD_{nH} -Wert keinen weiteren Abfall des Distanzwertes. Dies liegt begründet in der Tatsache, daß es sich bei diesem letzten Reaktionsschritt um eine Keto-Enol-Tautomerisierung handelt, die infolge ihrer ausschließlichen H-Verschiebung unter Beibehaltung des restlichen Bindungsskeletts nicht neutral bezüglich unterschiedlicher Bindungsordnungen (MCD) ist, aber in Bezug auf die Topologie (MTD) der Schweratome des Moleküls keine Veränderungen bewirkt.

4.3.2 Methode des ellipsoidalen Einschusses

Die Methode des ellipsoidalen Einschusses ist eine Erweiterung der oben vorgestellten Methode der monoton fallenden chemischen Distanzen. Sie erlaubt eine etwas flexiblere Festlegung des chemischen Ereignishorizonts bei der kontrollierten Reaktionsgenerierung im BE-Raum. Bei der Methode des ellipsoidalen Einschusses werden für jedes Intermediat zugleich sowohl die Distanz zum Edukt, als auch die Distanz zum Produkt berechnet. Strukturen von Zwischenstufen N, deren Summe aus $CD(N, \text{Edukt}) + CD(N, \text{Produkt})$ größer ist als ein gesetztes Limit, werden dann für eine weitere Generierung der Reaktionsbäume nicht mehr in Betracht gezogen.

Dies bedeutet, daß sich der chemische Ereignishorizont im BE-Raum auf ein L_1 -Ellipsoid beschränkt, in dem das Produktensemble und das Eduktensemble die beiden Brennpunkte bilden. Innerhalb dieses L_1 -Ellipsoids können alle beliebigen BE-Punkte zur Generierung von Reaktionspfaden herangezogen werden (solange diese allen anderen Randbedingungen gehorchen). Gültige Zwischenstufen N erfüllen demnach folgende Bedingung:

$$CD(N, \text{Edukt}) + CD(N, \text{Produkt}) \leq CD(\text{Edukt}, \text{Produkt}) + CD_{\text{tol}}$$

CD_{tol} stellt hierbei eine wählbare Größe dar und beschreibt die Ausdehnung des Ellipsoids um die beiden Brennpunkte (Edukt und Produkt). Da die chemische Distanz alle Bedingungen einer Metrik erfüllt, sind negative Werte für CD_{tol} unzulässig, denn es kann kein Intermediat existieren, dessen Distanzsumme zu Edukt und Produkt kleiner ist, als die Distanz zwischen Edukt und Produkt selber.

Zur weiteren Veranschaulichung des Konzepts wird eine graphische Darstellung gewählt, bei der eine formale Projektion aus dem L_1 -Hyperraum in eine gewöhnliche zweidimensionale L_2 -Fläche vorgenommen wird. CD_{tol} beschreibt in dieser Darstellungsweise eine echte Ellipse um Edukt und Produkt. Abbildung 4.12 zeigt diese Pseudoprojektion zusammen mit einigen zusätzlichen Kurven konstanter Distanzwerte.

Da bei dieser Projektion die geometrische Lage eines Ensemblebildpunktes lediglich durch die beiden Distanzen zum Edukt und zum Produkt definiert ist, würde die Darstellung einer Halbellipse (oberer oder unterer Teil) genügen. Da mitunter sehr

viele Zwischenstufen einzutragen sind, wird jedoch bei der Darstellung komplexer Reaktionsnetzwerke eine volle Ellipse gewählt, und die einzelnen Bildpunkte der Intermediatensembles willkürlich oder nach Bedarf auf die obere, bzw. untere Ellipsenhälfte verteilt.

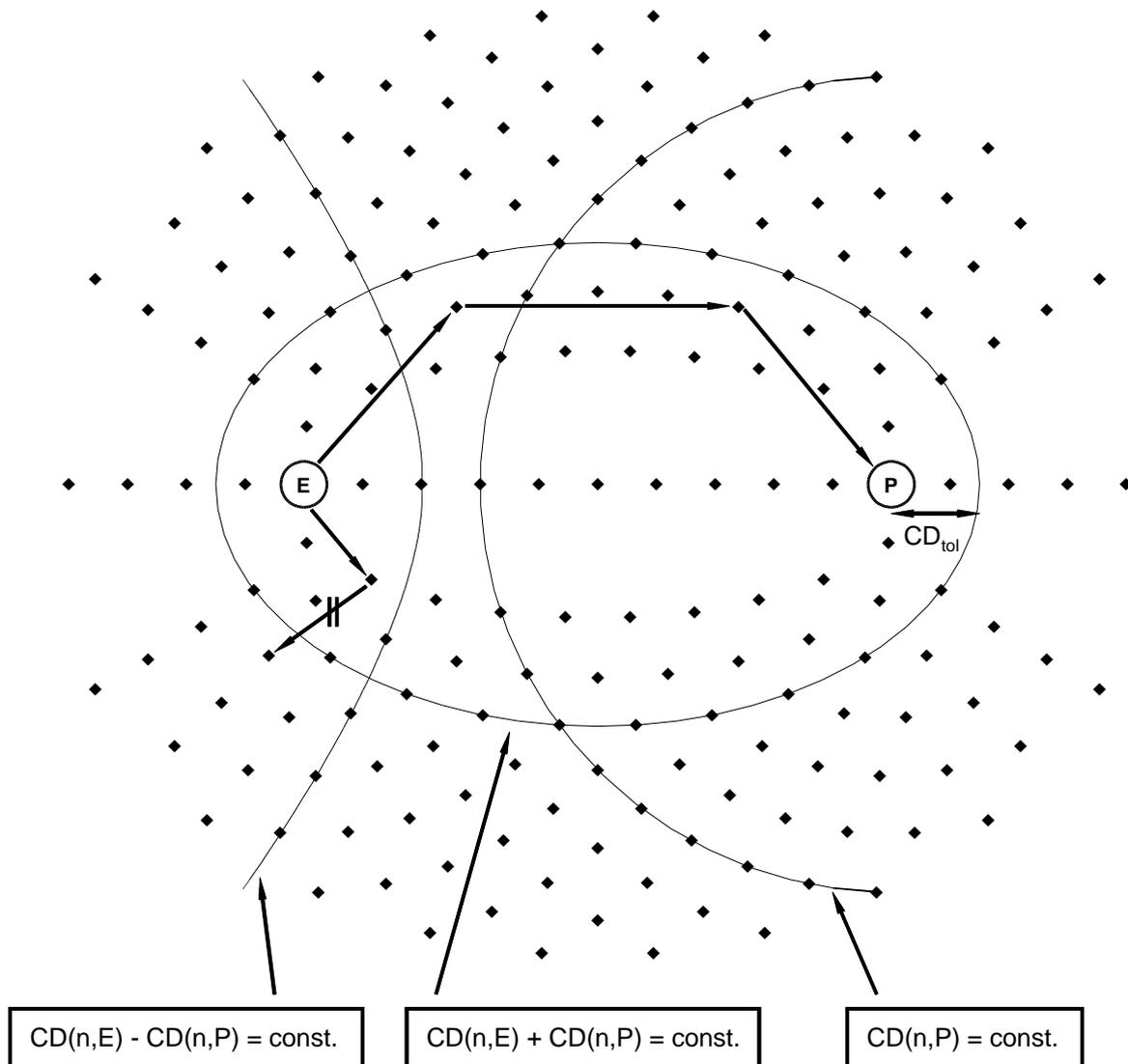
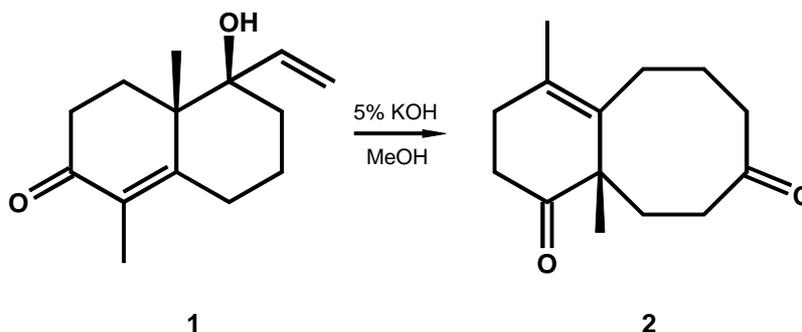


Abb. 4.12 Zweidimensionale Projektion eines L_1 -Ellipsoids um Edukt und Produkt einer chemischen Reaktion. Erlaubte Zwischenstufen von Reaktionspfaden liegen innerhalb der Ellipse, die durch CD_{tol} und der Distanz zwischen Edukt und Produkt definiert wird. Edukt und Produkt liegen in den Brennpunkten der Ellipse.

Im folgenden zeigen einige Anwendungsbeispiele, daß die Methode des ellipsoidalen Einschlusses den chemischen Ereignishorizont tatsächlich auf die Zwischenstufen interessierender Reaktionspfade begrenzt, und somit eine effiziente Technik zur kontrollierten Reaktionsgenerierung darstellt.

4.3.2.1 Beispiel: Oxy-Cope-Umlagerung

Als einführendes Beispiel für die Anwendung der Methode des ellipsoidalen Einschlusses sei eine basenkatalysierte Umlagerung eines Oxy-Cope-Systems gewählt ²¹². Hierbei reagiert das Vinylcarbinol **1** zum ringerweiterten 1,5-Dion **2**.



Die Autoren schlagen für den Reaktionsmechanismus eine konzertierte 3,3-sigmatrope Umlagerung der Oxy-Cope-Substruktur in Verbindung **1** vor. Die RAIN-Analyse dieser Reaktion wurde mit den Randbedingungen R04 in Tabelle 2.4 durchgeführt ²¹³.

Die bilaterale Generierung von je einem Reaktionsschritt ausgehend vom Edukt **1** bzw. vom Produkt **2** führte bereits zum geschlossenen Reaktionsweg **1** ® **6** ® **2** (siehe Abbildung 4.13). Hierbei wurde noch keine Limitierung erlaubter Strukturen hinsichtlich der MCD-Werte vorgenommen

!Ungültige

Zeicheneinstellung

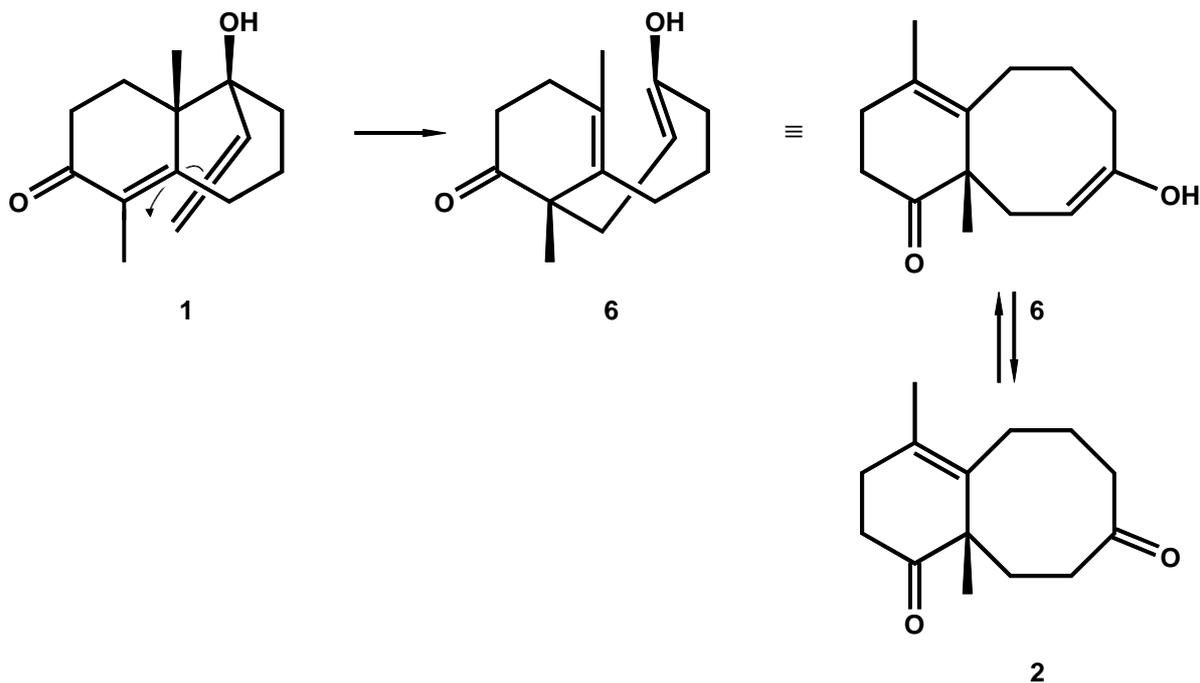


Abb. 4.13 Resultierender Reaktionsweg der Oxy-Cope-Umlagerung.

Außer dem Intermediat **6** werden bei der Pfadgenerierung jedoch noch 46 andere formal mögliche Folgeprodukte von **1** bzw. Vorläuferstrukturen von **2** erzeugt. Werden nun alle BE-Punkte dieser Strukturen in einer ellipsoidalen Pseudoprojektion um **1** und **2** als Brennpunkte dargestellt, so wird deutlich, daß nur wenige Verbindungen innerhalb des Toleranzbereichs liegen, der durch die Lage des Intermediats **6** definiert ist. Die meisten der anderen generierten Reaktionsschritte führen also zu Strukturen, die vom eigentlichen Reaktionsgeschehen abweichen und somit (zumindest bei der Konzentration auf kurze Reaktionssequenzen) nicht weiter verfolgt werden müssen.

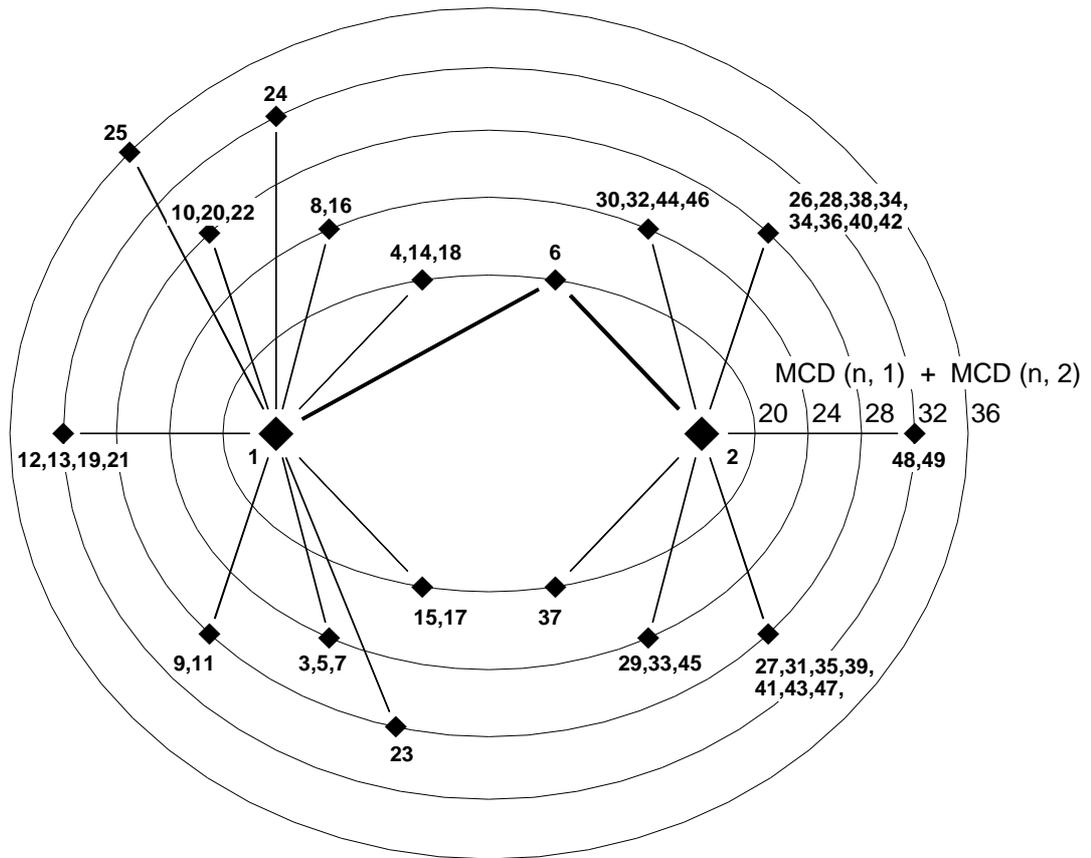
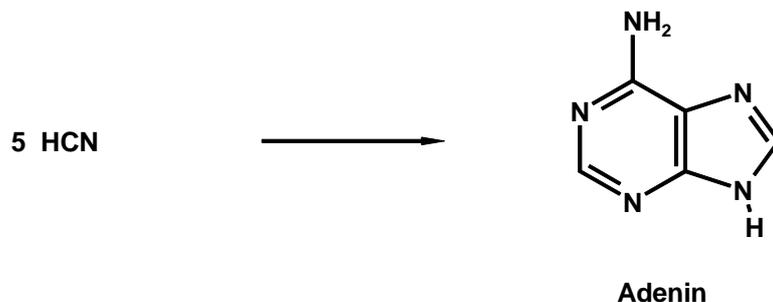


Abb. 4.14 Ellipsoidale Projektion des bilateralen Reaktionsnetzwerks der Oxy-Cope-Umlagerung.

4.3.2.2 Beispiel: Abiogenese von Adenin

Die grundlegenden Experimente von S.L. MILLER, in denen gezeigt werden konnte, daß durch elektrische Funkenentladungen in einer Atmosphäre aus Methan, Ammoniak und Wasser biologisch relevante Moleküle wie Aminosäuren in nennenswerten Mengen entstehen können^{214, 215, 216, 217}, bereiteten den Weg für ein äußerst interessantes Forschungsfeld, der präbiotischen organischen Chemie. Zielsetzung der Untersuchungen auf diesem Gebiet ist es, die Entstehungsmöglichkeiten der für eine evolutionäre Weiterentwicklung des Lebens unabdingbaren Grundbausteine (Aminosäuren, Nucleinsäuren etc.) aus kleinen, unter den Bedingungen der frühen Erde (oder eines anderen Himmelskörpers) existenzfähigen Molekülen (Methan, Ammoniak, Schwefelwasserstoff, Cyanwasserstoff, Wasser etc.) aufzuklären. Sowohl das Gebiet der präbiotischen Chemie, als auch das Geschehen bei den Anfängen der molekularen Evolution ist bis heute weitgehend ungeklärt und steht deutlich im Widerstreit unterschiedlicher Hypothesen und Auffassungen^{218, 219, 220, 221, 222}.

Cyanwasserstoff HCN wird im allgemeinen als eine der wichtigsten potentiellen Quellen für die Abiogenese stickstoffhaltiger Biomoleküle angesehen. J. ORÓ und A.P. KIMBALL²²³ berichteten zu Anfang der 60er Jahre von der Oligomerisierung von HCN zum Purin Adenin nach der Bruttogleichung:



Weitere experimentelle Untersuchungen^{224, 225, 226, 227, 228, 229} führten bis Ende der 70er Jahre zu einer ersten (recht groben) Vorstellung²³⁰ von dem Reaktionsgeschehen bei der Oligomerisierung von HCN zu Adenin (siehe Abbildung 4.15).

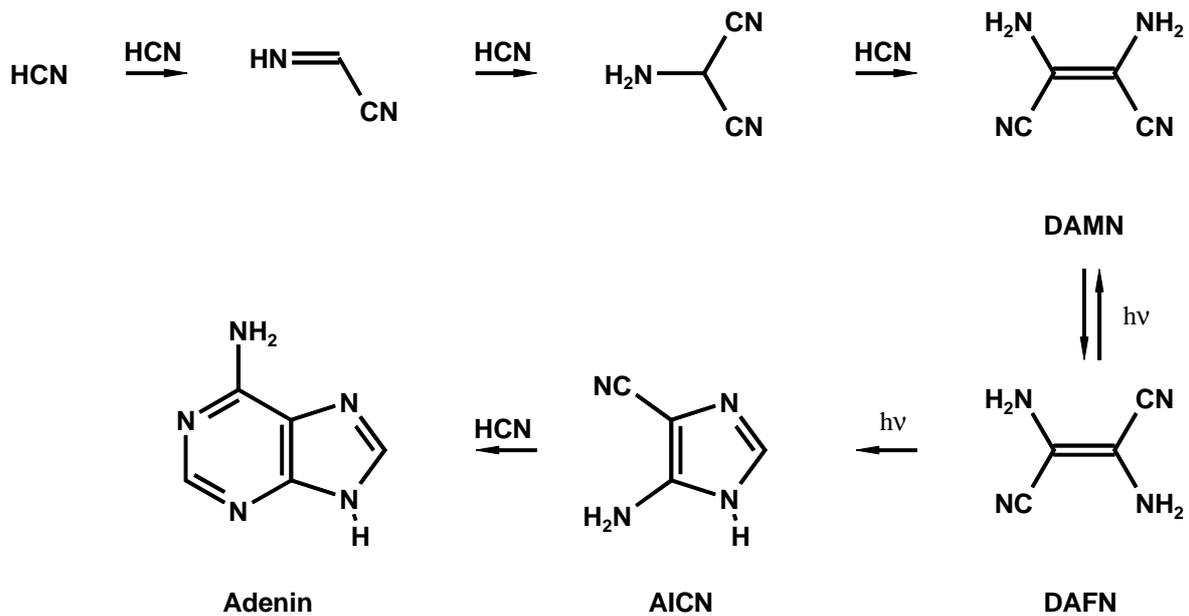
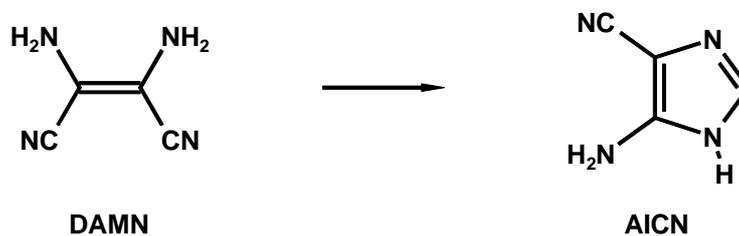


Abb. 4.15 Angenommenes Reaktionsschema der Pentamerisierung von HCN zum Purin Adenin, einer der Grundbasen des Genetischen Codes.

Neuere Arbeiten erweitern dieses Schema²³¹, bzw. bringen auch andere reaktive Stickstoffintermediate ins Spiel^{232, 233}.

Die Tetramerisierung von HCN zu Diaminomaleodinitril (DAMN) bzw. Diaminofumarodinitril (DAFN) war bereits Gegenstand einer intensiven Untersuchung¹¹¹ mit dem Programm RAIN. Ein weiterer Schlüsselschritt des kompletten Reaktionsschemas in Abbildung 4.15 ist jedoch die photoinduzierte Umwandlung von DAMN zum 4-Aminoimidazol-5-carbonitril (AICN):



Auch dieser Reaktionsschritt wurde einer Analyse mit dem Programm RAIN unterzogen²⁰⁴, wobei hier die zielgerichtete monolaterale Reaktionsnetzwerkgenerierung zum Einsatz kam. Um die Leistungsfähigkeit des ellipsoidalen Einschlusses zu testen, wird hier zunächst, ausgehend von DAMN, eine monolaterale Reaktionsnetzwerkgenerierung ohne Distanzbeschränkung durchgeführt. Es resultiert nach 3 Generierungsiterationen (entspr. Reaktionsschritten) das Netzwerk in

Abbildung 4.16. Das gewünschte Targetmolekül AICN ist nach dieser Anzahl von Schritten erreicht.

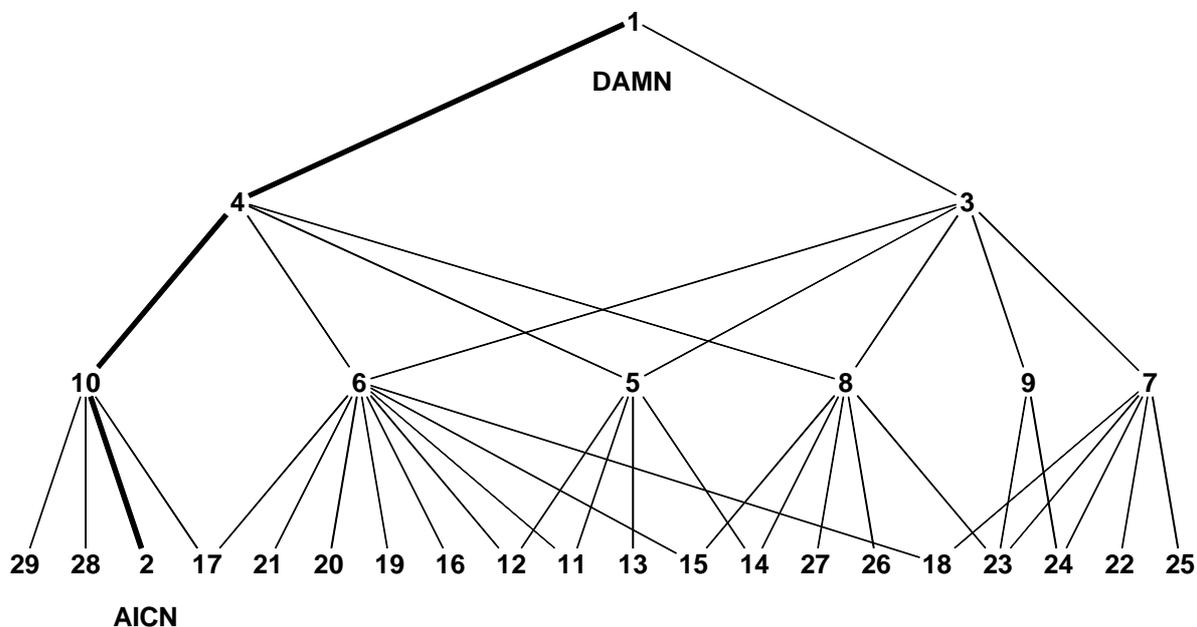
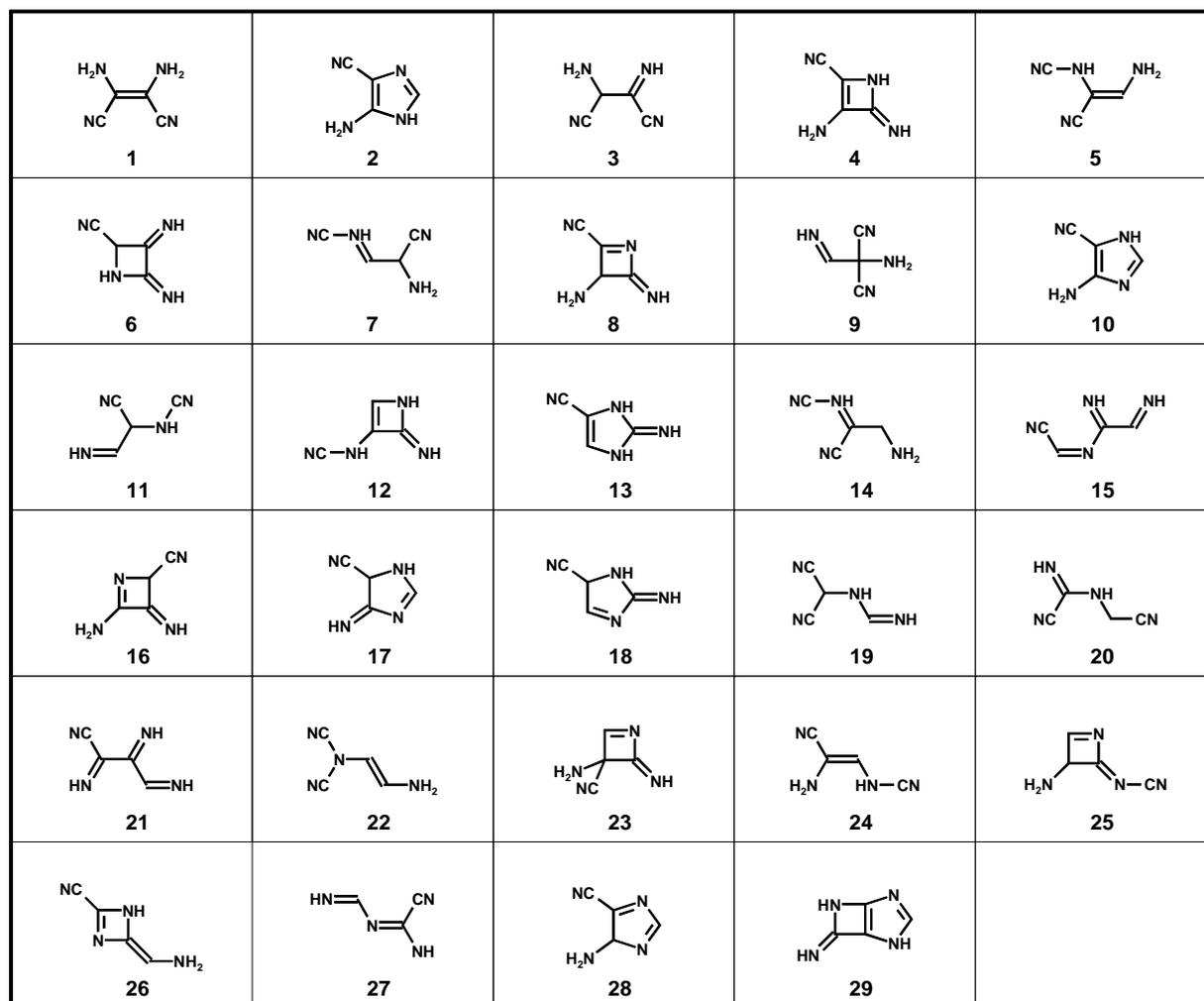


Abb. 4.16 Komplettes Reaktionsnetzwerk (monolateral), ausgehend von DAMN. Die Generierung (Randbedingungen R05 in Tabelle 2.4) wurde so viele Iterationen lang durchgeführt, bis das angestrebte Zielmolekül AICN erreicht war (dicke Verbindungslinie). Die entsprechenden Strukturen sind in Tabelle 4.2 aufgelistet.



Tab. 4.2 Strukturen zum Reaktionsnetzwerk in Abbildung 4.16

Für alle 29 Strukturen werden nun die Chemischen Distanzwerte (MCD-Werte) zum Edukt DAMN und zum Produkt AICN bestimmt. Anschließend wird das Reaktionsnetzwerk in der ellipsoidalen Projektion dargestellt (Abbildung 4.17).

Diese Darstellung veranschaulicht unmittelbar, daß der gefundene Reaktionsweg (**1 !Ungültige Zeicheneinstellung**) abgeschnitten werden. Letztlich würde sich damit das Reaktionsnetzwerk von Abbildung 4.16 auf *eine einzige* Reaktionssequenz (Abbildung 4.18) reduzieren.

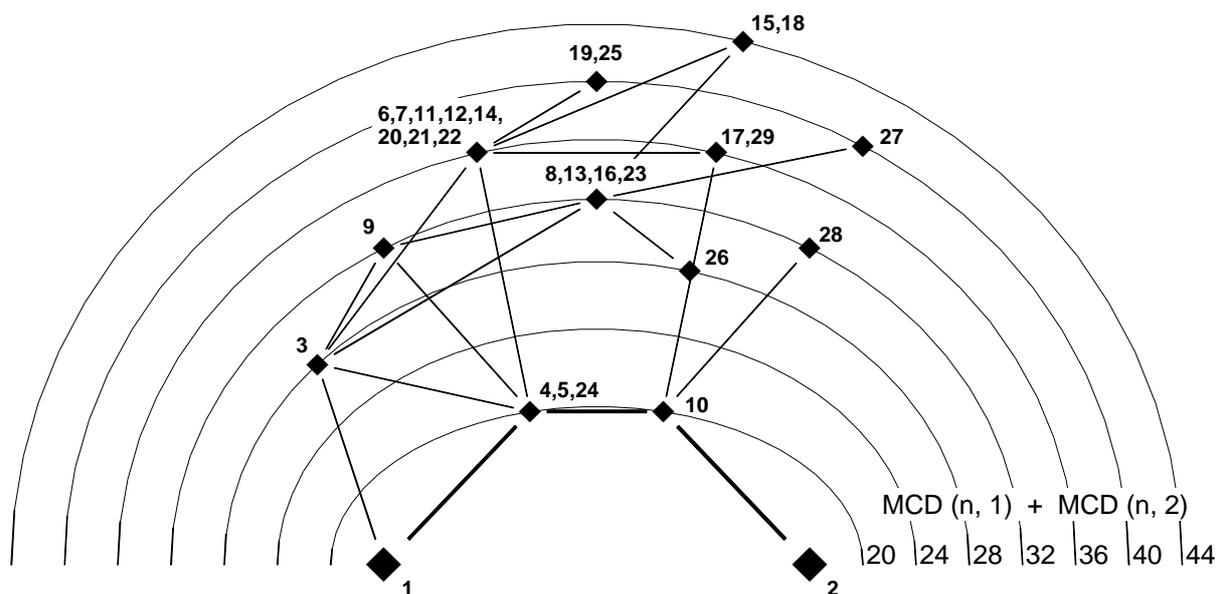


Abb. 4.17 Ellipsoidale Projektion des monolateralen Reaktionsnetzwerks von DAMN (1) nach AICN (2) (für Zuordnung der Nummern zu Strukturen, siehe Tabelle 4.2).

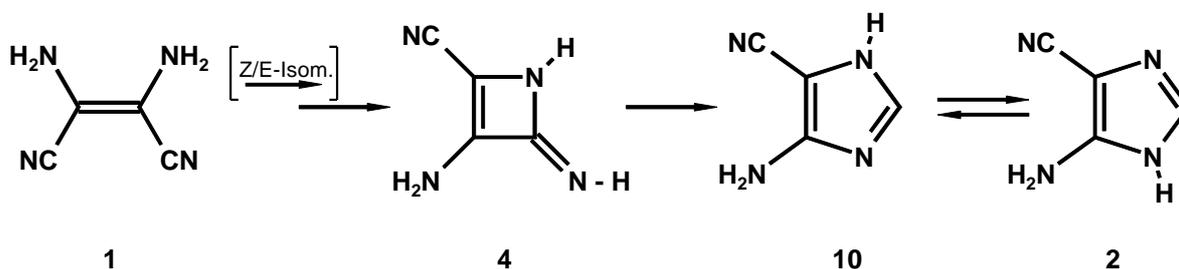


Abb. 4.18 Einziger resultierender Reaktionsweg von DAMN (1) nach AICN (2) unter einem ellipsoidalen Einschluß mit einem Toleranzwert $CD_{tol} = 4$. Der erste Schritt der Z/E-Isomerisierung zu Diaminofumarodinitril wird von RAIN nicht generiert, da die aktuelle Programmversion keine stereochemischen Aspekte berücksichtigt.

J.P. FERRIS und J.E. KUDER haben diesen kritischen Reaktionsschritt der Abiogenese von Adenin einer eingehenden mechanistischen Analyse unterzogen²³⁴. Ihre Ergebnisse stehen in gänzlicher Übereinstimmung mit der Reaktionssequenz von Abbildung 4.18.

Kapitel 5

Genetische Algorithmen

Zusammenfassung:

Genetische Algorithmen sind vielseitig einsetzbare Optimierungsverfahren, die innerhalb einer Population von Problemparametern mittels eines evolutionären Prozesses bestmöglich angepaßte Parametersätze finden können. Eine der ersten chemischen Anwendungen dieser Verfahren war die Berechnung der Chemischen Distanz. Hier wird die Fähigkeit der Genetischen Algorithmen deutlich, selbst aus riesigen Lösungsräumen mit relativ kleinem Rechenaufwand (sub)optimale Lösungen herauszufinden. Nach einer allgemeinen Einführung werden die speziellen Techniken der Genetischen Algorithmen, die zur Berechnung der Chemischen Distanz angewandt werden, vorgestellt.

5.1 Einführung

Genetische Algorithmen sind Optimierungsmethoden, die auf den Prinzipien der Darwinistischen Evolution beruhen. Die Grundidee besteht darin, eine künstliche Population von "Individuen" zu erzeugen, deren Mitglieder mittels genetischer Operatoren interagieren, und auf diese Weise den Optimierungsvorgang ausführen.

Ein Individuum ist in der Population repräsentiert durch sein "Chromosom", in den meisten Fällen ein linearer Bitstring, der durch eine Codierung der zu optimierenden Parameter definiert ist. Die Gesamtheit der Chromosomen der Population werden zu Anfang des Verfahrens mit zufälligen Bitsequenzen initialisiert.

Zur relativen Bewertung der unterschiedlichen Individuen einer Population wird eine "Fitnessfunktion" verwendet, die meist direkt das zu optimierende Problem repräsentiert. Die Fitnessfunktion ermittelt für einen Chromosomenstring die relative "Güte" der Parameterzusammenstellung dieses Individuums, im Vergleich zu allen anderen Mitgliedern der Population.

Ähnlich wie bei der natürlichen Vererbung genetischer Eigenschaften von den Eltern an ihre Kinder wird auch beim Genetischen Algorithmus eine Nachfolgeneration von Individuen erzeugt. Im allgemeinen werden die besten Individuen einer Population unverändert in diese Tochtergeneration aufgenommen ("Selektive Reproduktion"). Anschließend werden Paare von Individuen ausgewählt und deren Chromosomen nach speziellen Regeln gemischt ("Crossing over"). Die hierdurch geschaffenen neuen Individuen werden ebenfalls in die Tochtergeneration übernommen. Die Auswahl der Individuen für das Crossing over geschieht aufgrund deren relativer Fitness, d.h. Populationsmitglieder mit hoher Fitness werden mit größerer Wahrscheinlichkeit zum Crossing over herangezogen, als "schlechter angepasste" Individuen mit niedrigerer Fitness.

Zusätzlich zum Crossing over finden beim Vorgang der Reproduktion mit einer bestimmten (im allgemeinen sehr kleinen) Wahrscheinlichkeit zufällige Veränderungen der Chromosomen statt ("Mutationen"). Diese Mutationen sorgen dafür, daß eine gewisse genetische Variation innerhalb der Population erhalten bleibt. Ansonsten würden die Operatoren der selektiven Reproduktion und des Crossing over in einem meist raschen Konvergenzvorgang zu einer einheitlichen Population führen, die zu keiner weiteren Evolution fähig wäre.

Im wesentlichen besteht ein Genetischer Algorithmus nun aus einer kontinuierlichen Abfolge von Generationswechselln, wobei jede Nachfolgeneraion aus der Vorgän-

gergeneration durch die Operatoren Selektion, Crossing over und Mutation entsteht (siehe Abbildung 5.1). Da der Selektionsprozeß bevorzugt Individuen höherer Fitness auswählt, werden im Zuge der Iterationen Individuen erzeugt, die einer optimalen Lösung des zugrundeliegenden Problems immer näher kommen.

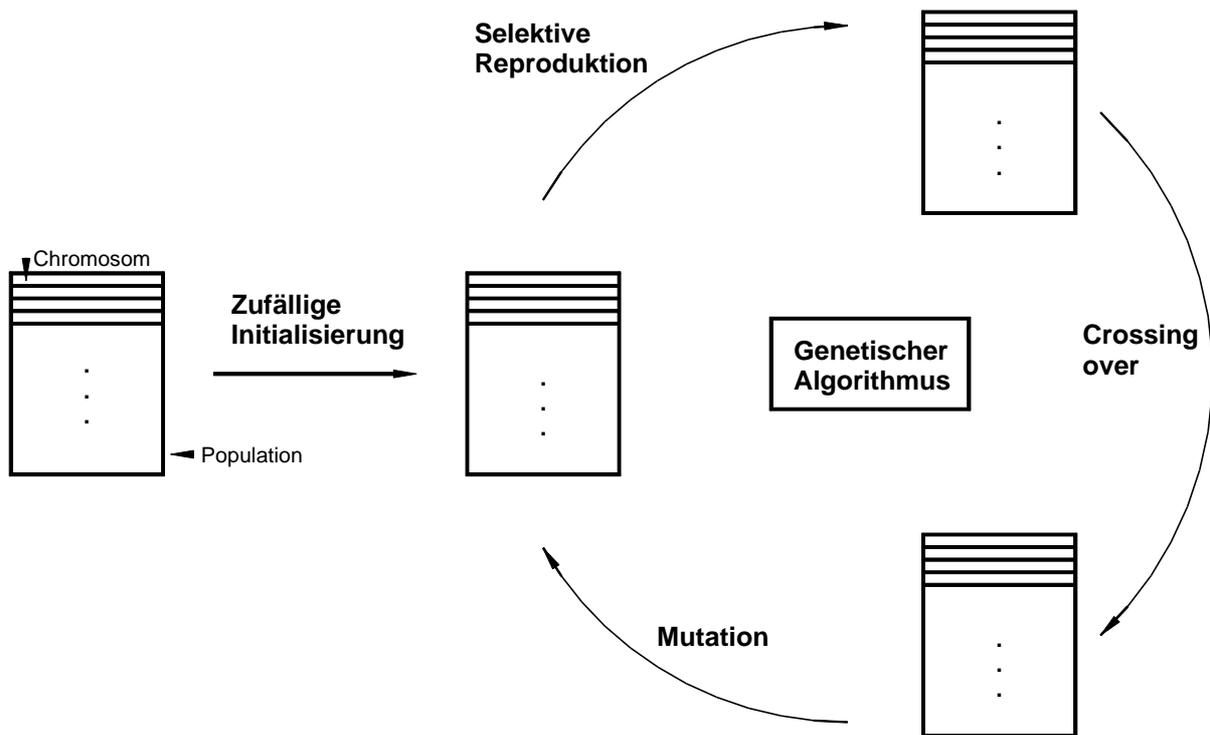


Abb. 5.1 Schematischer Ablauf eines Genetischen Algorithmus.

Nach einer vorher festgelegten Anzahl von Generationscyclen, oder bedingt durch ein anderes Abbruchkriterium, endet der Genetische Algorithmus. Das bis dahin beste gefundene Individuum (der Chromosomensatz mit der höchsten Fitness) stellt eine suboptimale Lösung des Problems dar. Diese Lösung liegt mit hoher Wahrscheinlichkeit sehr nahe bei dem Optimum, bzw. erreicht den bestmöglichen Wert, ohne daß hierzu letztendlich eine definitive Aussage möglich ist. Dies ist eine Eigenschaft die der Genetische Algorithmus mit allen stochastischen Optimierungsverfahren (siehe unten) gemein hat.

Die Genetischen Algorithmen wurden von J.H. HOLLAND an der Universität von Michigan bereits Mitte der 60er Jahre als Teil einer Arbeit über Adaptive Prozesse entwickelt ^{235, 236, 237}. Im Jahre 1992 wurde J.H. HOLLAND für diese wegweisenden Arbeiten mit dem MACARTHUR Award ausgezeichnet. Als vielseitige und äußerst ro-

buste Optimierungswerkzeuge werden die Genetischen Algorithmen seit dem in vielen Gebieten der Technik und Forschung eingesetzt, vor allem dann, wenn Komplexität oder Größe des Optimierungsraums andere Verfahren als ungeeignet erscheinen lassen. Einsatzgebiete für Genetische Algorithmen (außerhalb der Chemie) sind u.a.:

- Optimierung hochbeanspruchter Teile (z.B. Turbinenschaufeln)
- Bildverarbeitung (z.B. in der Medizin)
- Wettervorhersage
- Adaptives Filtern (z.B. zur Rauschunterdrückung)
- Operations Research (z.B. Ablauf- bzw. Kapazitätsplanung)
- Design von Computernetzwerken
- Pipeline-Optimierung
- Roboteranwendung (z.B. zur Trajektorienoptimierung)

Auch in der Chemie finden Genetische Algorithmen zunehmend Anwendung, vor allem im Bereich der chemischen Analytik. Aber auch in anderen Gebiete innerhalb der Chemie werden diese neuen Optimierungstechniken immer häufiger eingesetzt. Dies drückt sich in einer stark ansteigenden Anzahl von Publikationen über Anwendungen Genetischer Algorithmen aus (siehe Abbildung 5.2).

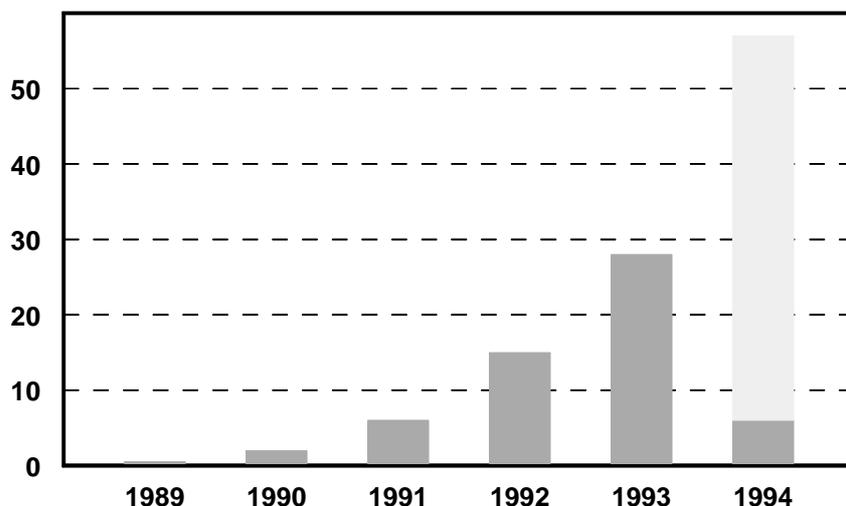


Abb. 5.2 Zunahme der Publikationen über die Anwendung Genetischer Algorithmen im Bereich der Chemie (Quelle: Chemical Abstract Services, Stand: Juni 1994).

Chemische Einsatzgebiete^{238, 239} für Genetische Algorithmen finden sich hauptsächlich dort, wo äußerst komplexe Suchräume mit einer großen Zahl von Parametern oder vielen lokalen Minima andere Optimierungstechniken scheitern lassen. Die im folgenden vorgestellte Berechnung der Chemischen Distanz war eine der ersten Anwendungen von Genetischen Algorithmen in der Chemie und ihre Publikation beschreibt den ersten Einsatz dieser Methode auf dem Gebiet der chemischen Konstitution. Andere chemische Anwendungsgebiete Genetischer Algorithmen sind u.a.:

- Konformationsanalyse kleinerer organischer Moleküle^{240, 241, 242, 243, 244}
- Konformationsanalyse von Biopolymeren^{245, 246, 247, 248, 249}
- Modelle zur Proteinfaltung^{250, 251, 252, 253, 254, 255}
- DNA-Mapping²⁵⁶
- Abschätzung reaktionskinetischer Parameter²⁵⁷
- Prozeßoptimierung^{258, 259, 260, 261, 262, 263, 264, 265, 266}
- Grafische Darstellung von Molekülen²⁶⁷
- Ermittlung der Minimalen Chemischen Distanz^{204, 205, 268}
- Ermittlung größter gemeinsamer Substrukturen²⁰⁶
- Bestimmung molekularer Ähnlichkeit^{204, 206, 269}
- Substruktursuche und Hyperstrukturermittlung^{270, 271}
- Optimierung spektroskopischer Parameter^{272, 273, 274, 275, 276, 277}
- NMR-Spektroskopie^{278, 279}
- Chemometrie^{280, 281, 282, 283, 284, 285}
- Ökologie und Umweltschutz^{286, 287, 288}

5.2 Optimierungsstrategien

Optimierungsaufgaben gehören heute zum Standardrepertoire jedes theoretischen und angewandten Wissenschaftszweiges. Es geht hierbei stets darum, unter mehreren möglichen alternativen Realisationen eine bessere, oder die beste (optimale) herbeizuführen. Definiert wird ein zu optimierendes Problem durch die Abhängigkeit einer Variablen o von einem oder mehreren Parametern (Merkmalen) x_j . Die Zielfunktion Φ beschreibt die "Güte" eines bestimmten Parametersatzes.

$$o = \Phi(x_1, x_2, \dots, x_n)$$

Ziel eines Optimierungsverfahrens ist es, einen Parametersatz zu finden, für den o einen maximalen (oder minimalen) Wert annimmt. In manchen Fällen kann es auch vorkommen, daß o kein einfacher Skalar ist, sondern wiederum ein Vektor von verschiedenen "Gütekriterien".

Jedes Minimierungsproblem kann (z.B. durch Inversion der Zielfunktion an einem Referenzwert) in ein Maximierungsproblem umgewandelt werden. Genetische Algorithmen optimieren einen Parametersatz immer in Richtung einer maximal günstigen Anpassung, sie können damit jedoch ohne Beschränkung der Allgemeingültigkeit auch zur Minimierung jeder beliebigen Zielfunktion herangezogen werden.

Die Komplexität des Optimierungsproblems wird nun hauptsächlich bestimmt durch die Struktur der Zielfunktion Φ und durch die Parameteranzahl n . Abbildung 5.3 gibt einen Überblick über verschiedene Möglichkeiten der Abhängigkeit der Variablen o von den Parametern x .

Im allgemeinen werden *direkte* und *indirekte* Optimierungsmethoden unterschieden. *Direkte* Optimierungsmethoden lösen das Problem durch eine mathematisch-analytische Vorgehensweise. Diese deduktiven Methoden sind jedoch nur anwendbar, wenn die Zielfunktion mathematisch definiert und differenzierbar ist (z.B. Abbildung 5.3 links oben).

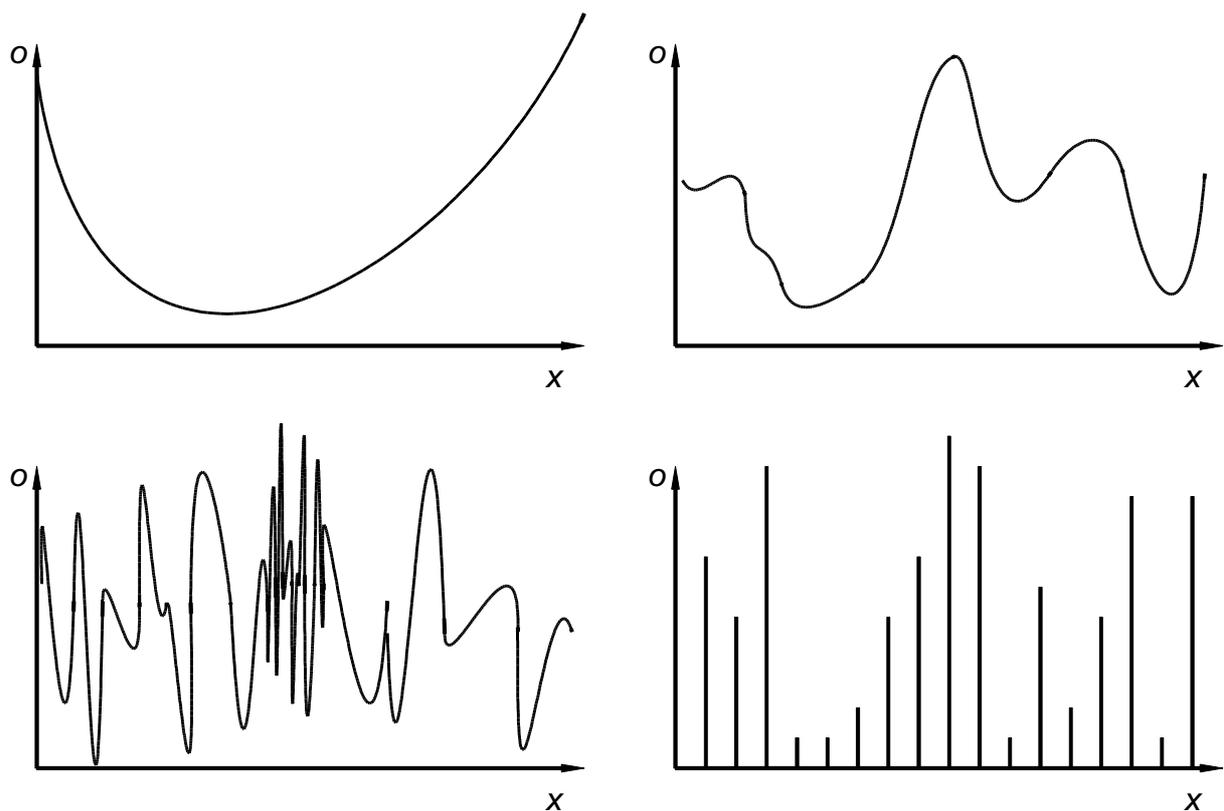


Abb. 5.3 Die Komplexität eines Optimierungsproblems hängt ab von der Struktur der Zielfunktion. Das Bild links oben zeigt eine einfach zu optimierende Zielfunktion mit nur einem Minimum im Definitionsbereich. Die Komplexität steigt mit der Zunahme der lokalen Minima (Bilder rechts oben und links unten). Den Extremfall bilden die diskontinuierlichen Zielfunktionen der kombinatorischen Optimierungsprobleme (Bild rechts unten).

Der weitaus größere Teil realer und anwendungsorientierter Optimierungsprobleme befaßt sich jedoch mit schlecht definierten, nicht ableitbaren, diskontinuierlichen, oder diskreten Zielfunktionen (siehe z.B. Abbildung 5.3 unten). In diesen Fällen werden *indirekte* (iterative) Optimierungsstrategien herangezogen. Hierbei wird, ausgehend von einem Startpunkt, der Lösungsraum in einer Folge einzelner Schritte entlang einer bestimmten Trajektorie durchmessen. Dies erfolgt mit dem Ziel, dem Optimum möglichst nahe zu kommen. Die einzelnen Schritte werden meist regelmäßig gesetzt oder nach dem Prinzip des "trial and error" gewählt, wobei hier verschiedene Heuristiken zum Einsatz kommen können.

Sogenannte "hill climbing" Strategien versuchen, durch Bildung lokaler Ableitungen der Zielfunktion die Richtung und die Größe des nächsten Schrittes zu bestimmen. Unabhängig von der Technik²⁸⁹, mit der dies erfolgt (z.B. Simplex, Steepest Descent, Fletcher-Reeves, Polak-Ribiere, Newton-Raphson, oder Block Diagonal Newton-Raphson), haben diese Strategien den Nachteil, daß sie lediglich das dem Ausgangs-

punkt nächstliegende lokale Optimum auffinden können, da sie nicht auf die Akzeptanz von lokalen Verschlechterungen ausgelegt sind. Wegen ihrer Zielstrebigkeit und ihres meist guten Zeitverhaltens werden diese Methoden in einer Vielzahl von Programmen zur Strukturoptimierung (Molecular Mechanics) eingesetzt. Für diskontinuierliche oder diskrete Zielfunktionen sind diese Methoden jedoch nicht einsetzbar, da hier keine lokale Differenzierbarkeit der Zielfunktion gegeben ist.

Optimierungsmethoden, die weniger starke Annahmen über die Struktur des Suchraumes machen, als die oben genannten "hill climbing" Strategien, müssen sich eines anderen Verfahrens bedienen, um ihre Trajektorie im Problemraum aufzubauen.

Eine komplette Durchmusterung des Suchraumes (*Enumerative Suche*) durch systematische Variation aller Parameter in einer vorgegebenen (möglichst kleinen) Schrittweite garantiert, daß das globale Optimum gefunden wird. Diese Methode verbietet sich jedoch meist von selbst, da die Suchräume realistischer Probleme mit einer Größenordnung von beispielsweise 10^{100} die Rechenkapazitäten der Computer weit überfordern. Hierbei sei angemerkt, daß das Vertrauen auf eine Weiterentwicklung in der Computertechnologie auch nicht weiterhilft, da selbst für moderate Probleme bei einer angenommenen und physikalisch begründeten²⁹⁰ höchstmöglichen Informationsübertragungsrate von etwa 10^{47} Bits pro Sekunde und Gramm Computermasse ein Rechner mit der Masse des gesamten Universum und eine Rechenzeit vom Urknall bis heute auch nicht andeutungsweise ausreichend wäre.

Dieselben Einschränkungen gelten auch für eine weitere "blinde" Suchmethode, der *Monte Carlo* Suche. Hier wird der Suchraum unvoreingenommen und zufällig abgetastet, das heißt es werden Stichproben durchgeführt und der bis zu jedem Schritt beste gefundene Parametersatz (oder eine Zusammenstellung mehrerer) stellt die gefundene suboptimale Lösung dar. Um für größere Suchräume auch nur mit einer mittleren Wahrscheinlichkeit in die Nähe des Optimums zu kommen, müssen Monte Carlo Verfahren eine große Anzahl von Stichproben durchführen, sodaß der Aufwand in die Größenordnung der einer enumerativen Suche gelangt.

Eine Verquickung einer stochastischen Methode mit einem lokalen Absuchen der Umgebung des aktuellen Punktes auf der Suchtrajektorie stellen die verschiedenen Evolutionsstrategien²⁹¹ dar (eine gute Übersicht über Zufalls- und Evolutionsstrategien im Vergleich zu anderen Optimierungsmethoden gibt H.P. SCHWEFEL²⁹²). Hierbei wird für ein Pseudogenom, das codiert die zu variierenden Parameter enthält (Vater), nach bestimmten Regeln ein mutierter (zufällig veränderter) Nachkomme erzeugt. Über eine Fitnessfunktion wird nun entschieden, welches dieser

beiden Individuen (Vater oder Nachkomme) in die nächste Generation übernommen wird. Für das Individuum der Folgegeneration wird im nächsten Generationszyklus nun wieder versucht, einen "besseren" Nachkommen zu erzeugen. Die Evolutionsstrategien und andere Zufallsstrategien²⁹³ (wie auch das im Abschnitt 4.2.3 erwähnte Simulated Annealing) bilden bereits einen Übergang zu den Genetischen Algorithmen.

Der wesentliche Unterschied zu den Genetischen Algorithmen besteht jedoch darin, daß die Zufalls- und Evolutionsstrategien nur *einen* Punkt im Suchraum behandeln und diesen mehr oder weniger gezielt bewegen. Genetische Algorithmen erfassen den Suchraum mit einer ganzen *Population* von Punkten (Größenordnung: 100 bis 1000 Individuen). Dies wäre bei den betrachteten Komplexitäten der Suchräume jedoch nur eine unwesentliche Verbesserung, wenn nicht die Effizienz der Suche durch einen Mechanismus des Informationsaustauschs zwischen den Individuen bei jedem Generationszyklus (Crossing over) zu einer exponentiell ansteigenden Zahl der durchmusterten Punkte im Suchraum führen würde. Die Technik des Crossing over unterscheidet somit die Genetischen Algorithmen *qualitativ* von anderen stochastischen Optimierungsmethoden. D.E. GOLDBERG gibt in seinem Buch²³⁷ über Genetische Algorithmen eine Analyse des informationstheoretischen Hintergrunds ihrer Funktionsweise (HOLLANDs Schematheorem²³⁵).

Unterscheidet man Optimierungstechniken hinsichtlich des Ausmaßes, in dem sie implizit oder explizit Annahmen über das zu optimierende Problem machen²³⁸, so finden sich die Genetischen Algorithmen in einem Mittelfeld zwischen den in großem Maße auf eine Problemkategorie zugeschnittenen Verfahren mit "starken" Annahmen und den allgemein anwendbaren Verfahren mit "schwachen" Annahmen. Im allgemeinen wird die generelle Anwendbarkeit einer Optimierungstechnik durch eine schlechte Effizienz bezahlt (wie z.B. bei der enumerativen Suche). Diese Verfahren zeichnen sich jedoch naturgemäß durch eine große Robustheit gegenüber veränderten Anforderungen und Spezialsituationen aus. Im Gegensatz hierzu führen die "starken" Annahmen der anderen Klasse von Optimierungsverfahren zu einer oft schnellen Konvergenz (hohe Effizienz), wobei allerdings die Anfälligkeit gegenüber lokalen Störungen im Problemraum und gegenüber Veränderungen der Problemdefinition hier sehr hoch ist. Da Genetische Algorithmen mit ihrer Technik der impliziten Parallelität und ihrer "schwachen" Problemkenntnisse die Vorteile beider Richtungen vereinen, kommt ihnen eine wichtige Rolle als sehr robustes und (meist) effizientes Optimierungsverfahren zu.

5.3 Techniken der Genetischen Algorithmen

Der Ablauf eines Genetischen Algorithmus und seine innere Struktur eignen sich hervorragend für eine streng modulare Konstruktion entsprechender Programme. Zumindest für einfach Genetische Algorithmen (Basic Genetic Algorithms) mit einer einheitlichen Kodierung der Parametersätze in Genome (meist auf der Basis von binären Bitstrings) lassen sich unabhängige Module für Selektion, Fitnessberechnung, Mutation, und Crossing over konstruieren. Der zunehmende Einsatz der Genetischen Algorithmen hat darüberhinaus dazu geführt, daß für verschiedenste Plattformen bereits vorgefertigte und vielfältig einsetzbare Programme angeboten werden, wie z.B. GATES²⁹⁴ (Genetic Algorithm Toolbox for Evolutionary Search), GENESIS²⁹⁵ und GENITOR²⁹⁶ (GENetic ImplementOR).

Für Genetische Algorithmen mit nicht einheitlich binärer Parameterkodierung oder für Genetische Algorithmen mit inhärenter Problemorientiertheit (Knowledge-augmented Genetic Algorithms) müssen jedoch spezifische Verfahren entwickelt und getestet werden, die sich dann aber als wesentlich effizienter erweisen als Standardmethoden. Die hier vorgestellte Methode der Bestimmung der Chemischen Distanz mittels Genetischer Algorithmen gehört zu dieser Kategorie.

Im folgenden werden die hierfür erarbeiteten Techniken und einige der Ergebnisse vorgestellt. In jedem Abschnitt wird neben einer kurzen Darlegung der entsprechenden Standardmethode die problemspezifische Technik erläutert.

5.3.1 Parameterrepräsentation

In vielen Anwendungen Genetischer Algorithmen werden die Genome der einzelnen Individuen als binäre Bitstrings repräsentiert. Diese entstehen aus einer mehr oder weniger direkten Abbildung der Problemparameter in binärer Form und durch anschließende Aneinanderreihung dieser Bitketten zu einem durchgehenden String (Abbildung 5.4).

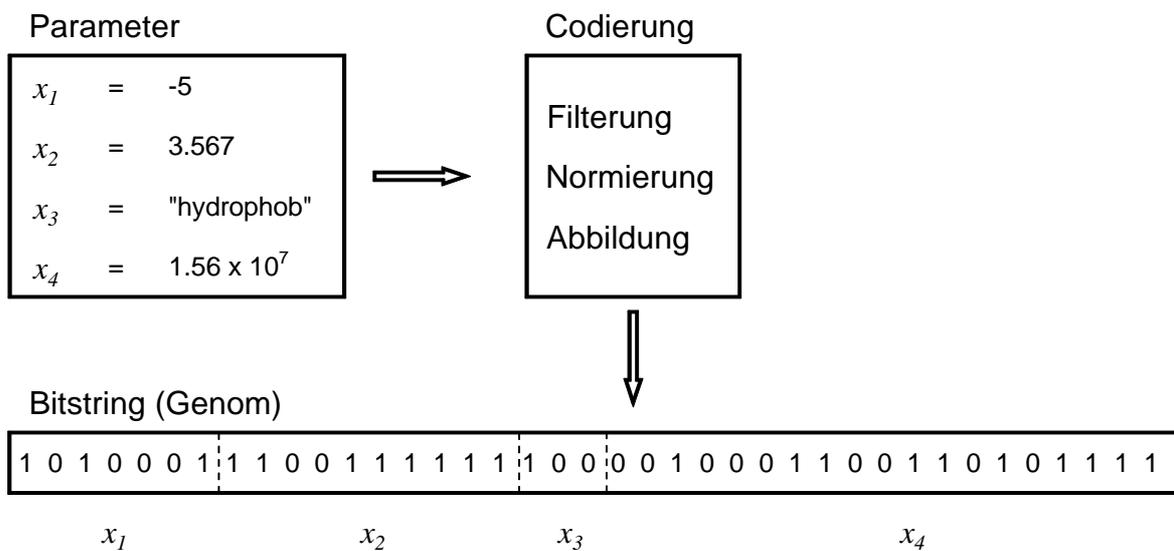


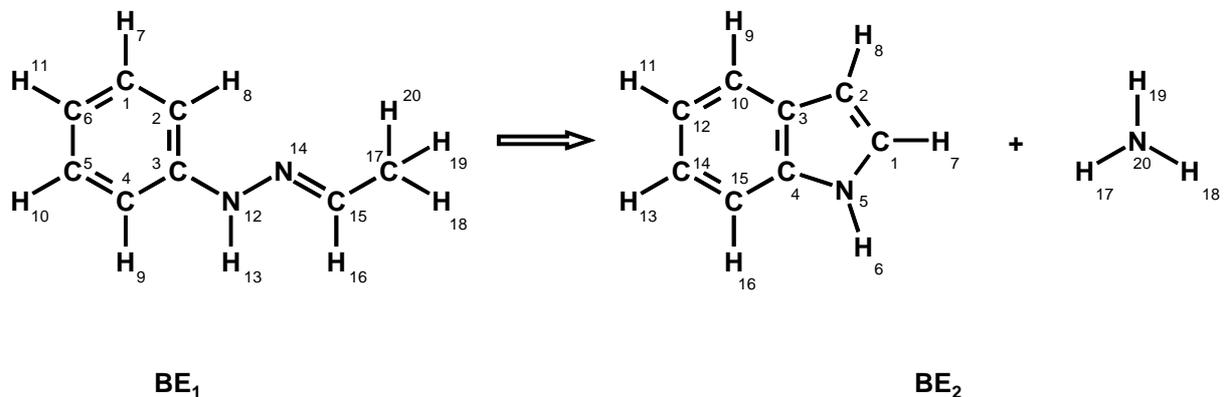
Abb. 5.4 Beispiel für die Abbildung von Problemparametern in einen linearen Bitstring eines Genoms.

Die "Trennstellen" der einzelnen Parameterabbildungen in den Bitstrings sind innerhalb der Strings nicht real vorhanden. Vielmehr muß die auswertende Funktion diese Trennstellen kennen, um aus einem beliebigen String der Population wieder einen Satz von Parametern herzustellen. Ein nicht zu unterschätzendes Problem stellt hierbei dar, daß jede auch nur denkbare Bitkombination (die ja durch Mutation oder Crossing over entstehen kann) immer in einen gültigen Parameterwert zurückübersetzt werden muß.

Die Bestimmung der Minimalen Chemischen Distanz gehört in die Kategorie der Reihenfolgenprobleme (zu der unter anderem auch klassische Aufgabenstellungen wie das Problem des Handlungsreisenden gehören). Eine direkte binäre Kodierung ist hier weniger vorteilhaft, da durch die Mechanismen der Mutation und des Crossing over beliebig veränderte Bitsequenzen meist zu nicht eindeutigen Reihenfolgenvektoren führen würden. Ein Genom besteht deshalb aus einem Permutati-

onsvektor P dessen Element P_i für ein Atom i im Ensemble BE_1 eindeutig das Bildatom dieses Atoms im Ensemble BE_2 bestimmt (Abbildung 5.5)

Die Permutationsvektoren, aus denen zu Anfang des Verfahrens eine Population aufgebaut wird, müssen bestimmten Grundregeln gehorchen. Zuerst muß für sie die Homologie-Regel gelten, d.h. ein Permutationsvektor der Länge N muß die Zahlen von 1 bis N exakt jeweils einmal enthalten. Dies ist eine selbstverständliche Voraussetzung für die ein-eindeutige Zuordnung der Atome.



Permutationsvektor (Genom)

10	3	4	15	14	12	9	17	16	13	11	5	6	20	1	7	2	8	18	19
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Position

Abb. 5.5 Beispiel für ein Genom des Genetischen Algorithmus bei der Berechnung der Minimalen Chemischen Distanz. Das Genom besteht aus einem Permutationsvektor, der die Atome des Ensembles BE_1 auf die Atome des Ensembles BE_2 abbildet (es handelt sich hierbei um die Nettoensembles der Fischerschen-Indolsynthese).

Desweiteren müssen Abbildungen ausgeschlossen werden, die Atome auf Atome mit einer anderen Ordnungszahl zuweisen. Aus einem Sauerstoffatom darf im Zuge der Abbildung kein Kohlenstoffatom werden. Dies kann erreicht werden, indem für jede Position des Permutationsvektors eine "Verträglichkeitsliste" erstellt wird, die die Atomnummern der Zielatome mit gleicher Ordnungszahl enthält. Einfacher ist es jedoch, die Atome der Ensembles nach Ordnungszahlen zu sortieren, und die Permutationsvektoren in Abschnitte zu zerteilen, wobei jeder Abschnitt für eine Elementsorte steht. Alle willkürlichen Veränderungen im Verlauf des Genetischen Algorithmus (Mutationen, etc.) dürfen dann nur innerhalb ein und desselben Permutationsvektorabschnittes erfolgen.

Dasselbe gilt auch für die zufällige Initialisierung der Permutationsvektoren zu Beginn des Genetischen Algorithmus. Es wird hierbei von einer natürlichen Folge der Permutationsvektorelemente von 1 bis N ausgegangen. Anschließend wird mittels einer (großen) Zahl von willkürlichen Zweivertauschungen innerhalb der Vektorabschnitte eine zufällige Abbildung erzeugt, die als Ausgangspunkt des evolutionären Prozesses dient. Dieses Verfahren garantiert Eindeutigkeit (Homologie) und verhindert elementübergreifende Atom-auf-Atom-Abbildungen.

Die Größe der Population (Anzahl M der Individuen) richtet sich am besten nach der Problemkomplexität. Große Populationen sind für komplexere Probleme (größere Ensemble) besser geeignet, da sie eine größere genetische Varianz zulassen. Andererseits bestimmt die Populationsgröße unmittelbar die Rechenzeit für einen Generationszyklus und sie sollte deshalb nicht größer sein als nötig. Für die Bestimmung der Chemischen Distanz innerhalb des Programms RAIN wird eine Populationsgröße von $M = 4 * N$ angenommen (N ist die Anzahl der Atome eines Ensembles BE_1 oder BE_2).

5.3.2 Selektive Reproduktion

Die Individuen der künstlichen Population werden im Verlauf des Genetischen Algorithmus einem Selektionsverfahren unterworfen, das einer Auslese im Sinne des DARWINSchen "survival of the fittest" ähnlich ist. Am Anfang hiervon steht eine Evaluierung jedes Populationsmitglieds. Dies erfordert eine Recodierung des Genoms in die einzelnen Problemparameter und die Berechnung eines Wertes der sogenannten "Fitnessfunktion" für die erhaltene Wertezusammenstellung. Meist ist die Fitnessfunktion direkt von dem zu optimierenden Problem ableitbar, oder gar mit diesem identisch. Anschließend wird für jedes Mitglied der Population entschieden, ob und in welcher Anzahl von Kopien dieses Individuum in der Folgegeneration erscheint. Hohe Werte der Fitnessfunktion sollen hierbei ein Überleben (mindestens eine Kopie) sicherstellen. Niedrige Werte sollen das Individuum nur mit geringer Wahrscheinlichkeit in die Tochtergeneration übernehmen oder es mangels Fitness "aussterben" lassen.

Die Auswahl der überlebenden Individuen kann auf vielfältige Weise geschehen. Im folgenden werden lediglich zwei Verfahren beschrieben, die im Rahmen der Arbeiten zur Minimierung der Chemischen Distanz eingesetzt wurden. In beiden Fällen

wurde als Fitnessfunktion indirekt die von einem Permutationsvektor induzierte Chemische Distanz verwendet. Da es sich bei der angestrebten Lösung um eine *Minimierung* handelt, das genetische Verfahren jedoch die *Maximierung* der Fitness anstrebt, muß bei der Berechnung der Fitness der berechnete Wert der Chemischen Distanz an einem Referenzwert invertiert werden. Für ein Mitglied i einer Population mit M Individuen (Permutationsvektoren $P(i)$ der Länge N) wird der Fitnesswert $F(i)$ gemäß Abbildung 5.6 berechnet:

$CD(i) = \sum_{k,n=1,N} be_{1kn} - P(i)(be_{2kn}) $	Chemische Distanz für Permutation $P(i)$
$CD_{SUM} = \sum_{j=1,M} CD(j)$	Summe aller CD-Werte in der Population
$CD_{AVE} = \frac{1}{M} CD_{SUM}$	Durchschnittliche Chemische Distanz
$CD_{MIN} = \min_{j=1,M} (CD(j))$	Kleinste Chemische Distanz
$CD_{REF} = 2 CD_{AVE} - CD_{MIN}$	Referenzwert für Invertierung
$F(i) = \begin{cases} (CD_{REF} - CD(i))^2 & \text{für } CD(i) \leq CD_{REF} \\ 0 & \text{für } CD(i) > CD_{REF} \end{cases}$	Fitness für Individuum i

Abb. 5.6 Schema der Berechnung des Fitnesswertes F für ein Individuum i einer Population.

Für den Referenzwert CD_{REF} wird ein "Fitnessfenster" angenommen, in dessen Mitte der Durchschnittswert aller Fitnesswerte liegt und das nach unten den besten Fitnesswert der Population (CD_{MIN}) einschließt (Abbildung 5.7)

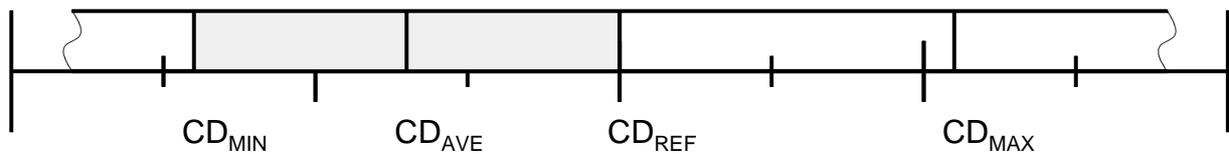


Abb. 5.7 "Fitnessfenster" zur Invertierung der CD-Werte bei der Berechnung der Fitness.

Zur Berechnung der Fitness $F(i)$ werden die Differenzen der entsprechenden Chemischen Distanzwerte zum Referenzwert CD_{REF} quadriert. Die Einführung dieser Nichtlinearität und die vorhergehende Methode der Berechnung des Referenzwertes sind das Ergebnis umfangreicher Optimierungsarbeiten am eingesetzten Verfahren und führen zu einer schnellen Konvergenz. Techniken, die hierbei eine zu starke Homogenisierung der Population verhindern können, werden in Abschnitt 5.3.6 besprochen.

Die selektive Reproduktion der Tochtergeneration berücksichtigt nun jedes Individuum mit einer Wahrscheinlichkeit, die seiner relativen Fitness $F_{REL}(i)$ entspricht.

$$F_{SUM} = \sum_{j=1, M} F(j) \qquad F_{REL}(i) = \frac{F(i)}{F_{SUM}}$$

Die sogenannte "roulette wheel selection" Methode²³⁷ (RWS Methode) benützt zur Auswahl der zu reproduzierenden Individuen eine Zufallszahl R im Bereich zwischen 0 und F_{SUM} . Anschließend werden die Fitnesswerte der Populationsmitglieder solange aufsummiert, bis der Wert R erreicht, bzw. gerade überschritten ist. Das Individuum, bei dem dies der Fall war, wird in die Tochtergeneration kopiert (und somit selektiv reproduziert). Der Vorgang wird M mal wiederholt, bis die Tochtergeneration komplett ist, und den weiteren Manipulationen, wie Mutation und Crossing over (siehe unten) unterzogen werden kann. Die RWS Methode hat den Vorteil, daß jedes Individuum, zumindest nominell, mit der Anzahl von Nachfolgern in der Tochtergeneration vertreten ist, die seiner relativen Fitness entspricht. Sind die Unterschiede zwischen den einzelnen Populationsmitgliedern (noch) nicht sehr groß, so kann es jedoch mitunter vorkommen, daß das Individuum mit der besten Fitness nicht "überlebt", da das RWS Verfahren keine Mindestanzahl von Kopien garantiert. Eine Möglichkeit, hier einzugreifen, besteht in der sogenannten "elitist" Methode, die bei jeder Generation das Individuum mit der höchsten Fitness auf jeden Fall

(unabhängig von anderen Auswahlmechanismen) mindestens einmal in die Tochtergeneration übernimmt.

Eine alternative Methode ²⁹⁷ der selektiven Reproduktion berechnet für jedes Individuum zunächst eine erwartete Kopienanzahl ("expected number count" $ENC(i)$).

$$ENC(i) = F_{REL}(i) * M$$

Der ganzzahlige Anteil von $ENC(i)$ bestimmt nun die Mindestanzahl von Kopien, die von einem Individuum angefertigt werden. Der Nachkommaanteil von $ENC(i)$ bestimmt dann die Wahrscheinlichkeit, mit der eine zusätzliche Kopie des Genoms in die Tochtergeneration aufgenommen wird. Ein Individuum mit durchschnittlicher Fitness wird demgemäß exakt einmal kopiert. Das Verfahren des ENC garantiert das Überleben des Individuums mit der höchsten Fitness und führt zu einem stärkeren Selektionsdruck als die RWS Methode. Die ENC Technik erfordert jedoch variable Populationsgrößen, da nach Berechnung der ENC-Werte mit einer angestrebten Populationsgröße M , die tatsächliche Individuenzahl etwas kleiner oder größer als M sein kann.

5.3.3 Crossing over

Zu den wichtigsten Mechanismen der biologischen Evolution gehört die effektive Vermischung genetischer Information und die dadurch bedingte große Vielfalt im Wettbewerb stehender Ausprägungen der Gene. Durch ständige Neukombination der Chromosomen infolge geschlechtlicher Vermehrung wird eine Variabilität und Anpassungsmöglichkeit garantiert, die durch Mutationsvorgänge alleine ohne weiteres nicht erreichbar wäre. Beim Vorgang des biologischen Crossing over wird während der Reifeteilung (Meiose) Genmaterial zwischen homologen Chromosomen umverteilt. In der frühen Prophase werden dabei durch Bruch und kreuzweises Wiedervereinigen gleichlange Abschnitte ausgetauscht und es entstehen sogenannte rekombinante Chromatiden mit gemischten Genen.

Der Vorgang des Crossing over ist ein wichtiger, wenn nicht der bedeutendste Unterschied von Genetischen Algorithmen im Vergleich zu anderen evolutionären Op-

timierungsverfahren (siehe Abschnitt 5.2). Es werden willkürlich zwei Elterngenome aus der Population entnommen und nach der Durchmischung der genetischen Information wieder in die Population zurückgespeichert. In einigen Varianten geschieht das Crossing over auch bereits beim Aufbau der Tochterpopulation während der Phase der selektiven Reproduktion.

Bei der Codierung in Form binärer Bitstrings werden im einfachsten Fall zwei zufällige Schnittstellen gewählt, die ein auszutauschendes Zwischen-, oder Endstück definieren (siehe Abbildung 5.8).

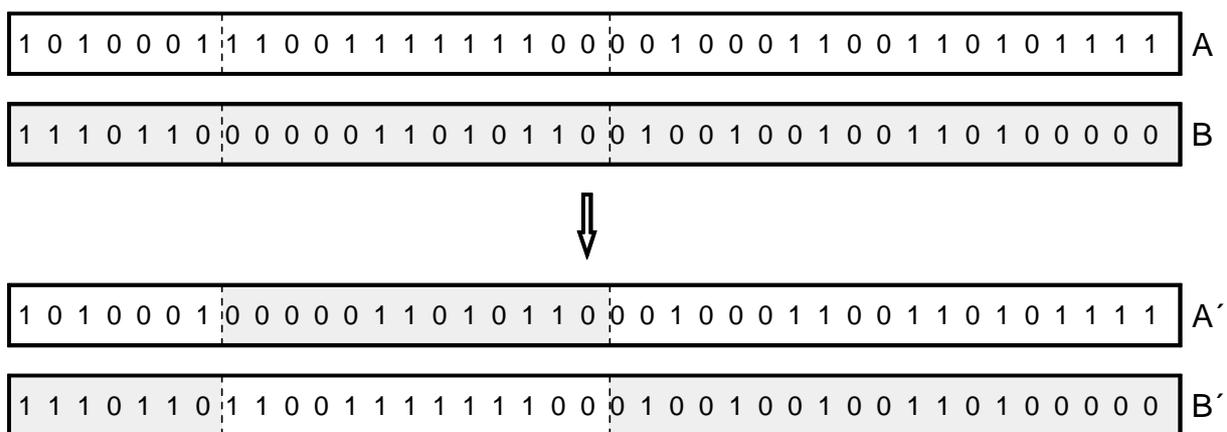


Abb. 5.8 Vorgang des Crossing over bei binären Bitstrings.

Damit die Randstücke der zu kreuzenden Bitstrings keine Sonderstellung einnehmen, werden die Genome als Ringvektoren betrachtet, das heißt, daß auch Tauschsegmente möglich sind, die über das Ende der Bitstrings hinaus in den Anfangsbereich hineinragen.

Bei einem Reihenfolgenproblem ist nun ein einfaches Austauschen von Genbereichen nicht möglich, da hier nahezu immer das Homologieprinzip verletzt werden würde. Es wäre sehr unwahrscheinlich, daß solch ein Crossing over in gültigen Permutationsvektoren resultieren würde. Nun könnte natürlich das Verfahren der selektiven Reproduktion über eine starke Abwertung solch ungültiger Permutationvektoren dafür sorgen, daß diese im Zuge des Generationswechsels eliminiert werden, aber dies würde die Effizienz des Algorithmus stark herabsetzen, da die Population ständig mit einer unnötig großen Zahl von chancenlosen Individuen "belastet" wäre.

Für die Bestimmung der Chemischen Distanz am besten geeignet hat sich ein Crossing-over Verfahren, daß als "partially matched crossover" bezeichnet wird, und wie

viele dieser Methoden als erstes bei der Behandlung des Problems des Handlungsreisenden eingesetzt wurde²⁹⁷. Bei der partially matched crossover Methode wird zunächst, wie bereits oben beschrieben, durch zufällige Wahl zweier Schnittpunkte in beiden zu kreuzenden Permutationsvektoren ein gleich langes auszutauschendes Segment definiert. Anschließend werden die positionsgleichen Elemente beider Segmente innerhalb der jeweiligen Permutationsvektoren ausgetauscht. Stehen sich also zum Beispiel innerhalb solch eines Segments zweier Permutationsvektoren *A* und *B* die Zahlen *x* und *y* gegenüber, so wechseln in *A* und *B* die Elemente *x* und *y* ihren Platz (siehe Beispiel in Abbildung 5.9).

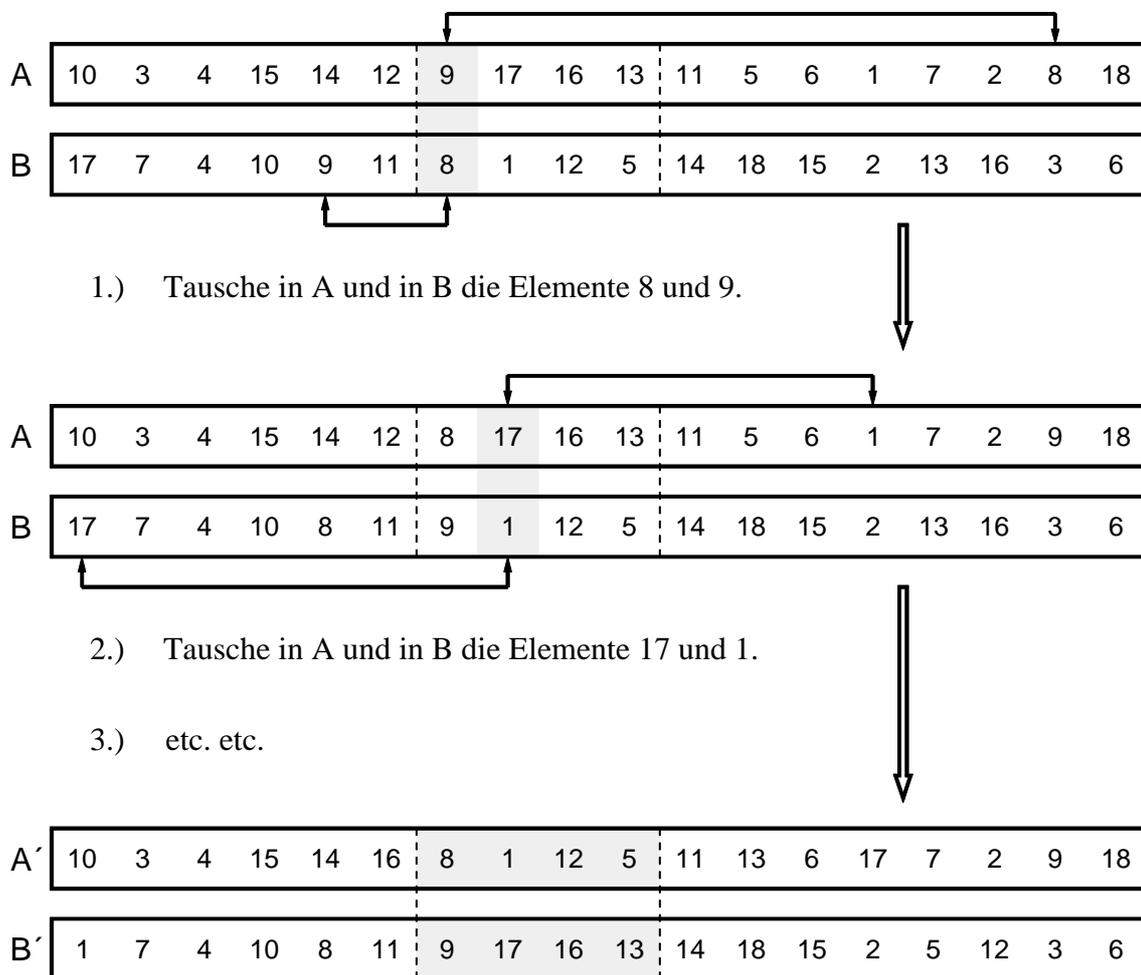


Abb. 5.9 Die Methode des Partially matched crossover gewährleistet, daß trotz des Segmentaustauschs zweier Genome *A* und *B* wieder gültige (homologe) Permutationsvektoren *A'* und *B'* entstehen.

Enthalten die auszutauschenden Segmente teilweise übereinstimmende Zahlen, so ist darauf zu achten, daß bei der Bestimmung der zu wechselnden Elemente die

Ausgangsfolge und nicht der aktuelle gerade im Wechsel befindliche Permutationsvektor zu berücksichtigen ist. Andernfalls kommt es außer zum Segmentaustausch noch zu zusätzlichen Elementpermutationen ²⁶⁸.

Zu bemerken ist hierbei, daß für die Ermittlung der Chemischen Distanz beim partially matched crossover keine Maßnahmen getroffen werden müssen, um elementübergreifende Abbildungen zu vermeiden. Dies geschieht implizit, da nur Vektorelemente miteinander vertauscht werden, die in den Ausgangsvektoren gültige Abbildungen bewirkten. Auf diese Weise ist es unmöglich, daß eine elementübergreifende Abbildung entsteht.

Das Crossing over wird für jedes Individuum der Population mit einer Wahrscheinlichkeit p_{CR} durchgeführt. Der Wert von p_{CR} ist, wie auch andere Operationsparameter das Ergebnis ausgiebiger Testreihen und Optimierungsvorgängen. Da bei jedem Vorgang des Crossing over noch ein zweiter Partner zufällig ausgewählt wird, und beide Genome nach dem Crossing over wieder in die Population (im allgemeinen verändert) zurückkopiert werden, sind für p_{CR} nur Werte unter 0,5 sinnvoll, da sonst in der Regel jeder Permutationsvektor in jeder Generation verändert wird. Für die Berechnung der Chemischen Distanz innerhalb des Programms RAIN wird für p_{CR} ein Wert von 0,4 angenommen.

Bemerkung:

Alle für die Anwendung des Genetischen Algorithmus relevanten Parameter (Crossing-over-Wahrscheinlichkeit, Mutationsrate etc.) sind im Programm RAIN mit Werten besetzt, die das Ergebnis ausgiebiger Studien mit Referenzproblemen waren. Sie sind somit für einen großen Bereich des Anwendungsfeldes von RAIN ohne weiteren Veränderungen zu übernehmen.

5.3.4 Mutation

Bei einer begrenzten Populationsgröße führt die selektive Reproduktion zwangsläufig zu einer immer weiter fortschreitenden Homogenisierung der Population. Da auch die Durchmischung durch das Crossing over hier keine Veränderungen herbeiführen kann (ein Crossing over von zwei ähnlichen oder identischen Permutationsvektoren führt wieder zu zwei ähnlichen oder identischen Permutationsvektoren), muß ein zusätzlicher stochastischer Prozeß die genetische Vielfalt in der Population erhalten.

In Genetischen Algorithmen mit bitcodierter Parameterrepräsentation wird bei der selektiven Reproduktion für jedes Bit mit einer (meist sehr kleinen Mutationswahrscheinlichkeit) eine Inversion (0!Ungültige Zeicheneinstellungach Position eine kleine oder eine große Veränderung desjenigen Parameters bewirkt, in dessen Codierungsbereiches dieses Bit liegt, können durch solch eine Mutation auch Genome erzeugt werden, die vollkommen neue Gebiete des Suchraumes erschließen.

Aus den gleichen Gründen, die für das Crossing over gelten, können bei Reihenfolgenproblemen auch keine willkürlichen Veränderungen (Mutationen) an einzelnen Stellen des Genoms vorgenommen werden, ohne die Homologie des betreffenden Permutationsvektor durch eine Korrektur an anderer Stelle wieder herzustellen. Aus diesen Gründen wird für eine Mutation ein einfacher Zweiertausch innerhalb des Permutationsvektors eingeführt (siehe Abbildung 5.10).

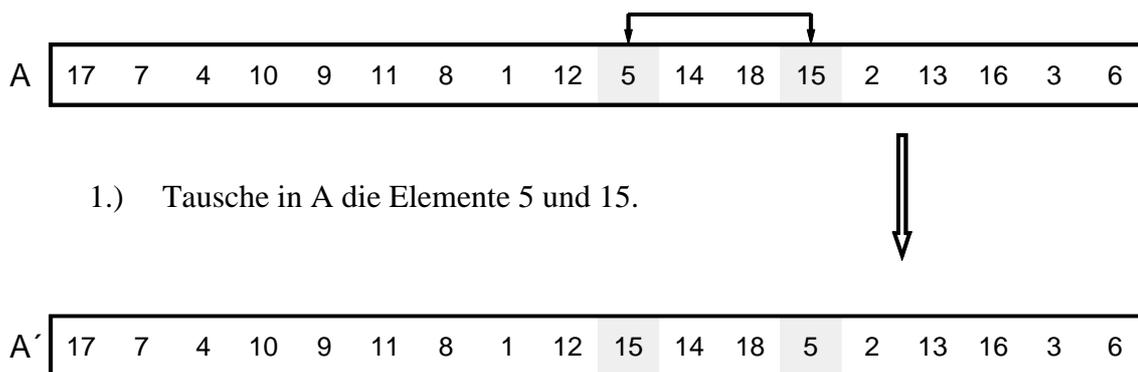


Abb. 5.10 Bei der Mutation eines Permutationsvektor werden zwei zufällig ausgewählte Vektorelemente miteinander vertauscht.

Bei der Auswahl der Mutationspositionen muß darauf geachtet werden, daß diese miteinander verträglich sind, das heißt, daß die zugehörigen Atompositionen in den Ensembles für Atome der selben Elementsorte stehen. Bei einer vorangegangenen Sortierung genügt es, die Mutationen auf einen willkürlich ausgewählten Permutationsvektorabschnitt zu begrenzen.

Die Mutation wird nach dem Crossing over für jedes Individuum der Population mit einer Wahrscheinlichkeit von p_{MU} durchgeführt. Für die Berechnung der Chemischen Distanz innerhalb des Programms RAIN wird für p_{MU} ein Wert von 0,2 angenommen. Dieser Wert liegt um ca. 2 Größenordnungen höher, als er für Mutationsraten von "klassischen" bitcodierten Genetischen Algorithmen in der Literatur angegeben wird. Da es sich hierbei jedoch nicht um die Wahrscheinlichkeit für die Mutation eines einzelnen Bits innerhalb der gesamten Population handelt, sondern um

eine auf die Veränderung eines Individuum bezogene Größe, führt dieser optimierte Wert in den meisten Fällen zu einem angemessenen Ausgleich der Konvergenzbestrebungen durch die selektive Reproduktion.

Die Mutationsrate kann auch dynamisch an die Homogenität der Population angepaßt werden, um Problemen einer vorzeitigen Konvergenz durch zu starken Selektionsdruck auszuweichen (siehe auch Abschnitt 5.3.6).

5.3.5 Wissensbasierte Operatoren

Gewöhnliche Genetische Algorithmen (sogenannte "basic genetic algorithms") besitzen als einzige Verbindung zur Problemstellung die Codierung der Fitnessfunktion, die das zu optimierende Problem in irgendeiner Form enthält. Die große Robustheit und die allgemeine Anwendbarkeit der Genetischen Algorithmen liegt gerade in dieser Beschränkung und der konsequenten Trennung zwischen Propagation und Evaluierung.

Die Effizienz der Genetischen Algorithmen kann jedoch, wie dies bei den meisten Optimierungsverfahren möglich ist, durch eine Implementierung spezifischer Kenntnisse über das zu optimierende Problem stark verbessert werden, wobei natürlich stets ein Verlust der allgemeinen Anwendbarkeit in Kauf genommen werden muß. Diese Implementierung kann durch veränderte Operatoren (wie zum Beispiel einer zielgerichteten Mutation) oder durch neu eingeführte zusätzliche Operatoren geschehen. Im folgenden wird ein wissensbasierter Operator vorgestellt, der zusammen mit den anderen Operatoren (Replikation, Mutation und Crossing over) einen sogenannten "knowledge augmented genetic algorithm" bildet.

Das Auffinden einer Atom-auf-Atom-Abbildung die die strukturellen Differenzen *minimiert*, impliziert, wie bereits erwähnt, eine *Maximierung* der strukturellen Überlappungen. Dies resultiert in den meisten Fällen in größeren zusammenhängenden Struktureinheiten, die ohne Änderung direkt aufeinander abgebildet werden können. Ein wissensbasierter genetischer Operator kann diese Tatsache ausnützen. Dieser Operator berücksichtigt bei der Atom-auf-Atom-Abbildung auch die unmittelbar Nachbarschaftssphären der involvierten Atome. Die Aktion des sogenannten "Reordering Operator" ist wie folgt definiert ²⁰⁴:

Wenn ein Atom **A** des Ensembles **BE**₁ auf ein Atom **B** des Ensembles **BE**₂ abgebildet wird, dann werden die Abbildungen der Nachbaratome von **A** so verändert, daß sie

auf Nachbaratome von **B** zeigen (siehe Abbildung 5.11). Die Bindungsordnungen der Bindungen zu den Nachbaratomen werden hierbei nicht betrachtet. Außerdem werden bei jedem Umordnungsprozeß die Nachbaratome zufällig sortiert, um die stochastische Natur des Gesamtprozesses zu erhalten.

Der Reordering Operator sorgt demnach dafür, daß ein gegebener Permutationsvektor nicht nur willkürlich einzelne Atome aufeinander abbildet, sondern daß dabei ganze Sphären des Radius 1 von einem Ensemble auf das andere erhalten bleiben. Dieser Operator wird in der RAIN-Implementierung für jeden Permutationsvektor mit der Wahrscheinlichkeit $p_{RE} = 0,2$ ausgeführt. Hierbei wird jedes Atom eines Permutationsvektors einmal als Zentralatom einer Abbildungssphäre herangezogen.

Der Einsatz des Reordering Operators wirkt sich sehr positiv auf die Effizienz des Gesamtverfahrens aus (siehe Abbildungen 5.12 und 5.13). Es wird hierdurch möglich, Minimale Chemische Distanzen für sehr viel größere Ensembles zu bestimmen, als es für andere Minimierungsverfahren möglich ist. Die Problemstellung in Abbildung 4.5 zeigt ein solches Beispiel. Der Genetische Algorithmus ist hierbei in der Lage, eine von zwei Abbildungen mit Minimaler Chemischer Distanz aus einem Lösungsraum der Größenordnung 10^{50} (!) in weniger als 30 Generationscyclen zuverlässig herauszufinden. Ein gewöhnlicher Personalcomputer benötigt hierzu lediglich einige wenige Sekunden.

Der Reordering Operator wird seit kurzem mit Erfolg auch in anderen Algorithmen eingesetzt, wie zum Beispiel in KVASNICKAS Verfahren zum Auffinden maximaler gemeinsamer Substrukturen mittels Simulated Annealing¹⁸⁴ (siehe Abschnitt 4.2.3).

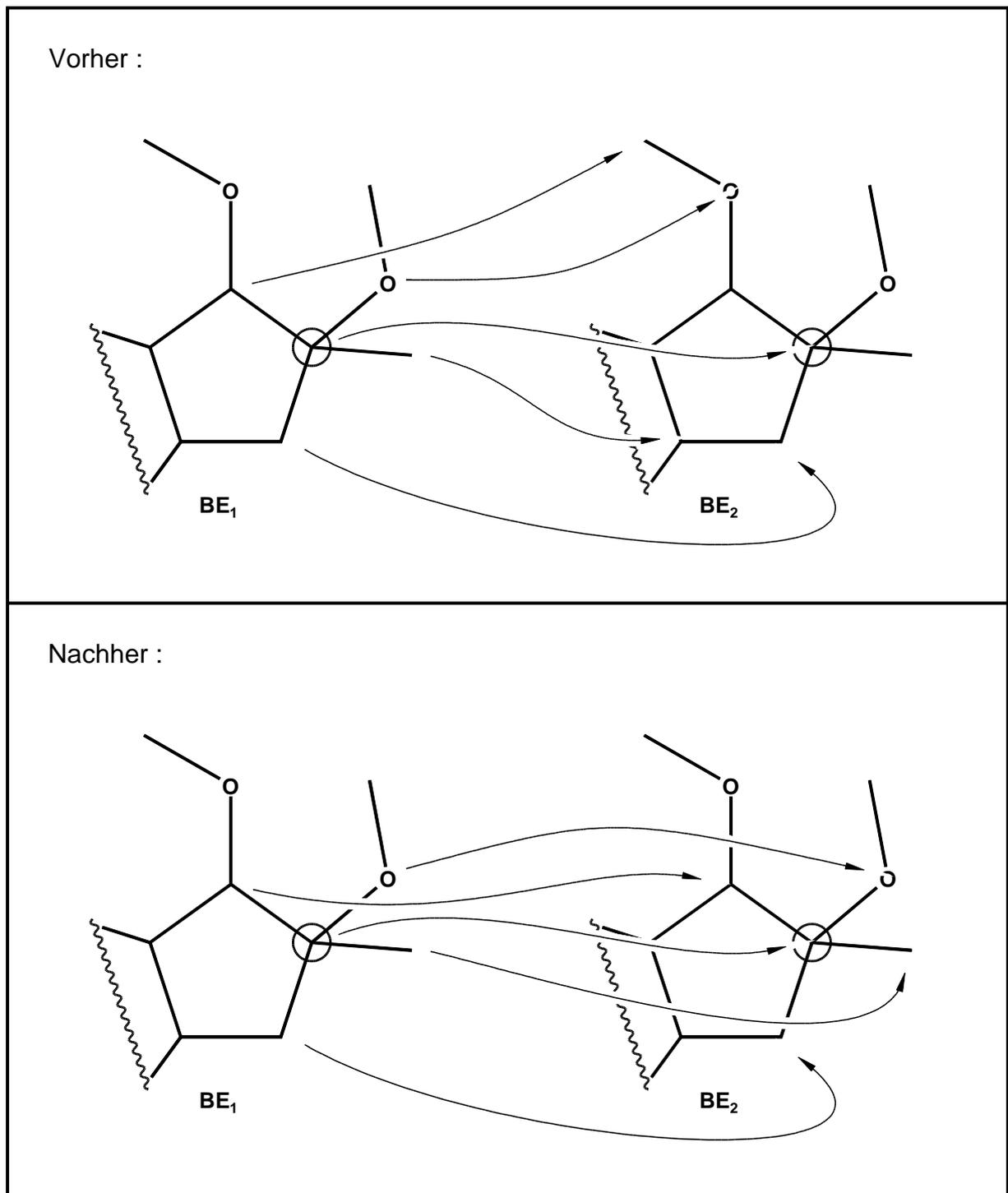


Abb. 5.11 Veränderungen einer Teilabbildung von BE_1 nach BE_2 , die durch den Reordering Operator verursacht werden. Der Permutationsvektor wird so verändert, daß die Bildatome der Nachbarn eines Zentralatoms von BE_1 (durch einen Kreis hervorgehoben) auf Nachbaratome des zentralen Bildatoms in BE_2 zu liegen kommen.

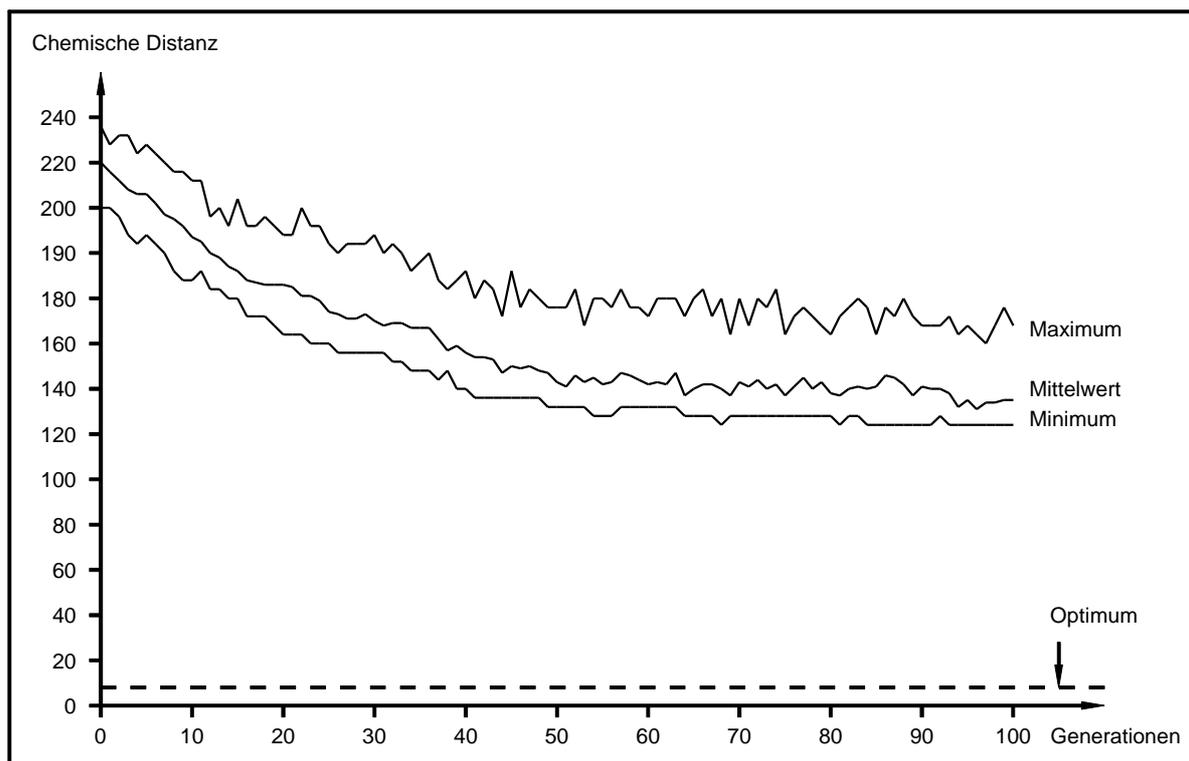


Abb. 5.12 Typischer Verlauf einer Bestimmung der Chemischen Distanz mit einem "Basic Genetic Algorithm" ohne Verwendung des Reordering Operators. Bei dem bearbeiteten Problem handelt es sich um die Reaktion in Abbildung 4.5 mit einem Lösungsraum von ca. 10^{50} Abbildungsmöglichkeiten. Angezeigt ist für die ersten 100 Generationen jeweils der beste Wert, der schlechteste Wert und das arithmetische Mittel der Chemischen Distanz innerhalb der Population. Der optimale Wert wird erst nach einigen tausend Generationscyclen erreicht.

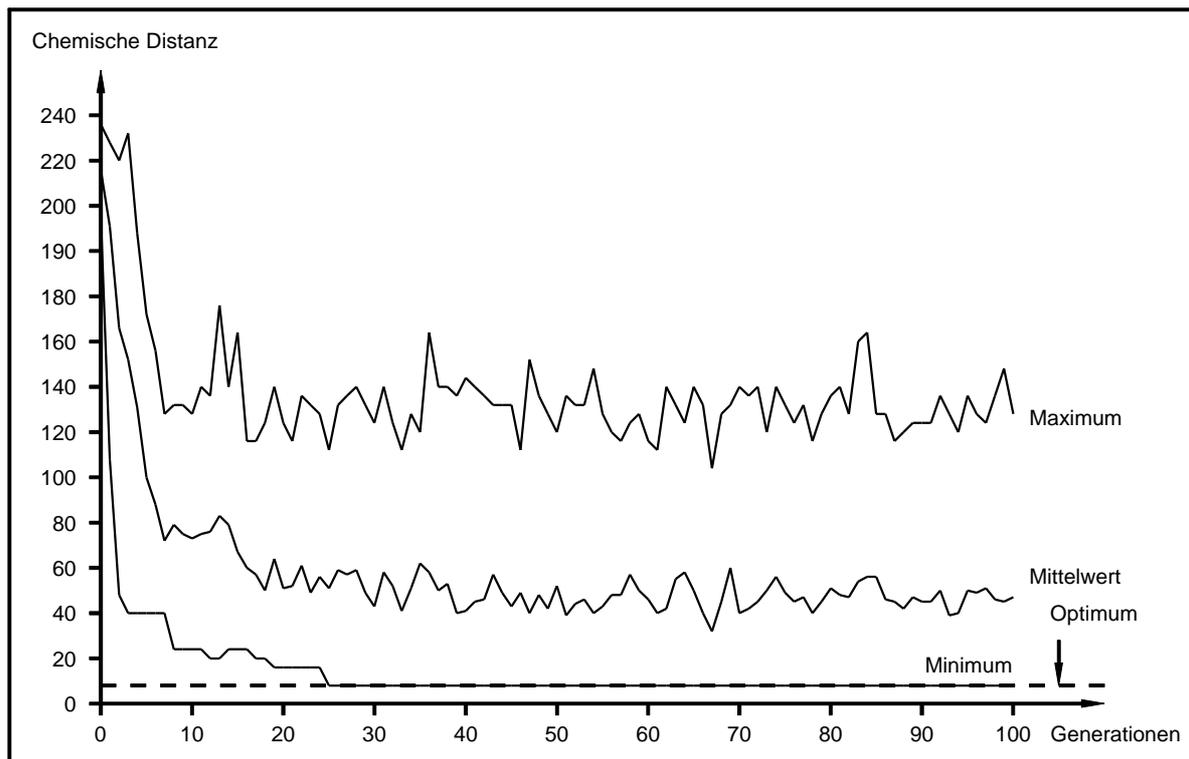


Abb. 5.13 Dieselbe Problemstellung wie in Abbildung 5.12, diesmal erfolgt die Bestimmung der Chemischen Distanz jedoch mit einem "Knowledge Augmented Genetic Algorithm" unter Verwendung des Reordering Operators. Das Optimum (eine bestimmte Abbildung aus ca. 10^{50} möglichen) wird bereits nach 25 Generationscyclen erreicht.

5.3.6 Vorzeitige Konvergenz

Bei der vorzeitigen Konvergenz handelt es sich um ein Problem, das bei Genetischen Algorithmen häufig auftritt, besonders bei großdimensionierten Suchräumen mit vielen lokalen Minima. Die Populationsgröße ist, bedingt durch praktische Erwägungen, normalerweise auf maximal einige hundert Individuen begrenzt. Diese Populationsgröße liegt weit unterhalb der von entwicklungsfähigen Populationen in der Natur. Eine selektive Reproduktion in solch einer kleinen Population führt häufig zu einer genetischen Verarmung (Inzest), was im Extremfall dazu führt, daß nahezu alle Populationsmitglieder dasselbe, oder ein sehr ähnliches Genom besitzen. Wenn nun der Suchraum wenig ausgeprägte Steigungen der Hyperfläche aufweist, oder wenn bei kombinatorischen Problemen ein starke Entartung auftritt²⁶⁸ (ein Fitnesswert kann durch viele verschiedene Parameterkombinationen gleichermaßen erreicht werden), so führen Mutationsereignisse nur mit einer sehr geringen Wahrscheinlichkeit zur dauerhaften Etablierung neuer Gene, da die durch sie veränderten Individuen im Vergleich zur Überzahl der gleichförmigen Populationsmitglieder meist "wegselektiert" werden.

Zur Verhinderung der vorzeitigen Konvergenz sind derzeit noch immer viele Methoden in Entwicklung und Erprobung. Der im Programm RAIN implementierte Genetische Algorithmus zur Bestimmung der Chemischen Distanz enthält zwei Verfahren, die eine zu große Gleichförmigkeit der Population verhindern sollen. Hierzu wird zunächst während jeder Generation die genetische Homogenität H der Population bestimmt. Die Homogenität ist definiert als relativer Anteil der Individuen N_{BEST} , die den aktuell höchsten Fitnesswert erreichen.

$$H = \frac{N_{BEST}}{N}$$

Wird für den Wert von H eine Homogenitätstoleranzgrenze von $H_L = 0,3$ erreicht, so wird die gesamte Population neu initialisiert. Dies bedeutet, daß alle Individuen ein neues, zufällig bestimmtes Genom erhalten, womit weitestgehende Maximierung der genetischen Varianz erreicht ist. Der Permutationsvektor mit der bis dahin

gefundenen höchsten Fitness wird jedoch separat gespeichert, damit dieses Genom nicht verloren geht. Insgesamt wird diese Reinitialisierung nicht beliebig oft, sondern maximal M_i mal durchgeführt. Für M_i gilt (N ist die Anzahl der Atome im Ensemble):

$$M_i = \max(4, N/6)$$

Eine zweite, etwas weniger stark in die Struktur der Gesamtpopulation eingreifende Methode zur Verhinderung vorzeitiger Konvergenz besteht in der Anpassung der Mutationsrate an die Homogenität der Genome. Ausgehend von einem minimalen Wert der Mutationswahrscheinlichkeit p_{MU}^0 wird ein linear von der Homogenität H abhängiger Term hinzuaddiert, um zur effektiven Mutationsrate p_{MU} zu gelangen.

$$p_{MU} = p_{MU}^0 + H * (1 - p_{MU}^0)$$

Das Inkrement der Mutationswahrscheinlichkeit wird so gewählt, daß bei einer Homogenität von $H = 1$ (vollständig gleichförmige Population) die effektive Mutationsrate ebenfalls den Wert 1 annimmt, das heißt, daß nominell jedes Individuum einer Mutation unterzogen wird.

Ein weiteres Verfahren zur Erhöhung der genetischen Vielfalt, das in chemisch orientierten Genetischen Algorithmen angewandt wird^{206, 269}, besteht in der Einführung von evolutionären Nischen mittels sogenanntem "Crowding"²⁹⁸ bzw. "Sharing"^{299, 300}.

5.3.7 Abbruchkriterien

Ein scharf definiertes Abbruchkriterium kann für Genetische Algorithmen nicht festgelegt werden, da weder der Suchraum erschöpfend durchmustert wird, noch ein Verfahren existieren kann, das die Qualität des bislang erreichten Ergebnisses absolut feststellt (aus diesem Grund würde sich eigentlich die Verwendung des Begriffs "Algorithmus" verbieten). Wie bei anderen stochastischen Verfahren muß auch hier ein formales Hilfskriterium, wie zum Beispiel die Anzahl der durchgeführten Generationscyclen, oder die verbrauchten Rechenressourcen (Zeit o.ä.) herangezogen werden.

Die Tatsache, daß Genetische Algorithmen generell nur suboptimale Lösungen finden können, und keine sichere Aussage darüber zu machen ist, ob das wirkliche Minimum der Chemischen Distanz erreicht wurde, stellt keinen wesentlichen Hinderungsgrund dar, da in den meisten Fällen nicht der eigentliche absolute Wert der Chemischen Distanz von Interesse ist, sondern die Tatsache, ob für ein gegebenes Molekül der bislang berechnete Wert für die Chemische Distanz *unter* einem gesetztem Limit liegt. Trotzdem muß bei der Aufstellung der Abbruchkriterien und bei der Einstellung der anderen Betriebsparameter des Genetischen Algorithmus darauf geachtet werden, daß die Wahrscheinlichkeit für das Erreichen des globalen Minimums nahe bei 1 liegt. Dies wurde durch eine große Anzahl von Referenzversuchen bei denen Ergebnisse, die mittels eines deterministischen Alternativverfahrens (Permutationsverfahren mit Branch and Bound) erhalten wurden, mit den Ergebnissen des Genetischen Algorithmus verglichen wurden, bestätigt.

Als Abbruchkriterium für den im Programm RAIN verwendeten Genetischen Algorithmus wird die Anzahl N_G der seit Start durchgeführten Generationscyclen verwendet. Wird für N_G ein Wert erreicht, der bei der vierfachen Zahl der Atome im Ensemble liegt, so bricht der Algorithmus ab. Zu beachten ist hierbei, daß bei jeder neuen Initialisierung der Population aufgrund einer zu hohen Homogenität (siehe Abschnitt 5.3.6) die Zählung von N_G ebenfalls neu beginnt.

5.4 Ergebnisse

Die zielgerichtete formale Entwicklung von Reaktionsbäumen im Rahmen von mechanistischen Studien ist auf eine praktikable und effiziente Methode zur Bestimmung der Chemischen Distanz angewiesen, die eine Orientierung der Propagation im Konstitutionsraum ermöglicht. Enumerative Verfahren und heuristische Näherungsmethoden haben sich bei den in Frage kommenden Größenordnungen der möglichen Abbildungsräume als nicht einsetzbar herausgestellt. Für das Wirkungsfeld des Programms RAIN stellen die Genetischen Algorithmen die zur Zeit einzig wirklich praktikable Möglichkeit dar, eine zielorientierte Generierung von Reaktionsbäumen durchzuführen.

Die Genetischen Algorithmen in der Form des "Knowledge Augmented Genetic Algorithm", wie sie im vorigen besprochen wurde, haben sich bei der Bestimmung dieser Chemischen Distanz als außerordentlich effizient erwiesen. In einer ausgedehnten Testreihe wurden 45 verschiedene Isomere Ensembles der Summenformel $C_5H_7ClO_3R_2$ erzeugt und die Chemischen Distanzen der sich hieraus ergebenden 1035 verschiedenen Paarungen von Ensembles jeweils mit einem enumerativen Verfahren (Branch-and-Bound-Algorithmus¹⁹¹) und einem Genetischen Algorithmus ermittelt²⁰⁵. Ein komplexeres Testbeispiel konnte leider nicht herangezogen werden, da die sehr langen Rechenzeiten des deterministischen Verfahrens wesentlich weniger verschiedene Testpaarungen zugelassen hätte. Der Vergleich der mittleren Rechenzeiten zeigt, daß der Genetische Algorithmus das gestellte Problem nahezu um zwei Größenordnungen effektiver löst, als das deterministische Verfahren (Abbildung 5.14)

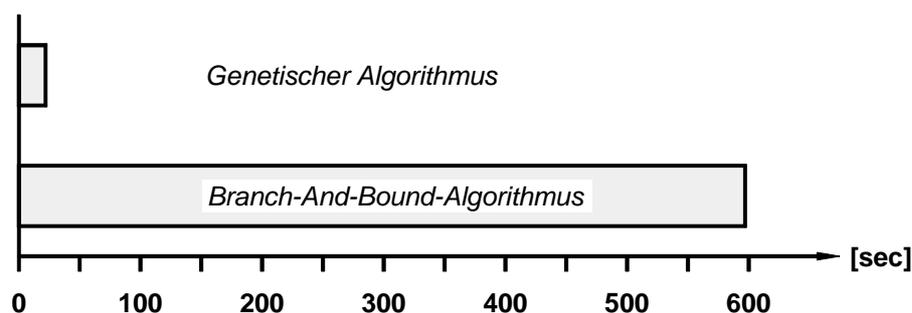


Abb. 5.14 Der Genetische Algorithmus erweist sich in einer Testreihe von 1035 Experimenten beim Vergleich der durchschnittlich benötigten Rechenzeiten (16MHz 80386 Prozessor) wesentlich effektiver, als das enumerative Branch-and-Bound-Verfahren.

Die oben angeführte Versuchsreihe konnte auch zeigen, daß Genetische Algorithmen trotz ihres stochastischen Charakters im Rahmen des Forschungsprojekts RAIN einsetzbar sind. Da die zum Vergleich herangezogene enumerative Berechnungsweise für alle in der Testreihe vorkommenden Ensemblepaare ebenfalls die absolut kleinsten Werte für die Chemische Distanz ermitteln, ist auch eine unmittelbare Überprüfung der mit dem Genetischen Algorithmus erhaltenen Distanzwerte möglich. Die hierbei gewonnenen Ergebnisse²⁰⁵ zeigen, daß die mittlere Abweichung der mit dem Genetischen Algorithmus gewonnenen Resultate von den realen Chemischen Distanzen klein genug ist, um bei der zielgerichteten Reaktionsgenerierung zuverlässige Ergebnisse zu liefern.

Illustrative Anwendungsbeispiele für den erfolgreichen Einsatz der Genetischen Algorithmen zur Bestimmung der Chemischen Distanz finden sich in den Abschnitten 4.1, 4.3.1 und 4.3.2. Alle dort angezeigten Werte für die Chemische Distanz, sowie alle Ellipsoidal-Projektionen von Reaktionsnetzwerken wurden mit Hilfe dieser Technik gewonnen.

Das Potential der möglichen Anwendungen Genetischer Algorithmen im Bereich der Chemie ist bei weitem noch nicht ausgeschöpft. Insbesondere auf dem Gebiet der Konformationsanalyse mittlerer und großer Moleküle, sowie auf dem Sektor der Proteinfaltungsproblematik besteht berechtigte Hoffnung, daß dieses elegante Verfahren zu interessanten Anwendungen führen wird.

Kapitel 6

Schlußbemerkung

6 Schlußbemerkung

Die Bewertung des heutigen Standpunktes der Computerchemie muß wohl in der Unterscheidung zwischen einem Dienstleistungsaspekt und einem Aspekt der Grundlagenforschung geschehen. Die Computerchemie übernahm naturgemäß die Funktion einer Quelle für spezialisierte Chemie-Software. Von diesen breit verfügbaren Programmen wird im allgemeinen erwartet, daß auch ein Nichtexperte mit ihrer Hilfe ein konkretes Problem in angemessener Zeit zu lösen vermag. Die Bewertungskriterien für diesen Dienstleistungsaspekt sind natürlich vollkommen andere, als sie im Bereich der Grundlagenforschung anzuwenden sind, bei dem die Programme der Computerchemie, nach entsprechender Anwendung, neue Erkenntnisse über die logische Struktur der Chemie und über die inneren Zusammenhänge dieser facettenreichen Wissenschaft bringen können.

Im Bereich der Dienstleistungen sind es wohl eher die eigentlichen "Nebenprodukte" der Entwicklungen auf dem Gebiet der Computerchemie, die innerhalb von Datenbank Anwendungen und Spektreninformationssystemen zum Erfolg gelangt sind. Heutzutage ist gespeicherte Information über chemische Strukturen, und mittlerweile auch über chemische Reaktionen, sowohl *in house*, als auch extern über Satellit, überall und auf Knopfdruck verfügbar. Hierzu haben eine ganze Reihe von Einzelentwicklungen, wie Techniken zur Kanonisierung, Substruktursuche, Atom-auf-Atom-Abbildung, graphischen Darstellung, Strukturbearbeitung, Speicher- und Retrievalsystemen usw. aus den "Labors" der verschiedenen Computerchemiearbeitsgruppen beigetragen. Viele dieser Programme sind heute an den Universitäten und in der chemischen Industrie als Hilfsmittel zur Informationsbeschaffung und zur Verwaltung des eigenen Know-Hows unverzichtbar.

Im Gegensatz hierzu haben es die eigentlichen Hauptziele der Computerchemie, die Syntheseplanungs- bzw. Vorhersageprogramme, bislang noch nicht geschafft, zu einer wirklich breiten Anwendungsreife zu gelangen. In den meisten Fällen sind die, bisher eher spärlichen Resultate dieser Programme das Ergebnis des *a posteriori* Einsatzes von Spezialisten, meist den Entwicklern dieser Programme selbst. Der Grund für diese ernüchternde Feststellung liegt einerseits wohl in einem nachlassenden Interesse an der an komplexen Naturstoffen orientierten Synthesechemie (Ausnahme: Taxol), als auch in der unterschätzten Komplexität des zugrundeliegenden Problems. Den Programmen ist es bisher nicht gelungen, die Gratwanderung zwischen der Trivialität der Ergebnisse und der "kombinatorischen Explosion" zu bewältigen. Ein Chemiker ist bei der Suche nach einer Antwort auf eine vom ihm ge-

stellte konkrete Frage weder bereit, sich nach vielen Bemühungen mit einer trivialen Antwort zufrieden zu geben, auf die er mit Leichtigkeit auch selber gekommen wäre, noch sieht er es als seine Aufgabe an, tausende von angebotenen Lösungsvorschlägen zu sichten, zu evaluieren, und bei jedem einzelnen Vorschlag die Entscheidung darüber zu treffen, und damit auch letztendlich die Verantwortung über die Entscheidung zu tragen, ob ein vom Programm gemachter Vorschlag sinnvoll ist oder nicht. Außerdem widerspricht es sicher auch der Eitelkeit vieler Menschen, sich bei einem intellektuellen Vorgang von einer Maschine helfen zu lassen. Ein einziger positiver Aspekt dieses Vorgehens ist es, daß der Anwender durch die immer wiederkehrende intensive Beschäftigung mit "naiven" Lösungsvorschlägen gezwungen ist, sich auch stets selbst zu hinterfragen und seinen Horizont zu erweitern, wobei er unter Umständen durch diese oder jene Anregung zu neuen Erkenntnissen gelangen kann.

Um eine größere Akzeptanz zu erlangen, müssen die Computerchemieprogramme jedoch mehr und mehr in der Lage sein, die generierten Lösungen nach den verschiedensten Kriterien zuverlässig zu bewerten, damit nach einer entsprechenden Vorklassifizierung die Ergebnisse dem Anwender in einer geordneten Reihe, zum Beispiel nach fallender Plausibilität, angeboten werden können. Dieses Unterfangen ist aber außerordentlich schwierig, da die hierzu nötigen Heuristiken noch lange nicht mit der erforderlichen Präzision arbeiten und die Anwendung quantenmechanischer Techniken in den hoch interaktiven Programmen aus rechenzeittechnischen Gründen wohl auch auf weiteres nicht in Frage kommt.

Eine im Gegensatz zum Bereich der breiten Anwendung eher positive Bilanz der bisherigen Bemühungen auf dem Gebiet der Computerchemie scheint auf dem Feld möglich zu sein, das sich mit der logischen Struktur der Chemie beschäftigt. Die intensive Arbeit an und mit diversen Methoden zur Einteilung, Klassifizierung und letztendlich auch Bewertung chemischer Transformationen hat zu einem Ordnungsprinzip geführt, das jeder bereits bekannten und jeder präzedenzlosen chemischen Reaktion den Platz in einer wohldefinierten Hierarchie zuweist.

Die strenge mathematische Organisation gibt den Programmen, die sich mit den rein formalen Aspekten der Chemie beschäftigen, wie zum Beispiel IGOR und RAIN, die Möglichkeit, dem Anwender jene Ästhetik zu offenbaren, die sich hinter der überwältigenden und facettenreichen Vielfalt verbirgt, mit der die Chemie aufwartet. Ob die Anwendung dieser Programme, über diesen ästhetischen Aspekt hinaus, einen wahren Erkenntnisgewinn leisten können, darüber entscheidet letztendlich der Benutzer.

Anhang

Literaturverzeichnis

Literaturverzeichnis

- [1] E.J. Corey, W.T. Wipke "Computer-Assisted Design of Complex Organic Syntheses" *Science* **1969**, *160*, 178-192.
- [2] E.J. Corey "General Methods for the Construction of Complex Molecules" *Pure Appl. Chem.* **1967**, *14*, 19-37.
- [3] D.J. Klein "Aromaticity via Kekule Structures and Conjugated Circuits" *J. Chem. Educ.* **1992**, *69*, 691-694.
- [4] E.J. Corey, X.-M. Cheng *The Logic of Chemical Synthesis*, Wiley, New York **1989**.
- [5] E.J. Corey, A.K. Long, S.D. Rubenstein "Computer-Assisted Analysis in Organic Synthesis" *Science* **1985**, *228*, 408-418.
- [6] W.T. Wipke, H. Braun, G. Smith, F. Choplin, W. Sieber "SECS - Simulation and Evaluation of Chemical Synthesis: Strategy and Planning" *Computer-Assisted Organic Synthesis*, ACS Symp. Ser., W.T. Wipke, W.J. Howe Eds., Washington DC **1977**, *61*, 97-127.
- [7] A.K. Long, S.D. Rubenstein, L.J. Joncas "A Computer Program for Organic Synthesis" *Chem. Eng. News* **1983**, *61*, 22-30.
- [8] A.P. Johnson, C. Marshall, P.N. Judson "Some Recent Progress in the Development of the LHASA Computer System for Organic Synthesis Design: Starting-Material-Oriented Retro-synthetic Analysis" *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 311-317.
- [9] E.L.M. v.Rozendaal, M.A. Ott, H.W. Scheeren "A LHASA Analysis of Taxol" *Recl. Trav. Chim. Pays-Bas* **1994**, *113*, 297-303.
- [10] F. Choplin, C. Laurencio, R. Marc, G. Kaufmann, W.T. Wipke "Synthese assistée par ordinateur en chimie des composés organophosphorés" *Nouv. J. Chim.* **1978**, *2*, 285-293.
- [11] F. Choplin, P. Bonnet, M.H. Zimmer, G. Kaufmann "Interactive Strategy in Computer Design of Synthesis" *Nouv. J. Chim.* **1979**, *3*, 223-230.
- [12] C. Laurencio, L. Villien, G. Kaufmann "Synthèse assistée par ordinateur de la Phospacarnéine - I Etablissement du plan de synthèse avec l'aide de PASCOP" *Tetrahedron* **1984**, *40*, 2721-2729.
- [13] W. Sieber "Reaction Retrieval and Synthesis Planning" *Chemical Structures: The International Language of Chemistry*, W.A. Warr (Ed.), Springer-Verlag, Berlin-Heidelberg **1988**, 361-366.
- [14] H.L. Gelernter, N.S. Sridharan, A.J. Hart, S.C. Yen, F.W. Fowler, H.J. Shue "The Discovery of Organic Synthetic Routes by Computer" *Top. Curr. Chem.* **1973**, *41*, 113-150.
- [15] H.L. Gelernter, A.F. Sanders, D.L. Larsen, K.K. Agarwal, R.H. Boivie, G.A. Spritzer, J.E. Searleman "Empirical Explorations of SYNCHEM" *Science* **1977**, *197*, 1041-1049.
- [16] K.K. Agarwal, D.L. Larsen, H.L. Gelernter "Application of Chemical Transforms in SYNCHEM2, a Computer Program for Organic Synthesis Route Discovery" *Comput. Chem.* **1978**, *2*, 75-84.
- [17] H.L. Gelernter, S.S. Bhagwat, D.L. Larsen, G.A. Miller "Knowledge-Based Enhancement via Training Sequence: The Education of SYNCHEM2" *Anal. Chem. Symp. Ser.* **1983**, *15*, 35-59.
- [18] M. Yanaka, K. Nakamura, A. Kurumisawa, W.T. Wipke "Automatic Knowledge Base Building for the Organic Synthesis Design Program (SECS)" *Tetrahedron Comput. Methodol.* **1990**, *3*, 359-375.
- [19] H. Gelernter, J.R. Rose, C. Chen "Building and Refining a Knowledge Base for Synthetic Organic Chemistry via the Methodology of Inductive and Deductive Machine Learning" *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492-504.

- [20] T.D. Salatin, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 1. Overview" *J. Org. Chem.* **1980**, *45*, 2043-2057.
- [21] A. Weise, H.G. Scharnow "Computergestützte Planung und Vorhersage von Synthesen der konstitutionellen organischen Chemie" *Z. Chem.* **1979**, *19*, 49-56.
- [22] F. Haase, K. Biedka "Heterocyclic Reaction Design with RDSS" *Tetrahedron Comput. Methodol.* **1990**, *3*, 461-467.
- [23] F.H. Allen, J.E. Davies, J.J. Galloy, O. Johnson, O. Kennard, C.F. Macrae, E.M. Mitchell, G.F. Mitchell, J.M. Smith, D.G. Watson "The Development of Versions 3 and 4 of the Cambridge Database System" *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187-204.
- [24] A. Weise "Derivation of Organochemical Reactions with the AHMOS [Automated Heuristic Modeling of Organochemical Syntheses] Simulation Program" *Z. Chem.* **1975**, *15*, 333-340.
- [25] R.G. Pearson *J. Am. Chem. Soc.* **1963**, *85*, 3533.
- [26] W.L. Jorgensen, E.R. Laird, A.J. Gushurst, J.M. Fleischer, S.A. Gothe, H.E. Helson, G.D. Paderes, S. Sinclair "CAMEO: a Program for the Logical Prediction of the Products of Organic Reactions" *Pure Appl. Chem.* **1990**, *62*, 1921-1932.
- [27] J.S. Burnier, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 7. Six-Electron Cycloadditions" *J. Org. Chem.* **1983**, *48*, 3923-3941.
- [28] G.D. Paderes, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 20. Ene and Retro-Ene Chemistry" *J. Org. Chem.* **1992**, *57*, 1904-1916.
- [29] G.D. Paderes, W.L. Jorgensen "Computer-Assisted Evaluation of Oxidation Reactions" *J. Org. Chem.* **1989**, *54*, 2058-2085.
- [30] T.D. Salatin, D. McLaughlin, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 3. Ylide Chemistry and the Organometallic Chemistry of Lithium, Magnesium, and Lithium Cuprates" *J. Org. Chem.* **1981**, *46*, 5284-5294.
- [31] T.H. Varkony, R.E. Carhart, D.H. Smith "Computer-Assisted Structure Elucidation: Modeling Chemical Reaction Sequences Used in Molecular Structure Problems" *Computer-Assisted Organic Synthesis*, ACS Symp. Ser., W.T. Wipke, W.J. Howe Eds., Washington DC **1977**, *61*, 188-216.
- [32] R.E. Carhart, T.H. Varkony, D.H. Smith "Computer Assistance for the Structural Chemist" *Computer-Assisted Structure Elucidation*, D.H. Smith (ed.), American Chemical Society, Washington, DC **1977**, 126.
- [33] T.H. Varkony, R.E. Carhart, D.H. Smith, C. Djerassi "Computer-Assisted Simulation of Chemical Reaction Sequences. Application to Problems of Structure Elucidation" *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 168.
- [34] T.H. Varkony, D.H. Smith, C. Djerassi "Computer-Assisted Structure Manipulation: Studies in the Biosynthesis of Natural Products" *Tetrahedron* **1978**, *34*, 841.
- [35] J.B. Hendrickson, D.L. Grier, A.G. Toczko "A Logic-Based Program for Synthesis Design" *J. Am. Chem. Soc.* **1985**, *107*, 5228-5238.
- [36] J.B. Hendrickson, A.G. Toczko "SYNGEN Program for Synthesis Design : Basic Computing Techniques" *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 137-145.
- [37] J.B. Hendrickson "Approaching the Logic of Synthesis Design" *Acc. Chem. Res.* **1986**, *19*, 274-281.
- [38] G. Moreau "MASSO - Un programme d'aide a la synthèse organique, utilisant des demi-reactions" *Nouv. J. Chim.* **1978**, *2*, 187-193.
- [39] R. Dönges, B.T. Gröbel, H. Nickelsen, J. Sander "TOSCA : A Topological Synthesis Design by Computer Application" *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 425-430.

- [40] D.A. Evans "Consonant and Dissonant Relationships. An Organizational Model for Organic Synthesis" *UCLA Physical Organic Chemistry Seminar*, May 6, 1971, unveröffentlichtes Manuskript **1971**.
- [41] D.A. Evans, G.C. Andrews "Allylic Sulfoxides: Usefule Intermediates in Organic Synthesis" *Acc. Chem. Res.* **1974**, 7, 147-155.
- [42] R. Barone, M. Arbelot, A. Baldy, M. Chanon, R. Gallo "RENE (REACTION GeNERator): a Program to Propose New Reactions by Construction of Isomers and Related Reactive Sets of Simple Functionalized Compounds. Application to Acids and Esters" *Rev. Roum. Chim.* **1991**, 36, 581-598.
- [43] R. Moreau "A Topological Code for Molecular Structures. A Modified Morgan Algorithm" *Nouv. J. Chim.* **1980**, 4, 17-22.
- [44] I. Ugi, P.D. Gillespie "Beschreibung chemischer Systeme und ihrer Umwandlungen durch BE-Matrizen und ihre Transformationen" *Angew. Chem.* **1971**, 83, 980-981.
- [45] I. Ugi, P.D. Gillespie "Stoffbilanz-erhaltende Synthesewege und semi-empirische Syntheseplanung mittels elektronischer Datenverarbeitung" *Angew. Chem.* **1971**, 83, 982-985.
- [46] J. Dugundji, I. Ugi "An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs" *Top. Curr. Chem.* **1973**, 39, 19-64.
- [47] J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie, I. Ugi "CICLOPS - A Computer Program for the Design of Syntheses on the Basis of a Mathematical Model" *Computer Representation and Manipulation of Chemical Information*, W.T. Wipke, S.R. Heller, R.J. Feldmann, E. Hyde Eds., Wiley, New York **1974**, 129-146.
- [48] J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie, I. Ugi "Representation of the Constitutional and Stereochemical Features of Chemical Systems in the Computer Assisted Design of Syntheses" *Tetrahedron* **1974**, 30, 1845-1859.
- [49] J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, I. Ugi "Computer Programs for the Deductive Solution of Chemical Problems on the Basis of a Mathematical Model of Chemistry" *Computer-Assisted Organic Synthesis*, ACS Symp. Ser., W.T. Wipke, W.J. Howe Eds., Washington DC **1977**, 61, 33-59.
- [50] J.C.J. Bart, E. Garagnani "Organic Reaction Schemes and General Reaction-Matrix Types, I Rearrangement Reactions" *Z. Naturforsch.* **1976**, 31b, 1646-1653.
- [51] J.C.J. Bart, E. Garagnani "Organic Reaction Schemes and General Reaction-Matrix Types, II Basic Types of Synthetic Transformations" *Z. Naturforsch.* **1977**, 32b, 455-464.
- [52] J.C.J. Bart, E. Garagnani "Organic Reaction Schemes and General Reaction-Matrix Types, III A Quantitative Analysis" *Z. Naturforsch.* **1977**, 32b, 465-468.
- [53] J.C.J. Bart, E. Garagnani "Organic Reaction Schemes and General Reaction-Matrix Types, IV Organic Name Reactions" *Z. Naturforsch.* **1977**, 32b, 678-683.
- [54] J. Mathieu, J. Weill-Raynal *Formation of C-C Bonds, Introduction of A Functional Carbon Atom*, G. Thieme Verlag, Stuttgart **1973**, 1.
- [55] J. Gasteiger, C. Jochum "EROS A Computer Program for Generating Sequences of Reactions" *Top. Curr. Chem.* **1978**, 74, 93-126.
- [56] J. Gasteiger, M.G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki "A New Treatment of Chemical Reactivity: Development of EROS (Elaboration of Reactions for Organic Synthesis), An Expert System for Reaction Prediction and Synthesis Design" *Top. Curr. Chem.* **1987**, 137, 19-73.
- [57] J. Gasteiger, M. Marsili, M.G. Hutchings, H. Saller, P. Löw, P. Röse, K. Rafeiner "Models for the Representation of Knowledge about Chemical Reactions" *J. Chem. Inf. Comput. Sci.* **1990**, 30, 467-476.

- [58] J. Gasteiger "A Case Study in Computer-Assisted Organic Synthesis Design" *Chim. Oggi* **1989**, 65-72.
- [59] P. RÖse, J. Gasteiger "EROS 6.0, a Knowledge Based System for Reaction Prediction - Application to the Regioselectivity of the Diels-Alder Reaction" *Software-Entwickl. Chem. 4*, Proc. Workshops 'Comput. Chem.' 4th 1989(1990), J. Gasteiger, Ed.; Springer: Berlin, FRG **1990**, 275-280.
- [60] P. RÖse, J. Gasteiger "Automated Derivation of Reaction Rules for the EROS 6.0 System for Reaction Prediction" *Anal. Chim. Acta* **1990**, 235, 163-168.
- [61] W. Schubert *Simulation und Vorhersage chemischer Reaktionen auf der Grundlage der Basistransformationen von Ensembles von Molekülen*, Dissertation, Technische Universität München **1978**.
- [62] W. Schubert "ASSOR - Allgemeines Simulationssystem Organischer Reaktionen" *Informal Commun. Math. Chem. (Match)* **1979**, 6, 213-255.
- [63] J. Brandt *Ein mathematisch begründetes hierarchisches Ordnungssystem chemischer Reaktionen und dessen theoretische und praktische Anwendungen*, Habilitationsschrift, Technische Universität München **1981**.
- [64] J. Brandt, J. Bauer, R.M. Frank, A. v.Scholley "Classification of Reactions by Electron Shift Patterns" *Chem. Scr.* **1981**, 18, 53-60.
- [65] J. Brandt, K. Stadler "A Recursive Reaction Generator" *Mod. Approaches Chem. React. Searching*, Proc. Conf., P. Willett, P. Gower, (Eds.): Aldershot, UK. **1986**, 221-239.
- [66] J. Bauer "Chemical Reactions and Structures without Precedent, Generated by Computer Program" *J. Chem. Res. (S)* **1982**, 11, 298.
- [67] J. Bauer "Chemical Reactions and Structures without Precedent, Generated by Computer Program" *J. Chem. Res. (M)* **1982**, 3101-3260.
- [68] J. Bauer, R. Herges, E. Fontain, I. Ugi "IGOR and Computer Assisted Innovation in Chemistry" *Chimia* **1985**, 39, 43-53.
- [69] J. Bauer "IGOR2 : A PC-Program for Generating New Reactions and Molecular Structures" *Tetrahedron Comput. Methodol.* **1989**, 2, 269-280.
- [70] I. Ugi, J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, R. Herges, M. Knauer, K. Reitsam, N. Stein "Die computerunterstützte Lösung chemischer Probleme - eine neue Disziplin der Chemie" *Angew. Chem.* **1993**, 105, 210-239.
- [71] E. Fontain, J. Bauer, I. Ugi "Computer Assisted Bilateral Generation of Reaction Networks from Educts and Products" *Chem. Lett.* **1987**, 37-40.
- [72] E. Fontain, J. Bauer, I. Ugi "Computerunterstützte mechanistische Analyse der Streith-Reaktion mit dem Programm RAIN" *Z. Naturforsch.* **1987**, 42b, 889-891.
- [73] E. Fontain, K. Reitsam "The Generation of Reaction Networks with RAIN. 1. The Reaction Generator" *J. Chem. Inf. Comput. Sci.* **1991**, 31, 96-101.
- [74] M. Knauer *Die Korrelation von Ausgangsstoffen und Synthesezielen über die größten gemeinsamen Substrukturen*, Dissertation, Technische Universität München **1992**.
- [75] A. Dengler *Algorithmen und Softwaremodule für die computerunterstützte Lösung chemischer Probleme*, Dissertation, Technische Universität München **1992**.
- [76] I. Ugi, A. Dengler "The Algebraic and Graph Theoretical Completion of Truncated Reaction Equations" *J. Math. Chem.* **1992**, 9, 1-10.
- [77] I. Ugi, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, P. Lemmen, W. Schubert "The Deductive Solution of Chemical Problems by Computer Programs on the Basis of a Mathematical Model of Chemistry" *Pure Appl. Chem.* **1978**, 50, 1303-1318.

- [78] J. Koca, M. Kratochvil, M. Kunz, V. Kvasnicka "Mathematical Model of Organic Chemistry. VI. Valence States of Atoms and Their Conversions" *Collect. Czech. Chem. Commun.* **1984**, *49*, 1247-1261.
- [79] J. Koca, M. Kratochvil, V. Kvasnicka, L. Matyska, J. Pospichal *Synthon Model of Organic Chemistry and Synthesis Design*, Lect. Notes Chem., Springer Verlag Berlin Heidelberg **1989**, 51.
- [80] I. Ugi, E. Fontain, J. Bauer "Transparent Formal Methods for Reducing the Combinatorial Abundance of Conceivable Solutions to a Chemical Problem - Computer-Assisted Elucidation of Complex Reaction Mechanisms" *Anal. Chim. Acta* **1990**, *235*, 155-161.
- [81] F.H. Westheimer "Studies on the Mechanism of Chromic Acid Oxidations" *Chem. Rev.* **1949**, *45*, 419-451.
- [82] E.J. Corey, J.P. Schaefer "Studies on the Mechanism of Oxidation of Ketones by Selenium Dioxide I." *J. Am. Chem. Soc.* **1960**, *82*, 918-929.
- [83] E. Fontain *BVTGEN, ein Programm zur erschöpfenden Erzeugung von BR-Transformatoren*, unveröffentlicht.
- [84] H.L. Morgan "The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstract Service" *J. Chem. Doc.* **1965**, *5*, 107-113.
- [85] J. Figueras "Morgan Revisited" *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 717-718.
- [86] W. Schubert, I. Ugi "Darstellung chemischer Strukturen für die computergestützte deduktive Lösung chemischer Probleme" *Chimia* **1979**, *33*, 183-191.
- [87] W. Schubert, I. Ugi "Constitutional Symmetry and Unique Descriptors of Molecules" *J. Am. Chem. Soc.* **1978**, *100*, 37-41.
- [88] M.L. Bender "Oxygen Exchange as Evidence for the Existence of an Intermediate in Ester Hydrolysis" *J. Am. Chem. Soc.* **1951**, *73*, 1626-1629.
- [89] E.H. Sussenguth "A Graph-Theoretical Algorithm for Matching Chemical Structures" *J. Chem. Doc.* **1965**, *5*, 36-43.
- [90] J.M. Barnard "Substructure Search Methods: Old and New" *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 532-538.
- [91] A. Bartmann, H. Maier, D. Walkowiak, B. Roth, M.G. Hicks "Substructure Search on Very Large Files by Using Multiple Storage Techniques" *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 539-541.
- [92] A. Dengler, I. Ugi "A Central Atom Based Algorithm and Computer Program for Substructure Search" *Comput. Chem.* **1991**, *15*, 103-107.
- [93] B.L. Roos-Kozel, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 2. Perception of Rings, Aromaticity, and Tautomers" *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 101-111.
- [94] C.E. Peishoff, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 9. Reactions of Unsaturated Electrophiles Including Nucleophilic Aromatic Substitution" *J. Org. Chem.* **1985**, *50*, 1056-1068.
- [95] C.E. Peishoff, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 10. Stereochemistry" *J. Org. Chem.* **1985**, *50*, 3174-3184.
- [96] M.G. Bures, B.L. Roos-Kozel, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 11. Electrophilic Aromatic Substitution" *J. Org. Chem.* **1985**, *50*, 4490-4498.
- [97] A.J. Gushurst, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 12. pKa Predictions for Organic Compounds in Me₂SO" *J. Org. Chem.* **1986**, *51*, 3515-3522.

- [98] M.G. Bures, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 15. Heterocycle Synthesis" *J. Org. Chem.* **1988**, *53*, 2504-2520.
- [99] E.R. Laird, W.L. Jorgensen "Computer-Assisted Mechanistic Evaluation of Organic Reactions. 17. Free Radical Chain Reactions" *J. Org. Chem.* **1990**, *55*, 9-27.
- [100] J. Gasteiger, P. Röse, H. Saller "Multidimensional Explorations into Chemical Reactivity: The Reactivity Space" *J. Mol. Graphics* **1988**, *6*, 87-97.
- [101] J. Gasteiger, M.G. Hutchings, H. Saller, P. Löw "Prediction of Chemical Reactivity and Design of Organic Synthesis" *Chemical Structures*, W.A. Warr (Ed.), Springer: Berlin **1988**, 343-359.
- [102] J. Gasteiger, H. Saller, P. Löw "Elucidating Chemical Reactivity by Pattern Recognition Methods" *Anal. Chim. Acta* **1986**, *1991*, 111-123.
- [103] J. Gasteiger "Empirical Methods for the Calculation of Physicochemical Data of Organic Compounds" *Physical Property Prediction in Organic Chemistry*, C. Jochum, M.G. Hicks, J. Sunkel (Eds.), Springer: Berlin, FRG **1988**, 119-137.
- [104] A. Lapworth "Latent Polarities of Atoms and Mechanism of Reaction, with Special Reference to Carbonyl Compounds" *Mem. Proc. Manchester Lit. Philos. Soc.* **1920**, *64(III)*, 1-16.
- [105] A. Lapworth "Theoretical Derivation of the Principle of Induced Alternate Polarities" *J. Chem. Soc.* **1922**, *121*, 416-440.
- [106] R. Robinson "Obituary Notice. Arthur Lapworth. 1872-1941." *J. Chem. Soc.* **1947**, 989.
- [107] M.D. Saltzman "Sir Robert Robinson. A Centennial Tribute" *Chem. Br.* **1986**, 543-548.
- [108] M. Wagener *Konsonante und dissonante Beziehungen in der computerunterstützten Syntheseplanung*, Diplomarbeit, Technische Universität München **1988**.
- [109] P. Schenk *Die relative Komplexität formaler Übergangszustände in Reaktionsnetzwerken*, Diplomarbeit, Technische Universität München **1989**.
- [110] G.M. Downs, V.J. Gillet, J.D. Holliday, M.F. Lynch "Review of Ring Perception Algorithms for Chemical Graphs" *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 172-187.
- [111] E. Fontain *Die bilaterale Generierung von Reaktionsnetzwerken*, Dissertation, Technische Universität München **1987**.
- [112] S. Fujita "Description of Organic Reactions Based on Imaginary Transitions Structures. 1. Introduction of New Concepts" *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 205.
- [113] S. Fujita "Description of Organic Reactions Based on Imaginary Transitions Structures. 3. Classification of One-String Reactions Having an Odd-Membered Cyclic Reaction Graph" *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 224-230.
- [114] S. Fujita "Description of Organic Reactions Based on Imaginary Transitions Structures. 2. Classification of One-String Reactions Having an Even-Membered Cyclic Reaction Graph" *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 212-223.
- [115] S. Fujita "Description of Organic Reactions Based on Imaginary Transitions Structures. 4. Three-Nodal and Four-Nodal Subgraphs for a Systematic Characterization of Reactions" *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 231-237.
- [116] S. Fujita "Description of Organic Reactions Based on Imaginary Transitions Structures. 6. Classification and Enumeration of Two-String Reactions with One Common Node" *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 99.
- [117] J. Mockus, R.E. Stobaugh "The Chemical Abstracts Service Chemical Registry Service. VII. Tautomerism and Alternating Bonds" *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 18-22.

- [118] S.M. Welford "Die Beilstein Datenstruktur für organische Verbindungen" *Software-Entwickl. Chem. 1*, Proc. Workshops 'Comput. Chem.' 1st 1986(1987), J. Gasteiger, Ed.; Springer: Berlin, FRG **1987**, 1, 5-11.
- [119] N. Stein *Das sXBE- und sXR-Modell der konstitutionellen Chemie*, Dissertation, Technische Universität München **1993**.
- [120] N. Stein "New Perspectives in Computer-Assisted Formal Synthesis Design - Treatment of Delocalized Electrons" *J. Chem. Inf. Comput. Sci.* **submitted**.
- [121] S.M. Welford "Tautomer Processing in the Beilstein Registry System" *Software-Entwickl. Chem. 2*, Proc. Workshops 'Comput. Chem.' 2nd 1987(1988), J. Gasteiger, Ed.; Springer: Berlin, FRG **1988**, 2, 35-43.
- [122] W. Setzer *Computerunterstützte Erkennung von Mesomerie und Tautomerie*, Diplomarbeit, Technische Universität München **1987**.
- [123] W. Ellenrieder *Reaktionssimulation - Methoden zur Behandlung von Substrukturen, mesomeren und tautomeren Strukturen*, Dissertation, Technische Universität München **1983**.
- [124] W. Schubert, W. Ellenrieder "Automatic Generation of Relevant Mesomeric Structures for Computer-Oriented Representation of pi-Electron Systems" *J. Chem. Res. (S)* **1984**, 256-257.
- [125] J.R. Dias "Benzenoids Having a Constant Number of Isomers" *J. Chem. Inf. Comput. Sci.* **1990**, 30, 61-64.
- [126] J.R. Dias "Isomer Enumeration and Topological Characteristics of Benzenoid Quinones" *J. Chem. Inf. Comput. Sci.* **1990**, 30, 53-61.
- [127] J. Brunvoll, B.N. Cyvin, S.J. Cyvin "Enumeration and Classification of Coronoid Hydrocarbons" *J. Chem. Inf. Comput. Sci.* **1987**, 27, 14-21.
- [128] R. Tosic, M. Kovacevic "Generating and Counting Unbranched Catacondensed Benzenoids" *J. Chem. Inf. Comput. Sci.* **1988**, 28, 29-31.
- [129] A.T. Balaban "Applications of Graph Theory in Chemistry" *J. Chem. Inf. Comput. Sci.* **1985**, 25, 334-343.
- [130] B. Dzonova-Jerman-Blazic, N. Trinajstic "Application of Reduced Graph Model to the Enumeration of Kekulé Structures and Conjugated Circuits of Benzenoid Hydrocarbons" *Croat. Chim. Acta* **1982**, 55, 347.
- [131] W.C. Herndon "Structure-Resonance Theory. A Review of Applications to pi-Hydrocarbon Systems" *Isr. J. Chem.* **1980**, 20, 270.
- [132] M. Randic "Aromaticity and Conjugation" *J. Am. Chem. Soc.* **1977**, 99, 444.
- [133] E.C. Kirby "A General Method for Finding Principal Resonance Structures for Conjugated Systems by Semi-random Searching of an Adjacency Matrix" *Comput. Chem.* **1985**, 9, 155-163.
- [134] L.A. Zlatina, M.E. Elyashberg "Generation of Stereoisomers and their Spatial Models Corresponding to their Given Molecular Structure" *Commun. Math. Chem. (Match)* **1992**, 27, 191-297.
- [135] M. Razinger, K. Balasubramanian, M. Perdih, M.E. Munk "Stereoisomer Generation in Computer-Enhanced Structure Elucidation" *J. Chem. Inf. Comput. Sci.* **1993**, 33, 812-825.
- [136] T. Wieland "Erzeugung, Abzählung und Konstruktion von Stereoisomeren" *Commun. Math. Chem. (Match)* **1994**, 31, 153-203.
- [137] J.G. Nourse "Specification and Enumeration of Conformations of Chemical Structures for Computer-Assisted Structure Elucidation" *J. Chem. Inf. Comput. Sci.* **1981**, 21, 168.
- [138] J. Koca "Potential Energy Hypersurface and Molecular Flexibility" *J. Mol. Struct.* **1993**, 291, 255-269.

- [139] Y.K. Kim "Conformational Analysis in Molecular Modeling" *Hwahak Sekye* **1993**, 33, 7-10.
- [140] H.R. Henze, C.M. Blair "The Number of Isomeric Hydrocarbons of the Methane Series" *J. Am. Chem. Soc.* **1931**, 53, 3077.
- [141] J. Lederberg, G.L. Sutherland, B.G. Buchanan, E.A. Feigenbaum, A.V. Robertson, A.M. Duffield, C. Djerassi "Applications of Artificial Intelligence for Chemical Inference. I. The Number of Possible Organic Acyclic Structures Containing C, H, O and N" *J. Am. Chem. Soc.* **1969**, 91, 2973.
- [142] R.K. Lindsay, B.G. Buchanan, E.A. Feigenbaum, J. Lederberg *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*, McGraw-Hill, New York **1980**.
- [143] N.A.B. Gray "Dendral and Meta-Dendral - The Myth and the Reality" *Chemom. Intell. Lab. Syst.* **1988**, 5, 11-32.
- [144] B.G. Buchanan, E.A. Feigenbaum, J. Lederberg "On Gray's Interpretation of the Dendral Project and Programs: Myth or Mythunderstanding?" *Chemom. Intell. Lab. Syst.* **1988**, 5, 33-35.
- [145] N.A.B. Gray "Response to Comments by Buchanan, Feigenbaum and Lederberg" *Chemom. Intell. Lab. Syst.* **1988**, 5, 37-38.
- [146] J. Lederberg "Topological Mapping of Organic Molecules" *Proc. Nat. Acad. Sci. (USA)* **1965**, 53, 134.
- [147] L.M. Masinter, N.S. Sridharan, J. Lederberg, D.H. Smith "Applications of Artificial Intelligence for Chemical Inference. XII. Exhaustive Generation of Cyclic and Acyclic Isomers" *J. Am. Chem. Soc.* **1974**, 96, 7702.
- [148] R.E. Carhart, D.H. Smith, H. Brown, C. Djerassi "Applications of Artificial Intelligence for Chemical Inference. XVII. An Approach to Computer-Assisted Elucidation of Chemical Structure" *J. Am. Chem. Soc.* **1975**, 97, 5755.
- [149] Y. Kudo, S. Sasaki "Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information" *J. Chem. Inf. Comput. Sci.* **1976**, 16, 43.
- [150] S. Sasaki, H. Abe, Y. Hirota, Y. Ishida, Y. Kudo, S. Ochiai, K. Saito, T. Yamasaki "CHEMICS-F: A Computer Program System for Structure Elucidation of Organic Compounds" *J. Chem. Inf. Comput. Sci.* **1978**, 18, 211.
- [151] H. Abe, T. Okuyama, I. Fujiwara, S. Sasaki "A Computer Program for Generation of Constitutionally Isomeric Structural Formulas" *J. Chem. Inf. Comput. Sci.* **1984**, 24, 220-229.
- [152] K. Funatsu, N. Miyabayashi, S. Sasaki "Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS" *J. Chem. Inf. Comput. Sci.* **1988**, 28, 18-28.
- [153] Y. Kudo, S. Sasaki "The Connectivity Stack, a New Format for Representation of Organic Structures" *J. Chem. Doc.* **1974**, 14, 200.
- [154] H. Abe, Y. Kudo, T. Yamasaki, K. Tanaka, M. Sasaki, S.I. Sasaki "A Convenient Notation System for Organic Structure on the Basis of Connectivity Stack" *J. Chem. Inf. Comput. Sci.* **1984**, 24, 212-216.
- [155] C.A. Shelley, T.R. Hays, M.E. Munk, R.V. Roman "An Approach to Automated Partial Structure Expansion" *Anal. Chim. Acta* **1978**, 103, 121.
- [156] C.A. Shelley, M.E. Munk "CASE, A Computer Model of the Structure Elucidation Process" *Anal. Chim. Acta* **1981**, 133, 507-516.
- [157] V.V. Serov, M.E. Elyashberg, L.A. Gribov "Mathematical Synthesis and Analysis of Molecular Structures" *J. Mol. Struct.* **1976**, 31, 381.

- [158] R.E. Carhart, D.H. Smith, N.A.B. Gray, J.G. Nourse "GENOA: A Computer Program for Structure Elucidation Utilizing Overlapping and Alternative Substructures" *J. Org. Chem.* **1981**, *46*, 1708-1718.
- [159] A. Kerber, R. Laue, D. Moser "Ein Strukturgenerator für molekulare Graphen" *Anal. Chim. Acta* **1990**, *235*, 221-228.
- [160] R. Grund, A. Kerber, R. Laue "MOLGEN, ein Computeralgebra-System für die Konstruktion molekularer Graphen" *Commun. Math. Chem. (Match)* **1992**, *27*, 87-131.
- [161] C. Benecke, R. Grund, A. Kerber, R. Laue, T. Wieland "News from MOLGEN" *Commun. Math. Chem. (Match)* **1994**, *31*, 229-230.
- [162] B. Glock "Strukturgenerator - preisgekrönt" *Nachr. Chem. Tech. Lab.* **1994**, *42*, 400-403.
- [163] H.J. Luinge, J.H. v.d.Maas "AEGIS, an Algorithm for the Exhaustive Generation of Irredundant Structures" *Chemom. Intell. Lab. Syst.* **1990**, *8*, 157-165.
- [164] H.J. Luinge "AEGIS, a Structure Generation Program in Prolog" *Commun. Math. Chem. (Match)* **1992**, *27*, 175-189.
- [165] H.J. Luinge *EXSPEC, a Knowledge-Based System for Structure Analysis of Organic Molecules from Combined Spectral Data*, Dissertation, University of Utrecht, Utrecht **1989**.
- [166] S. Bohanec, J. Zupan "Structure Generator GEN" *Commun. Math. Chem. (Match)* **1992**, *27*, 49-85.
- [167] J. Bauer, E. Fontain, I. Ugi "IGOR and RAIN - the First Mathematically Based Multipurpose Problem-solving Computer Programs for Chemistry and their Use as Generators of Constitutional Formulas" *Commun. Math. Chem. (Match)* **1992**, *27*, 31-47.
- [168] N.A.B. Gray *Computer-Assisted Structure Elucidation*, Wiley, New York **1986**.
- [169] I. Kubb, Y.W. Lee, M. Pettei, F. Pilkiewicz, K. Nakanishi "Potent Army Worm Antifeedants from the East African Warburgia Plants" *J. Chem. Soc., Chem. Commun.* **1976**, 1013.
- [170] T.K. Devon, A.I. Scott *Handbook of Naturally Occurring Compounds, Volume II, Terpenes*, Academic Press, New York **1972**, *58*, 58-72.
- [171] A. Stock *Hydrides of Boron and Silicon*, Cornell University Press, Ithaca, New York **1933**.
- [172] W.N. Lipscomb *Boron Hydrides*, Benjamin, New York **1963**.
- [173] K. Wade "The Structural Significance of the Number of Skeletal Bonding Electron-pairs in Carboranes, the Higher Boranes and Borane Anions, and Various Transition-metal Carbonyl Cluster Compounds" *J. Chem. Soc., Chem. Commun.* **1971**, 792.
- [174] K. Wade "Structural and Bonding Patterns in Cluster Chemistry" *Adv. Inorg. Chem. Radiochem.* **1976**, *18*, 1-66.
- [175] L. Barton "Systematization and Structures of the Boron Hydrides" *Top. Curr. Chem.* **1982**, *100*, 169-206.
- [176] R.W. Rudolph, D.A. Thompson "Systematics in Boron Hydride Reactivities. Acceptable Valence Structures and Rearrangements in Unimolecular and Bimolecular Nucleophilic and Electrophilic Reactions" *Inorg. Chem.* **1974**, *13*, 2780.
- [177] B. Brellochs, H. Binder "B₆H₁₄, ein dimeres B₃H₇?" *Angew. Chem.* **1988**, *100*, 270-271.
- [178] H. Horn, R. Ahlrichs, C. Kölmel "The Boranes B₃H₇ and B₆H₁₄. A Theoretical Investigation" *Chem. Phys. Lett.* **1988**, *150*, 263-268.
- [179] M.L. McKee "Estimation of Heats of Formation of Boron Hydrides from ab Initio Energies" *J. Phys. Chem.* **1990**, *94*, 435-440.

- [180] H. Binder, B. Brellochs, B. Frei, A. Simon, B. Hettich "Über die ersten monosubstituierten Derivate von Triboran(7).- Kristall- und Molekülstruktur von Benzoyloxytriboran(7)" *Chem. Ber.* **1989**, *122*, 1049-1056.
- [181] M. Bühl, P.v.R. Schleyer, M.L. McKee "The Structures of the hypno-Compounds $B_5H_{12}^-$ and B_6H_{14} : Application of the Combined Ab Initio/IGLO/NMR Method" *Heteroat. Chem.* **1991**, *2*, 499-506.
- [182] E. Fontain "The B_6H_{14} -Problem: Generation of a Catalogue of Conceivable Isomers" *Heteroat. Chem.* **1994**, *5*, 61-64.
- [183] C. Jochum *Algorithmen zur Auswertung konstitutioneller Information organisch-chemischer Strukturen*, Dissertation, Technische Universität München **1978**.
- [184] J. Pospichal, V. Kvasnicka "Fast Evaluation of Chemical Distance by Simulated-Annealing Algorithm" *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 879-885.
- [185] W. Hüchel *Theoretische Grundlagen der Organischen Chemie*, Akademische Verlagsgesellschaft, Leipzig, 8.Aufl. **1956**, *2*, 216-219.
- [186] C. Jochum, J. Gasteiger, I. Ugi, J. Dugundji "The Principle of Minimum Chemical Distance and the Principle of Minimum Structure Change" *Z. Naturforsch.* **1982**, *37b*, 1205-1215.
- [187] J.B. Hendrickson "A Systematic Organization of Synthetic Reactions" *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 129-136.
- [188] J.B. Hendrickson, C.A. Parks "A Program for the Forward Generation of Synthetic Routes" *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 209-215.
- [189] J.B. Hendrickson, E. Braun-Keller "Systematic Synthesis Design. 8. Generation of Reaction Sequences" *J. Comput. Chem.* **1980**, *1*, 323-333.
- [190] V. Kvasnicka, J. Pospichal "Graph-theoretical Interpretation of Ugi's Concept of the Reaction Network" *J. Math. Chem.* **1990**, *5*, 309-322.
- [191] E.L. Lawler "Branch and Bound Method" *J. Oper. Res. Soc. Am.* **1966**, *14*, 217-235.
- [192] J.J. McGregor "Backtrack Search Algorithm and the Maximal Common Subgraph Problem" *Software Pract. Exp.* **1982**, *12*, 23-34.
- [193] V. Kvasnicka, J. Pospichal "Maximal Common Subgraphs of Molecular Graphs" *Rep. Mol. Theory* **1990**, *1*, 99-106.
- [194] V. Nicholson, C.C. Tsai, M. Johnson, M. Naim "A Subgraph Isomorphism Theorem for Molecular Graphs" *Graph Theory and Topology in Chemistry*, R.B. King, D. Rouvray (Eds.), Elsevier, Amsterdam **1987**, *51*, 226-230.
- [195] C. Bron, J. Kerbosch *Comm. ACM* **1973**, *16*, 575.
- [196] M.F. Lynch, P. Willett "The Automatic Detection of Chemical Reaction Sites" *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 154-159.
- [197] W. Schubert, W. Ellenrieder "Automatic Determination of Common Substructures" *J. Chem. Res. (S)* **1984**, 258-259.
- [198] M. Wochner *Die exakte Lösung der Zuordnungsaufgabe für chemische Reaktionen*, Dissertation, Technische Universität München **1985**.
- [199] M. Wochner, J. Brandt, A. v.Scholley, I. Ugi "Chemical Similarity, Chemical Distance, and its Exact Determination" *Chimia* **1988**, *42*, 217-225.
- [200] M.H. Kalos, D.A. Whitlock *The Basic of Monte Carlo Methods*, J. Wiley, New York **1986**.
- [201] P.M.J. v.Laarhoven, E.H.L. Aarts *Simulated Annealing: Theory and Applications*, Reidel: Dordrecht, The Netherlands **1987**.

- [202] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller "Equation of State Calculation for Fast Computing Machines" *J. Chem. Phys.* **1953**, *21*, 1087-1092.
- [203] M.E. Snow "Powerful Simulated Annealing Algorithm Locates Global Minimum of Protein-Folding Potentials from Multiple Starting Conformations" *J. Comput. Chem.* **1992**, *13*, 579-584.
- [204] E. Fontain "Application of Genetic Algorithms in the Field of Constitutional Similarity" *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 748-752.
- [205] E. Fontain "The Problem of Atom-to-Atom Mapping. An Application of Genetic Algorithms" *Anal. Chim. Acta* **1992**, *265*, 227-232.
- [206] M. Wagener, J. Gasteiger "Die Bestimmung größter deckungsgleicher Teilstrukturen mit einem genetischen Algorithmus: Anwendung in der Syntheseplanung und zur strukturellen Analyse biologischer Aktivität" *Angew. Chem.* **1994**, *106*, 1245-1248.
- [207] T. Schmid, M. Hanack, C. Maichle, J. Strähle "Cyansulfine: Eine unerwartete Reaktion von *N,N*-Bis(trimethylsilyl)inaminen" *Angew. Chem.* **1993**, *105*, 300-302.
- [208] I. Ugi, A. Dömling, W. Hörl "Multikomponentenreaktionen" *GIT Fachzeitschrift für das Laboratorium* **1994**, *38*, 430-437.
- [209] A. Dömling, I. Ugi "Die Siebenkomponenten-Reaktion" *Angew. Chem.* **1993**, *105*, 634-635.
- [210] I. Ugi "Mit Sekundär-Reaktionen gekoppelte Alpha-Additionen von Immonium-Ionen und Anionen an Isonitrile" *Angew. Chem.* **1962**, *74*, 9-22.
- [211] F. Asinger, M. Thiel "Einfache Synthesen und chemisches Verhalten neuer heterocyclischer Ringsysteme" *Angew. Chem.* **1958**, *70*, 667-683.
- [212] R. Uma, S. Swaminathan, K. Rajagopalan "Base-Catalyzed Rearrangement of Oxy-Cope Systems" *Tetrahedron Lett.* **1984**, *25*, 5825-5828.
- [213] I. Ugi, J. Bauer, C. Blomberger, J. Brandt, A. Dietz, E. Fontain, B. Gruber, A. v.Scholley-Pfab, A. Senff, N. Stein "Models, Concepts, Theories, and Formal Languages in Chemistry and Their Use as a Basis for Computer Assistance in Chemistry" *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 3-16.
- [214] S.L. Miller "Production of Some Organic Compounds under Possible Primitive Earth Conditions" *J. Am. Chem. Soc.* **1955**, *77*, 2351-2361.
- [215] S.L. Miller "The Mechanism of Synthesis of Amino Acids by Electric Discharges" *Biochim. et Biophys. Acta* **1957**, *23*, 480-489.
- [216] S.L. Miller "The Formation of Organic Compounds on the Primitive Earth" *Ann. N. Y. Acad. Sci.* **1957**, *69*, 260-275.
- [217] S.L. Miller *Science* **1959**, *130*, 245.
- [218] G. Waechtershaeuser "Biomolecules: the Origin of Their Optical Activity" *Med. Hypotheses* **1991**, *36*, 307-311.
- [219] G. Waechtershaeuser "Groundworks for An Evolutionary Biochemistry: the Iron-Sulfur World" *Prog. Biophys. Mol. Biol.* **1992**, *58*, 85-201.
- [220] W. Gilbert "The RNA World" *Nature* **1986**, *319*, 618-618.
- [221] E. Bloechl, M. Keller, G. Waechtershaeuser, K.O. Stetter "Reactions Depending on Iron Sulfide and Linking Geochemistry with Biochemistry" *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 8117-8120.
- [222] G. Waechtershaeuser "The Cradle Chemistry of Life: on the Origin of Natural Products in Pyrite-Pulled Chemoautotrophic Origin of Life" *Pure Appl. Chem.* **1993**, *65*, 1343-1348.
- [223] J. Oro, A.P. Kimball "Synthesis of Purines under Possible Primitive Earth Conditions. I. Adenine from Hydrogen Cyanide" *Arch. Biochem. Biophys.* **1961**, *94*, 217-227.

- [224] H. Wakamatsu, Y. Yamada, T. Saito, I. Kumashiro, T. Takenishi "Synthesis of Adenine by Oligomerization of Hydrogen Cyanide" *J. Org. Chem.* **1966**, 31, 2035-2036.
- [225] J.P. Ferris, L.E. Orgel "Aminomalononitrile and 4-Amino-5-cyanoimidazole in Hydrogen Cyanide Polymerization and Adenine Synthesis" *J. Am. Chem. Soc.* **1965**, 87, 4976-4977.
- [226] J.P. Ferris, L.E. Orgel "An Unusual Photochemical Rearrangement in the Synthesis of Adenine from Hydrogen Cyanide" *J. Am. Chem. Soc.* **1966**, 88, 1074.
- [227] J.P. Ferris, L.E. Orgel "Studies in Prebiotic Synthesis. I. Aminomalononitrile and 4-Amino-5-cyanoimidazole" *J. Am. Chem. Soc.* **1966**, 88, 3829-3831.
- [228] H. Yamada, M. Hirobe, K. Higashiyama, H. Takahashi, K.T. Suzuki "Reaction Mechanism for Purine Ring Formation as Studied by ^{13}C - ^{15}N Coupling" *Tetrahedron Lett.* **1978**, 42, 4039-4042.
- [229] O.W. Webster, D.R. Hartter, R.W. Begland, W.A. Sheppard, A. Cairncross "Hydrogen Cyanide Chemistry. III. Synthesis of Diiminosuccinonitrile and Its Conversion to Diaminomaleonitrile" *J. Org. Chem.* **1972**, 37, 4133-4155.
- [230] J.P. Ferris, P.C. Joshi, E.H. Edelson, J.G. Lawless "HCN: A Plausible Source of Purines, Pyrimidines and Amino Acids on the Primitive Earth" *J. Mol. Evol.* **1978**, 11, 293-311.
- [231] A.B. Voet, A.W. Schwartz "Prebiotic Adenine Synthesis from HCN - Evidence for a New Discovered Major Pathway" *Bioorg. Chem.* **1983**, 12, 8-17.
- [232] S. Drenkard, J. Ferris, A. Eschenmoser "Aziridin-2-carbonitril: photochemische Bildung aus 2-Aminopropennitril" *Helv. Chim. Acta* **1990**, 73, 1373-1390.
- [233] E. Wagner, Y.-B. Xiang, K. Baumann, J. Glück, A. Eschenmoser "Aziridin-2-carbonitril, ein Vorläufer von *rac*- O^3 -Phosphoserinnitril und Glycolaldehyd-phosphat" *Helv. Chim. Acta* **1990**, 73, 1391-1409.
- [234] J.P. Ferris, J.E. Kuder "Chemical Evolution. III. The Photochemical Conversion of Enaminonitriles to Imidazoles" *J. Am. Chem. Soc.* **1970**, 92, 2527-2533.
- [235] J.H. Holland *Adaptation in Natural and Artificial Systems*, Ann Arbor: The University of Michigan Press **1975**.
- [236] J.H. Holland "Genetische Algorithmen" *Spektr. Wiss.* **1992**, 7, 44-51.
- [237] D.E. Goldberg *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading Massachusetts **1989**.
- [238] C.B. Lucasius, G. Kateman "Understanding and Using Genetic Algorithms. Part 1. Concepts, Properties and Context" *Chemom. Intell. Lab. Syst.* **1993**, 19, 1-33.
- [239] D.B. Hibbert "Genetic Algorithms in Chemistry" *Chemom. Intell. Lab. Syst.* **1993**, 19, 277-293.
- [240] R.S. Judson, E.P. Jaeger, A.M. Treasurywala, M.L. Peterson "Conformational Searching Methods for Small Molecules. II. Genetic Algorithm Approach" *J. Comput. Chem.* **1993**, 14, 1407-1414.
- [241] M. Heilingbrunner *Optimierung molekularer Konformation mit einem Genetischen Algorithmus*, Diplomarbeit, Technische Universität München **1993**.
- [242] A.W.R. Payne, R.C. Glen "Molecular Recognition Using a Binary Genetic Search Algorithm" *J. Mol. Graphics* **1993**, 11, 74-91.
- [243] B. Hartke "Global Geometry Optimization of Clusters Using Genetic Algorithms" *J. Phys. Chem.* **1993**, 97, 9973-9976.
- [244] Y. Xiao, D.E. Williams "Genetic Algorithm: a New Approach to the Prediction of the Structure of Molecular Clusters" *Chem. Phys. Lett.* **1993**, 215, 17-24.

- [245] M.J.J. Blommers, C.B. Lucasius, G. Kateman, R. Kaptein "Conformational Analysis of a Dinucleotide Photodimer with the Aid of the Genetic Algorithm" *Biopolymers* **1992**, 32, 45-52.
- [246] C.B. Lucasius, S. Werten, A.H. vanAert, G. Kateman, M.J. Blommers "Conformational Analysis of DNA using Genetic Algorithms" *Lect. Notes Comput. Sci., 'Parallel Problem Solving in Nature'*, H.P. Schwefel, R. Männer (Eds.), Springer-Verlag, Berlin Heidelberg **1990**, 496, 90-97.
- [247] D.B. McGarrah, R.S. Judson "Analysis of the Genetic Algorithm Method of Molecular Conformation Determination" *J. Comput. Chem.* **1993**, 14, 1385-1395.
- [248] R.S. Judson "Teaching Polymers to Fold" *J. Phys. Chem.* **1992**, 96, 10102-10104.
- [249] C.B. Lucasius, M.J. Blommers, L.M. Buydens, G. Kateman "A Genetic Algorithm for Conformational Analysis of DNA" *Handbook of Genetic Algorithms*, L. Davis (Ed.), Van Nostrand Reinhold, New York London **1991**, 251-281.
- [250] T. Dandekar, P. Argos "Potential of Genetic Algorithms in Protein Folding and Protein Engineering Simulations" *Protein Eng.* **1992**, 5, 637-645.
- [251] T. Dandekar, P. Argos "Folding the Main Chain of Small Proteins with the Genetic Algorithm" *J. Mol. Biol.* **1994**, 236, 844-861.
- [252] S. Sun "Reduced Representation Model of Protein Structure Prediction: Statistical Potential and Genetic Algorithms" *Protein Sci.* **1993**, 2, 762-785.
- [253] R. Unger, J. Moult "Genetic Algorithms for Protein Folding Simulations" *J. Mol. Biol.* **1993**, 231, 75-81.
- [254] R. Unger, J. Moult "Effects of Mutations on the Performance of Genetic Algorithms Suitable for Protein Folding Simulations" *Comput. Aided Innovation New Mater.* 2, Proc. Int. Conf. Exhib. Comput. Appl. Mater. Mol. Sci. Eng., 2nd, Meeting Date 1992, Issue Pt. 2, 1283-6. Edited by: Doyama, Masao. North-Holland: Amsterdam **1993**, 2, 1283-1286.
- [255] S.M. Le Grand "The Application of the Genetic Algorithm to Protein Tertiary Structure Prediction" *Diss. Abstr. Int. B* **1993**, 54, 3603-3604.
- [256] W. Cedeno, V. Vemuri "An Investigation of DNA Mapping with Genetic Algorithms: Preliminary Results" *Proc. SPIE-Int. Soc. Opt. Eng.* **1993**, 2204, 133-140.
- [257] D.B. Hibbert "A Hybrid Genetic Algorithm for the Estimation of Kinetic Parameters" *Chemom. Intell. Lab. Syst.* **1993**, 19, 319-329.
- [258] C.L. Karr, S.K. Sharma, W. Hatcher, T.R. Harper "Fuzzy Logic and Genetic Algorithms for the Control of An Exothermic Chemical Reaction" *Simul. Control Hydrometall. Processes, Proc. Int. Symp.* Edited by: Papangelakis, Vladimiro G.; Demopoulos, George P. Can. Inst. Min. Metall. Pet.: Montreal, Canada **1993**, 227-236.
- [259] S.K. Sharma, D.A. Stanley, D.R. Spears "Mathematical Modeling of Shear Flocculation Using a Genetic Algorithm" *Fluid/Part. Sep. J.* **1993**, 6, 10-14.
- [260] H.M. Cartwright, R.A. Long "Simultaneous Optimization of Chemical Flowshop Sequencing and Topology Using Genetic Algorithms" *Ind. Eng. Chem. Res.* **1993**, 32, 2706-2713.
- [261] K. Matsuura, H. Shiba, Y. Nunokawa, H. Shimizu, S. Shioya, K. Suga "Calculation of Optimal Trajectories for Fermentation Processes by Genetic Algorithm" *Seibutsu Kagaku Kaishi* **1993**, 71, 171-178.
- [262] M. Karutz, B. Joksch, S. Freyer, D. Weuster-Botz, C. Wandrey "Fermentative L-Isoleucine Production with Optimized Concentrations of Trace Elements - Application of Genetic Algorithms" *BioEngineering* **1992**, 8, 26-34.
- [263] S. Freyer, D. Weuster-Botz, C. Wandrey "Medium Optimization by Genetic Algorithms" *BioEngineering* **1992**, 8, 16-25.

- [264] D.S. Szarkowicz "A Multi-Stage Adaptive-Coding Genetic Algorithm for Design Applications" *Proc. Summer Comput. Simul. Conf.* **1991**, 138-144.
- [265] P.W. Poon "Genetic Algorithms and Fuel Cycle Optimization" *Nucl. Eng. (Inst. Nucl. Eng.)* **1990**, 31, 173-178.
- [266] I.P. Androulakis, V. Venkatasubramanian "A Genetic Algorithmic Framework for Process Design and Optimization" *Comput. Chem. Eng.* **1991**, 15, 217-228.
- [267] D.B. Hibbert "Generation and Display of Chemical Structures by Genetic Algorithms" *Chemom. Intell. Lab. Syst.* **1993**, 20, 35-43.
- [268] M. Freymann *Parameteroptimierung eines Genetischen Algorithmus zur Berechnung der Minimalen Chemischen Distanz*, Diplomarbeit, Technische Universität München **1992**.
- [269] M. Wagener *Bestimmung molekularer Ähnlichkeit mit Hilfe eines Genetischen Algorithmus*, Dissertation, Technische Universität München **1993**.
- [270] R.D. Brown, G. Jones, P. Willett, R.C. Glen "Matching Two-Dimensional Chemical Graphs Using Genetic Algorithms" *J. Chem. Inf. Comput. Sci.* **1994**, 34, 63-70.
- [271] R.D. Brown, G.M. Downs, G. Jones, P. Willett "Hyperstructure Model for Chemical Structure Handling: Techniques for Substructure Searching" *J. Chem. Inf. Comput. Sci.* **1994**, 34, 47-53.
- [272] C.B. Lucasius, M.L.M. Beckers, G. Kateman "Genetic Algorithms in Wavelength Selection: a Comparative Study" *Anal. Chim. Acta* **1994**, 286, 135-153.
- [273] A.P. De Weijer, C.B. Lucasius, L. Buydens, G. Kateman, H.M. Heuvel, H. Mannee "Curve Fitting Using Natural Computation" *Anal. Chem.* **1994**, 66, 23-31.
- [274] C.B. Lucasius, A.P. de Weijer, L.M.C. Buydens, G. Kateman "CFIT: a Genetic Algorithm for Survival of the Fitting" *Chemom. Intell. Lab. Syst.* **1993**, 19, 337-341.
- [275] D. Wienke, C. Lucasius, M. Ehrlich, G. Kateman "Multicriteria Target Vector Optimization of Analytical Procedures Using a Genetic Algorithm. Part II. Polyoptimization of the Photometric Calibration Graph of Dry Glucose Sensors for Quantitative Clinical Analysis" *Anal. Chim. Acta* **1993**, 271, 253-268.
- [276] D. Wienke, C. Lucasius, G. Kateman "Multicriteria Target Vector Optimization of Analytical Procedures Using a Genetic Algorithm. Part I. Theory, Numerical Simulations and Application to Atomic Emission Spectroscopy" *Anal. Chim. Acta* **1992**, 265, 211-225.
- [277] M. Bos, H.T. Weber "Comparison of the Training of Neural Networks for Quantitative X-Ray Fluorescence Spectrometry by a Genetic Algorithm and Backward Error Propagation" *Anal. Chim. Acta* **1991**, 247, 97-105.
- [278] R. Wehrens, C. Lucasius, G. Kateman "Sequential Assignment of 2D-NMR Spectra of Proteins Using Genetic Algorithms" *J. Chem. Inf. Comput. Sci.* **1993**, 33, 245-251.
- [279] R. Wehrens, C. Lucasius, L. Buydens, G. Kateman "HIPS, a Hybrid Self-Adapting Expert System for Nuclear Magnetic Resonance Spectrum Interpretation Using Genetic Algorithms" *Anal. Chim. Acta* **1993**, 277, 313-324.
- [280] R. Leardi "Application of a Genetic Algorithm to Feature Selection Under Full Validation Conditions and to Outlier Detection" *J. Chemom.* **1994**, 8, 65-79.
- [281] C.B. Lucasius, A.D. Dane, G. Kateman "On k-Medoid Clustering of Large Data Sets with the Aid of a Genetic Algorithm: Background, Feasibility and Comparison" *Anal. Chim. Acta* **1993**, 282, 647-669.
- [282] A.P. de Weijer, C.B. Lucasius, L. Buydens, G. Kateman, H.M. Heuvel "Using Genetic Algorithms for An Artificial Neural Network Model Inversion" *Chemom. Intell. Lab. Syst.* **1993**, 20, 45-55.

- [283] R. Leardi, R. Boggia, M. Terrile "Genetic Algorithms As a Strategy for Feature Selection" *J. Chemom.* **1992**, 6, 267-281.
- [284] C.B. Lucasius, G. Kateman "Genetic Algorithms for Large-scale Optimization in Chemometrics: an Application" *Trends Anal. Chem.* **1991**, 10, 254-261.
- [285] G. Kateman "Evolutions in Chemometrics" *Analyst* **1990**, 115, 487-493.
- [286] H.M. Cartwright, S.P. Harris "Analysis of the Distribution of Airborne Pollution Using Genetic Algorithms" *Atmos. Environ.* **1993**, 27A, 1783-1791.
- [287] M. Furukawa "Simulating Tree Formulation Under Specified Climatic Environment by Use of Genetic Algorithm" *Kenkyu Hobun - Asahikawa Kogyo Koto Senmon Gakko* **1993**, 30, 71-81.
- [288] Z. Guo "Nuclear Power Plant Fault Diagnostics and Thermal Performance Studies Using Neural Networks and Genetic Algorithms" *Diss. Abstr. Int. B* 1993, 53(7), 3741 **1992**.
- [289] *Computational Chemistry, HyperChem, Practical Guide, Theory and Methods* Autodesk Inc., Sausalito, CA. **1992**.
- [290] H.J. Bremermann "Optimization through Evolution and Recombination" *Self-Organizing Systems*, M.C. Yovits, G.T. Jacobi, D.G. Goldstein (Eds.), Spartan, Washington D.C. **1962**, 93-106.
- [291] I. Rechenberg *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Frommann-Holzboog, Stuttgart **1973**.
- [292] H.P. Schwefel *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*, Birkhäuser Verlag, Basel und Stuttgart **1977**.
- [293] G. Dueck, T. Scheuer, H.M. Wallmeier "Toleranzschwelle und Sintflut: neue Ideen zur Optimierung" *Spektr. Wiss.* **1993**, 3, 42-51.
- [294] C.B. Lucasius, G. Kateman *GATES: Genetic Algorithm Toolbox for Evolutionary Search*, Software library in ANSI C, Laboratory for Analytical Chemistry, Katholieke Universiteit Nijmegen **1991**.
- [295] J.J. Grefenstette "A User's Guide to GENESIS" *Technical Report No. CS-84-11*, Computer Science Department, Vanderbilt University, Nashville, Tennessee **1984**.
- [296] D. Whitley "The GENITOR Algorithm and Selection Pressure: Why Ranked-Based Allocation of Reproductive Trials is Best" *Proceedings of the Third International Conference on Genetic Algorithms*, J.D. Schaffer (Ed.), Morgan Kaufmann, San Mateo, CA. **1989**, 116-121.
- [297] D.E. Goldberg, R. Lingle "Alleles, Loci, and the Traveling Salesman Problem" *Proceedings of an International Conference on Genetic Algorithms and Their Applications* **1985**, 154-159.
- [298] K.A. DeJong *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*, Dissertation, University of Michigan **1975**.
- [299] D.E. Goldberg "Genetic Algorithms with Sharing for Multimodal Function Optimization" *Proceedings of the Second International Conference on Genetic Algorithms*, J.J. Grefenstette (Ed.), Lawrence Erlbaum Associates **1987**, 41-49.
- [300] K. Deb, D.E. Goldberg "An Investigation of Niche and Species Formation in Genetic Function Optimization" *Proceedings of the Third International Conference on Genetic Algorithms*, J.D. Schaffer (Ed.), Morgan Kaufmann Publishers **1989**, 42-50.