

# Die computerunterstützte Lösung chemischer Probleme – eine neue Disziplin der Chemie

Von Ivar Ugi\*, Johannes Bauer, Klemens Bley, Alf Dengler, Andreas Dietz,  
Eric Fontain, Bernhard Gruber, Rainer Herges, Michael Knauer,  
Klaus Reitsam und Natalie Stein

*Professor Karl-Heinz Büchel gewidmet*

Die 25jährige Entstehungsgeschichte und der gegenwärtige Entwicklungsstand der Computerchemie, der computerunterstützten Lösung chemischer Probleme, ist Gegenstand dieser Übersicht. Anfänglich war die Computerchemie auf die Strukturermittlung anhand spektroskopischer Daten und auf die auf Reaktionsdateien basierende Syntheseplanung für eine relativ enge Auswahl von Verbindungsklassen beschränkt, während nun Computerprogramme für die Lösung einer großen Vielfalt chemischer Probleme verfügbar sind oder entwickelt werden. Früher war man froh, mit Computerunterstützung überhaupt Lösungen chemischer Probleme zu erhalten, heute ist die effiziente, transparente und möglichst willkürfreie Auswahl sinnvoller Lösungen aus der großen Menge nunmehr beliebig generierbarer potentieller Problemlösungen gefragt. Zwei Aspekte der Chemie, die konstitutive Chemie und die Stereochemie, hängen zwar zusammen, erfordern aber doch recht unterschiedliche Betrachtungsweisen. Vor rund zwanzig Jahren wurde deshalb ein zweiteiliges algebraisches Modell der logischen Struktur der Chemie vorgestellt, die konstitutionsbezogene Algebra der be- und r-Matrizen und die stereochemiebezogene Theorie der chemischen Identitätsgruppe. Neue chemische Konzepte, Definitionen und Betrachtungsweisen sowie die direkte mathematische Repräsentation chemischer Prozesse kennzeichnen dieses Logikmodell. Es ermöglicht die Implementierung formaler Reaktionsgeneratoren, die ohne detaillierte empirische chemische Informationen die denkbaren Lösungen chemischer Probleme, einschließlich der präzedenzlosen, erzeugen können. Auch neue formale Selektionsverfahren für computergenerierte chemische Informationen, die zweckmäßig mit interaktiven Auswahlmethoden kombiniert werden, ermöglicht dieses Modell. Schwerpunkt dieser Übersicht ist das Münchner Projekt, das sowohl die Weiterentwicklung und Erweiterung des mathematischen Modells der logischen Struktur der Chemie und die Implementierung darauf basierender Computerprogramme als auch die experimentelle Realisierung der mit diesen Computerprogrammen generierten Vorschläge umfaßt. Es wird eine Übersicht über neue Reagentien, Reaktionen und Reaktionsmechanismen gegeben, die mit den inzwischen bei uns verfügbaren interaktiven PC-Programmen IGOR und RAIN gefunden wurden.

## 1. Einleitung

Die Chemie befaßt sich mit der Analytik und Synthese chemischer Stoffe sowie der Ermittlung ihrer Molekülstrukturen, Reaktionsweisen und Stoffeigenschaften. Vom Erkenntnisgewinn abgesehen, rechtfertigen auch die vielfältigen Nutzanwendungen der Chemie den Aufwand der chemischen Forschung. Die Suche nach Stoffen mit bestimmten erwünschten Wirkungen, Funktionen und Eigenschaften sowie die Entwicklung optimaler Methoden zu deren Synthese und Qualitätsprüfung kennzeichnen die Forschung in der angewandten Chemie.

Der Fortschritt in der Chemie kommt im allgemeinen durch systematische Forschung zustande und vollzieht sich evolutionär in kleinen Schritten<sup>[1]</sup>.

Die seltenen revolutionären Innovationsschübe ergeben sich durch neue Denkweisen sowie sachkundig erkannte und konsequent genutzte zufällige Beobachtungen. Beispiele hierfür sind die Erkenntnis der Bedeutung von Konformationen<sup>[2]</sup> durch Sachse<sup>[3]</sup> und Mohr<sup>[4]</sup>, Meerweins Konzept der ionisch ablaufenden Reaktionen<sup>[5]</sup>, die Entdeckung der ersten übergangsmetallorganischen Katalysatoren durch Reppe et al.<sup>[6a]</sup> und Roelen<sup>[7a]</sup>, die Lösung des Problems der absoluten Konfigurationen durch Bijvoet et al.<sup>[8]</sup> und die Erschließung der Chemie supersaurer Medien durch Olah et al.<sup>[9]</sup>.

Auch Glück und bisweilen sogar anfängliche Mißverständnisse spielen eine wichtige Rolle, vor allem in der angewandten Chemie. Spektakuläre Beispiele sind die von Domagk, Mietsch und Klarer<sup>[10]</sup> eingeführten Sulfonamide, welche die Chemotherapie der bakteriellen Infektionen revolutionierten, und die von Büchel und Plempel<sup>[11]</sup> entdeckten Azol-Fungizide, durch die Pilzinfektionen erstmals effizient und in aller Breite bekämpfbar wurden.

Die mannigfaltigen Einsatzmöglichkeiten für Computer in der chemischen Forschung entsprechen der Vielfalt der Forschungsaktivitäten. Computer dienen der chemischen Forschung, seitdem sie verfügbar sind. Den ersten käufli-

[\*] Prof. Dr. I. Ugi, Dr. J. Bauer, Dr. K. Bley, Dr. A. Dengler,  
Dipl.-Chem. A. Dietz, Dr. E. Fontain, Dr. B. Gruber, Dr. M. Knauer,  
Dipl.-Chem. K. Reitsam, Dipl.-Chem. N. Stein  
Organisch-chemisches Institut der Technischen Universität München  
Lichtenbergstraße 4, W-8046 Garching  
Dr. R. Herges  
Organisch-chemisches Institut der Universität Erlangen-Nürnberg  
Henkestraße 42, W-8520 Erlangen

chen Computer, einen Magnettrommelrechner Z22, verkaufte Konrad Zuse<sup>[12]</sup>, der Erfinder des Computers, 1959 an die Bayer AG, welche den Z22 auch für wissenschaftliche Zwecke einsetzte.

Anfangs wurden Computer in der Chemie ausschließlich für numerische Berechnungen genutzt, und diese Nutzung nimmt immer noch zu. Für diese Anwendungen der Computer, unter denen die Berechnung der Energieniveaus und Geometrien molekularer Systeme eine dominierende Rolle spielt, verwendet man den kollektiven Terminus *Computational Chemistry*<sup>[13–15]</sup>, für den es bisher noch keine akzeptable deutsche Übersetzung gibt.

Computer sind ein unentbehrlicher Bestandteil des Instrumentariums der Quantenchemie und der modernen Analytischen Chemie, einschließlich der Strukturermittlung mit röntgenographischen und spektroskopischen Methoden. Darüber hinaus helfen sie bei der Bestimmung von Stoffeigenschaften, der Planung von Experimenten und Messungen, der Erfassung, Speicherung, statistischen Auswertung, graphischen Darstellung und Interpretation von Meßdaten sowie der visuellen Darstellung von Molekülen, deren Atomkoordinaten aus Meßdaten oder durch numerische Quantenchemie und Kraftfeldrechnungen gewonnen wurden.

Dieser Gebrauch von Computern als Rechenhilfe in der Chemie – meist sind es Rechnungen nach festgelegten „Rezepten“ – ist weitgehend automatisierbar. Abgesehen vom molecular modeling, bei dem sowohl die numerischen als auch die graphischen Fähigkeiten der Computer genutzt werden, kann auf eine interaktive Beteiligung des Benutzers bei den genannten numerischen Anwendungen der Computer weitgehend verzichtet werden.

Es gibt jedoch auch viele Arten chemischer Probleme, die zwar mit Computerunterstützung, nicht aber durch numerische Rechnungen lösbar sind. In diesen Fällen werden vor allem die logischen und kombinatorischen Fähigkeiten der

Computer benötigt. Man bezeichnet die entsprechende Disziplin heute als computerunterstützte Chemie oder kurz *Computerchemie*<sup>[16–20]</sup>. Letzterer Begriff ist zwar nicht optimal gewählt, hat sich aber eingebürgert.

Die ältesten und bekanntesten Computerprogramme zur Lösung chemischer Probleme dienen der Syntheseplanung sowie der Strukturermittlung anhand spektroskopischer Daten. Die Suche nach Synthesewegen oder Reaktionsmechanismen, d. h. Folgen von Zwischenstufen oder Reaktionsschritten, und die Ermittlung von Molekülstrukturen anhand spektroskopischer Daten gehören zu den „klassischen“ chemischen Problemen, bei deren Lösung es darum geht, ein oder mehrere Moleküle oder chemische Reaktionen zu finden, die bestimmte chemische Merkmale haben. Klassische chemische Probleme haben einen erheblichen kombinatorischen Aspekt und viele denkbare Lösungen, die nicht durch numerische Rechnungen erhältlich sind; letztere können aber mittelbar von Nutzen sein<sup>[21]</sup>. Die besten Lösungen werden oft durch intuitive und kreative Denkprozesse nebst „trial and error“ gefunden. Die computerunterstützte Lösung chemischer Probleme bedarf der interaktiven Mitwirkung des Anwenders, denn es gibt meist sehr viele Lösungen und keine allgemeingültigen Regeln, die eine willkürfreie Auswahl der brauchbarsten Lösungen ermöglichen. Die Methoden zur Lösung solcher Probleme sind im allgemeinen nichtdeterministisch.

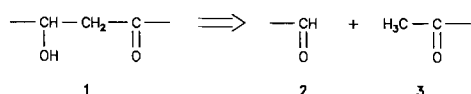
Joshua Lederberg und der 1990 verstorbene George Vleducs sind die geistigen Väter der Computerchemie. Das von Lederberg<sup>[22]</sup> initiierte DENDRAL-Projekt<sup>[23]</sup> hatte ein Computerprogramm zum Ziel, das anhand spektroskopischer Daten die signifikanten Strukturmerkmale eines Moleküls ermittelt und daraus dessen Struktur zusammensetzt<sup>[24]</sup>. Das DENDRAL-Projekt wurde zwar eingestellt, und seine Ziele sind nur teilweise erreicht, aber DENDRAL repräsentiert dennoch einen bedeutenden Meilenstein in der



*Ivar Ugi wurde am 5. September 1930 zu Arensburg in Estland geboren. Im Jahr 1941 siedelte die Familie Ugi nach Dillingen an der Donau um, wo Ivar Ugi 1949 sein Abitur machte. Im Herbst desselben Jahres begann er das Studium der Chemie an der Universität Tübingen, das er nach dem Diplom 1952 und einem Wechsel an die Ludwig-Maximilian-Universität in München 1954 mit der Promotion bei Prof. Huisgen über „Die sterische Hinderung der Mesomerie in mittleren Ringen“ abschloß. An derselben Universität habilitierte sich Ugi 1959 mit der Arbeit „Pentazole und Isonitrile“ und widmete sich danach noch zwei Jahre der Hochschulforschung. 1962 wechselte er zur Fa. Bayer nach Leverkusen, wo er zum Direktor und Vorsitzenden der Kommission für Grundlagenforschung aufstieg. Nach sechsjähriger Firmenzugehörigkeit zog es den forschenden Geist nach Amerika: 1968 nahm Ugi die Professur als Full Professor of Chemistry an der University of Southern California in Los Angeles an. Im Juli 1971 folgte er dem Ruf als ordentlicher Professor auf den traditionsreichen Lehrstuhl I für Organische Chemie der Technischen Universität München. Ivar Ugi arbeitet in den Komitees mehrerer Stiftungen mit und ist seit 1987 ordentliches Mitglied der schwedischen königlichen Gesellschaft der Wissenschaften in Uppsala und seit 1990 Mitglied der Estnischen Akademie der Wissenschaften. Seine wissenschaftlichen Leistungen wurden mit mehreren Auszeichnungen gewürdigt: 1964 erhielt er den Forschungspreis der Akademie der Wissenschaften in Göttingen und 1988 den Philip-Morris-Forschungspreis für das zusammen mit dem Topologen Dugandji entwickelte Dugandji-Ugi-Modell. 1992 folgte die Ehrung seines Lebenswerkes mit der Emil-Fischer-Medaille der Gesellschaft Deutscher Chemiker, zum einen für die Entwicklung mathematischer Modelle der logischen Strukturen der Chemie, zum anderen für eine Fülle neuer präparativer Methoden, die aus der Isonitrilchemie, der Vier-Komponenten-Kondensation und dem stereochemischen Strukturmodell resultieren und zahlreiche neue Arbeitsgebiete eröffnen. Diese Synthesemethoden geben starke Impulse insbesondere der Chemie der Peptide, Nucleotide und  $\beta$ -Lactam-Antibiotica.*

Entwicklung der Computerchemie<sup>[13]</sup>. Es war das erste Computerprogramm, das zur Lösung chemischer Probleme auch formale Mittel einsetzte. DENDRAL ist außerdem mit der Entstehung des Begriffs artificial intelligence eng verbunden<sup>[23]</sup>.

Bereits 1963 wies Vleduts<sup>[25]</sup> den Weg zu Computerprogrammen für die Planung mehrstufiger Synthesen (CAOS: Computer Assisted Organic Synthesis)<sup>[26]</sup>, d. h. für die Ermittlung einer nach gewählten Kriterien optimalen Möglichkeit, eine gegebene Zielverbindung aus verfügbaren Stoffen aufzubauen. Als einer der Pioniere von Reaktionsdatenbanken<sup>[27]</sup> erörterte Vleduts als erster eingehend, wie man durch Umkehr gespeicherter Reaktionen (Retroreaktionen) retrosynthetisch Synthesewege erzeugen kann, die vom Synthesziel über Vorstufen zu verfügbaren Ausgangsstoffen führen. Beispielsweise führt das Strukturelement **1** im Synthesziel zum Kapitel „Aldolreaktion“ in der Reaktionsbibliothek und zu den Vorstufen **2** und **3**.



Aus den Vorstufen des Synthesziels werden wiederum deren Vorstufen erhalten, und das so lange, bis verfügbare Ausgangsmaterialien erreicht werden. So entsteht ein Synthesebaum mit im allgemeinen vielen Synthesewegen, unter denen einer ausgewählt werden muß. Für diese Vorgehensweise führte Corey später den Begriff retrosynthetische Analyse ein.<sup>[28]</sup>

Als Alternative dazu gibt es die bilaterale Syntheseplanung, bei der man sowohl das Synthesziel als auch alle Ausgangsstoffe kennt. Man muß lediglich Sequenzen von Reaktionen finden, die von den Ausgangsstoffen – die im allgemeinen zunächst separat ermittelt werden müssen – zum Synthesziel führen.

Computer lassen sich bei der Lösung chemischer Probleme nur dann erfolgreich verwenden, wenn bereits so viel Wissen und Erfahrung vorliegen, daß es Modellvorstellungen gibt und Analogieschlüsse möglich sind. Bei der Exploration von chemischem Neuland, den ersten Anfängen völlig neuer Arbeitsgebiete und bei der Gewinnung neuartiger Einsichten sind Visionen, Ideen und nicht zuletzt auch glückliche Zufälle nach wie vor entscheidend. Computer sind hierbei allenfalls von indirektem Nutzen, denn ein Chemiker könnte beispielsweise einen vom Computer generierten, durchaus realisierbaren, jedoch „phantasievollen“ und völlig neuartigen Vorschlag als unsinniges Artefakt verwerfen, statt zu versuchen, ihn experimentell zu realisieren.

Da der Computer keine eigene kreative Intelligenz besitzt, kann er keine revolutionären Ideen produzieren und wissenschaftliche Pionierleistungen hervorbringen, aber er kann als „Intelligenzverstärker“ die Evolution einer Wissenschaft im Rahmen des gegebenen Wissensstandes erheblich beschleunigen und auch Forschungsarbeiten ermöglichen, die ohne Computer nicht denkbar wären. Selbst wenn ein Computer in der Lage wäre, für ein gegebenes chemisches Problem alle denkbaren Lösungen zu generieren, könnten die sinnvollen, kreativen Lösungen wegen des Auswahlproblems nicht ohne die Beteiligung eines qualifizierten Benutzers gefunden werden.

## 2. Syntheseplanungsprogramme

Syntheseplanungsprogramme lassen sich einteilen in empirische, semiformale und formale. Empirische Programme zur Syntheseplanung basieren auf Reaktionsbibliotheken, semiformale Systeme haben meist formale Reaktionsgeneratoren, die mit heuristischen Selektionsverfahren arbeiten. Rein formale Syntheseplanungsprogramme verzichten weitgehend auf Heuristik und übertragen wichtige Selektionsentscheidungen dem Anwender. Hierzu sind sie mit einem reichen Repertoire an hierarchischen Klassifikationsmöglichkeiten und Auswahlmethoden versehen. Über neueste Entwicklungen auf diesem Gebiet wird in *Recl. Trav. Chim. Pays-Bas* vom Juni 1992 berichtet.

### 2.1. Empirische Syntheseplanung mit Reaktionsbibliotheken

Die auf Reaktionsbibliotheken beruhenden und retrosynthetisch operierenden Syntheseplanungsprogramme LHASA<sup>[29, 30]</sup>, Wipkes SECS<sup>[31]</sup> sowie Gelernters SYNCHM<sup>[32]</sup> und das bilateral arbeitende Leverkusener Computerprogramm zur Planung von Peptidsynthesen<sup>[33, 34]</sup> gehören zu den ersten Computerprogrammen für die Lösung chemischer Probleme. Inzwischen gibt es viele Arbeitsgruppen, die Computerprogramme zur Syntheseplanung entwickeln oder bereits implementiert haben<sup>[18, 35]</sup>.

Die Retrosyntheseprogramme ahmen das seit der Frühzeit der organischen Synthese übliche retrosynthetische Denken der Chemiker<sup>[36]</sup> nach, die das Synthesziel unter Berücksichtigung der verfügbaren Synthesereaktionen im Gegensatz der Syntheserichtung gedanklich zerlegen. Retrosynthetisches Denken wurde von Sir Robert Robinson besonders konsequent praktiziert<sup>[37a]</sup> und in einer Monographie<sup>[37b]</sup> eingehend dargelegt. Auch die von Warren<sup>[38]</sup> empfohlene disconnection method entspricht der retrosynthetischen Denkweise.

Die frühen Synthetiker hatten ihre Vorgehensweise allerdings noch nicht mit einem Namen belegt. Seit der Einführung der Begriffe Synthon, Transform und Retrosynthese, auch Antithese genannt, sowie der entsprechenden Transform-Pfeile ( $\Rightarrow$ ) durch Corey<sup>[28]</sup> wird in Publikationen über Synthesen häufig nachträglich die retrosynthetische Analyse angegeben. Den vielstufigen Synthesen lag natürlich ein Plan zugrunde, aber die experimentell realisierten vielstufigen, komplizierten Synthesewege ließen sich nicht von vornherein planen, sondern sie wurden aus der Menge der erdachten Synthesen anhand des experimentell Gelingenden im Trial-and-error-Verfahren durch laufende Anpassung der Planung an das Realisierbare ausgewählt.

Anhand der chemischen Konstitution des Synthesziels erzeugen die empirischen Retrosyntheseprogramme Synthesewege gemäß eingegebener Transforms, Regeln und Schemata. An einem gegebenen Synthesziel werden anhand gespeicherter Reaktionsinformationen und unter Anwendung heuristischer Regeln Bindungen gebrochen und geknüpft und so Vorstufen des Synthesziels generiert (vgl. **1–3**), die wiederum in Vorstufen bis hin zu verfügbaren Ausgangsstoffen zerlegt werden. Die empirischen Retrosyntheseprogramme sind typische Expertensysteme<sup>[39]</sup> mit einer Wissensbasis und einem Satz von Regeln.

Für ein Retrosyntheseprogramm, das nicht auf enge Anwendungsgebiete und ein kleines Repertoire beschränkt ist, benötigt man eine umfangreiche Reaktionsdatei und einen leistungsfähigen Rechner. Die Erstellung und Pflege der fehleranfälligen großen Reaktionsdateien sowie die Implementierung der Synthetiker-Erfahrungen in die Strategie- und Selektionsmoduln empirischer Syntheseprogramme erfordern großen Arbeitsaufwand. Besondere Schwierigkeiten treten bei der Behandlung von Reaktionen auf, die aus mehreren Einzelschritten bestehen, denn diese führen oft zur Speicherung redundanter Information.

Es ist ein Vorzug der empirischen Retrosyntheseprogramme, daß sie die vorhandene chemische Erfahrung nutzen und daher mit hoher Wahrscheinlichkeit Synthesevorschläge liefern, die im Laboratorium realisierbar sind, denn sie entsprechen Kombinationen der gespeicherten bekannten Reaktionen. Dies definiert aber auch zugleich die Grenzen solcher Programme; Neuland ist mit ihnen nicht erreichbar. Da es zudem kaum möglich ist, alle bekannten Reaktionen abzuspeichern, beruhen die Reaktionsdateien von Retrosyntheseprogrammen immer auf einer mehr oder weniger willkürlichen Auswahl von Synthesereaktionen.

Da es auf jedem Vorstufenniveau viele denkbare Vorstufen gibt und ein rasch expandierender Baum von Synthesewegen entsteht, muß man – wie beim Schachspiel – unter allen Möglichkeiten eine besonders vielversprechende aussuchen, wobei man bereits auch an deren weitere Vorstufen denken muß. In den genannten Retrosyntheseprogrammen versucht man dies mit heuristischen Strategien und Auswahlregeln zu erreichen.

Die empirisch-heuristischen Retrosyntheseprogramme haben bereits vor ca. zehn Jahren ein sehr hohes Niveau erreicht, insbesondere in Gestalt des Projekts CASP<sup>[40]</sup>, das große deutsche und Schweizer Chemiefirmen ausgehend von SECS<sup>[31]</sup>, dem seinerzeit leistungsfähigsten Programm zur empirischen Syntheseplanung, mit erheblichem Aufwand gemeinsam entwickelten. Es ist kaum vorstellbar, daß in der auf Reaktionsbibliotheken basierenden retrosynthetischen Syntheseplanung, von Erweiterungen und Detailverbesserungen abgesehen, noch nennenswerte Weiterentwicklungen möglich sind.

CASP konnte sich bei den potentiellen Benutzern bisher nicht durchsetzen. Der Hauptgrund hierfür dürfte sein, daß CASP mit seiner umfangreichen Reaktionsdatei einen verhältnismäßig großen Rechner erfordert und nicht vom Chemiker direkt eingesetzt werden kann, sondern einen dazwischengeschalteten Spezialisten benötigt. Es zeichnet sich darüber hinaus ab, daß die Chemiker bei der Planung von Synthesen die direkte Nutzung der in Reaktionsdatenbanken gespeicherten Informationen ihrem mittelbaren Gebrauch unter Manipulation durch ein empirisches Syntheseprogramm vorziehen.

### 2.1.1. Was sind strategische Bindungen in Ringsystemen?

Viele interessante Syntheseeziele enthalten komplizierte polycyclische Einheiten, vor deren Synthese zunächst die geeignete Strategie festgelegt werden muß<sup>[41]</sup>. Die von Corey et al.<sup>[42, 43]</sup> ausgearbeiteten heuristischen Synthesestrategien für carbocyclische Ringsysteme sind ein besonders wichtiger Beitrag zur computerunterstützten retrosynthetischen Ana-

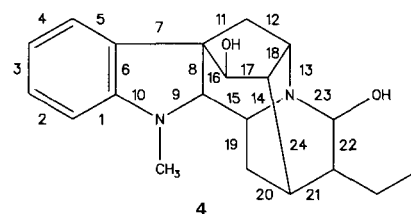
lyse. Ihre Regeln beziehen sich auf „strategische“ Bindungen, die vorzugsweise gebrochen oder geknüpft werden, und lassen sich wie folgt zusammenfassen (vgl. Abschnitt 4.3.2):

1. Strategische Bindungen befinden sich in einem vier- bis siebengliedrigen primären Ring, wobei ein Ring primär ist, wenn er kein Üerring von zwei oder mehr kleineren Ringen ist.
2. Eine strategische Bindung muß mit einem anderen Ring (exo-ständig zu einem anderen Ring) direkt verbunden sein, es sei denn, es handelt sich um eine Bindung in einem Dreiring.
3. Eine strategische Bindung sollte sich in dem Ring befinden, der die meisten überbrückten Stellen (Brückenkopf-atome) enthält.
4. Bindungen, bei deren Bruch ein Ring mit mehr als sieben Gliedern entsteht, sind nicht strategisch.
5. Bindungen, die an aromatischen Systemen teilnehmen, werden nicht als strategisch angesehen.
6. Enthält eine Brücke ein chirales Zentrum, so ist keine Bindung dieser Brücke strategisch, es sei denn, sie ist direkt mit dem chiralen Zentrum verbunden.

Kohlenstoff-Heteroatom-Bindungen werden gesondert behandelt. Für solche Bindungen werden nur die Regeln 4–6 sowie Regel 2, soweit sie Dreiringe betrifft, beachtet.

### 2.1.2. Schwierigkeiten bei der Identifizierung strategischer Bindungen

Es hat sich herausgestellt, daß die Anwendung der Regeln aus vorigem Abschnitt bei komplexen verbrückten Ringsystemen mit vier und mehr Ringen (gemäß Frèrejacques-Zählung<sup>[44]</sup>) gelegentlich zu unbefriedigenden Lösungen führt, vor allem wenn Heteroatome vorkommen, weil diese Regeln nicht hierarchisch gegliedert sind und ohne hinreichende Berücksichtigung von Heteroatomen konzipiert wurden. Dadurch werden in manchen Fällen mehrere Bindungen, die mit gleicher Priorität gebrochen werden können, als strategisch erkannt. Ein Beispiel ist Ajmalin **4**<sup>[45]</sup>. Im Ajmalinmo-



lekül werden mit diesen Regeln sechs strategische Bindungen (Bindungen 9, 10, 15, 16, 23 und 24) gefunden. Der Bruch der Bindungen 9 und 10 trägt jedoch nicht nennenswert zur Vereinfachung der Struktur bei. Ein einzelner Bindungsbruch erzeugt bei **4** noch keine synthetisch zugängliche Vorstufe, d. h. es müssen weitere strategische Bindungen ermittelt werden. Dies führt aber nicht zwangsläufig zu einer Vereinfachung des Problems, da die Wahrscheinlichkeit zunimmt, daß immer mehr Bindungen die Regeln für strategische Bindungen erfüllen, wenn die Komplexität des Ringsystems abnimmt. Dadurch steigt die Zahl der zu bearbeitenden Vorstufen erheblich an. So entstehen durch den Bruch der Bindung 16 in **4** sieben neue strategische Bindungen.

Den Coreyschen Regeln fehlt auch ein Abbruchkriterium für die Beendigung der Ringfragmentierung. Wie bereits erwähnt, muß ein einzelner Bindungsbruch nicht zu einem „vereinfachten“ Molekül führen. Ein Abbruch des Verfahrens wäre nach der Zerlegung aller Verbrückungen sinnvoll; dann müßten Bindungen, die zu keinen Brücken gehören, nicht berücksichtigt werden, d. h. die Bindungen 9 und 10 von **4** fielen ganz aus der Bewertung.

Es gibt keine heuristischen Regeln, die in allen Bereichen der Chemie gleichermaßen gelten, und ihr Gültigkeitsbereich läßt sich meist auch nicht genau definieren. Die automatisierte Anwendung heuristischer Regeln ist daher willkürlich und kann zu Fehlbewertungen führen<sup>[21]</sup>.

Bei moderaten Datenmengen ist eine interaktive Selektion durch den Benutzer möglich. Sie hat den Vorteil der unbegrenzten Anwendungsbreite, und eine von einem kundigen Chemiker durchgeführte interaktive Auswahl von Synthesewegen dürfte bei der computerunterstützten Syntheseplanung, von wenigen Ausnahmefällen abgesehen, bessere Ergebnisse liefern als jegliche automatisierte Heuristik, denn bei der retrosynthetischen, computerunterstützten Syntheseplanung werden beispielsweise Synthesewege bevorzugt, deren letzte Stufen besonders attraktiv sind, und Synthesen mit ungünstigen letzten Stufen werden unterbewertet<sup>[46]</sup>.

## 2.2. Semiformale Syntheseplanung und Reaktionsvorhersage

Die „Kreativität“ der heutigen semiformalen Syntheseprogramme wird vor allem dadurch eingeschränkt, daß die Selektion der Problemlösungen im allgemeinen automatisch und heuristisch erfolgt und somit weder transparent noch willkürfrei ist<sup>[21]</sup>. Deshalb kann die Gesamtheit aller denkbaren Lösungen eines Problems kaum sinnvoll berücksichtigt werden.

Die semiformalen Syntheseprogramme können zwar grundsätzlich beliebige retrosynthetische Prozesse durchführen, verfügen aber nicht über die für die Bearbeitung sehr großer Lösungsmengen notwendigen Auswahlverfahren auf einer formalen Grundlage; ihre Anwendung ist daher aus kombinatorischen Gründen auf relativ „kleine“ Syntheseprobleme beschränkt, die dem Gültigkeitsbereich der eingegebenen Regeln und der zugrundeliegenden schematischen Vorgehensweise entsprechen; ohne einen formalen Rahmen ist darüber hinaus auch kein interaktives Ordnen und Klassifizieren der denkbaren Lösungen möglich. Bezüglich der Generierung der Synthesewege dürfte bei der semiformalen Syntheseplanung kein nennenswerter Fortschritt mehr möglich sein, wohl aber hinsichtlich der Auswahl der Retroreaktionen und der Synthesestufen.

Die Programme zur semiformalen und vor allem zur formalen Syntheseplanung haben im Gegensatz zu denen der rein empirischen beachtliche „erfinderische“ Fähigkeiten. Die semiformale Syntheseplanung kann allerdings nicht gezielt (wie z. B. IGOR, siehe Abschnitt 4.4.3) zur „Entdeckung“ neuer Reaktionen eingesetzt werden. Sie erfindet neue Molekülarten und chemische Reaktionen nur, wenn diese im Rahmen eines zu generierenden Syntheseweges zufällig benötigt werden.

Unseres Wissens sind bisher noch keine Verbindungen oder Reaktionen publiziert worden, die mit einem Pro-

gramm zur empirischen oder semiempirischen Syntheseplanung vorausgesagt und anschließend experimentell realisiert worden sind.

Das auch bei der Syntheseplanung einsetzbare Programm zur semiempirischen Vorhersage von Reaktionsprodukten CAMEO von Jorgensen et al.<sup>[47]</sup> basiert auf der Annahme, daß sich die meisten organischen Reaktionen als Kombinationen weniger reaktionsmechanistischer Elementarschritte darstellen lassen (vgl. Lit.<sup>[48]</sup>). Das erforderliche chemische Wissen ist in Form individueller Programm-Moduln implementiert. Da CAMEO mechanistische Schritte chemischer Reaktionen zur Generierung von Reaktionen nutzt, ist es innovationsfähiger als Retrosyntheseprogramme auf der Basis von Reaktionsbibliotheken<sup>[49]</sup>.

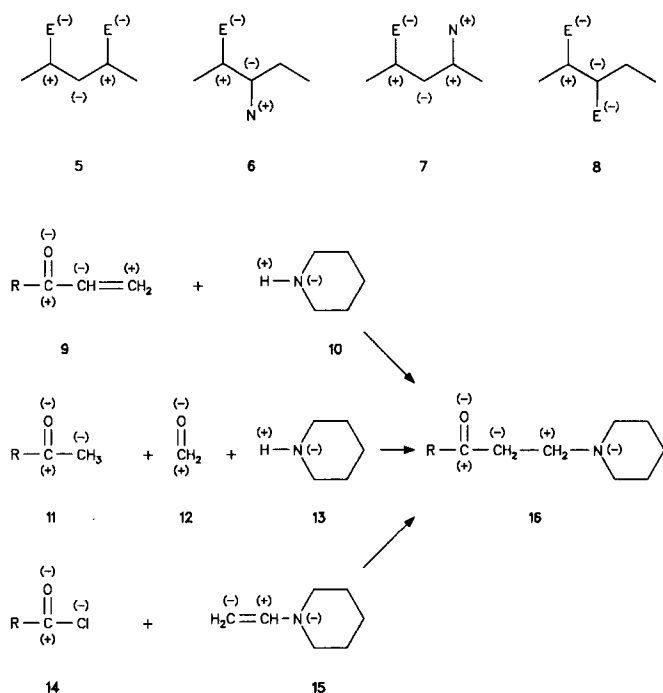
Hendrikson<sup>[50]</sup> führte die sogenannte Halbreaktionen ein, aus denen chemische Reaktionen unter Beachtung valenzchemischer Regeln zusammengesetzt werden, ohne Reaktionsmechanismen zu implizieren. Auf dieser Grundlage entwickelte Moreau<sup>[51]</sup> das Syntheseprogramm MASSO. Hendrikson et al.<sup>[52]</sup> implementierten etwas später SYNGEN, ein ähnliches Programm, das konvergente Synthesewege aus maximal vier Untereinheiten des Syntheseziels erzeugt.

Die ersten Versionen des aus der Programmstudie CICLOPS<sup>[53]</sup> entstandenen Programms EROS<sup>[54]</sup> führten noch eine formale Syntheseplanung aus, während die späteren Versionen von EROS<sup>[55]</sup> semiformalen Charakter haben. Die verwendeten Verfahren zur Selektion von Synthesewegen, die auf Annahmen über die Brechbarkeit von Bindungen und auf physikalischen Daten beruhen, haben heuristischen Charakter, und die kombinatorische Vielfalt der Reaktionsmöglichkeiten wird dadurch eingeschränkt, daß bestimmte Bindungen als „brechbar“ ausgewählt werden. Im Verlauf mehrstufiger Synthesen werden im allgemeinen mehr als zehn Bindungen gebrochen oder geknüpft. Nimmt man an, daß pro Reaktionsschritt ein bis drei Bindungen gebrochen oder geknüpft werden, so gibt es mehr als  $7 \times 10^6$  Synthesewege, die sich bezüglich der Folge dieser Operationen unterscheiden und die keineswegs immer äquivalent sind; äquivalent sind sie nur, wenn alle Operationen völlig unabhängig voneinander und somit vertauschbar sind.

Dem Programm TOSCA zur semiformalen Syntheseplanung<sup>[56]</sup> liegt als besonders vielversprechender Ansatz Evans' Konzept der konsonanten und dissonanten Moleküle zugrunde. D. A. Evans hat in einem Vortrag am 6. Mai 1971 an der University of California at Los Angeles die Wichtigkeit konsonanter und dissonanter Strukturelemente in Molekülen aufgezeigt. Er legte seine Vorstellungen in einem unveröffentlichten Manuskript<sup>[57]</sup> mit dem Titel „Consonant and Dissonant Relationships. – An Organizational Model for Organic Synthesis“ dar, das in der Fachwelt einen hohen Bekanntheitsgrad erlangte, da Evans 1972 eine große Zahl von Vorabdrucken an interessierte Kollegen versandte.

Evans' Konzept ist eine Weiterentwicklung der Lapworthschen Polaritätsmuster heteroatomhaltiger organischer Moleküle<sup>[58]</sup>, die von Evans „Ladungsaffinitätsmuster“ genannt werden. Als konsonant werden Moleküle wie **5** und **6** bezeichnet. Sie enthalten alternierend Zentren positiver und negativer Ladungsaffinität, die potentieller nucleophiler bzw. elektrophiler Reaktivität entsprechen, während Moleküle wie **7** und **8**, für die es nicht möglich ist, ein solches Muster anzugeben, dissonant genannt werden. Die Synthe-

sen von 16 aus 9 + 10, 11 + 12 + 13 und 14 + 15 sind Beispiele für konsonante Synthesewege.



Evans zeigte, daß bei Synthesen über ionische Reaktionen für eine Umwandlung konsonanter Moleküle in dissonante und umgekehrt „Inversionsoperationen“ notwendig sind. Die von Evans geforderte Inversionsoperation entspricht der von Seebach<sup>[59]</sup> definierten Umpolung. Die Ermittlung konsonanter und dissonanter Strukturelemente in den Ausgangsstoffen und Produkten vielstufiger Synthesen ist für die Syntheseplanung sehr wichtig, damit die gegebenenfalls erforderlichen Inversionsoperationen angemessen berücksichtigt werden können. Wir berücksichtigen das Evans-Lapworth-Schema bei der bilateralen Syntheseplanung mit dem Computerprogramm RAIN (Reaction and Intermediates Network)<sup>[60–63]</sup>. Evans hat mit seinem Konsonanz/Dissonanz-Konzept einen Beitrag von großer Tragweite zur computerunterstützten Syntheseplanung geliefert.

## 2.3. Syntheseplanung auf formaler Basis

### 2.3.1. Was ist formal?

Die geschilderten Nachteile der Programme zur empirischen und semiformalen Syntheseplanung lassen sich nur mit Hilfe einer umfassenden mathematischen Theorie der Chemie vermeiden, durch die alle für die Lösung eines Problems benötigten molekularen Systeme und Reaktionen mit formalen Mitteln generiert und ausgewählt werden können<sup>[21]</sup>.

Der Begriff formal wird je nach Disziplin unterschiedlich verwendet. D. Hilbert und P. Bernays<sup>[64]</sup> nannten die Mathematik die Lehre von den formalen Systemen. In der Informatik arbeitet man mit formalen Beschreibungen. Eine solche formale Beschreibung setzt eine Sprache (Syntax) voraus, welche auch die Mathematik sein kann, sowie eine

mathematisch beschriebene Bedeutung (Semantik) und die Festlegung von Umformungs- und Beweisregeln. Für die Gültigkeit und den Geltungsbereich eines formalen Algorithmus besteht, genau wie für ein Theorem, Beweispflicht. In der Chemie ist das Wort formal nicht genau definiert, und es wird bisweilen recht wahllos für alle Arten von Verallgemeinerungen, Klassifizierungen, mehr oder weniger abstrakte Darstellungen und Algorithmen verwendet. Die Chemie als empirische Wissenschaft läßt sich zwar nicht streng formal im Sinne des Sprachgebrauches der Mathematiker und Informatiker behandeln, aber wir versuchen die Computerunterstützung in der Chemie so weit wie möglich auf eine formale Basis zu stellen und, wo dies nicht möglich ist, heuristische Regeln mit schwer zu überschauenden Auswirkungen zu vermeiden sowie die jeweils notwendigen Entscheidungen interaktiv vom Benutzer fällen zu lassen, wenn sie nicht aufgrund formaler Kriterien getroffen werden können. Die formale Beschreibung und Behandlung der Chemie erfordert eine adäquate, detailgetreue Übersetzung der Chemie, einschließlich ihres dynamischen Aspekts, in die Sprache der Mathematik und eine transparente Rückübersetzung der mathematischen Darstellung chemischer Objekte und Sachverhalte in die Sprache der Chemie. Dies läßt sich durch ein mathematisches Modell der logischen Struktur der Chemie erreichen, das sowohl molekulare Systeme als auch deren chemische Veränderungen in der erforderlichen Weise repräsentiert (siehe Abschnitt 3).

Ein solches mathematisches Modell ermöglicht die Implementierung von Computerprogrammen, die nicht auf einen einzigen Problemtyp beschränkt sind und ohne gespeichertes Detailwissen mit rein formalen Mitteln Problemlösungen generieren und sichten können<sup>[21]</sup>. Solche Computerprogramme eignen sich auch für die Erschließung chemischen Neulandes, da sie auch präzedenzlose Moleküle und chemische Reaktionen liefern können.

Mit geeigneten Computerprogrammen kann man grundsätzlich die Gesamtheit der kombinatorisch möglichen Lösungen chemischer Probleme – einschließlich der präzedenzlosen – berücksichtigen, denn formale Reaktionsgeneratoren (siehe Abschnitt 4.4.1) kennen die chemische Literatur nicht. Dies nützt aber wenig, wenn es nicht gelingt, die erzeugten Informationen in sinnvoller Weise zu ordnen und eine zweckentsprechende Auswahl durchzuführen. Derlei ist nur zum Teil mit formalen Mitteln möglich und erfordert deshalb die Mitwirkung eines qualifizierten Benutzers, dessen Erfahrung, Intuition und Sachkunde nicht nur bei der Formulierung der Probleme, sondern vor allem bei der Selektion der brauchbaren Lösungen zum Tragen kommen.

Die weitverbreitete Meinung, aus einem Computer könne nichts „herauskommen“, was man nicht hineingesteckt habe, klingt einleuchtend. Sie ist jedoch strenggenommen nur bei Datenbanksystemen und den im wesentlichen darauf beruhenden einfachen Expertensystemen richtig.

### 2.3.2. Die Mathematisierung der Chemie

Die Quantenchemie hat als ein Zweig der Theoretischen Physik der Moleküle auf dem Umweg über die Physik erheblich zur Mathematisierung der Chemie beigetragen und ist als Verständnisgrundlage sowie für die computerunterstützte numerische Berechnung meßbarer Eigenschaften wohldefi-

nierter Systeme unerlässlich. Die mathematische Chemie dagegen hat die Mathematisierung der Chemie ohne den Umweg über die Physik zum Ziel sowie die Lösung chemischer Probleme durch qualitative mathematische Methoden. Die traditionellen Ansätze zur Behandlung chemischer Probleme mit qualitativen Formalismen der diskreten Mathematik beschränken sich auf eine statische Betrachtungsweise. Eine solche mathematische Chemie geht über die graphentheoretische Veranschaulichung und Interpretation der chemischen Konstitution von Molekülen und Familien von Isomeren<sup>[65–68]</sup> sowie die gruppentheoretische Klassifizierung und Abzählung von Isomeren und Produkten von Isomerisierungen nicht hinaus<sup>[69–77]</sup>.

Weder die vielfach auch als Theoretische Chemie bezeichnete Quantenchemie noch die traditionelle mathematische Chemie eignen sich als theoretische Basis der computerunterstützten Syntheseplanung oder der Lösung anderer chemischer Probleme, die im wesentlichen kombinatorischer Art sind, wie die Suche nach Molekülen oder Reaktionen, welche bestimmten chemischen Forderungen genügen. Die computerunterstützte Lösung chemischer Probleme mit formalen Mitteln erfordert vielmehr eine globale mathematische Modellierung der Chemie, welche über die Behandlung chemischer Einzelobjekte hinausgeht und vor allem die Beziehungen zwischen den Objekten zum Gegenstand hat und die logische Struktur der Chemie wiedergibt.

Die Chemie hat einen konstitutiven und einen stereochemischen Aspekt, die zwar zusammenhängen, aber doch so verschiedenartig sind, daß sie eine unterschiedliche mathematische Betrachtungsweise erfordern. Ein mathematisches Modell der Chemie muß daher aus zwei Teilen bestehen, einem, der die konstitutive Chemie wiedergibt, und einem, der die Stereochemie repräsentiert.

Die Theorie der be- und r-Matrizen<sup>[78]</sup> ist ein algebraisches Modell der logischen Struktur der konstitutiven Chemie (siehe Abschnitt 3.2), das in der Literatur<sup>[79–81]</sup> auch „Dugundji-Ugi-Modell oder -Theorie“ oder kurz „DU-Modell“ genannt wird. Neben dem algebraischen DU-Modell wird auch das von Kvasnička et al.<sup>[80, 81]</sup> später vorgestellte, äquivalente graphentheoretische Modell der Chemie als mathematische Grundlage von Programmen zur formalen Syntheseplanung verwendet. Die logische Struktur der Stereochemie läßt sich mit der Theorie der Chemischen Identitätsgruppe (CIG)<sup>[82]</sup> beschreiben, die auf dem Konzept der Permutationsisomerie basiert<sup>[76]</sup>.

Im Gegensatz zur traditionellen mathematischen Chemie beziehen sich diese beiden Modelle auch und vor allem auf den dynamischen Aspekt der Chemie. Im Rahmen des DU-Modells werden chemische Reaktionen durch Transformationen von be-Matrizen repräsentiert, welche den Graphen der chemischen Konstitution entsprechen, und in der CIG-Theorie werden stereochemische Prozesse durch mengenwertige Abbildungen dargestellt. Diese mathematischen Methoden zur direkten Modellierung chemischer Vorgänge sind die theoretische Grundlage der computerunterstützten deduktiven Lösung chemischer Probleme, für die nicht auf detaillierte empirische chemische Daten zurückgegriffen werden muß.

Das Wesentliche am DU-Modell und an der CIG-Theorie ist nicht die darin enthaltene Mathematik, sondern die direkte und detaillierte Übersetzung von Chemie in Mathematik und umgekehrt. Die Mathematik ist damit zu einer natürlichen Sprache der Chemie geworden, einschließlich einer ei-

genen Semantik und Syntax, ebenso wie es die Sprache der chemischen Formeln schon lange ist.

## 2.4. Programme zur formalen Syntheseplanung

Das DU-Modell ermöglicht es, chemische Objekte wie Moleküle, Ensembles von Molekülen (EM) und auch chemische Reaktionen durch Matrizen, d. h. durch mathematische Objekte, zu repräsentieren und somit chemische Probleme in mathematische Probleme zu übersetzen, deren Lösungen den Lösungen der chemischen Probleme entsprechen. Dieses Modell leitete die Entwicklung der formalen Algorithmen, Reaktionsgeneratoren und Computerprogramme zur deduktiven Lösung chemischer Probleme mit mathematischen Mitteln ein. Damit können nun nicht nur Lösungen für viele Arten chemischer Probleme gefunden, sondern auch deren Sichtung und Auswahl erleichtert werden<sup>[21]</sup>.

Die Entwicklung von Programmen für die formale Syntheseplanung begann mit der Pilotstudie CICLOPS<sup>[53]</sup>. Durch CICLOPS sollte vor allem nachgewiesen werden, daß die computerunterstützte Lösung chemischer Probleme auf der Grundlage des DU-Modells prinzipiell möglich ist. In diesem Sinne war die Studie erfolgreich, sie zeigte jedoch auch, daß CICLOPS in der Syntheseplanungspraxis an kombinatorischen Problemen scheitern würde. Ausgehend von CICLOPS entwickelten Gasteiger und Jochum<sup>[54]</sup> die ersten Versionen des Syntheseplanungsprogramms EROS, indem sie den partiell implementierten Reaktionsgenerator (siehe unten) von CICLOPS auf die drei (von ca. 100 denkbaren) wichtigsten r-Klassen (siehe Abschnitt 3.2.2) reduzierten und die von Stevens und Brownscombe<sup>[83]</sup> eingeführten Abschätzungen von Reaktionsenthalpien zur Auswahl von Reaktionsschritten heranzogen.

Die Tatsache, daß Reaktionsmatrizen aus Basiselementen zusammengesetzt sind, welche den Elementarschritten chemischer Reaktionen entsprechen<sup>[78]</sup>, ermöglicht Reaktionsgeneratoren, die chemische Reaktionen aus ihren Elementarschritten aufbauen. Das Syntheseplanungsprogramm ASSOR<sup>[48]</sup> enthält einen solchen Reaktionsgenerator und berücksichtigt daher den mechanistischen Aspekt chemischer Reaktionen.

Das DU-Modell eignet sich auf vielfältige Weise als theoretische Grundlage von Programmen zur monolateralen Syntheseplanung, die vom Zielmolekül retrosynthetisch oder von den Ausgangsstoffen vorwärtssynthetisch Synthesewege generieren, doch die Möglichkeiten, die es der bilateralen Syntheseplanung<sup>[84, 85]</sup> eröffnet, erscheinen noch interessanter. Die kombinatorischen Vorzüge der bilateralen Syntheseplanung haben in letzter Zeit auch in anderen Gruppen die Entwicklung entsprechender Programme ausgelöst. Johnson et al.<sup>[86]</sup> entwickeln auf der Grundlage von LHASA ein Programm zur bilateralen Syntheseplanung, und kürzlich haben Hendrickson und Parks<sup>[87]</sup> SYNGEN mit dem entsprechenden Programm FORWARD zur Vorwärtssyntheseplanung kombiniert.

Bei der bilateralen Syntheseplanung werden die Synthesewege von ihren beiden Enden – den Ausgangsstoffen und dem Synthesziel – aus ermittelt. Gegenwärtig wird von uns ein System von Computerprogrammen zur bilateralen Syntheseplanung entwickelt, das dreistufig arbeitet<sup>[85, 88, 89]</sup>. Ein solches System hat schon wegen der Partitionierung des Pro-

blems gegenüber monolateralen Ansätzen erhebliche kombinatorische Vorteile. Als erstes wird mit dem Substrukturkorrelationsprogramm CORREL-S (siehe Abschnitt 4.3.2) aus einer Liste verfügbarer Stoffe für das Synthesziel ein Sortiment geeigneter Ausgangsstoffe ausgewählt. Anschließend wird das Synthesziel mit dem Programm STOECH (siehe Abschnitt 4.3.3) durch seine Coprodukte stöchiometrisch ergänzt, so daß ein Ziel-EM resultiert, das zum EM der Ausgangsstoffe isomer ist. Aus diesen beiden EM generiert das Programm RAIN (siehe Abschnitt 4.4) ein Netzwerk von Synthesewegen, die die beiden Enden verbinden.

Als weitere Programme zur formalen Syntheseplanung sind FLAMINGOES<sup>[90]</sup>, PEGAS<sup>[91]</sup> und MAPOS<sup>[92]</sup> zu nennen. Hippe<sup>[93]</sup> versucht in seinen Syntheseprogrammen SCANSYNTH, SCANMAT und SCANPHARM die Vorteile des empirischen und des formalen Ansatzes zu kombinieren, desgleichen Johnson et al.<sup>[94]</sup> im System SYNLMA.

### 3. Die Chemie und ihre logische Struktur

#### 3.1. Die Hierarchie der Isomerietypen

Die Chemie ist eine empirische Wissenschaft, doch ihre Objekte lassen sich in eine konsistente logische Struktur einordnen, weil Moleküle nach einheitlichen Prinzipien aufgebaut sind<sup>[95]</sup>. Diese logische Struktur ist ein System von Äquivalenzrelationen, deren wichtigste die Isomerietypen und die wechselseitige Umwandelbarkeit isomerer Moleküle und Ensembles von Molekülen sind, und sie ist durch die valenzchemischen Eigenschaften der chemischen Elemente und einige allgemeine Naturgesetze festgelegt. Vergleicht man die Chemie mit einer Sprache, so entsprechen die chemischen Fakten, Objekte, Phänomene und Ereignisse dem Vokabular, und die logische Struktur repräsentiert die strukturgebende und ordnende Grammatik.

Das Rückgrat dieser logischen Struktur ist die hierarchische Klassifizierung der Moleküle nach Isomerietypen<sup>[96]</sup>. Isomere enthalten die gleiche Zahl und Art von Atomen und haben folglich die gleiche Bruttoformel. Der Begriff „isomer“ bezieht sich sowohl auf die Moleküle als auch auf die daraus bestehenden Stoffe. A. von Humboldt folgerte aus der Existenz von Isomeren, daß Moleküle eine innere Struktur haben, denn sonst könnten sie bei gleicher Elementarzusammensetzung nicht unterscheidbar sein<sup>[97]</sup>.

Diejenigen molekularen Systeme, die alle Atome eines gegebenen Sortiments von Atomen – genauer: das gleiche Sortiment von Atomrümpfen (Atomkerne + Elektronen der inneren Schale) und die gleiche Zahl von Valenzelektronen – genau einmal enthalten, sind die isomeren EM, die jeweils aus einem oder mehreren Molekülen bestehen können. Diese Erweiterung des Isomeriebegriffes von Molekülen auf EM gehört zu den begrifflichen Grundlagen der mathematischen Darstellung der logischen Struktur der Chemie<sup>[78]</sup>.

Die chemische Konstitution von Molekülen entspricht den durch kovalente Bindungen festgelegten Nachbarschaftsbeziehungen der im EM enthaltenen Atome. Isomere, die sich bezüglich ihrer chemischen Konstitution unterscheiden, nennt man Konstitutionsisomere.

Laut ihrer klassischen Definition sind Stereoisomere konstitutionsgleiche Moleküle, die sich in der relativen räumli-

chen Anordnung ihrer Atome unterscheiden<sup>[82, 98]</sup>. Diese Definition ist nur bedingt brauchbar, da ein großer Teil der Stereoisomere flexible Moleküle sind, die nicht durch starre geometrische Modelle adäquat darstellbar sind. Eine allgemeingültige Definition der Stereoisomere, die auch bewegliche Moleküle einschließt, ist mit dem im nächsten Absatz definierten Begriff der chemischen Identität möglich. Die Äquivalenzrelation der chemischen Identität ermöglicht unter anderem die Definition eines Chiralitätsbegriffes, der auch für Moleküle gilt, die sich nicht durch ein starres geometrisches Modell repräsentieren lassen<sup>[73, 98]</sup>.

Moleküle sind chemisch identisch, wenn sie sich unter den gegebenen Beobachtungsbedingungen spontan ineinander umwandeln und zur selben einheitlichen chemischen Verbindung gehören. Ein Film über die zeitlichen Veränderungen der Geometrie eines Moleküls würde sich zwar vom entsprechenden Film über ein chemisch identisches Molekül in der Abfolge der Bilder unterscheiden, jedoch insgesamt die gleichen Momentaufnahmen enthalten, und eine Momentaufnahme eines hinreichend großen Ensembles von Molekülen der entsprechenden Verbindung würde ebenfalls die gleiche Kollektion von Molekülbildern enthalten wie die beiden Filme.

Sind Moleküle gleicher chemischer Konstitution nicht chemisch identisch, so sind sie Stereoisomere<sup>[73, 82, 98]</sup>. Die so definierte Stereoisomerie stellt ein eigenständiges Klassifizierungsprinzip dar, das der Konstitutionsisomerie untergeordnet ist.

#### 3.2. Ein algebraisches Modell der logischen Struktur der konstitutiven Chemie

Die logische Struktur der konstitutiven Chemie ergibt sich aus folgenden Aussagen, die J. Dugundji und I. Ugi<sup>[78]</sup> gleichsam als axiomatische Grundlage der Theorie der Bindungs- und Elektronen- sowie der Reaktionsmatrizen (be- bzw. r-Matrizen) verwendeten:

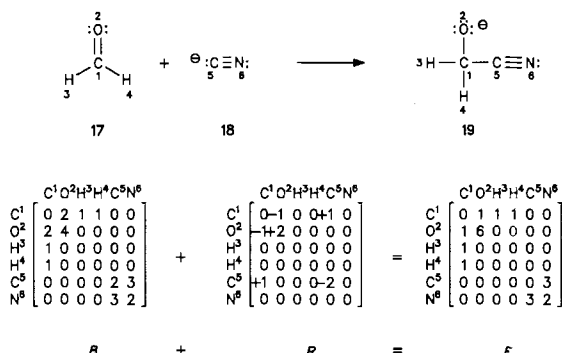
Moleküle sind aus Atomrümpfen und Valenzelektronen aufgebaut und werden durch kovalente Bindungen zusammengehalten. Eine kovalente Bindung entspricht einem Valenzelektronenpaar, das zwei benachbarten Atomen gemeinsam angehört. Eine chemische Reaktion ist die Umwandlung eines EM in ein isomeres EM durch Verschiebung von Valenzelektronen, wobei die Atomrümpfe und die Zahl der Valenzelektronen erhalten bleiben.

Die logische Struktur der konstitutiven Chemie läßt sich anhand der Chemie einer festgelegten Kollektion  $A = \{A_1, \dots, A_n\}$  von Atomen besonders leicht erkennen. Da die Atome in  $A$  beliebig wählbar sind, gilt ein Modell für die logische Struktur der Chemie von  $A$  für die gesamte Chemie. Die Chemie von  $A$  ist durch die Familie aller isomeren EM( $A$ ) gegeben, wobei ein EM( $A$ ) ein EM ist, das jedes Atom aus  $A$  genau einmal enthält<sup>[93]</sup>.

Im Rahmen des DU-Modells werden EM aus  $n$  Atomen durch ihre ganzzahligen symmetrischen  $n \times n$  be-Matrizen repräsentiert, deren  $i$ -te Zeilen/Spalten dem Atom  $A_i$  ( $1 \leq i \leq n$ ) zugeordnet sind. Die Eintragungen einer be-Matrix geben die Verteilung der Valenzelektronen im EM an: Die außerdiagonalen Eintragungen  $b_{ij}$  ( $=b_{ji}$ ) sind die formalen Bindungsordnungen der kovalenten Bindungen zwischen den Atomen  $A_i$  und  $A_j$ , und die Diagonaleintragungen  $b_{ii}$  geben an, wie viele freie Valenzelektronen an den



Atomen  $A_i$  vorhanden sind. Die Eintragung  $b_{12} = b_{21} = 2$  der be-Matrix des EM (17, 18) bedeutet beispielsweise, daß das C-Atom Nr. 1 mit dem O-Atom Nr. 2 durch eine kovalente Doppelbindung verknüpft ist, und  $b_{22} = 4$  bedeutet, daß am O-Atom Nr. 2 vier freie Valenzelektronen existieren (Schema 1).



Schema 1. Eine Reaktion und ihre Beschreibung mit Matrizen.

Die Zeilen/Spalten von be-Matrizen geben die Valenzelektronenzahl um das jeweilige Atom wieder („Valenzschemata“). Daraus folgt, daß sämtliche Zeilen/Spalten der be-Matrizen stabiler EM erlaubte Valenzschemata der zugeordneten chemischen Elemente darstellen. Dies sind die valenzchemischen Randbedingungen der be-Matrizen.

Die in der chemischen Dokumentation verwendeten Adjazenz- und Konnektivitätsmatrizen<sup>[99]</sup>, von denen sich die be-Matrizen in den Diagonaleintragungen unterscheiden, sind lediglich tabellarische Darstellungen der chemischen Konstitution einzelner Moleküle. Im Gegensatz dazu sind be-Matrizen echte mathematische Objekte mit wohldefinierten mathematischen Eigenschaften, und sie beziehen sich nicht nur auf Einzelmoleküle, sondern auch auf EM, die aus mehreren Molekülen bestehen können. Die be-Matrizen und die r-Matrizen (siehe unten) sind Elemente der freien abelschen Gruppen<sup>[78]</sup>.

Die Verschiebungen von Valenzelektronen, durch welche die Edukte chemischer Reaktionen  $EM_B$  in deren Produkte  $EM_E$  umgewandelt werden, beschreibt man üblicherweise durch Elektronenverschiebungspfeile und repräsentiert sie durch r-Matrizen  $R$ . Deren außerdiagonale Eintragungen  $r_{ij}$  ( $=r_{ji}$ ) geben die Veränderungen der formalen Ordnung der Bindungen zwischen  $A_i$  und  $A_j$  an, die Diagonaleintragungen  $r_{ii}$  die Veränderungen der Zahl freier Valenzelektronen. Die Summe der Eintragungen einer r-Matrix ist Null, weil die Gesamtzahl der Valenzelektronen während chemischer Reaktionen konstant bleibt. Die be-Matrizen  $B$  der Edukte werden durch additive Transformationen mit den Reaktionsmatrizen  $R$  in be-Matrizen  $E$  der Produkte gemäß  $B + R = E$ , der Grundgleichung des DU-Modells, überführt (Schema 1). Die Addition von Matrizen erfolgt Eintragung für Eintragung, d.h.  $B + R = E$  bedeutet, daß für alle Eintragungen  $b_{ij} + r_{ij} = e_{ij}$  gilt. Die Gleichung  $B + R = E$  repräsentiert nur dann eine denkbare chemische Reaktion, wenn  $B$  und  $E$  definitionsgemäße be-Matrizen sind und  $R$  eine definitionsgemäße r-Matrix ist. Da es keine negativen formalen Bindungsordnungen und keine negativen Valenzelektronenzahlen gibt, müssen die negativen Eintragungen von  $R$

so platziert sein, daß sie betragsmäßig mindestens ebenso großen positiven Eintragungen von  $B$  entsprechen und damit  $E$  von negativen Eintragungen frei ist.

### 3.2.1. Chemische Distanz

Faßt man die  $n^2$  Eintragungen von be-Matrizen als Koordinaten von Punkten in einem  $n^2$ -dimensionalen euklidischen Raum auf, dann läßt sich das algebraische Modell der logischen Struktur der konstitutiven Chemie durch ein geometrisches Modell veranschaulichen<sup>[78, 100]</sup>, in welchem die EM be-Punkten und die chemischen Reaktionen verbindenden r-Vektoren entsprechen. Die  $L_1$ -Länge  $d(B, E)$  eines r-Vektors ist die Summe der Absolutbeträge der Koordinatendifferenzen der be-Punkte  $P(B)$  und  $P(E)$  [Gl. (1)]. Man

$$d(B, E) = \sum_{ij} |b_{ij} - e_{ij}| = \sum_{ij} |r_{ij}| \quad (1)$$

nennt die  $L_1$ -Distanz („Taxifahrerdistanz“)<sup>[101]</sup> zwischen den be-Punkten  $P(B)$  und  $P(E)$  die chemische Distanz (CD) von  $EM_B$  und  $EM_E$ ; sie ist zweimal die Zahl der bei der wechselseitigen Umwandlung von  $EM_B$  und  $EM_E$  verschobenen Valenzelektronen<sup>[100]</sup> (siehe Tabelle 1).

Betrachtet man das Punktgitter einer Familie aller isomeren EM mit gepaarten Valenzelektronen von einem beliebigen EM aus, so befinden sich die be-Punkte auf konzentrischen  $L_1$ -Hyperkugelschalen, deren Radien sich um jeweils vier Einheiten unterscheiden. Der Radius der größten dieser Kugelschalen entspricht der Gesamtvalenzelektronenzahl der zugrundeliegenden Kollektion von Atomen. Dieses geometrische Bild der logischen Struktur der Chemie, eine „Landkarte“ der Minima der Energiehyperfläche einer gegebenen Kollektion von Atomen<sup>[102]</sup>, besticht durch seine Ästhetik.

Die Lösungen der Grundgleichung  $B + R = E$  des DU-Modells entsprechen den Lösungen einer Vielzahl chemischer Probleme. Somit kann das DU-Modell als universelle theoretische Grundlage von Computerprogrammen zur deduktiven Lösung konstitutionschemischer Probleme dienen. Darüber hinaus eignet es sich aber auch als Basis streng formaler, transparenter Verfahren zur Auswahl chemisch sinnvoller Problemlösungen aus einer meist sehr großen Zahl prinzipiell möglicher Lösungen<sup>[121]</sup>. Hierbei spielt das Prinzip der minimalen chemischen Distanz<sup>[78, 103–107]</sup> eine entscheidende Rolle. Dieses Prinzip, eine quantitative Version des qualitativen klassischen Prinzips der minimalen Strukturänderung<sup>[108]</sup>, besagt: Die wechselseitige Umwandlung isomerer EM durch chemische Reaktionen läuft vorzugsweise unter Verschiebung einer minimalen Zahl von Valenzelektronen ab.

Die minimale CD entspricht einer (oder mehreren) Atom-auf-Atom-Abbildung(en) der EM, da die CD zwischen isomeren EM von der Korrelation ihrer Atome abhängt. Dies gilt auch cum grano salis für Reaktionsfolgen, d.h. die CD von Reaktionsfolgen liegt selten um mehr als vier Einheiten über der minimalen CD.

Mit dem Prinzip der minimalen CD lassen sich nicht nur kürzeste Reaktionswege auswählen, sondern auch reaktive Zentren, d.h. diejenigen Atome, deren Bindungen gebrochen/geknüpft werden und deren Valenzelektronen an der

Tabelle 1. Korrespondierende Darstellungen der logischen Struktur der konstitutiven Chemie.

Chemische Darstellung	Algebraische Darstellung	Geometrische Darstellung
Konstitutionsformel von $EM_B(A)$ ( $A = \{A_1, \dots, A_n\}$ )	Symmetrische $n \times n$ be-Matrix $B = \langle b_{ij} \rangle$ von $EM_B(A)$	be-Punkt $P(B)$ mit den Koordinaten $(b_{11}, \dots, b_{1n}, \dots, b_{n1}, \dots, b_{nn})$ im $n^2$ -dimensionalen euklidischen Darstellungsraum
Bis zu $n!$ Permutationen der Atomindizes in $EM_B(A)$	Bis zu $n!$ äquivalente be-Matrizen $B'$ , die durch Zeilen/Spaltenpermutationen aus $B$ gemäß $B' = P \cdot B \cdot P^{-1}$ erhalten werden, wobei $P$ eine $n \times n$ Permutationsmatrix ist.	b-Cluster der bis zu $n!$ äquivalenten be-Punkte $P(B')$ von $EM_B(A)$
Chemische Reaktion $EM_B(A) \rightarrow EM_E(A)$ durch Verschiebung von Valenzelektronen	Die additive Transformation von $B$ in $E$ . Die Eintragungen $r_{ij}$ von $R$ zeigen die Verschiebung der Valenzelektronen an.	$r$ -Vektoren im $n^2$ -dimensionalen Raum, welche die be-Punkte $P(B)$ und $P(E)$ der Reaktionsteilnehmer $EM_B(A)$ und $EM_E(A)$ verbinden.
Zahl der im Zuge einer Reaktion $EM_B(A) \rightarrow EM_E(A)$ verschobenen Valenzelektronen (entspricht der CD)	$d(B, E) = \sum  r_{ij} $ , d. h. Summe der Absolutbeträge der Eintragungen der $r$ -Matrix $R = E - B$ .	$L_1$ -Distanz („Taxifahrerdistanz“) zwischen den be-Punkten $P(B)$ und $P(E)$
Die Chemie von $A$ , die durch die Familie aller isomeren $EM(A)$ und die wechselseitige Umwandlung der $EM(A)$ gegeben ist.	Die Mengen der be-Matrizen der $EM(A)$ nebst ihrer Transformationen unter gegebenen Randbedingungen	Gitter von Punkten, welche den Elementen der Familie aller isomeren $EM(A)$ entsprechen, nebst den verbindenden $r$ -Vektoren
Chemische Reaktionen verlaufen bevorzugt unter Verschiebung einer Minimalzahl von Valenzelektronen.	Diejenigen Atom-Atom-Korrelationen $EM_B(A)$ und $EM_E(A)$ werden bevorzugt, welche einen Minimalwert von $d(B, E)$ aufweisen.	Bevorzugte $r$ -Vektoren für $EM_B(A) \rightarrow EM_E(A)$ führen von einem Punkt des $B$ -Clusters zum „nächstgelegenen“ Punkt des $E$ -Clusters.

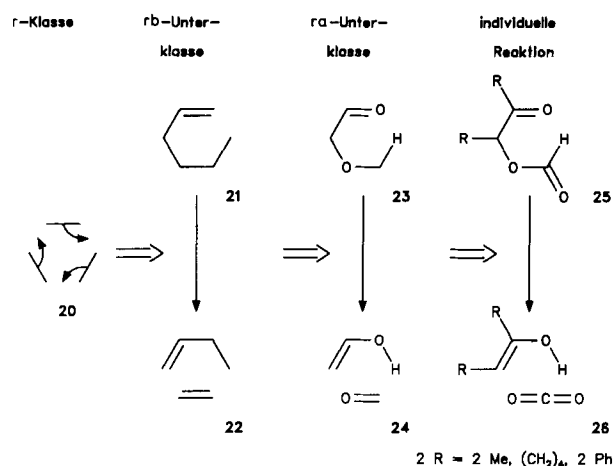
Reaktion direkt beteiligt sind, bestimmen. Die reaktiven Zentren werden als Reaktionskern zusammengefaßt. Aus den Atom-auf-Atom-Abbildungen isomerer EM ergibt sich eine Festlegung der Bindungen, die im Zuge der wechselseitigen Umwandlung dieser EM gebrochen und geknüpft werden müssen.

Inzwischen sind etliche Abstandsfunktionen und Ähnlichkeitsmaße für molekulare Systeme publiziert worden, die davon Gebrauch machen, daß chemische Reaktionen Vektoren entsprechen (siehe Tabelle 1) und metrische Eigenschaften haben<sup>[78]</sup>. Dies sind neben der CD<sup>[100]</sup> die von Kvasnička et al.<sup>[80, 81, 109]</sup> und Hendrickson et al.<sup>[87]</sup> unterschiedlich definierte Reaktionsdistanz, die synthetic proximity von Johnson et al.<sup>[110]</sup> und der Adjazenzabstand von Fontain<sup>[111]</sup>, welcher der Summe der Absolutbeträge der Eintragungen von  $t$ -Matrizen (Differenz zweier Adjazenzmatrizen) entspricht.

### 3.2.2. Die hierarchische Klassifizierung chemischer Reaktionen

Aus dem Prinzip der minimalen CD folgt eine hierarchische Klassifizierung chemischer Reaktionen nach Ähnlichkeitsklassen<sup>[100]</sup> (siehe Schema 2).

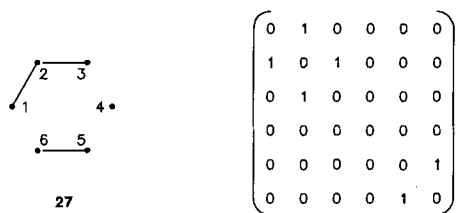
Die Einordnung einer chemischen Reaktion, z.B. **25**  $\rightarrow$  **26**, in dieses hierarchische Klassifizierungssystem erfolgt durch ein stufenweises Weglassen der individuellen Merkmale. Als erstes läßt man in den Reaktionsteilnehmern alle Molekülteile weg, die an der Reaktion nicht beteiligt sind, d. h. die Reaktionen werden auf den Reaktionskern reduziert. Dies führt zur  $ra$ -Unterklasse der betrachteten Reaktion, deren Mitglieder sich in den nicht zum Reaktionskern gehörenden reaktionsinvarianten Molekülteilen unterscheiden.



Schema 2. Das Vorgehen zur hierarchischen Klassifizierung chemischer Reaktionen, von der konkreten Reaktion **25**  $\rightarrow$  **26** zum Elektronenverschiebungsschema **20**.

In der nächsten Stufe des Abstraktionsprozesses wird die Unterscheidung der Atomarten an den reaktiven Zentren aufgegeben. Dies führt zur  $rb$ -Unterklasse (auch Basisreaktion genannt), deren Mitglieder durch die gleiche Anordnung von reaktionsinvarianten kovalenten Bindungen im Reaktionskern gekennzeichnet sind (vgl. **27**). Diese werden durch die sogenannten intakten be-Matrizen repräsentiert (Schema 3)<sup>[79]</sup>.

Ist Reaktionen nur noch das gleiche Elektronenverschiebungsschema und somit die gleiche irreduzible  $r$ -Matrix<sup>[100, 112, 113]</sup> gemeinsam, so sind sie Elemente derselben  $r$ -Klasse. Eine beliebige  $r$ -Matrix kann durch Wegstreichen der Nullzeilen/-spalten reduziert werden, wobei eine irreduzible  $r$ -Matrix resultiert, die in jeder Zeile und Spalte minde-



Schema 3. Intakte be-Matrix der rb-Unterklasse **21** → **22**.

stens eine von Null verschiedene Eintragung enthält. Die Zeilen/Spalten einer irreduziblen r-Matrix sind den reaktiven Zentren zugeordnet.

Die hierarchische Klassifizierung chemischer Reaktionen endet bei den CD-Klassen<sup>[79, 100, 112–114]</sup>. Reaktionen gehören zur gleichen CD-Klasse, wenn in ihrem Verlauf die gleiche Zahl von Valenzelektronen verschoben und somit die gleiche CD durchlaufen wird.

In Schema 2<sup>[100, 114]</sup> ist am Beispiel der Retro-En-Reaktion **21** → **22** die hierarchische Klassifizierung chemischer Reaktionen erläutert. Die rb-Unterklasse wird hier durch das Bindungsschema **27** und die zugehörige 6 × 6 intakte be-Matrix von Schema 3 repräsentiert.

Das oben beschriebene Klassifizierungssystem eröffnet nicht nur neue Möglichkeiten der Reaktionsdokumentation<sup>[79, 100, 112, 113]</sup>, sondern es spielt auch im Computerprogramm IGOR (Interactive Generation of Organic Reactions)<sup>[114–116]</sup> bei der Auswahl und Bestimmung des Neuheitsgrades der mit Computerunterstützung generierten chemischen Reaktionen eine wichtige Rolle<sup>[88, 117]</sup>. Dabei wird für die betrachtete Reaktion die zugehörige Hierarchie ihrer Klassen und Unterklassen gebildet und ermittelt, ab welchem Niveau der Hierarchie sich keine literaturbekannten Reaktionen mehr zuordnen lassen. Je „höher“ dieses Niveau ist, umso größer ist der Neuheitsgrad der betrachteten Reaktion<sup>[88, 113, 114, 117]</sup>. Die mit Computerunterstützung entdeckte Reaktion **25** → **26** in Schema 2 gehört zu den Retro-En-Reaktionen **21** → **22**<sup>[114]</sup> (siehe Abschnitt 4.4.3), die durch das rb-Unterlassenmerkmal **27** gekennzeichnet sind und zur r-Klasse von **20** mit insgesamt 13 Elementen gehören.

Für die ra-Unterklasse **23** → **24** gab es bis 1985 keine publizierten Beispiele. Somit war die Reaktion **25** → **26**<sup>[114, 118]</sup> bis zum Niveau der ra-Unterklassen präzedenzlos.

Bereits vor rund zwanzig Jahren entwickelten Brownson und Stevens<sup>[83]</sup> ein Computerprogramm, das durch Permutieren der chemischen Elemente des Reaktionskerns beliebige Elemente der rb-Unterklasse der Retro-En-Reaktionen erzeugen konnte. Grundsätzlich hätte schon mit diesem Computerprogramm die erwähnte Reaktion entdeckt werden können.

## 4. Das Münchner Projekt

Die Anwendungen des DU-Modells sind nicht auf die Syntheseplanung beschränkt. Nach Abschluß der CLOPS-Studie<sup>[53]</sup> (siehe Abschnitt 2.4) wurde daher 1974 ein umfassendes Arbeitsprogramm, das Münchner Projekt, entworfen, das die Erweiterung des DU-Modells und seine möglichst vielseitige Verwendung bei der computerunter-

stützten Lösung chemischer Probleme zum Gegenstand hatte<sup>[119]</sup>.

Dieses Projekt hat sich im Laufe seiner Durchführung so stark verändert, daß von seinem ursprünglichen Inhalt nur noch die Grundgedanken übriggeblieben sind, aber es erfüllte eine wichtige Funktion: Es setzte die Entwicklung von Computerprogrammen zur deduktiven Lösung chemischer Probleme auf der Grundlage der logischen Struktur der Chemie auf breiter Front in Gang<sup>[119]</sup>. Das Münchner Projekt, dessen allgemeine Ziele wir bereits weitgehend erreicht haben, besteht aus folgenden Teilprojekten:

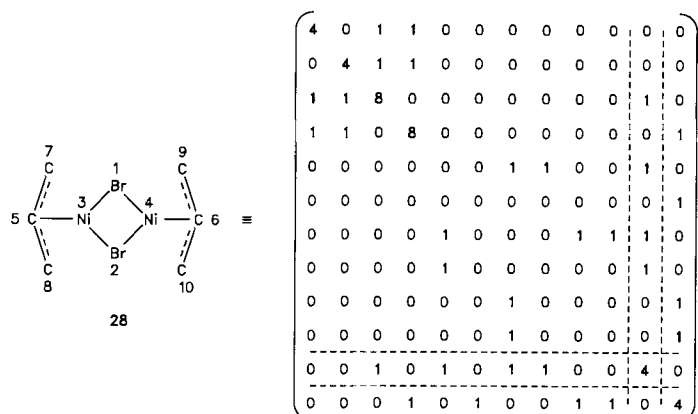
- a) Weiterentwicklung und Erweiterung des DU-Modells insbesondere für EM mit Mehrzentrenbindungen und delokalisierten Elektronensystemen
- b) Entwicklung einer computerorientierten mathematischen Theorie der Stereochemie, die auch den stereochemischen Aspekt konstitutionsverändernder Reaktionen berücksichtigt
- c) Entwicklung einer Software-Infrastruktur für die nachstehenden Programme gemäß (d)
- d) Entwicklung von Computerprogrammen zur deduktiven Lösung chemischer Probleme. Diese Programme sollen das DU-Modell nicht nur bei der Generierung von Problemlösungen nutzen, sondern auch bei deren Sichtung und Selektion.
- e) Erprobung und Verwendung der Programme gemäß (d) und Verbesserung der Programme aufgrund gewonnener Erfahrungen
- f) Die experimentelle Realisierung der gemäß (e) erhaltenen Reaktionsvorschläge.

Über die Realisierung einzelner Teile des Münchner Projekts ist bereits eingehend berichtet worden, wobei wir uns stets bemüht haben, auch die Details der beschriebenen Programme nachvollziehbar zu veröffentlichen, häufig einschließlich der Quellversionen der Programme. Nachstehend wird, der Liste der Teilprojekte a)–f) folgend, ein Überblick über die historische Entwicklung und den gegenwärtigen Stand des Münchner Projekts gegeben.

### 4.1. Die Erweiterung des DU-Modells

Lange Zeit gelang es weder anderen noch uns, das inzwischen gut zwanzig Jahre alte DU-Modell im Sinne von (a) (Abschnitt 4) zu erweitern. Kürzlich jedoch wurde ein Formalismus gefunden, durch den es möglich wird, auch EM mit Mehrzentrenbindungen und Systemen von delokalisierten Valenzelektronen (DE-Systemen) durch erweiterte be-Matrizen, die xbe-Matrizen (x = extended) mit ganzzahligen Eintragungen, darzustellen<sup>[95]</sup>. Die xbe-Matrix eines EM besteht aus seiner be-Matrix und zusätzlichen Zeilen/Spalten. Die be-Matrix mit  $n$  Zeilen/Spalten bezieht sich auf die ganzzahlig repräsentierbaren kovalenten Bindungen des EM und die einzelnen Atomen zugeordneten freien Valenzelektronen. Die zusätzlichen Zeilen/Spalten mit Indizes  $n + k$  gehören jeweils zu einem DE-System, z.B. einem System mit Mehrzentrenbindungen oder mit delokalisierten  $\pi$ -Elektronen. Die außerdiagonalen Eintragungen  $b_{i, n+k} = b_{n+k, i} = 1$  der  $k$ -ten Zusatzzeilen/-spalten zeigen an, daß die Atome  $A_i$  am  $(k - n)$ -ten DE-System beteiligt sind, und die Diagonaleintragungen  $b_{(n+k), (n+k)}$  geben an, wie viele Valenzelektronen zum  $(k - n)$ -ten DE-System gehören. Die

xbe-Matrix des  $\pi$ -Allylnickelbromids **28**<sup>[120]</sup> in Schema 4 möge der Erläuterung dienen; einfachheitshalber wurden die C-H-Bindungen und die zugehörigen Eintragungen weggelassen.



Schema 4. Strukturformel von **28** mit Numerierung und zugehöriger xbe-Matrix. Die Zeilen/Spalten zu den beiden DE-Systemen sind durch gestrichelte Linien abgetrennt.

In Analogie zum ursprünglichen DU-Modell werden chemische Reaktionen als Transformationen der xbe-Matrizen durch Addition von xr-Matrizen beschrieben. Wenn man die außerdiagonalen Eintragungen der Zusatzzeilen/-spalten vernachlässigt, gelten die 18 Theoreme des DU-Modells<sup>[78]</sup> auch für die Algebra der xbe- und xr-Matrizen<sup>[95]</sup>, die „Modell der Chemie delokalisierten Elektronensysteme“ genannt wird.

Zur computergerechten Aufbereitung dieses Formalismus wurde eine neuartige Datenstruktur eingeführt. Ein Reaktionsgenerator zur automatischen Generierung chemischer Reaktionen, an denen auch Systeme mit delokalisierten Elektronen beteiligt sind, wird zur Zeit entwickelt<sup>[121]</sup>. Dies erfordert vor allem eine Berücksichtigung der besonderen Eigenschaften von DE-Systemen durch die Formulierung geeigneter Randbedingungen.

## 4.2. Computerorientierte Formalisierung der Stereochemie – Theorie der chemischen Identitätsgruppe und Akkumulationen

Die Stereochemie ist die Lehre vom räumlichen Bau der Moleküle und dessen beobachtbaren Auswirkungen. Die Unterschiede in den Bildungsweisen, den Reaktionen und den Eigenschaften von Stereoisomeren sind zentrale Anliegen der Stereochemie. Stereoselektive Synthesen gehören zu den aktuellsten Themen der Organischen Chemie<sup>[1]</sup>. Daher besteht großes Interesse an einer Berücksichtigung des stereochemischen Aspekts bei der computerunterstützten Lösung chemischer Probleme. Die immense Datenmenge, die im allgemeinen bei der computerunterstützten Lösung stereochemischer Probleme anfällt, läßt sich nur dadurch bewältigen, daß die den Problemen zugrundeliegenden mathematischen Strukturen voll genutzt werden. Es ist vorteilhaft, hierfür neue Begriffe und Denkweisen einzuführen.

### 4.2.1. Bisherige Ansätze

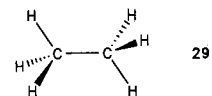
Corey et al.<sup>[122]</sup> ergänzen die Transforms einzelner Reaktionen durch detaillierte empirische stereochemische Angaben. Dies ist ein sehr aufwendiges Verfahren; allein die Berücksichtigung der Stereochemie sechsgliedriger Ringe stellt ein umfangreiches Projekt dar. Wipke und Dyott<sup>[123]</sup> schlugen 1974 ein Verfahren zur Behandlung stereochemischer Probleme vor, das auf der Raumerfüllung von hinreichend starren Reaktionsteilnehmern beruht. Das Programm CHIRON von Hanessian et al.<sup>[124]</sup> ordnet chiralen Synthesezielen chirale Ausgangsstoffe zu. In der Frühzeit des Münchner Projekts wurde die Stereochemie von EM mit stereogenen Untereinheiten der Koordinationszahl  $\leq 4$  durch Paritätsvektoren repräsentiert<sup>[125]</sup>.

Keiner dieser Ansätze hat bisher zu einer allgemein anwendbaren Methode der computerunterstützten Lösung stereochemischer Probleme geführt. Die umfassende computerorientierte formale Behandlung der Stereochemie setzt eine grundsätzlich andere Vorgehensweise voraus.

### 4.2.2. Eine nicht-geometrische Alternative

In der traditionellen Stereochemie werden Fakten und Phänomene anhand starrer geometrischer Molekülmodelle interpretiert und vorausgesagt. Die Stereochemie wird gleichsam auf elementare Geometrie reduziert; molekulare Geometrien und Punktgruppensymmetrien werden hierbei drastisch überbetont. Für breit anwendbare computerunterstützte Methoden zur Lösung stereochemischer Probleme wird ein Ansatz benötigt, der eine einheitliche theoretische Behandlung statischer und dynamischer Objekte der Stereochemie, einschließlich der Relationen zwischen diesen Objekten ermöglicht und der chemischen Realität besser Rechnung trägt als die üblichen stereochemischen Methoden.

Ein großer Teil der Moleküle ist nicht starr, sondern ihre Geometrien sind zeitlich veränderlich; solche Moleküle sind durch geometrische Modelle nicht adäquat repräsentierbar. Beispielsweise ist ein Ethanmolekül **29**, dessen Methylgruppen sich bei 20 °C mit einer Frequenz von ca.  $10^9 \text{ s}^{-1}$  gegeneinander drehen, durch kein starres geometrisches Modell repräsentierbar<sup>[126, 127]</sup>.



Als erste Voraussetzung für eine umfassende, einheitliche theoretische Behandlung der Stereochemie muß die traditionelle Definition von Stereoisomeren durch eine Definition ersetzt werden, die auch flexible Stereoisomere einschließt, eine Definition, die auf dem Begriff der chemischen Identität (siehe Abschnitt 3.1) basiert und sich nicht auf explizite geometrische Vorstellungen stützt.

Das 1970 eingeführte Prinzip der Permutationsisomerie<sup>[76]</sup> und die 1984 publizierte Theorie der chemischen Identitätsgruppe (CIG)<sup>[82]</sup> spielten eine entscheidende Rolle bei der angestrebten Formalisierung der Stereochemie. Die CIG-Theorie<sup>[82]</sup> ist ein gruppentheoretisches Modell der logischen Struktur der Stereochemie, das sowohl energiebezo-

gene chemische und geometrische Moleküleigenschaften als auch die jeweils geltenden Beobachtungsbedingungen erfaßt und den direkten Gebrauch starrer geometrischer Modelle und ihrer Punktgruppensymmetrien vermeidet. Im Gegensatz zu den traditionellen stereochemischen Anwendungen der Gruppentheorie (siehe Abschnitt 2.3.2) bezieht sich die CIG-Theorie nicht nur auf molekulare Objekte, sondern sie hat insbesondere stereochemische Vorgänge und Relationen zwischen molekularen Systemen zum Gegenstand.

#### 4.2.3. Grundgedanken der CIG-Theorie

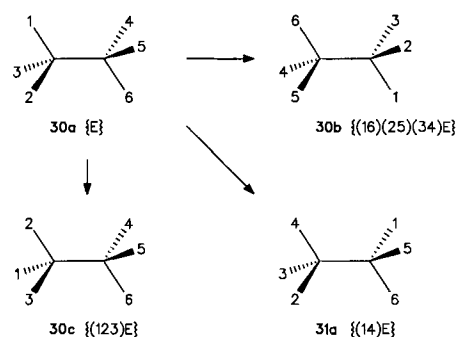
Zerlegt man ein Molekül gedanklich in ein Molekülgerüst und ein Sortiment von Liganden, so entsprechen die distinkten Moleküle, die sich durch permutierte Plazierungen der Liganden am Gerüst unterscheiden, den Permutationsisomeren. Die Menge aller aus einem Bezugsisomer erhältlichen Permutationsisomere wird dessen Familie von Permutationsisomeren genannt.

Sei  $m$  die Momentaufnahme eines Moleküls. Die gedankliche Zerlegung von  $m$  in ein Molekülgerüst und ein Sortiment  $L$  von Liganden heißt Bezugsmodell. Das Bezugsmodell gehört zu einem Isomer  $X$ , das im folgenden als Bezugsisomer dient. Ein Molekülmodell ist im Gegensatz zu einem Molekül hinsichtlich seiner räumlichen Orientierung fixiert. Die chemische Identität eines Moleküls dagegen ist von dessen Orientierung im Raum unabhängig. Es existieren daher Ligandenvertauschungen, die ein Molekülmodell in ein anderes Modell überführen, welches dasselbe Molekül, jedoch in räumlich gedrehter Orientierung repräsentiert (siehe Schema 5). Alle Ligandenpermutationen, durch welche Molekülmodelle von  $X$  in andere Modelle von  $X$  überführt werden – beispielsweise Permutationen, die als Rotationen des Gesamtmoleküls oder als dessen spontane innere Bewegungen interpretiert werden können –, bewahren die chemische Identität von  $X$ . Sind alle Liganden des Sortiments  $L$  unterscheidbar, so bilden die die chemische Identität von  $X$  während Ligandenpermutationen im allgemeinen eine Gruppe  $S(E)$ , die CIG des durch  $E$  repräsentierten Bezugsisomers  $X$ , eine Untergruppe der symmetrischen Permutationsgruppe  $Sym(L)$  der Liganden aus  $L$ . Die distinkten Permutationsisomere von  $X$  werden durch die Linksnebenklassen  $\lambda S(E)$  ( $\lambda \in Sym(L)$ ) von  $S(E)$  in  $Sym(L)$  repräsentiert<sup>[\*]</sup>. Trägt ein Modell  $n$  Liganden, so besteht  $Sym(L)$  aus  $n!$  Permutationen. Dies ist auch die Kardinalität der Familie permutierter Modelle  $P(E)$ . Die Zahl der chemisch verschiedenen Permutationsisomere beträgt maximal  $n!/|S(E)|$ , da die Nebenklassen von  $S(E)$  die gleiche Kardinalität  $|S(E)|$  haben. Die Linksnebenklassen der CIG in  $Sym(L)$  oder jedes ihrer Elemente können als Nomenklaturdeskriptoren von Permutationsisomeren dienen<sup>[76]</sup>. Eine Linksnebenklasse der CIG entspricht demjenigen Permutationsisomer des Bezugsisomers, dessen Modelle aus dem Bezugsmodell durch Ligandenpermutationen aus der CIG und anschließende Anwendung einer Permutation aus der betrachteten Linksnebenklasse entstehen.

[\*] Für Leser, die mit mathematischen Begriffen und Symbolen weniger vertraut sind, werden diese im Anhang kurz erklärt. Eine Kurzdarstellung eines Großteils der hier verwendeten Mathematik findet sich im Anhang von Lit. [82].

Permutationen werden üblicherweise in Cyclenschreibweise, z.B. (123), angegeben<sup>[128]</sup>. Der Zahlenvektor [1 2 3] wird durch (123) im Sinne der Abbildung  $1 \rightarrow 2; 2 \rightarrow 3; 3 \rightarrow 1$  zu [2 3 1] umgeordnet. Die Verknüpfung (12)(123) der Permutationen (12) und (123) entspricht der Anwendung der Permutation (12) auf das Ergebnis der Permutation (123), d.h. (12)(123) überführt [1 2 3] in [1 3 2] und entspricht somit (23).

Das Bezugsmodell **30a** des Ethanderivats **30** wird beispielsweise durch die Ligandenpermutation (16)(25)(34), die einer 180°-Drehung von **30a** um eine senkrecht zur Papierebene stehende Achse entspricht, in das Modell **30b** umgewandelt. Man bemerke, daß diese Drehung des Modells **30a** keiner Deckoperation aus der  $D_{3d}$ -Punktgruppensymmetrie des Ethangerüsts entspricht, denn etwaige Verschiedenheit der Liganden zerstört die Gerüstsymmetrie (siehe Kap. 3 in Lit. [82]).



Schema 5. Ligandenpermutationen ausgehend von **30a**, dem Bezugsmodell des beliebig substituierten Ethanmoleküls.

Die chemische Identität von Molekülen ist von ihrer räumlichen Orientierung unabhängig; daher sind **30a** und **30b** chemisch identisch. Auch das Modell **30c** gehört zu **30**, wenn die Ligandenpermutation (123), die als innere Rotation um die C-C-Achse interpretierbar ist und **30a** in **30c** überführt, unter den gegebenen Beobachtungsbedingungen zu den spontanen inneren Bewegungen von **30** gehört.

$$S(30a) = \{(), (123), (132), (456), (465), (123)(456), (123)(465), (132)(456), (132)(465), (14)(26)(35), (15)(24)(36), (16)(25)(34), (2)(142635), (143526), (152436), (153624), (162534), (163425)\}$$

Die CIG (2) enthält alle Ligandenpermutationen, welche die chemische Identität von **30** erhalten. Die Ligandenpermutationen aus  $S(30a)$  erzeugen aus beliebigen Modellen aus  $S(30a) \cdot \mathbf{30a}$  wieder Modelle aus  $S(30a) \cdot \mathbf{30a}$ . Die Ligandenpermutation (14) verwandelt **30a** in **31a**, das ein Permutationsisomer von **30** ist. Ebenso überführt die Ligandenpermutation (14) beliebige Modelle von **30** in Modelle von **31**, die allesamt aus **30a** durch Anwendung von Permutationen aus  $S(30a)$  und eine anschließende beliebige Permutation aus (14) ·  $S(30a)$  erhalten werden. Nur bei Verschiedenheit der Liganden 1 und 4 sind **30** und **31** unterscheidbare Permutationsisomere.

Werden einzelne Liganden als chemisch äquivalent betrachtet, oder existieren Isomerisierungsreaktionen zwischen Permutationsisomeren, so sind die im folgenden Abschnitt beschriebenen mengenwertigen Abbildungen und das darauf basierende Verfahren der Akkumulation von großem Nutzen.

#### 4.2.4. Implementierung und Anwendung der CIG-Theorie

Stereochemische Äquivalenzrelationen, einschließlich der wechselseitigen Umwandelbarkeit von Permutationsisomeren und der denkbaren gleichzeitigen Bildung von Stereoisomeren im Zuge konstitutionsverändernder Reaktionen<sup>[77, 98, 100]</sup>, werden im Rahmen der CIG-Theorie durch die surjektiven und totalen mengenwertigen Abbildungen (MA) von Permutationen auf Mengen von Permutationen, die als Vereinigungen von Nebenklassen und Wignerschen Unterklassen<sup>[129]</sup> von Untergruppen von  $Sym(L)$  vorliegen, repräsentiert. Die direkte Berücksichtigung des dynamischen Aspekts ist der wichtigste Beitrag, den die CIG-Theorie zur mathematischen Behandlung der Stereochemie geleistet hat.

Mengenwertige Abbildungen wurden in der Chemie erstmals im Rahmen der CIG-Theorie als vielseitiges Mittel zur Lösung von Problemen, denen als Äquivalenzrelationen formulierbare Beziehungen zwischen Permutationsisomeren zugrundeliegen, eingesetzt. Äquivalenzrelationen lassen sich durch Äquivalenzklassenräume repräsentieren, die durch MA erzeugt werden.

Eine Menge  $M$  sei auf zwei unterschiedliche Weisen in Äquivalenzklassen partitioniert (Abb. 1). Die MA einer

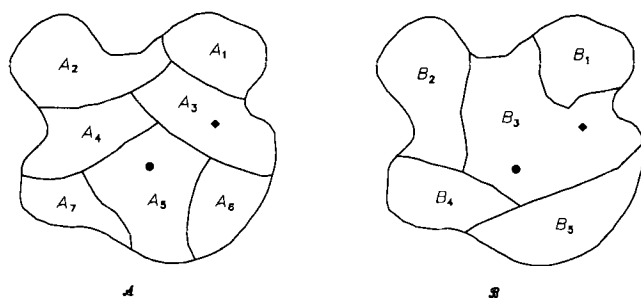


Abb. 1. Zwei Partitionen  $\mathcal{A}$  und  $\mathcal{B}$  einer Menge  $M$ .

Klasse  $B$  aus Partition  $\mathcal{B}$  liefert die Vereinigung derjenigen Äquivalenzklassen  $A$  aus Partition  $\mathcal{A}$ , die einen nichtleeren Schnitt mit  $B$  bilden [Gl. (3)]. Die Klasse  $A_3$  enthält unter

$$MA(\mathcal{A}, B) = \bigcup_{A \cap B \neq \{\}, A \in \mathcal{A}} A \quad (3)$$

anderem das Element  $\blacklozenge$ , die Klasse  $A_5$  das Element  $\bullet$ . Diese beiden Elemente sind in einer Klasse  $B \in \mathcal{B}$  (in  $B_3$ ) enthalten. Alle Elemente der Klasse  $A_3$  sind zu  $\blacklozenge$  und alle Elemente der Klasse  $A_5$  zu  $\bullet$  äquivalent. Zugleich sind  $\blacklozenge$  und  $\bullet$  äquivalent, da sie sich in derselben Klasse  $B_3$  befinden.

Äquivalenzrelationen sind reflexiv, symmetrisch und transitiv (siehe S. 14 in Lit.<sup>[128]</sup>). Alle Elemente der Klasse  $A_3$  sind zu  $\blacklozenge$  und, aufgrund der Transitivität, zu allen Elementen aus  $B_3$  äquivalent, also insbesondere zu  $\bullet$ . Somit sind auch alle Elemente der Klasse  $A_3$  zu allen Elementen der Klasse  $A_5$  äquivalent. Die Äquivalenzklassen  $A_3$  und  $A_5$  (und alle weiteren Äquivalenzklassen aus  $\mathcal{A}$ , die sich mit  $B_3$  schneiden) bilden eine einzige Äquivalenzklasse. Es ergibt sich die MA gemäß (4).

$$MA(\mathcal{A}, B_3) = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6 \quad (4)$$

In den letzten Jahren ist die CIG-Theorie modifiziert und erweitert worden und eignet sich jetzt besser für die computerunterstützte Behandlung einer großen Vielfalt stereochemischer Probleme. Insbesondere stellt das kürzlich eingeführte Verfahren der (Äquivalenz-)Akkumulation<sup>[130, 131]</sup> für die Praxis der computerunterstützten Stereochemie einen entscheidenden Fortschritt dar.

Die CIG nebst ihren Nebenklassen in  $Sym(L)$  spannt einen Äquivalenzklassenraum auf. Da Nebenklassen per definitionem disjunkt oder identisch sind, kann somit jede Nebenklasse auch als Äquivalenzklasse betrachtet werden, die genau ein Permutationsisomer repräsentiert. Um die gesamte Familie der Permutationsisomere zu beschreiben, genügt es daher, aus jeder Äquivalenzklasse eine Permutation als Repräsentant für jeweils ein Permutationsisomer zu wählen. Bei der Behandlung stereochemischer Probleme verringert sich die Datenmenge dadurch auf einen Bruchteil.

Ausgehend von einem Äquivalenzklassenraum  $A$  von Modellen und einer weiteren, davon unabhängigen Information  $I$  über Äquivalenzbeziehungen zwischen Modellen, die sich beispielsweise aufgrund einer Ligandenäquivalenz oder einer Isomerisierungsmöglichkeit ergibt, läßt sich ein Äquivalenzklassenraum  $U$  schaffen, der sich von  $A$  darin unterscheidet, daß einige der Äquivalenzklassen von  $A$  in ihm vereinigt sind. Die Information  $I$  ist eine Äquivalenzrelation, die sich aus gruppentheoretischen Beziehungen der Modelle ergibt und die der Anwender der CIG-Theorie daher mit wenigen Erzeugenden, die sich unmittelbar aus den chemischen Verhältnissen ergeben, definieren kann. Ein Beispiel für eine solche Information ist der Nebenklassenraum der in Abschnitt 4.2.3 erwähnten CIG  $S(30a)$  in  $Sym(L)$ : Für jede Nebenklasse  $\lambda S(30a)$  ist  $\lambda S(30a) \cdot 30a$  eine Äquivalenzklasse von chemisch identischen Modellen. Alle Paare von Modellen dieser Äquivalenzklasse sind Elemente der zugehörigen Äquivalenzrelation  $I$ , d. h. für alle Modelle  $m, n \in \lambda S(30a) \cdot 30a$  gilt:  $(m, n) \in I$ .

Für eine Menge von Modellen,  $M$ , die in den Äquivalenzklassenraum  $A$  partitioniert ist, und eine Information  $I$  ist die Akkumulation  $Akk(A, I)$  gemäß (5)–(7) definiert (+ symbolisiert die transitive Hülle<sup>[132]</sup> einer Relation).

$$Akk(A, I) := \{[m]_R \mid m \in M\} \quad (5)$$

$$R = (I \cup a \times a \mid a \in A)^+ \quad (6)$$

$$[m]_R = \{n \in M \mid (m, n) \in R\} \quad (7)$$

Liganden, die die gleiche chemische Interpretation aufweisen und deren Vertauschung die chemische Identität von Molekülen nicht verändert, heißen äquivalent. Ihre Vertauschungen bilden die Ligandenäquivalenzgruppe  $\Sigma$ . Die Rechtsnebenklassen der Ligandenäquivalenzgruppe in  $Sym(L)$  symbolisieren Modelle, die bei Verschiedenheit aller Liganden möglicherweise unterscheidbar wären, aber wegen der Äquivalenz eines Teils der Liganden chemisch identisch sind.

Eine Ligandenpermutation, die ein Bezugsmodell in ein Modell eines distinkten Permutationsisomers umwandelt, und anschließend durchgeführte Ligandenpermutationen aus der Ligandenäquivalenzgruppe  $\Sigma$  entsprechen einer Ligandenpermutation aus einer Rechtsnebenklasse von  $\Sigma$ .

Der Rechtsnebenklassenraum der Ligandenäquivalenzgruppe  $\Sigma$  in  $Sym(L)$  enthält also folgende Information: Die

Modelle  $\lambda E$  einer jeden Äquivalenzklasse  $\Sigma\mu$  sind jeweils chemisch identisch. Schneidet eine Äquivalenzklasse  $\Sigma\mu$  beispielsweise die Äquivalenzklassen  $\lambda S(E)$  und  $\sigma S(E)$  aus dem Nebenklassenraum der CIG eines Modells  $E$ , bedeutet dies, daß Permutationen  $\varrho$  und  $\tau$  mit  $\varrho \in \Sigma\mu \cap \lambda S(E)$  und  $\tau \in \Sigma\mu \cap \sigma S(E)$  existieren. Wegen  $\varrho \in \lambda S(E)$  ist das Modell  $\varrho E$  chemisch identisch mit allen Modellen  $\lambda S(E) \cdot E$ . Analog ist  $\tau E$  wegen  $\tau \in \sigma S(E)$  chemisch identisch zu allen Modellen  $\sigma S(E) \cdot E$ . Zugleich sind  $\varrho E$  und  $\tau E$  chemisch identisch, weil  $\varrho$  und  $\tau$  zur selben Äquivalenzklasse  $\Sigma\mu$  gehören, d. h. die durch  $\varrho E$  und  $\tau E$  repräsentierten Moleküle sind chemisch identisch, weil sich die Modelle  $\varrho E$  und  $\tau E$  nur durch die Vertauschung äquivalenter Liganden unterscheiden.

Da die chemische Identität als Äquivalenzrelation reflexiv, symmetrisch und transitiv ist, gilt: Alle Modelle aus  $\lambda S(E) \cdot E$  sind ebenfalls chemisch identisch mit  $\tau E$  (weil  $\varrho E$  mit  $\tau E$  chemisch identisch ist) und somit auch mit allen Modellen  $\sigma S(E) \cdot E$  (weil  $\tau E$  mit all diesen Modellen chemisch identisch ist). Unter Beachtung der in Form der Äquivalenzklassen  $\Sigma\mu$  vorliegenden Information, daß alle Modelle  $\Sigma\mu \cdot E$  chemisch identisch sind, „verschmelzen“ also die Repräsentanten  $\lambda S(E) \cdot E$  und  $\sigma S(E) \cdot E$  (und alle weiteren Äquivalenzklassen  $\xi S(E) \cdot E$  mit  $\xi S(E) \cap \Sigma\mu \neq \{\}$ ) von ursprünglich als unterscheidbar betrachteten Permutationsisomeren zu der durch (8) beschriebenen Klasse permutierter Modelle, die ein einziges Permutationsisomer repräsentiert.

$$\bigcup_{\xi S(E) \cap \Sigma\mu \neq \{\}} \xi S(E) \cdot E = \text{MA}(\{\lambda S(E) \cdot E \mid \lambda \in \text{Sym}(L)\}, \Sigma\mu \cdot E) \quad (8)$$

Die Tatsache, daß bei einer solchen MA nicht der gesamte Rechtsnebenklassenraum von  $\Sigma$  in Betracht gezogen wird, kann in manchen Fällen zu unzureichenden Ergebnissen führen. Bisher wurde nur die MA aller Modelle  $\nu E$  ( $\nu \in \Sigma\mu$ ) betrachtet. Es kann daher der Fall eintreten, daß eine Nebenklasse  $\lambda S(E)$  existiert, die eine nichtleere Schnittmenge sowohl mit  $\Sigma\mu$  als auch mit einer weiteren Rechtsnebenklasse  $\Sigma\omega$  aufweist. In diesem Fall liefert die MA von  $\Sigma\mu \cdot E$  nicht die Gesamtheit der chemisch identischen Modelle: Der Ausdruck (9) bedeutet, daß Modelle  $\lambda S(E) \cdot E$  sowohl zu allen Modellen aus (10) als auch zu allen Modellen aus (11) chemisch identisch sind. Folglich sind auch alle Modelle dieser beiden Vereinigungen chemisch identisch. Derartige Probleme vermeidet die Akkumulation (12) durch Bildung der transitiven Hülle (13).

$$\lambda S(E) \cap \Sigma\mu \neq \{\} \wedge \lambda S(E) \cap \Sigma\omega \neq \{\} \quad (9)$$

$$\bigcup_{\xi S(E) \cap \Sigma\mu \neq \{\}} \xi S(E) \cdot E \quad (10)$$

$$\bigcup_{\xi S(E) \cap \Sigma\omega \neq \{\}} \xi S(E) \cdot E \quad (11)$$

$$\text{Akk}(\{\lambda S(E) \cdot E \mid \lambda \in \text{Sym}(L)\}, \{\Sigma\mu \cdot E \times \Sigma\mu \cdot E \mid \Sigma\mu \cdot E \in \{\Sigma\mu \cdot E \mid \mu \in \text{Sym}(L)\}\}) \quad (12)$$

$$R = (\{\Sigma\mu \cdot E \times \Sigma\mu \cdot E \mid \Sigma\mu \cdot E \in \{\Sigma\mu \cdot E \mid \mu \in \text{Sym}(L)\}\} \cup \{\lambda S(E) \cdot E \times \lambda S(E) \cdot E \mid \lambda S(E) \cdot E \in \{\lambda S(E) \cdot E \mid \lambda \in \text{Sym}(L)\}\})^+ \quad (13)$$

Die Familie der Permutationsisomere mit teilweise äquivalenten Liganden ergibt sich schließlich als Menge der Doppelnebenklassen  $\Sigma\lambda S(E)$  ( $\lambda \in \text{Sym}(L)$ ). In solchen Fällen ist bei der Anwendung der üblichen Methoden zur gruppentheoretischen Abzählung und Klassifizierung von Isomeren<sup>[70–74]</sup> Vorsicht geboten. Im Gegensatz zu diesen Metho-

den gibt die CIG-Theorie auch an, um welche Isomere es sich jeweils handelt und welche Isomere durch Äquivalenz von Liganden zu einem einzigen Isomer verschmelzen sowie welche chiralen Isomere ein chirales Isomer ergeben und welche ein achirales.

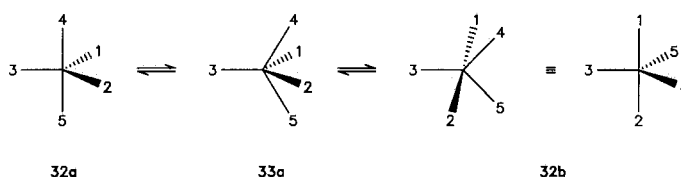
Die Akkumulation ist insofern eine Verallgemeinerung der MA, als bei der Akkumulation die gesamte vorhandene Information  $I$ , in diesem Fall der gesamte Rechtsnebenklassenraum von  $\Sigma$ , herangezogen wird, um die wirklich verschiedenen (distinkten) Permutationsisomere zu bestimmen.

Neben Ligandenäquivalenzen gibt es weitere Gründe, die dazu führen, daß ursprünglich als unterscheidbar betrachtete Permutationsisomere als chemisch identisch einzustufen sind.

Von besonderem Interesse sind Isomerisierungen eines Permutationsisomers  $X$  zu anderen Permutationsisomeren der gleichen Familie, die über ein Permutationsisomer  $Y$  einer weiteren Familie verlaufen (siehe Schema 6). Es seien  $E$  und  $F$  Bezugsmodelle von  $X$  und  $Y$ . Man wählt die Reaktion  $E \rightleftharpoons F$  als Bezugsreaktion, d. h. es wird angenommen, daß eine Umwandlung von  $\lambda E$  in  $\lambda F$  ( $\forall \lambda \in \text{Sym}(L)$ ) und umgekehrt stattfindet. Ist man an den Permutationsisomeren von  $X$  interessiert, so bildet man zunächst die Linksnebenklassenräume der beiden CIG  $S(E)$  und  $S(F)$ . Der Linksnebenklassenraum von  $S(E)$  entspricht den distinkten Permutationsisomeren von  $X$ , wenn keine Isomerisierung stattfindet. Die Information, welche Permutationsisomere von  $X$  durch Isomerisierung ineinander umwandelbar und daher als chemisch identisch zu behandeln sind, liegt in Form des Nebenklassenraums von  $S(F)$  in  $\text{Sym}(L)$  vor. Das Verfahren zur Bestimmung der distinkten Isomere von  $X$  unter Berücksichtigung der Isomerisierung entspricht genau dem Verfahren, mit dem die Verschmelzung von Äquivalenzklassen von  $S(E)$  aufgrund von Ligandenäquivalenzen untersucht wird.

Aus den MA der Linksnebenklassen  $\lambda S(E)$  und  $\lambda S(F)$  zweier CIG  $S(E)$  und  $S(F)$ , die Bezugsreaktionen repräsentieren, können Reaktionsschemata (oder Reaktionsgraphen)<sup>[82]</sup> erzeugt werden, welche die Gesamtheit aller wechselseitigen Umwandlungen der entsprechenden Permutationsisomere wiedergeben.

Die Generierung des Graphen aller Berry-Pseudorotationen<sup>[133]</sup> und Turnstile-Rotationen<sup>[82, 134, 135]</sup> der Familie **32** von Permutationsisomeren aus der Bezugsreaktion **32a**  $\rightarrow$  **32b** durch MA der Linksnebenklassen von  $S(32a)$  und  $S(32b)$  möge dies illustrieren (Schema 6). Hierbei werden die



Schema 6. Die Bezugsreaktion für die Erzeugung des Graphen der Berry-Pseudo- und der Turnstile-Rotation.

einzelnen Mitglieder der Familie **32** von je einer Linksnebenklasse von  $S(32a)$  und  $S(32b)$  repräsentiert. Isomerisierungsprozesse zwischen mehreren Familien von Permutationsisomeren wie **32** und **33** können durch (Äquivalenz-) Akkumulationen beschrieben werden, wenn Permutations-

isomere mit dem gleichen Ligandensortiment an der Bezugsreaktion beteiligt sind. Verwendet man  $32a \rightleftharpoons 33a$  als Bezugsreaktion, dann resultiert ein Graph der Berry-Pseudorotation, in dem die Reaktionsteilnehmer  $32$  über die Übergangszustände  $33$  verbunden sind (siehe S. 132 in Lit.<sup>[82]</sup>).

Ligandenpermutationen, die der gleichen Wignerschen Unterklasse<sup>[129]</sup> der CIG angehören, entsprechen Permutationsisomerisierungen mit analogen (bezüglich der Gerüstsymmetrie äquivalenten) Bewegungsabläufen, d. h. dem gleichen Umordnungsmechanismus<sup>[96, 134]</sup>.

Werden Linksnebenklassen einer CIG von einer ihrer Wignerschen Unterklassen geschnitten, so repräsentiert ihre Vereinigung einen Musher-Mode<sup>[82, 136, 137]</sup>, d. h. diejenigen Permutationsisomere, die aus einem Bezugsisomer durch einen gegebenen Umordnungsmechanismus (oder einen äquivalenten Mechanismus) unmittelbar hervorgehen. Die Berry-Pseudorotation und die Turnstile-Rotation sind bezüglich ihrer Ergebnisse äquivalent und gehören daher dem gleichen Musher-Mode an. Sie werden durch den gleichen Isomerisierungsgraphen repräsentiert (siehe S. 132 in Lit.<sup>[82]</sup>).

Da das Resultat einer (Äquivalenz-)Akkumulation wieder ein Äquivalenzklassenraum ist, kann man mit weiteren Informationen (z. B. zusätzlichen Ligandenäquivalenzen oder Isomerisierungsmechanismen) erneut das Verfahren der Akkumulation anwenden und erhält schließlich einen Äquivalenzklassenraum, dessen Mächtigkeit Aussagen über die Zahl chemisch unterscheidbarer Permutationsisomere zulässt und außerdem über Inklusionsbeziehungen auch Aussagen über Chiralität und Hyperchiralität<sup>[138]</sup> enthält. Der schrittweise Ablauf des Akkumulationsalgorithmus erlaubt zudem die Aufstellung von Isomerisierungsgraphen und eignet sich daher vorzüglich zur Analyse und Beschreibung stereoselektiver Reaktionen. Sind die Liganden 1 und 2 sowie 3 und 4 von  $32$  chemisch äquivalent, so wird der Äquivalenzklassenraum (14) einer Akkumulation mit dem Rechtsnebenklassenraum der entsprechenden Ligandenäquivalenzgruppe (15) unterworfen, wobei ein neuer Äquivalenzklassenraum (16) resultiert. Der Graph der Berry-Pseudo/Turnstile-Rotation ändert sich entsprechend (siehe S. 133 in Lit.<sup>[82]</sup>).

$$A = \text{Akk}(\{\lambda S(32a) \cdot 32a | \lambda \in \text{Sym}(L)\}, \{\lambda S(32b) \cdot 32b \times \lambda S(32b) \cdot 32b | \lambda S(32b) \cdot 32b \in \{\lambda S(32b) \cdot 32b | \lambda \in \text{Sym}(L)\}\}) \quad (14)$$

$$\Sigma = \{(), (12), (34), (12)(34)\} \quad (15)$$

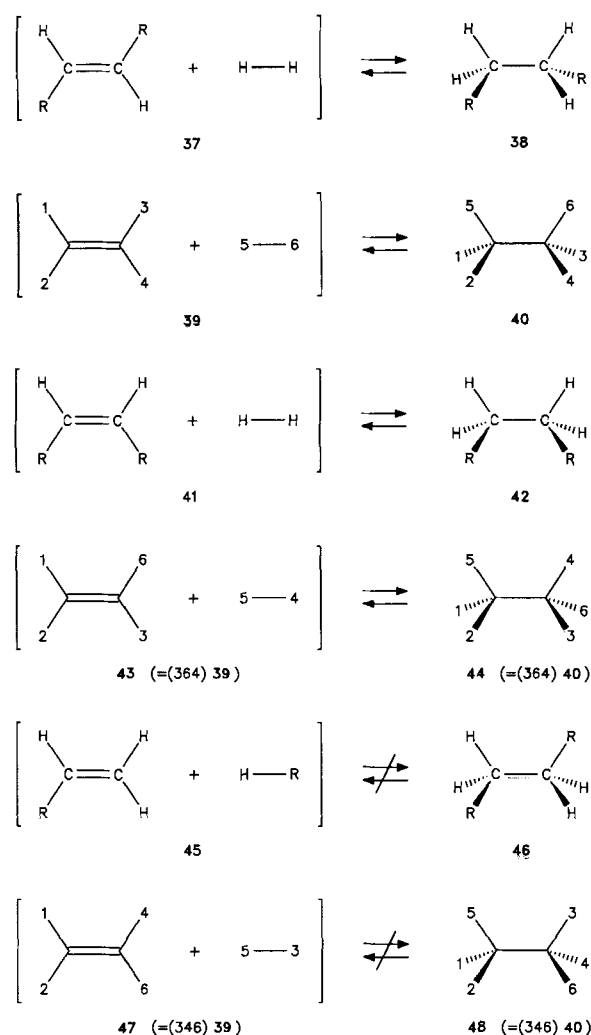
$$B = \text{Akk}(A, \{\Sigma\mu \cdot E \times \Sigma\mu \cdot E | \Sigma\mu \cdot E \in \{\Sigma\mu \cdot E | \mu \in \text{Sym}(L)\}\}) \quad (16)$$

Analog ist es auch möglich festzustellen, welche Stereoisomere im Rahmen einer gegebenen konstitutionsverändernden Reaktion, bei der das Ligandensortiment erhalten bleibt, gebildet werden können. Die Bildung eines der Stereoisomere dient hierbei als Bezugsreaktion, und die Konstitutionsänderung wird als Übergang von einer Familie

von Permutationsisomeren zu einer anderen behandelt. Auf diese Art können beispielsweise alle Diels-Alder-Reaktionen, welche zwischen  $34$  und  $35$  analog der Bezugsreaktion zu  $36$  stattfinden können, dargestellt werden<sup>[98]</sup>.

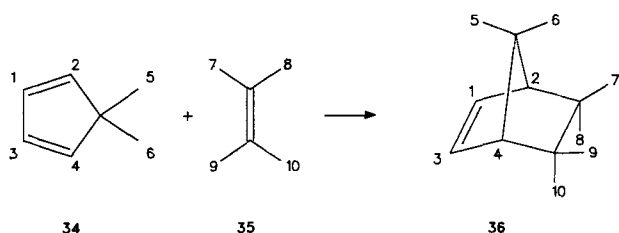
Die geschilderten Anwendungen der Akkumulation zeigen die Bandbreite der mit MA und Reaktionsschemata behandelbaren Problemstellungen<sup>[94, 95, 97, 124]</sup>.

Mit Hilfe der Akkumulation können aber auch Isomerisierungsprozesse zwischen zwei Familien von Permutationsisomeren  $X$  und  $Y$  mit verschiedenen Ligandensätzen (mit Bezugsmodellen  $E$  bzw.  $F$ ) behandelt werden, bei denen aus einer Referenzisomerisierung  $E \rightleftharpoons F$  nicht automatisch folgt, daß alle Prozesse  $\lambda E \rightleftharpoons \lambda F$  ebenfalls stattfinden. Die über ein Alkan verlaufende *cis-trans*-Isomerisierung eines Alkens gemäß Schema 7 möge als einfaches Beispiel dienen: Zunächst wird eine unter den Beobachtungsbedingungen freie innere Drehbarkeit um die C-C-Achse der Alkane angenommen. Daraus folgt, daß die Modelle (346) $40$  (=  $48$ ) und (364) $40$  (=  $44$ ) mit  $40$  identisch sind. Die CIG  $S(40)$  enthält unter anderem die Permutation (346) und (364). Diese befinden



Schema 7. Zur *cis-trans*-Isomerisierung von Alkenen.

sich in unterschiedlichen Linksnebenklassen von  $S(39)$ , so daß die Schnitte  $S(39) \cap (346)S(39)$ ,  $S(39) \cap (364)S(39)$  und  $(346)S(39) \cap (364)S(39)$  leer sind. Jedoch schneidet  $S(40)$  jede der genannten Linksnebenklassen von  $S(39)$ :  $() \in S(40) \cap$



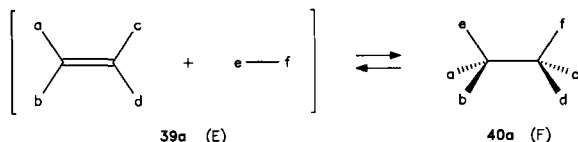


$S(39)$ ,  $(346) \in S(40) \cap (346)S(39)$  und  $(364) \in S(40) \cap (364)S(39)$ .

Nach den bisherigen Ausführungen folgt, daß **39**, **(346)39** ( $=$  **47**) und **(364)39** ( $=$  **43**) äquivalent sein müssen. Für **39** und **(364)39** trifft dies sicher zu, denn **37** läßt sich auf dem durch  $39 \rightarrow 40 \rightarrow (364)40 \rightarrow (364)39$  repräsentierten Weg der Hydrierung mit nachfolgender innerer Rotation des Hydrierungsprodukts und anschließende Dehydrierung in **41** umwandeln. Die beiden ursprünglich (aufgrund des Linksnebenklassenraums von  $S(39)$ ) als unterscheidbar eingestuft *cis-trans*-Isomere **37** und **41** (die innere Rotation um die C-C-Bindung ist hier nicht erlaubt) werden daher äquivalent.

Im Gegensatz zu  $39 \rightleftharpoons 40$  und  $(364)39 \rightleftharpoons (364)40$  repräsentiert jedoch  $(346)39 \rightleftharpoons (346)40$  keine Hydrierung/Dehydrierung, sofern  $R \neq H$  ist. Folglich kann **46**, auch wenn es mit **38** und **42** äquivalent ist, nicht in **45** umgewandelt werden, und  $(346)39 \rightleftharpoons (346)40$  entspricht nicht der Modellreaktion einer Hydrierung/Dehydrierung. Eine einfache Bestimmung von Schnittmengen ist in diesem Falle nicht zur Feststellung der wechselseitigen Umwandelbarkeit geeignet.

Intermolekulare Relationen lassen sich durch Angabe von „Filtern“ realitätsgerecht beschreiben, die dazu dienen, aus allen denkbaren Umwandlungen  $\lambda E \rightleftharpoons \lambda F$  ( $\lambda \in \text{Sym}(L)$ ) diejenigen herauszufiltern, welche der chemischen Realität entsprechen. Diese „erlaubten“ individuellen Isomerisierungen  $\lambda E \rightleftharpoons \lambda F$  zeichnen sich dadurch aus, daß die beteiligten permutierten Modelle eine bestimmte Plazierung des Ligandensortiments an den Gerüstplätzen aufweisen. Die Gerüstplätze werden folgendermaßen festgelegt: In einem Bezugsmodell  $M$  ist der Ligand 1 am Gerüstplatz  $a(M)$  platziert, Ligand 2 am Gerüstplatz  $b(M)$  etc. Die Zuordnung der Gerüstplätze der Hydrierung/Dehydrierung von Schema 7 entspricht der in Schema 8 mit  $39a = E$  und  $40a = F$  als Bezugsmodellen.



Schema 8. Gerüstplatzbezeichnung zur *cis-trans*-Isomerisierung gemäß Schema 7.

Ein *syn*-Mechanismus für die Hydrierung/Dehydrierung setzt voraus, daß die Gerüstplätze  $e(F)$  und  $f(F)$  ebenso wie die Gerüstplätze  $e(E)$  und  $f(E)$  von H-Atomen besetzt sind. Die übrigen Gerüstplätze können prinzipiell beliebig belegt sein. Es wird lediglich gefordert, daß sich entsprechende Gerüstplätze  $a(E)/a(F)$ ,  $b(E)/b(F)$  etc. jeweils dieselben Liganden tragen. Der Filter für diese Isomerisierung ist also gemäß (17) zu formulieren.

$$\Phi: = \{a(E) = a(F) \wedge b(E) = b(F) \wedge c(E) = c(F) \wedge d(E) = d(F) \wedge e(E) = e(F) = H \wedge f(E) = f(F) = H\} \quad (17)$$

Die Formulierung „Gerüstplatz“ = „Gerüstplatz“ drückt aus, daß die beiden spezifizierten Gerüstplätze äquivalente Liganden tragen müssen (Beispiel:  $a(E) = a(F)$ ). Mit „Gerüstplatz“ = „Gerüstplatz“ = „Ligand“ wird darüber hinaus anhand von „Ligand“ festgelegt, welcher Ligand dies sein muß (Beispiel:  $e(E) = e(F) = H$ ). Man kann erkennen,

daß sowohl die individuelle Isomerisierung  $E \rightleftharpoons F$  als auch die Isomerisierung  $(364)E \rightleftharpoons (364)F$  den Ansprüchen dieses Filters genügen, nicht aber die Isomerisierung  $(346)E \rightleftharpoons (346)F$ .

Es gibt auch Isomerisierungen, die Mitglieder von Familien mit unterschiedlichen Ligandensätzen ineinander umwandeln. Es seien  $E$  und  $F$  die Bezugsmodelle der beteiligten Verbindungen  $X$  und  $Y$ . Der Ligandensatz von  $E$  sei  $L(E) = \{1,2,3,4\}$ , der von  $F$   $L(F) = \{1,2,3,4,5\}$ . Offensichtlich läßt sich die Permutation (15) nur auf  $F$  sinnvoll anwenden, denn das Modell  $E$  besitzt keinen fünften Liganden. Außerdem kann der Ligand 4 an  $E$  und  $F$  verschiedene chemische Reste darstellen (z.B.  $4(E) = \text{Me}$ ,  $4(F) = \text{Et}$ ). Daher kann die chemische Bedeutung der Permutation (14) je nach Modell verschieden sein. Auch in solchen Fällen eignet sich die Bestimmung der Schnittmengen der CIG  $S(E)$  und  $S(F)$  und ihrer Nebenklassen nicht, um festzulegen, welche Permutationsisomere  $X$  und  $Y$  durch die Isomerisierung verbunden sind; es müssen Filter eingesetzt werden, um die erlaubten Isomerisierungen auszuwählen. Der Vorzug der Akkumulation<sup>[130, 131]</sup> gegenüber den bisher üblichen MA und Reaktionsschemata<sup>[82, 100]</sup> besteht unter anderem in dieser Selektion erlaubter Isomerisierungen. Die automatische Ermittlung von Relationen (z.B. wechselseitiger Umwandelbarkeit) zwischen Permutationsisomeren (die verschiedenen Familien angehören können) ausschließlich anhand der nichtleeren Schnittmengen von Äquivalenzklassen schränkt die Anwendungsbreite der Reaktionsschemata und MA erheblich ein. Diese Einschränkung ist durch den Akkumulationsformalismus beseitigt worden, der auf der Auswertung beliebiger Relationen zwischen Modellen in Familien von Permutationsisomeren basiert.

Seit vier Jahren entwickeln wir auf der Grundlage der hier besprochenen mathematischen Modellvorstellungen Computerprogramme zur Lösung stereochemischer Probleme<sup>[130, 131]</sup>.

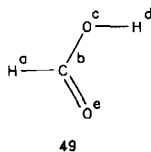
### 4.3. Die Software-Infrastruktur

Für die Computerprogramme zur deduktiven Lösung chemischer Probleme auf der Grundlage des DU-Modells und der CIG-Theorie mußte zunächst eine geeignete Software-Infrastruktur geschaffen werden.

#### 4.3.1. Die kanonische Bezifferung von Atomen

Für die interne Dokumentation wurde der Algorithmus und das Computerprogramm CANON<sup>[139, 140]</sup> entwickelt, das sich auf die Ordnungszahlen und die Koordinationszahlen der Atome sowie deren durch die kovalenten Bindungen gegebene Nachbarschaftsverhältnisse bezieht. Ein einfaches Beispiel möge das Prinzip von CANON erläutern (Schema 9).

CANON beginnt mit einer beliebig gewählten Markierung der Atome; im Beispiel **49** die Buchstaben a–e. Dann werden den Atomen die ersten Atomindizes zugeordnet, und zwar nach fallender Ordnungszahl (O-Atome > C-Atome > H-Atome). Aus den ersten Atomindizes eines Atoms und seiner kovalent gebundenen unmittelbaren Nachbaratome werden die ersten Atomdeskriptoren gebildet (z.B. für b:



1. Atomindizes : a: 3 b: 2 c: 1 d: 3 e: 1  
 1. Atomdeskriptoren : a: 3.2 b: 2.113 c: 1.23 d: 3.1 e: 1.2

2. Atomindizes : a: 5 b: 3 c: 2 d: 4 e: 1  
 2. Atomdeskriptoren : a: 5.3 b: 3.125 c: 2.34 d: 4.2 e: 1.3

Schema 9. Beispiel zur Erläuterung von CANON.

2.113). Anschließend wird die lexikographische Ordnung der ersten Atomdeskriptoren zur Erzeugung der zweiten Atomindices genutzt (z.B.  $1.2 < 1.23 < 2.113$ , daher e: 1, c: 2, b: 3). Da die lexikographische Ordnung der zweiten Atomdeskriptoren bereits identisch mit den zweiten Atomindices ist, sind letztere die endgültigen Atomindizes nach CANON. Wäre dies nicht der Fall, so würde ein weiterer Cyclus mit der Berechnung der nächsten Atomdeskriptoren beginnen. Bei Vorliegen konstitutiv äquivalenter Atome wird ohne Verlust der Eindeutigkeit des Verfahrens ein Vertreter der Äquivalenzklasse willkürlich ausgezeichnet und das Verfahren fortgesetzt.

Mit CANON lassen sich Atome in Molekülen und die zugehörigen Zeilen/Spalten von be-Matrizen eindeutig und unter Berücksichtigung konstitutiver Äquivalenzen beziffern und so Redundanzen vermeiden<sup>[21]</sup>. Für unsere Zwecke ist CANON der vom Chemical Abstracts Service verwendeten Morgan-Bezifferung<sup>[141]</sup> überlegen. Wegen der Erkennung konstitutiv äquivalenter Atome eignet sich CANON auch für die Ligandenbezifferung in der Stereochemie<sup>[140, 142]</sup> sowie für Aussagen über das Muster der chemischen Verschiebungen in NMR-Spektren. Dem System der CIP-Regeln<sup>[143]</sup>, das im Gegensatz zu CANON die Liganden einzeln vergleicht und nicht die gesamte chemische Konstitution berücksichtigt und sich auf Bindungsordnungen statt auf Koordinationszahlen stützt, ist die Liganden-Bezifferung durch CANON bei der computerunterstützten Lösung stereochemischer Probleme unter anderem deswegen vorzuziehen, weil die Atom- und Ligandenindizes nach CANON die konstitutive Chemie mit der Stereochemie verknüpfen<sup>[140]</sup>. Demnächst werden wir CANON interessierten Benutzern zur Verfügung stellen<sup>[144]</sup>.

#### 4.3.2. Erkennung und Korrelation von Substrukturen

Bei der computerunterstützten Lösung chemischer Probleme ist es häufig notwendig, die An- oder Abwesenheit bestimmter Substrukturen in generierten Molekülen zu fordern. Hierfür wurde der zentralatombezogene Substruktur-suchalgorithmus CABASS<sup>[145]</sup> und ein entsprechendes Computerprogramm entwickelt. CABASS ist unter anderem in IGOR2 eingebaut<sup>[115]</sup>.

Substruktursuchsysteme wie das DARC-System von Du-bois et al.<sup>[146]</sup> arbeiten häufig mit einem Screening-Verfahren und einer nachgeschalteten Atom-auf-Atom-Abbildung. Die ersten Verfahren zur Erstellung von Atom-auf-Atom-Abbildungen wurden vom Chemical Abstracts Service<sup>[147]</sup>

und von Sussenguth<sup>[148]</sup> publiziert. Das DARC-System verwendet beim Screening Sphärenfragmente, die in einer Datenbank als Bytefolge codiert und in einer Baumstruktur gespeichert werden. In dieser Baumstruktur werden die aus der gesuchten Substruktur erzeugten Fragmente gesucht. Die Methode der Sphärenfragmentierung wird auch von Lynch et al.<sup>[149]</sup> bei der Generierung von Screening-Verfahren für die Substruktursuche in Strukturdatensätzen angewendet.

Friedrich<sup>[150]</sup> analysierte die üblichen Methoden der Substruktursuche<sup>[146, 151]</sup> und fand, daß sie sich nicht für die von uns angestrebte Korrelation von Molekülen anhand von gemeinsamen Substrukturen eignen. Dies gilt vor allem, wenn große Datenmengen zu bearbeiten sind, denn der Speicherplatz- und Rechenzeitbedarf steigt mit dem Umfang des Problems steiler als polynomisch an<sup>[152]</sup>. Ihlenfeldt<sup>[153]</sup> entwickelte kürzlich ein Computerprogramm zur substruktur-bezogenen Korrelation von Molekülen, dem zwar die erwähnten Nachteile der gängigen Verfahren ebenfalls anhaften, das aber immerhin mäßige Datenmengen recht effizient bearbeitet.

Glücklicherweise sind Molekülgraphen von mäßiger Komplexität, so daß es gelingt, Algorithmen zur Korrelation von Molekülen über ihre Substrukturen zu konzipieren, deren Rechenzeit- und Speicherplatzbedarf mit der Größe und Zahl der Moleküle allenfalls polynomisch ( $\text{Grad} \leq 5$ ) ansteigt. Es sind Algorithmen, die auf der stufenweisen Fragmentierung der Moleküle und einer Einbettung der Stufenfragmente in ein Netzwerk von Vater-Sohn-Relationen beruhen, welches alle aufgelisteten Moleküle umfaßt. Auf der Grundlage dieses Konzepts implementierte J. Friedrich<sup>[150]</sup> CORREL, ein Substrukturkorrelationsprogramm für die bilaterale Syntheseplanung, aus dem das routinemäßig einsetzbare Programm CORREL 2<sup>[154]</sup> hervorging. Für die bilaterale Syntheseplanung wurde kürzlich CORREL-S implementiert<sup>[155]</sup>, das auf einer Auswahl synthesesrelevanter Substrukturen basiert (siehe später).

Für die Dokumentation und den Vergleich von Peptidsequenz ist eine Sondervision von CORREL implementiert worden<sup>[156]</sup>, die nun für die Einbeziehung von Nucleotidsequenzen und ihre Korrelation mit Peptidsequenzen ausgebaut wird.

Das Grundkonzept von CORREL, dessen Publikation<sup>[150]</sup> auch den Quellcode enthält, stimulierte die Implementierung etlicher analog aufgebauter Substruktursuch- und -korrelationsprogramme wie RESY<sup>[157]</sup>, KOWIST<sup>[158]</sup>, HTSS<sup>[159]</sup>, S4<sup>[151, 160]</sup> und des Programms von Klopman<sup>[161, 162]</sup>. Letzteres Programm sowie RESY und KOWIST sind besonders für Struktur-Wirkungs-Studien geeignet<sup>[158]</sup>.

#### Strategische Bindungen und synthesesrelevante Substrukturen

Zur bilateralen Syntheseplanung müssen in Datenbanken mit für gegebene Syntheseeziele verfügbaren Ausgangsstoffen anhand der jeweils größten gemeinsamen Substruktur geeignete Kandidaten gefunden werden. Hierfür wurde CORREL-S entwickelt<sup>[89, 155]</sup>.

Bei der Erzeugung einer Datei synthesesrelevanter Substrukturen reduziert der nachstehende Satz von Regeln über zu brechende Bindungen den Umfang der Substrukturdatei

in sinnvoller Weise<sup>[163]</sup>. Hierfür wurden Coreys Regeln über strategische Bindungen modifiziert, hierarchisch geordnet und mit einem Abbruchkriterium versehen; das Verfahren wird abgebrochen, wenn alle nichtaromatischen Ringe gespalten sind.

1. Die Bindung gehört zu keinem carbocyclischen aromatischen System. Für Heteroarene gilt dies nicht, da diese oft während der Synthese erzeugt werden.
2. Bindungen werden um so höher bewertet, je stärker ihre Spaltung die Zahl der Ringe reduziert; das Abzählen der Ringe erfolgt dabei mit dem Programm CANON<sup>[139]</sup> nach der Frèrejaques-Formel<sup>[44]</sup>.
3. Existieren mehrere Bindungen, die der Regel 2 entsprechen, so werden diejenigen bevorzugt, deren Bruch die wenigsten neuen Ringe erzeugt.
4. Bindungen, an denen Heteroatome beteiligt sind, werden als strategisch betrachtet.
5. Eine Bindung, die bezüglich anderer Ringe exo-ständig ist, wird bevorzugt.
6. An der Bindung ist mindestens ein Kohlenstoffatom mit mehr als zwei Nichtwasserstoffatomen als Nachbarn beteiligt. Dies entspricht im wesentlichen Regel 5, wenn polycyclische Strukturen vorliegen. Regel 6 kommt allerdings zum Tragen, wenn nur noch einzelne isolierte Ringe im Molekül vorhanden sind.
7. Die Bindung ist eine Mehrfachbindung.
8. Die Bindung ist unmittelbare oder übernächste Nachbarbindung einer Bindung gemäß Regel 4 oder 7.

Mit Hilfe dieser Regeln können sowohl einfache als auch komplizierte Ringsysteme bearbeitet werden und dank ihrer Hierarchie kann die Zahl der strategischen Bindungen beschränkt und in eine Prioritätenfolge eingeordnet werden, nach der sie gebrochen werden müssen, um möglichst schnell zu synthesesrelevanten Substrukturen verfügbarer Vorstufen zu gelangen.

#### 4.3.3. Stöchiometrischer Ausgleich unvollständiger Reaktionsgleichungen

Die Coprodukte des Synthesziels, sein stöchiometrisches Komplement im Ziel-EM, werden mit dem Computerprogramm STOECH<sup>[164]</sup> ermittelt. Dieses basiert auf einem Matrizenformalismus, der bereits in den sechziger Jahren bei der Aufklärung des Reaktionsmechanismus<sup>[165]</sup> der Vierkomponentenkondensation (4CC) mit Sekundärreaktionen gekoppelte  $\alpha$ -Additionen von Iminium-Ionen und Anionen an Isocyanide<sup>[166]</sup>, Ugi-Reaktion<sup>[167]</sup> verwendet wurde. Der stöchiometrische Ausgleich unvollständiger Reaktionsgleichungen<sup>[168]</sup> ist auch im Hinblick auf eine systematische Reaktionsdokumentation mit hierarchischer Klassifizierung der Reaktionen aktuell geworden (siehe Abschnitt 4.3.5).

#### 4.3.4. Ermittlung der minimalen chemischen Distanz (CD) und der zugehörigen Atom-auf-Atom-Abbildungen

Die laufende Bestimmung der CD zwischen isomeren EM ist Voraussetzung für die Optimierung von Reaktionsnetzwerken<sup>[21, 100, 169]</sup>. Die zugehörigen Atom-auf-Atom-Ab-

bildungen werden für die hierarchische Reaktionsdokumentation<sup>[79, 100, 112]</sup> benötigt. Da das auf einem heuristischen „Branch-and-bound“-Algorithmus<sup>[170]</sup> beruhende Programm PMCD<sup>[105]</sup> zur näherungsweise Ermittlung der Minima der CD nicht immer befriedigende Resultate liefert, wurde das Programm PEMCD<sup>[106]</sup> zur exakten Ermittlung der CD-Minima implementiert, das derzeit Reaktionskerne mit maximal 16 zusammenhängenden reaktiven Zentren behandeln kann. Daher entwickeln wir gegenwärtig ein leistungsfähigeres Nachfolgeprogramm für PEMD.

Die CD von EM mit bis zu einhundert reaktiven Zentren läßt sich mit einem kürzlich von Fontain<sup>[107]</sup> implementierten Computerprogramm ermitteln, das auf „Genetic Algorithms“<sup>[171]</sup> beruht. Allerdings liefert dieses CD-Minimierungsprogramm nicht immer alle Atom-auf-Atom-Abbildungen, die den CD-Minima entsprechen.

#### 4.3.5. Reaktionsdokumentation

Die handelsüblichen Reaktionsdokumentationssysteme, die Reaktionen lediglich durch ihre Edukte und Produkte repräsentieren, sind keine befriedigende Lösung des Reaktionsdokumentationsproblems<sup>[172]</sup>. Unter den verfügbaren Reaktionsdokumentationssystemen ragte bis vor kurzem ORAC<sup>[173]</sup> durch seine wohlgedachte Struktur und Effizienz hervor, aber es ist leider nicht mehr verfügbar. Da die Unzulänglichkeiten der traditionellen Reaktionsdokumentation bekannt sind und die Reaktionsdokumentation zu den wichtigsten noch nicht befriedigend gelösten Problemen der Computerchemie gehört, gibt es auf diesem Gebiet viele neue Ansätze<sup>[174, 175]</sup>. Daher wurde im Rahmen des Münchner Projekts auch ein hierarchisch geordnetes System geschaffen<sup>[79, 112]</sup>, das sich auf Reaktionsvorgänge als solche bezieht; bisher hat sich dieses System noch nicht durchsetzen können. Der Hauptgrund dafür ist, daß die zu speichernden Reaktionen stöchiometrisch ausgeglichen sein müssen, d. h. sie müssen als Isomerisierungen von EM darstellbar sein. Chemische Reaktionen werden in der Literatur aber oft in „verstümmelter“ Form unter Weglassung von Reaktionsteilnehmern beschrieben und müssen daher stöchiometrisch komplettiert werden. Dies ist nun mit dem erst seit kurzem verfügbaren Programm STOECH (siehe Abschnitt 4.3.3) möglich.

#### 4.3.6. Graphische Ausgabe von Ergebnissen

Da die Ergebnisse der computerunterstützten deduktiven Lösung chemischer Probleme in Form von Konnektivitäts- oder be-Matrizen anfallen, wurde das graphische Ausgabeprogramm MDRAW<sup>[176]</sup> entwickelt, das diese Matrizen in die anschaulicheren und dem Chemiker vertrauteren graphischen Konstitutionsformeln umwandelt. ARGOS, ein ähnliches von den Coautoren dieses Beitrags J. Bauer und E. Fontain implementiertes Programm, ist Bestandteil der Benutzeroberflächen von IGOR und RAIN und wird unter anderem seit mehreren Jahren vom Beilstein-Institut routinemäßig bei der Erzeugung von Konstitutionsformeln aus Konnektivitätslisten eingesetzt.

## 4.4. Die Mehrzweckprogramme IGOR und RAIN<sup>[\*]</sup>

### 4.4.1. Formale Reaktionsgeneratoren

Die Fundamentalgleichung  $B + R = E$  des DU-Modells kann von einer be-Matrix, z.B.  $B$ , aus gelöst werden, indem man diejenigen Paare  $(R, E)$  sucht, welche unter den geltenden Randbedingungen die Gleichung  $B + R = E$  erfüllen. Diese Lösungen nennen wir die b-Lösungen. Sie werden von Reaktionsgeneratoren (RG) vom Typ RGB erzeugt<sup>[169, 177, 178]</sup>. Die Gleichung  $B + R = E$  kann aber auch von einer gegebenen r-Matrix  $R$  aus gelöst werden. Diese r-Lösungen entsprechen Paaren  $(B, E)$ , für die  $B + R = E$  gilt und werden von Reaktionsgeneratoren des Typs RGR produziert<sup>[114–116]</sup>.

Den beiden komplementären Lösungstypen der Gleichung  $B + R = E$  und den zugehörigen RG entsprechen die beiden sich ergänzenden Grundtypen von Computerprogrammen zur Lösung chemischer Probleme auf der Basis des DU-Modells. Die jetzige Version IGOR2<sup>[115]</sup> des Computerprogramms IGOR<sup>[114–116]</sup> enthält einen RGR als Kernstück. Das seit kurzem ebenfalls allgemein verfügbare RAIN<sup>[60–62, 169, 177, 178]</sup> ist mit einem RGB als „Triebwerk“ ausgestattet.

Die RG von IGOR und RAIN stützen sich auf Übergangstabellen (Transition Table, TT)<sup>[62, 63, 177, 178]</sup>. Jedem zu berücksichtigenden chemischen Element wird eine Standard-TT zugeordnet oder eine TT, die ad hoc vom Benutzer für den gegebenen Fall definiert wird. In einer TT sind die jeweils erlaubten Valenzschemata und deren Übergänge verzeichnet (siehe Abb. 2).

C	$\begin{array}{c}   \\ -C- \\   \end{array}$	$\begin{array}{c} \diagup \\ C= \\ \diagdown \end{array}$	$\begin{array}{c} \diagup \\ C^{\oplus} \\ \diagdown \end{array}$	$=C=$	$-C\equiv$
$\begin{array}{c}   \\ -C- \\   \end{array}$	•	•	•	o	o
$\begin{array}{c} \diagup \\ C= \\ \diagdown \end{array}$	•	•	•	•	•
$\begin{array}{c} \diagup \\ C^{\oplus} \\ \diagdown \end{array}$	•	•	o	o	o
$=C=$	o	•	o	•	•
$-C\equiv$	o	•	o	•	•

N	$\begin{array}{c} \diagup \\ N- \\ \diagdown \end{array}$	$\begin{array}{c} \diagup \\ N^{\oplus} \\ \diagdown \end{array}$	$\begin{array}{c} \diagup \\ N^{\oplus} \\ \diagdown \end{array}$
$\begin{array}{c} \diagup \\ N- \\ \diagdown \end{array}$	•	•	•
$\begin{array}{c} \diagup \\ N^{\oplus} \\ \diagdown \end{array}$	•	o	o
$\begin{array}{c} \diagup \\ N^{\oplus} \\ \diagdown \end{array}$	•	o	o

H	H-
H-	•

o	$\begin{array}{c} \diagup \\ O- \\ \diagdown \end{array}$	$\begin{array}{c} \diagup \\ O= \\ \diagdown \end{array}$
$\begin{array}{c} \diagup \\ O- \\ \diagdown \end{array}$	•	•
$\begin{array}{c} \diagup \\ O= \\ \diagdown \end{array}$	•	•

Abb. 2. Beispiele für Übergangstabellen. • = Übergang erlaubt, o = Übergang verboten.

Die Computerprogramme IGOR und RAIN enthalten streng formale Reaktionsgeneratoren und sind daher nicht für die Lösung eines speziellen Problems, z.B. der retrosynthetischen Analyse, zuständig. Beide können unter Verwendung ergänzender Computerprogramme eine breite Palette chemischer Fragestellungen bearbeiten, so daß man sie als Mehrzweckprogramme zur Lösung chemischer Probleme bezeichnen kann. Wegen ihrer Unabhängigkeit von detaillierten empirischen Daten generieren sie Problemlösungen, die der bereits bekannten Chemie angehören oder auch völlig präzedenzlos sein können.

Der erste TT-gestützte RG (TRG) war der TRGR von IGOR, der die Gleichung  $B + R = E$  ausgehend von einer

r-Matrix  $R$  und einer Liste der TT der relevanten chemischen Elemente löst. Für jedes Zeilen-Spalten-Paar der potentiellen Lösungen  $(B, E)$  wählt der Benutzer die ihm geeignet erscheinenden Sätze von TT. Der TRGR kontrolliert, ob die Eintragungen der TT mit  $R$  kompatibel sind. So resultiert ein reduzierter Satz von TT, die verwendet werden, um die be-Matrix  $B$  Zeile für Zeile, Spalte für Spalte zu generieren. Aus den Zeilen/Spalten von  $B$  werden mit  $R$  die Zeilen/Spalten von  $E$  erhalten. Nun werden anhand der TT alle Zeilen-Spalten-Kombinationen  $(B, E)$  verifiziert. Die Neuheit jedes so fertiggestellten Paares  $(B, E)$  wird mit CANON<sup>[139]</sup> überprüft, so daß die Redundanzfreiheit der Paare sichergestellt ist. Anschließend wird mit CABASS<sup>[145]</sup> analysiert, ob die Paare verbotene oder geforderte Substrukturen enthalten. Die akzeptablen Lösungen werden schließlich graphisch am Bildschirm dargestellt oder ausgedruckt.

IGOR erzeugt mit seinem TRGR aus einer irreduziblen r-Matrix  $R$  (siehe Abschnitt 3.2.2), die einem Elektronenverschiebungsschema entspricht, Paare von Edukten und Produkten, die durch die von  $R$  repräsentierte Elektronenverschiebung ineinander umgewandelt werden, d.h. es nutzt angenommene Elektronenverschiebungsschemata, um chemische Reaktionen zu erfinden. Diese Elektronenverschiebungsschemata können in Form ihrer zugehörigen r-Matrizen durch ein von IGOR unabhängiges Computerprogramm erzeugt werden<sup>[179]</sup>. Die generierten Reaktionen werden anschließend gemäß der hierarchischen Klassifizierung chemischer Reaktionen einem interaktiven Auswahlverfahren unterworfen<sup>[114]</sup>.

Gibt man eine Nullmatrix  $R = \{\}$  als r-Matrix in das Programm IGOR ein, so produziert IGOR im Strukturgenerierungsmodus be-Matrizen  $B (= E)$ , wegen  $R = \{\}$  von Molekülen oder EM. Schränkt man IGOR durch die Angabe molekularer Strukturmerkmale ein, so werden nur Moleküle mit den eingegebenen Merkmalen erhalten<sup>[114, 177, 180, 181]</sup>.

Ein RGB erzeugt von einem gegebenen EM aus alle EM, in die ersteres direkt umwandelbar ist. Da RGB auch  $-R$  einbezieht, erzeugt er auch diejenigen EM, aus denen das eingegebene EM erhalten werden kann. Als Teil eines retrograden Syntheseprogramms kann ein RGB die gleiche Rolle spielen wie eine Reaktionsbibliothek in Kombination mit einem Strukturerkennungsmodul, d.h. der RGB findet vom Synthesziel aus dessen Vorstufen und dann deren Vorstufen usw.

Ein TRGB arbeitet in zwei Schritten: Zuerst werden anhand der Zeilen/Spalten von  $B$  unter Berücksichtigung der TT die erlaubten Valenzschemata für diejenigen chemischen Elemente ermittelt, die den Zeilen/Spalten von  $E$  zugeordnet sind. Dann werden alle  $E$  generiert, für die  $B + R = E$  unter den geltenden Randbedingungen zutrifft.

Das Programm RAIN erzeugt mit einem TRGB<sup>[60, 63, 178]</sup> aus einem durch seine be-Matrix  $B$  repräsentierten EM<sub>B</sub> isomere EM, die entweder Produkten aus dem EM<sub>B</sub> oder Edukten für das EM<sub>B</sub> entsprechen, und kann aus zwei EM gleichzeitig Reaktionsbäume, d.h. Folgen isomerer EM, generieren. Durch Nutzen der „geometrischen“ Eigenschaften der Familie aller isomeren EM werden die beiden von EM<sub>B</sub> und EM<sub>E</sub> aus entstehenden Bäume so dirigiert, daß sie möglichst rasch zu einem Netzwerk von Reaktionswegen zusammenwachsen; dazu werden sie anhand der CD der intermediären EM zu EM<sub>B</sub> oder EM<sub>E</sub> zusammengeführt<sup>[21, 100, 107, 182, 183]</sup>. Die bilaterale Generierung von Re-

[\*] IGOR und RAIN sind auf Anfrage von den Autoren erhältlich.

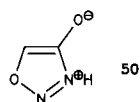
aktionspfaden endet jeweils, sobald ein EM von beiden Seiten aus erreicht ist. RAIN erzeugt so Reaktionsnetzwerke, welche isomere EM verbinden<sup>[60, 63, 88, 177, 184, 185]</sup>. Die Art des entstehenden Reaktionsnetzwerks läßt sich vom Benutzer durch interaktive Festlegung von unteren und oberen Schranken für die Kennzahlen der Reaktionswege (Zahl und Strukturmerkmale der Zwischenstufen, Zahl der verschobenen Valenzelektronen pro Reaktionsschritt etc.) beeinflussen<sup>[21]</sup>. Je nachdem, ob nur Valenzschemata stabiler Verbindungen oder auch Valenzschemata labiler Zwischenstufen in den TT zugelassen werden, enthalten die von RAIN erzeugten Netzwerke nur stabile Moleküle oder auch kurzlebige Intermediate.

RAIN kann somit bei der mono- und der bilateralen Syntheseplanung<sup>[84, 88, 89]</sup>, der Reaktionsplanung und -prognose sowie der Aufklärung von Biosynthesen und komplexen Reaktionsmechanismen eingesetzt werden.

Vor wenigen Jahren hat Valdés-Pérez<sup>[186]</sup> das System MECHEM vorgestellt, das durch symbolisches Rechnen ebenfalls Netzwerke von Reaktionswegen generiert. Dieses Programm berücksichtigt jedoch lediglich die Summenformeln und nicht die Konstitutionsformeln der Moleküle.

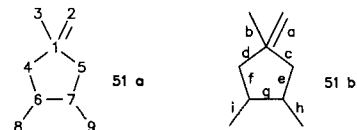
#### 4.4.2. Strukturgenerierung mit IGOR und RAIN

Im Strukturgenerierungsmodus erzeugen IGOR und RAIN alle Konstitutionsformeln, die den jeweils zu befolgenden Bedingungen gehorchen. So lassen sich beispielsweise alle 23 EM der Formel  $(\text{CH})_8$  und alle 18 1,3-Dipole<sup>[187]</sup> aus den Elementen C, N und O produzieren<sup>[114]</sup>. Um zu zeigen, wie sehr die Zahl der Isomere ansteigt, wenn man in Kohlenwasserstoffen die C- und H-Atome teilweise durch Heteroatome ersetzt, sei die mit IGOR und RAIN ermittelte Zahl der Konstitutionsformeln zur Summenformel  $\text{C}_2\text{H}_2\text{N}_2\text{O}_2$  (nichtionisch; keine Dreifachbindungen und kumulierten Doppelbindungen in Ringen; Ammonium- und Iminium- $\text{N}^+$ , Amid- $\text{N}^-$  und  $\text{O}^-$  als geladene Atome zugelassen) von 1806 mit der zur Summenformel  $\text{C}_4\text{H}_4$  von 6 verglichen. Bei der Strukturgenerierung mit RAIN ging man von je zwei freien Atomen der Elemente C, H, N und O unter den genannten Bedingungen aus und spannte das Netzwerk aller Konstitutionsisomere auf. Zum Vergleich wurden außerdem diejenigen nichtionischen  $\text{C}_2\text{H}_2\text{N}_2\text{O}_2$ -Isomere erzeugt, die einen fünfgliedrigen Ring enthalten und je eine positive und negative Formalladung tragen. Es resultieren das Sydnon **50** und seine 51 Analoga.

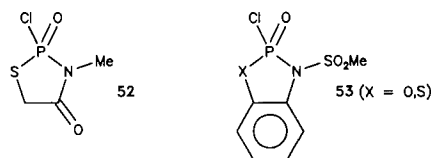


Die heute gebräuchlichen Reagentien zur Oligonucleotidsynthese sind Derivate der phosphorigen Säure<sup>[188]</sup>, weil die alternativ in Frage kommenden Phosphorsäurederivate im allgemeinen nicht genügend rasch mit Nucleosidderivaten reagieren. Fünfgliedrige cyclische Phosphorsäurederivate setzten sich jedoch mit Nucleophilen oder Apicophilen<sup>[189]</sup>  $10^5$  bis  $10^7$ mal rascher um als ihre acyclischen Analoga<sup>[189, 190]</sup>.

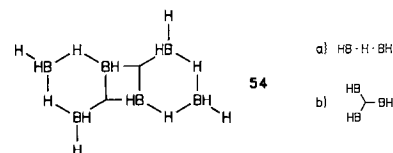
Seit 1973 haben die Arbeitskreise von F. Ramirez und I. Ugi sich bemüht, diese Tatsache präparativ zu nutzen, indem sie hochreaktive fünfgliedrige cyclische Phosphorylierungsmittel mit  $\text{P}^V$  für Oligonucleotidsynthesen zu entwickeln versuchten<sup>[191]</sup>, die gegenüber den üblichen  $\text{P}^{\text{III}}$ -Reagentien einige Vorteile hätten<sup>[192]</sup>. Nach einigen Anfangserfolgen<sup>[193]</sup> stagnierte die Entwicklung, weil die erforderlichen neuen Ideen fehlten. Erst nachdem Ugi et al. sich entschlossen<sup>[180]</sup>, mit IGOR die Gesamtheit der in Frage kommenden, gemäß **51 a** und **51 b** definierten  $\text{P}^V$ -Verbindungen Revue passieren zu lassen, wurde rascher Fortschritt erzielt.



Für **51 a** wurden folgende Atomarten zugelassen: 1 = P; 2 = O; 3 = X; 4, 5 = O, N, S; 6 = C, N, O, S; 7 =  $\text{sp}^2$ -C; 8, 9 = C, Cl, N, O, S. H-Atome wurden gemäß den Valenzregeln aufgefüllt. Als Bindungen wurden für **51 b** erlaubt: a: Doppelbindung; b, c, d: Einfachbindung; e, f, g: Einfach- oder Doppelbindung; h, i: Einfachbindung oder keine Bindung. Mit diesen einschränkenden Vorgaben lieferte IGOR 278 Formeln gemäß **51**<sup>[180]</sup>, von denen fünf als Kandidaten für präparative Studien ausgewählt wurden<sup>[194]</sup>. Diese haben bereits zu einigen für Oligonucleotidsynthesen brauchbaren Phosphorylierungsmitteln wie **52**<sup>[192]</sup> und **53**<sup>[195]</sup> geführt.



Lipscomb postulierte anhand der von ihm erkannten strukturellen Gesetzmäßigkeiten von Borwasserstoffverbindungen für  $\text{B}_6\text{H}_{14}$  die Struktur **54** mit offenen (a) und geschlossenen Dreizentrenbindungen (b)<sup>[196]</sup>. Laut neueren NMR-Daten enthält  $\text{B}_6\text{H}_{14}$  zwei Arten von B-Atomen (in erster Sphäre)<sup>[197]</sup>. RAIN generierte davon ausgehend für  $\text{B}_6\text{H}_{14}$  die Strukturvorschläge von Abbildung 3.



Eine besonders vielversprechende Anwendungsmöglichkeit von IGOR und RAIN ist die computerunterstützte Strukturermittlung à la DENDRAL anhand instrumentell-analytischer Daten. Letztere können bei der Strukturgenerierung als einschränkende Bedingungen dienen. Dann erhält man alle Konstitutionsformeln, die mit den vorliegenden Daten kompatibel sind.

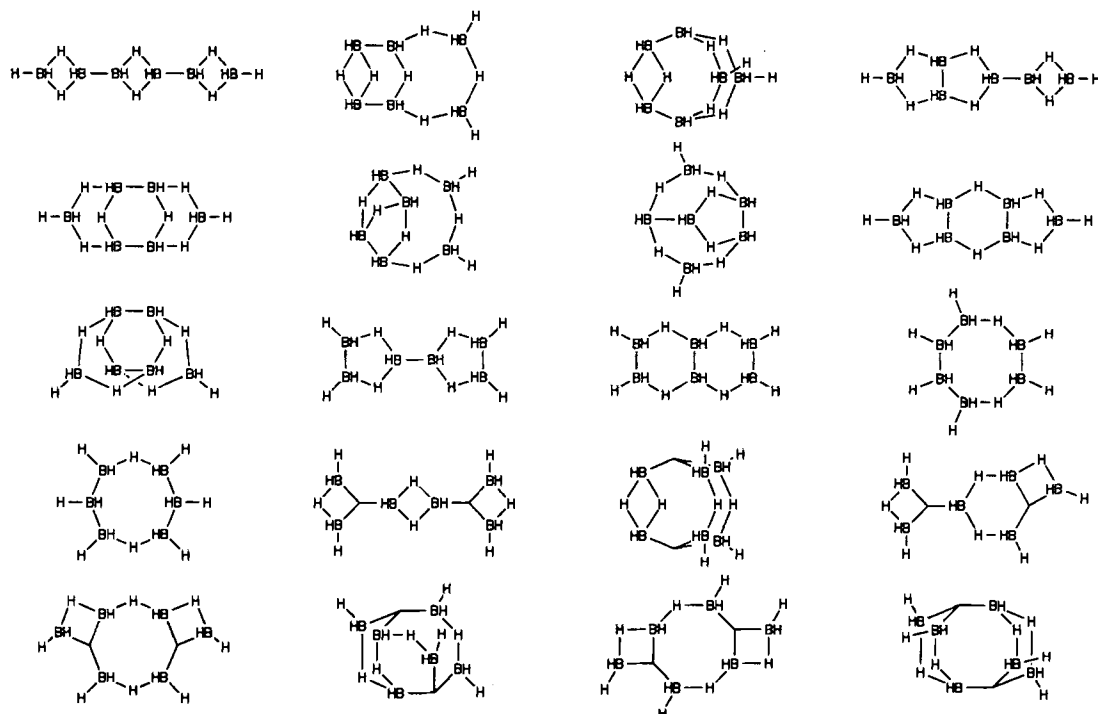
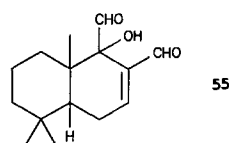
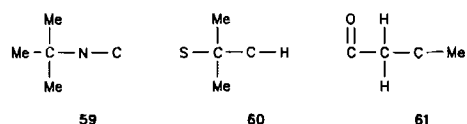
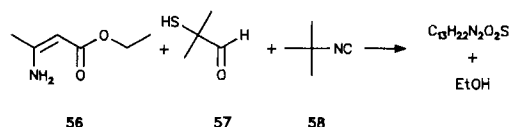


Abb. 3. Mit RAIN generierte Strukturvorschläge für  $B_6H_{14}$ .

Die graphentheoretische Darstellung der chemischen Konstitution hat sich bereits im Rahmen des DENDRAL-Projekts<sup>[23]</sup> als mathematische Grundlage der Computerprogramme CONGEN<sup>[198]</sup> und GENOA<sup>[199]</sup> bewährt, ebenso wie später bei den entsprechenden Programmen RASTR<sup>[200]</sup>, MOLGRAPH<sup>[201]</sup>, ESSESA<sup>[202]</sup> und GENI<sup>[203]</sup>. Mit diesen Programmen lassen sich Familien von Konstitutionsformeln erzeugen. GENOA war das erste leistungsfähige Programm dieser Kategorie. Die computerunterstützte Konstitutionsaufklärung von Walbarganal **55** mit ihm gehört zu den eindrucksvollsten Anwendungsmöglichkeiten des DENDRAL-Projekts<sup>[24]</sup>.



RAIN erzeugt denselben Satz von 42 Konstitutionsvorschlägen wie GENOA, wenn die Bruttoformel  $C_{15}H_{22}O_3$  von **55** und die aufgrund spektroskopischer Daten geforderten und verbotenen Substrukturen berücksichtigt werden.



Aus Ethyl- $\beta$ -aminocrotonat **56**,  $\alpha$ -Mercaptoisobutyraldehyd **57** und *tert*-Butylisocyanid **58** lässt sich eine Substanz der Bruttoformel  $C_{13}H_{22}N_2O_2S$  synthetisieren [Gl. (18)]<sup>[204]</sup>. Unter der Annahme, daß die Substrukturen **59–61** der Ausgangsstoffe auch im Produkt vorhanden sind, erzeugt RAIN die Strukturvorschläge in Abbildung 4.

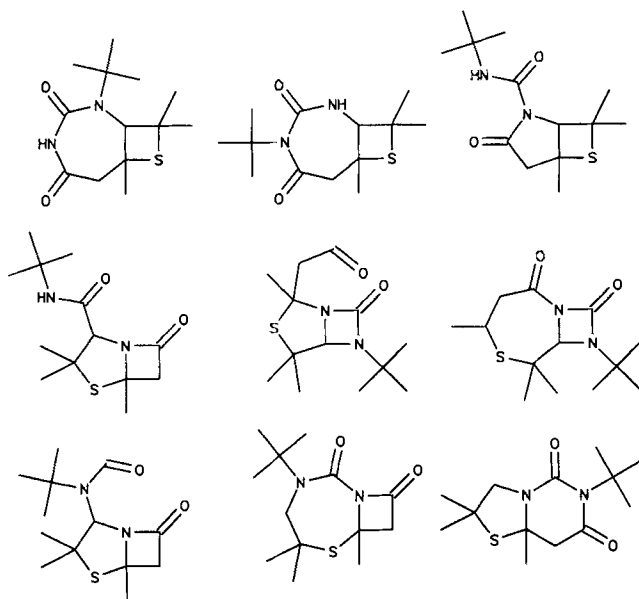


Abb. 4. Mit RAIN erzeugte Strukturvorschläge für das Produkt von Reaktion (18).

Im Routinebetrieb erkennt RAIN automatisch die chemische Äquivalenz oder Identität von prototropen Tautomeren bzw. mesomeren Grenzstrukturen. Für die Nucleobase Cytosin beispielsweise wurden die Tautomere und mesomeren Grenzstrukturen in Abbildung 5 erhalten.

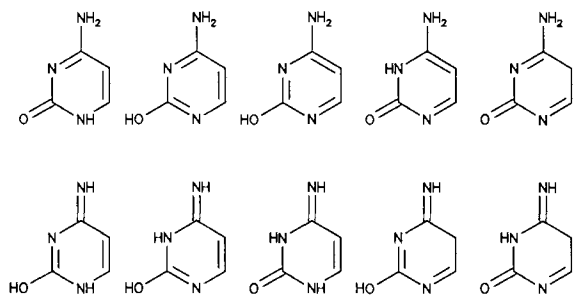


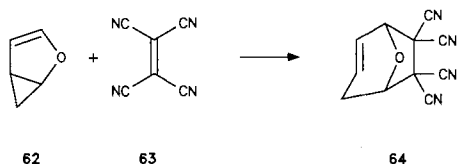
Abb. 5. Mit RAIN ermittelte Tautomere und mesomere Grenzstrukturen von Cytosin.

#### 4.4.3. Reaktionsgenerierung mit IGOR

Vor wenigen Jahren hat Barton<sup>[205]</sup> eine sehr bemerkenswerte Arbeit über die systematische Suche nach neuen Reaktionen ohne Computerunterstützung publiziert. Wir beschreiben im folgenden die computerunterstützte Entdeckung neuer Reaktionen.

Vor rund zehn Jahren sollte R. Herges in seiner Doktorarbeit<sup>[118]</sup> mit dem damals noch in Entwicklung befindlichen Computerprogramm IGOR<sup>[116]</sup> präzedenzlose Reaktionen finden und im Laboratorium realisieren. Die dabei gesammelten Erfahrungen und Anregungen haben sowohl die Effizienz als auch den Benutzerkomfort von IGOR erhöht und dazu beigetragen, das Programm zur Anwendungsreife zu führen.

Unter anderem wurde festgestellt, daß der Anteil des Chemikers an der computerunterstützten Entdeckung neuer Reaktionen stark variieren kann<sup>[206]</sup>. Bei der Entwicklung der Cycloadditionen von Homodienen<sup>[207]</sup> gab der Computer nur einen Denkanstoß. Die systematische Untersuchung der r-Klasse der pericyclischen Sechs-Zentren-Sechs-Elektronen-Reaktionen zeigte, daß in dieser 6,6-r-Klasse von 13 Basisreaktionen nur **21** → **22** (Schema 2) mit C-Atomen im Reaktionskern wenig untersucht, aber gleichzeitig plausibel war. Als zu realisierendes Beispiel wurde die Reaktion **62** + **63** → **64** unter heuristischen Gesichtspunkten (Ringspannung, Polarisierung, Molekülgeometrie etc.) ausgewählt und erfolgreich experimentell durchgeführt<sup>[114, 118]</sup>.



Hierbei machten wir die Erfahrung, daß es sehr schwierig ist, die Präzedenzlosigkeit einer Reaktion anhand der Fachliteratur sicherzustellen. Es existieren zwar Strukturdatenbanken (CAS-online, Beilstein-online), in denen fast alle jemals synthetisierten Verbindungen abgespeichert und bequem recherchierbar sind. Die kommerziell verfügbaren Reaktionsdatenbanken dagegen sind nicht hinreichend strukturiert und bei weitem nicht vollständig. Die Präzedenz oder Präzedenzlosigkeit einer Reaktion muß daher auf dem Umweg über alle „potentiellen“ Edukte oder Produkte überprüft werden. Einer von uns (I.U.) wurde von H. Prinzbach in der

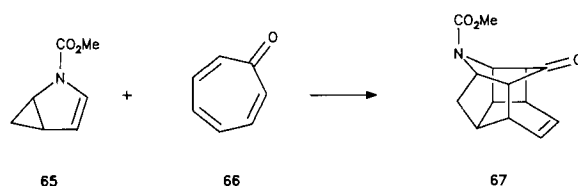
Folgezeit darauf aufmerksam gemacht, daß Fowler<sup>[208]</sup> bereits 1971 *N*-Methoxycarbonylhomopyrrol **65** an Acetylenedicarbonsäuredimethylester cycloaddiert hatte. Diese Reaktion steht unserer nahe, denn die Reaktionsteilnehmer **62** und **65** ähneln einander.

Die Retro-En-Reaktion **25** → **26** (Schema 2) ist ein weiteres Beispiel für eine mit IGOR anhand der hierarchischen Klassifizierung gefundenen und experimentell realisierten Reaktion. Als zu übertragende Gruppe wurde hier Wasserstoff gewählt, da H-Übertragungen besonders leicht stattfinden; als  $sp^2$ -Zentrum erschien ein Carbonylsauerstoffatom günstig. Alle übrigen Atome durften C oder O sein gemäß **23** → **24**. Die chemisch sinnvolle zusätzliche Bedingung, daß  $CO_2$  das extrudierte Molekül sein sollte, wurde festgelegt. Dies führte zu vier Reaktionstypen. Von diesen waren drei bereits bekannt, einer jedoch nicht, nämlich die Pyrolyse von  $\alpha$ -Formyloxyketonen **25** zu **26**. Diese wurde im Laboratorium anhand dreier Beispiele erfolgreich erprobt<sup>[114, 118]</sup> (siehe Schema 2).

Der Neuheitsgrad auch dieser Reaktion ist mäßig, denn wir hatten die rb-Unterklasse der Retro-En-Reaktionen vorgegeben und konnten somit als „neue Reaktion“ bestenfalls eine weitere Retro-En-Reaktion erwarten. Die Reaktion **25** → **26** ist dennoch interessant, neuartig und sogar von präparativem Wert, weil sie unter anderem bei der Synthese cyclischer Ketone durch Acyloinkondensation<sup>[209]</sup> der üblichen Reduktion der Acyloine mit  $Zn/HCl$ <sup>[210]</sup> überlegen ist. Diese Reaktion hätte man selbstverständlich auch ohne Computerhilfe finden können. Offenkundig und erstaunlicherweise wurde hier eine naheliegende Reaktion auf dem Gebiet der intensiv bearbeiteten Retro-En-Reaktionen lange übersehen. Dies zeigt, daß man bei der systematischen Suche nach neuen Reaktionen mit Computerhilfe mehr erreichen kann als ohne sie.

Denkbare Reaktionen, die bis zum Niveau der ra-Unterklassen neuartig sind, gibt es jedoch bereits ab der CD-Klasse 8 (siehe Abschnitt 3.2.2) in großer Menge<sup>[114]</sup>.

Vor einigen Jahren wurde mit IGOR eine systematische Suche nach pericyclischen Reaktionen durchgeführt, zu deren Basisreaktion es keine literaturbekannte Präzedenz gibt. Die CD-Klasse 20 (Reaktionen mit fünf verschobenen Elektronenpaaren) war die erste, in der eine neue experimentell realisierbare Reaktion (**65** + **66** → **67**) gefunden wurde<sup>[88, 117]</sup>.



Weitere von Herges mit IGOR gefundene Reaktionen sind eine neue Carben-Umlagerung<sup>[211]</sup> sowie eine Methode zur Synthese von Isocyaniden mit elektronenziehenden Gruppen<sup>[206]</sup>.

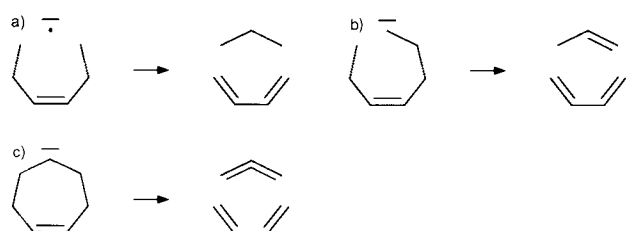
Neben dem systematischen Screening von Reaktionsklassen nach neuartigen Reaktionen ist es auch möglich, systematisch nach Reaktionen zur Synthese bestimmter Verbindungsklassen zu suchen. Herges und Hook<sup>[212, 213]</sup> haben die pericyclischen Sieben-Zentren-Acht-Elektronen-Reak-

tionen auf eine neue Methode zur Synthese von 1,3-Dienen hin untersucht. Randbedingungen für diese Suche lieferten die irreduzible r-Matrix und die valenzchemischen Eigenschaften der an der Reaktion beteiligten Atome (Schema 10).



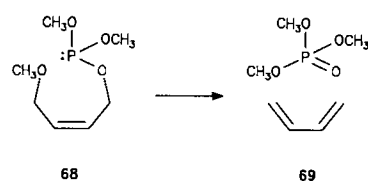
Schema 10. Pericyclische Sieben-Zentren-Acht-Elektronen-Reaktionen.

Von den insgesamt generierten 72 Basisreaktionen (unter Zulassung von Dreifachbindungen werden 470 Basisreaktionen erzeugt) sind lediglich drei (Schema 11) für die Synthese von 1,3-Dienen geeignet (es wird jeweils ein Butadien-Fragment erzeugt). Nur Reaktion a) ist mit einigen Hetero-

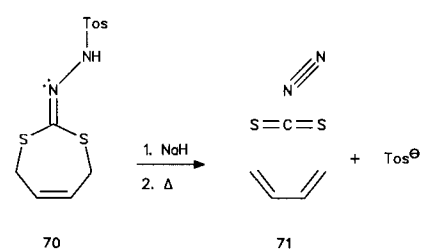


Schema 11. Potentiell zur Synthese von 1,3-Dienen geeignete Reaktionen gemäß Schema 10. Der Pfeil markiert ein Atom mit freiem Elektronenpaar (als Strich dargestellt).

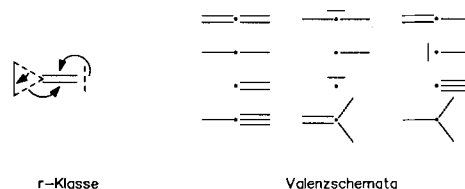
varianten bekannt. Ein Beispiel ist die Reduktion von 1,4-Dichlor-2-buten mit Zink. Für die Basisreaktionen b) und c) gibt es unseres Wissens keine literaturbekannten Beispiele. Die Variante **68** → **69** der Basisreaktion b) wurde kürzlich experimentell verifiziert. Sie liefert in der Gasphase



bei 0.4 Torr und 350 °C Butadien in 40% Ausbeute und ist somit nicht von präparativem Interesse, stellt aber immerhin einen neuen Typ einer 1,4-Eliminierung dar. Auch für die Basisreaktion c) wurde ein Beispiel, **70** → **71**, experimentell realisiert<sup>[212]</sup>.

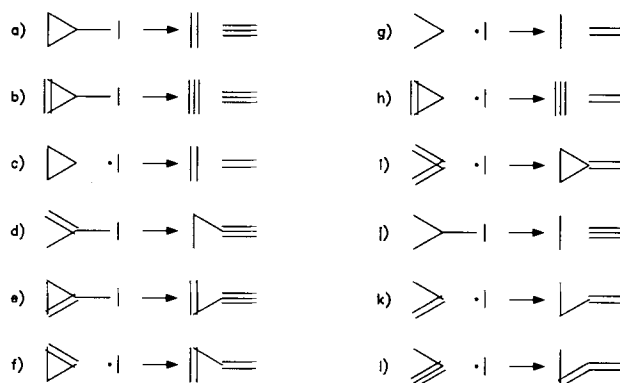


Es können jedoch nicht nur pericyclische Reaktionen generiert werden, sondern auch Reaktionen aus anderen Bereichen der Organischen Chemie, wie das folgende Beispiel der Suche nach einer neuartigen Carbenreaktion (Schema 12) zeigt<sup>[214]</sup>.



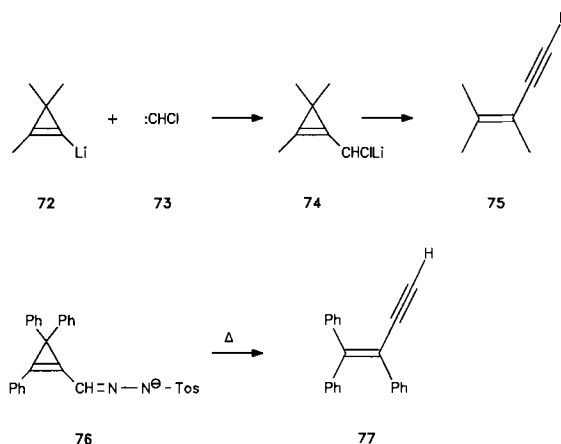
Schema 12. Carbenreaktionen.

Für die ersten vier Basisreaktionen von Schema 13 gibt es bereits publizierte Beispiele. Die Fragmentierungen von Cyclopropyl- und 2-Cyclopropenylcarbenen a und b) sind sehr gut untersucht<sup>[215, 216]</sup>. Auch die Reaktion von Carbenen mit gespannten Dreiringen (c)<sup>[217]</sup> und die Carben-Umlagerung (d)<sup>[218]</sup> sind bekannt. Von den präzedenzlosen Basisre-



Schema 13. Potentielle Carbenreaktionen. Freie Elektronenpaare sind durch Striche symbolisiert.

aktionen wurde die Reaktion (e) von Herges<sup>[214]</sup> experimentell mit zwei verschiedenen Carbenvorläufern verifiziert (**72** + **73** → **74** → **75**; **76** → **77**).



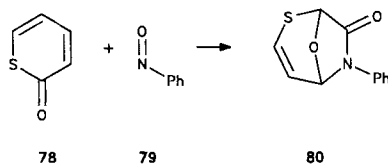
Kürzlich haben Fisher, Juarez-Brambila, Goralski, Wipke und Singaram<sup>[219]</sup> eine neuartige Umlagerung von  $\alpha$ -Aminoalkylboranen zu den entsprechenden  $\beta$ -Dialkylaminoal-



kylboranen mit IGOR2 gefunden und experimentell realisiert.

#### 4.4.4. Generierung von Reaktionsnetzwerken mit RAIN

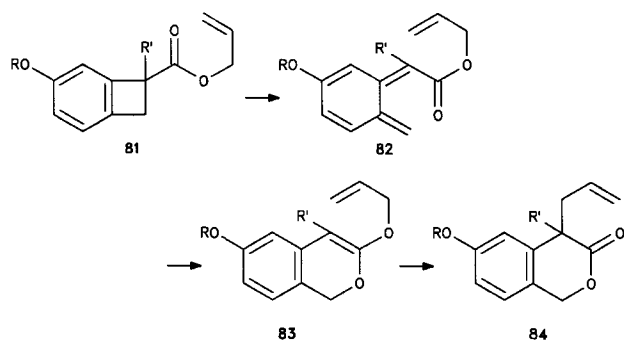
An der University of California at Los Angeles wurde im Sommer 1982 ein reaction mechanism contest durchgeführt, bei dem für die damals gerade veröffentlichte Streith-Defoin-Reaktion  $78 + 79 \rightarrow 80$  ein plausibler Reaktionsmechanismus vorgeschlagen werden sollte.



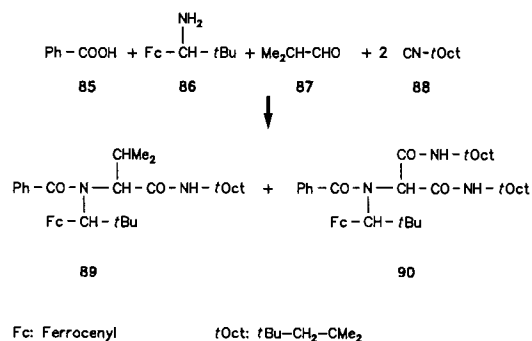
M. Jung, der Veranstalter dieses Wettbewerbs, forderte I. Ugi auf, auch mit Computerunterstützung mechanistische Vorschläge zur Streith-Defoin-Reaktion zu machen. Dies gelang allerdings erst, nachdem RAIN zur Verfügung stand<sup>[21, 60, 62, 63]</sup>. Die Streith-Defoin-Reaktion<sup>[220, 221]</sup> spielte bei der Entwicklung von RAIN als „Fahrrad zum Üben“ eine wichtige Rolle. Das von RAIN unter geeigneten Randbedingungen generierte Reaktionsnetzwerk<sup>[63]</sup> enthält die vier besten Vorschläge des erwähnten Wettbewerbs; einer davon entspricht dem inzwischen ermittelten Reaktionsmechanismus<sup>[221]</sup>.

Die Indazolsynthese von Ried et al.<sup>[222]</sup>, die Thiazolsynthese nach Hantsch<sup>[223]</sup> und die Favorskii-Umlagerung<sup>[224]</sup> gehören zu den ersten Anwendungsbeispielen von RAIN<sup>[60]</sup>. Für die Favorskii-Umlagerung wurde zunächst ein überraschend großes Netzwerk von denkbaren Reaktionsmechanismen mit 116 Intermediaten auf sechs Niveaus gefunden, das an der „breitesten“ Stelle 43 Alternativ-Intermediate aufweist. Dieses Netzwerk enthält neben dem publizierten Reaktionsmechanismus<sup>[224]</sup> noch viele weitere, die ebenfalls mit den vorliegenden experimentellen Ergebnissen vereinbar sind. Mit der neuesten Version von RAIN läßt sich bei Eingabe eines geeigneten Satzes von Einschränkungen auch ein einziger Reaktionsmechanismus für die Favorskii-Umlagerung erzeugen, und das ist der literaturbekannte.

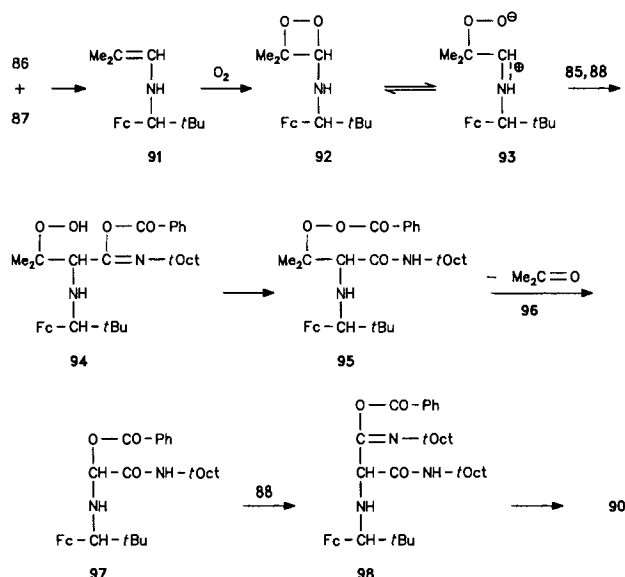
Für die von Kametani et al.<sup>[225]</sup> studierte Umlagerung eines Benzocyclobutens **81** in ein Isochroman **84** findet RAIN einen einzigen Reaktionsmechanismus und zwar über **82** und **83**<sup>[62]</sup>.



Kürzlich wurde der Reaktionsmechanismus einer störenden Nebenreaktion der Vierkomponentenkondensation<sup>[166, 167]</sup> mit Unterstützung von RAIN aufgeklärt<sup>[184, 185]</sup>. Als Modellreaktion diente die Bildung von **89** und dem Malonamidderivat **90** aus **85–88**. Der Mechanismus der Bildung von **90**



entspricht laut unseren experimentellen und den mit RAIN erhaltenen Ergebnissen dem in Schema 14 gezeigten über die Zwischenstufen **91–98**.



Schema 14. Mechanismus der Bildung von **90**.

Die Aufklärung des Mechanismus der hypothetischen präbiotischen Synthese von Adenin aus fünf Molekülen Cyanwasserstoff<sup>[226]</sup> mit RAIN<sup>[88]</sup> ist vorerst nicht abschließend möglich, weil die experimentelle Information nicht ausreicht, um zwischen den mit RAIN generierten Möglichkeiten zu entscheiden.

## 5. Perspektiven

Der Computer gehört heute zum Instrumentarium der chemischen Forschung, und es ist sicher, daß er immer mehr und auf immer vielfältigere Weise zum Fortschritt in der Chemie beitragen wird. Die Bedeutung der numerischen Anwendungen der Computer in der Chemie, vor allem in der Quantenchemie, in der Chemometrie sowie in der Planung,

Durchführung und Auswertung von Messungen und im Molecular Modeling, wird noch zunehmen. Die Effizienz und der Umfang des Gebrauchs von Computern in der Dokumentation chemiebezogener Daten und Informationen wird weiterhin stark wachsen.

Es ist vorhersehbar, daß auch Computerprogramme zur Lösung chemischer Probleme bald routinemäßig in der chemischen Forschung eingesetzt werden. Dabei werden formale Algorithmen eine zunehmend wichtigere Rolle spielen, auch bei Expertensystemen, die sich auf gespeicherte detaillierte empirische Informationen stützen. Empirische Daten werden in erster Linie zum Ähnlichkeitsvergleich mit Ergebnissen verwendet werden, die mit formal operierenden Computerprogrammen erzeugt wurden.

Die Neugier der potentiellen Benutzer von Computerprogrammen für chemische Fragestellungen wird zur Folge haben, daß innovationsfähige Programme bevorzugt werden. Voraussetzung ist allerdings, daß die angebotene Software eine komfortable Benutzeroberfläche hat und für breit zugängliche Hardware geeignet ist. Wichtig ist auch, daß solche Computerprogramme interaktiv arbeiten, denn ein intelligenter und erfahrener Benutzer möchte am Prozeß der Problemlösung aktiv teilnehmen. Man kann zwar mit einem geeigneten Programm grundsätzlich alle denkbaren Lösungen eines chemischen Problems, einschließlich der phantasievollsten, erzeugen, aber sie nützen wenig, wenn sie unter Unmengen belangloser Daten versteckt sind; die sinnvolle und willkürfreie Selektion von Lösungen ist nicht ohne die Beteiligung eines qualifizierten Benutzers möglich. Der Computer kann zwar helfen, zu neuen chemischen Ideen zu gelangen, aber der Chemiker muß sie als solche erkennen und aus der Vielzahl der denkbaren Alternativen auswählen. Das heißt, die Innovation verlagert sich von der Generierung von Vorschlägen zu deren Bewertung und Auswahl. Neue Chemie kann nach wie vor nicht ohne Beteiligung menschlicher Kreativität geschaffen werden.

Der Fortschritt in der Computerchemie beruht auf neuen Einsichten, Denkansätzen, Formalismen, Algorithmen, Software-Techniken und auch Hardware-Neuheiten. Da sich die Entwicklung der Computerchemie noch sehr im Fluß befindet, ist es verfrüht, die Bedeutung einzelner Projekte und Tendenzen zu bewerten. Vieles, was heute eindrucksvoll erscheint, wird morgen vergessen sein, und manches, was man gegenwärtig für ausgefallen oder belanglos hält, wird sich später vielleicht als der Beginn einer nützlichen und bedeutenden Entwicklung erweisen.

*Wir danken der Deutschen Forschungsgemeinschaft, der Volkswagen-Stiftung, der Alexander-von-Humboldt-Stiftung, der Kommission der Europäischen Gemeinschaft, dem Bundesministerium für Forschung und Technologie und dem Fonds der Chemischen Industrie für die Förderung unserer Arbeiten. Frau Natalia Berryman sind wir für ihre Hilfe bei der Abfassung dieser Übersicht dankbar.*

Eingegangen am 6. Juli 1991,  
veränderte Fassung am 14. Juli 1992 [A 894]

## Mathematische Begriffe und Notationen

**Boolesche Algebra** (a und b seien im folgenden Boolesche Variable)

$\mathbb{B}$	Menge der Booleschen Werte $\{0, 1\}$ , Wahrheitswerte									
$\wedge$	Konjunktion (logisches Und): Verknüpfung $\wedge$ : $\mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ gemäß der Tabelle									
	<table><tr><td><math>\wedge</math></td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>1</td></tr></table>	$\wedge$	0	1	0	0	0	1	0	1
$\wedge$	0	1								
0	0	0								
1	0	1								
$\vee$	inklusive Disjunktion (logisches Oder): Verknüpfung $\vee$ : $\mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ gemäß der Tabelle									
	<table><tr><td><math>\vee</math></td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td></tr></table>	$\vee$	0	1	0	0	1	1	1	1
$\vee$	0	1								
0	0	1								
1	1	1								
$\Rightarrow$	Implikation ( $a \Rightarrow b$ : „aus a folgt b“): Verknüpfung $\Rightarrow$ : $\mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ gemäß der Tabelle									
	<table><tr><td><math>\Rightarrow</math></td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>1</td></tr></table>	$\Rightarrow$	0	1	0	1	1	1	0	1
$\Rightarrow$	0	1								
0	1	1								
1	0	1								
$\Leftrightarrow$	Äquivalenz ( $a \Leftrightarrow b$ : „a ist äquivalent zu b“, „a hat denselben Wahrheitswert wie b“): Verknüpfung $\Leftrightarrow$ : $\mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ gemäß der Tabelle									
	<table><tr><td><math>\Leftrightarrow</math></td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>1</td></tr></table>	$\Leftrightarrow$	0	1	0	1	0	1	0	1
$\Leftrightarrow$	0	1								
0	1	0								
1	0	1								
$\neg$	Negation (logisches Nicht): $\neg$ : $\mathbb{B} \rightarrow \mathbb{B}$ mit $\neg(0) = 1$ und $\neg(1) = 0$ .									

## Quantoren

$\exists$	Existenzquantor („es existiert wenigstens ein ...“)
$\forall$	Allquantor („für alle ...“)

## Mengen (A, B, M und N seien Mengen)

$\in$	Der Ausdruck $m \in M$ besagt, daß m ein Element von M ist.
$\cap$	Der Schnitt $A \cap B$ enthält diejenigen Elemente, die A und B angehören.
$\cup$	Die Vereinigung $A \cup B$ enthält diejenigen Elemente, die A oder B angehören.
$\setminus$	Die Mengendifferenz $A \setminus B$ enthält diejenigen Elemente aus A, die nicht B angehören.
$\{m: P(m)\}$	Die Menge aller Objekte, für die das Prädikat P gilt. Beispielsweise bezeichnet der Ausdruck $\{n: n = 2m; m \in \mathbb{N}\}$ die Menge aller geradzahigen natürlichen Zahlen.
$ M $	Die Zahl der Elemente in einer Menge M. Zahl der Elemente einer Menge nennt man deren Mächtigkeit oder Kardinalität.
$M \times N$	Die Menge aller geordneten Paare (m, n), die sich aus den Elementen $m \in M$ und $n \in N$ bilden lassen.
$MA(\mathcal{A}, B) = \bigcup_{A \cap B \neq \{\}, A \in \mathcal{A}} A$	Die Vereinigung aller Mengen A der Mengenfamilie $\mathcal{A}$ , die eine nichtleere Schnittmenge mit der Menge B haben.

## Relationen, Abbildungen

Binäre Relation	Eine binäre Relation R assoziiert paarweise Elemente zweier Mengen X und Y. Es handelt sich also um eine Menge von geordneten Paaren $R \subseteq X \times Y$ , ausgedrückt durch $(x, y) \in R$ oder in Infixschreibweise $xRy$ . Zur Definition der binären Relation R verwendet man ein Prädikat P, das alle Paare erfüllen müssen, $R = \{(x, y) \in X \times Y: P(x, y)\}$ , oder man gibt alle Paare explizit an.
Transitive Hülle	Die transitive Hülle $R^+$ einer (binären) Relation R ist wie folgt definiert: (1) $(a, b) \in R \Rightarrow (a, b) \in R^+$ (2) $(a, b) \in R^+ \wedge (b, c) \in R \Rightarrow (a, c) \in R^+$ (3) $R^+$ enthält nichts, was nicht aus (1) und (2) folgt. $R^+$ ist die kleinste transitive Relation, die R umfaßt.

Äquivalenzrelation	<p>Spezialfall einer binären Relation <math>R \subseteq X \times Y</math>, wobei die Mengen <math>X</math> und <math>Y</math> identisch sind und die zudem drei besondere Eigenschaften aufweist:</p> <ol style="list-style-type: none"> <li>(1) Reflexivität: <math>\forall x \in X: xRx</math> (d.h. jedes Element ist mit sich selbst in Relation)</li> <li>(2) Symmetrie: <math>\forall x, y \in X: xRy \Rightarrow yRx</math> (d.h. wenn Element <math>x</math> mit Element <math>y</math> in Relation steht, dann auch <math>y</math> mit <math>x</math>)</li> <li>(3) Transitivität: <math>\forall x, y, z \in X: xRy \wedge yRz \Rightarrow xRz</math> (d.h. wenn <math>x</math> in Relation mit <math>y</math> steht und dieses mit <math>z</math>, dann steht auch <math>x</math> in Relation mit <math>z</math>).</li> </ol> <p>Ist eine Äquivalenzrelation <math>R</math> auf <math>X</math> gegeben, wird wegen der Symmetrie für ein Paar <math>(x, y) \in R</math> die Sprechweise „<math>x</math> steht in Relation <math>R</math> zu <math>y</math>“ meist durch „<math>x</math> und <math>y</math> stehen in Relation <math>R</math>“ ersetzt; für ein <math>x \in X</math> bezeichnet man die Menge <math>[x] = \{y: xRy\}</math> als die Äquivalenzklasse von <math>x</math> in <math>X</math> bezüglich <math>R</math>. Für ein <math>z \in X</math> ist die Äquivalenzklasse <math>[z]</math> entweder identisch zu <math>[x]</math>, falls <math>z \in [x]</math>, oder die beiden Äquivalenzklassen sind disjunkt (elementfremd, d.h. ohne gemeinsame Elemente). Die durch die Menge aller Äquivalenzklassen induzierte Partition <math>\{[x]: x \in X\}</math> von <math>X</math> nennt man einen Äquivalenzklassenraum oder eine Quotientenmenge.</p>	
$[m]_R$	Die Äquivalenzklasse von $m$ bezüglich der Äquivalenzrelation $R$ , d.h. alle Elemente, die gemäß der durch $R$ gegebenen Eigenschaft mit $m$ äquivalent sind. Äquivalenzrelationen spielen in der Theorie der CIG immer dann eine Rolle, wenn es gilt, permutierte Modelle einer Familie bezüglich einer Eigenschaft als „gleichwertig“ zu definieren.	
Permutationen	Im vorliegenden Beitrag werden Permutationen durch kleine griechische Buchstaben beschrieben. Unter einer Permutation von $n$ Elementen versteht man eine bijektive Abbildung $\lambda: L \rightarrow L$ einer Menge $L$ auf sich selbst. Eine Permutation $\lambda(I_a) = I_b$ , $\lambda(I_b) = I_c$ , ... $\lambda(I_g) = I_a$ wird in Cyclenschreibweise durch $(I_a I_b \dots I_g)$ dargestellt. Alle weiteren Elemente $I \in L \setminus \{I_a, I_b, \dots, I_g\}$ werden durch $\lambda$ auf sich selbst abgebildet.	
Gruppen		
Gruppe	<p>Eine Gruppe <math>G = (M, \cdot)</math> besteht aus einer Menge <math>M</math> und einer Verknüpfung <math>\cdot</math> auf <math>M</math> und hat folgende Eigenschaften:</p> <ol style="list-style-type: none"> <li>(1) Abgeschlossenheit: Alle Elemente <math>a, b \in M</math> ergeben verknüpft wieder ein Element der Menge <math>M</math>: <math>a \cdot b \in M</math>.</li> <li>(2) Assoziativität: Für alle Elemente <math>a, b, c \in M</math> gilt: <math>a \cdot (b \cdot c) = (a \cdot b) \cdot c</math>.</li> <li>(3) Existenz eines neutralen Elements: Es existiert ein <math>e \in M</math>, so daß für alle Elemente <math>a \in M</math> gilt: <math>a \cdot e = e \cdot a = a</math>.</li> <li>(4) Existenz eines Inversen: Für jedes Element <math>a \in M</math> existiert ein Element <math>a^{-1}</math>, so daß gilt: <math>a \cdot a^{-1} = a^{-1} \cdot a = e</math>.</li> </ol> <p>Im Kontext einer Gruppe vernachlässigt man das Symbol <math>\cdot</math> für die Verknüpfung und schreibt „<math>ab</math>“ statt „<math>a \cdot b</math>“. Die Kardinalität einer Gruppe heißt auch ihre Ordnung. Eine Untergruppe der Gruppe <math>G = (M, \cdot)</math> ist eine Gruppe <math>U = (N, \cdot)</math> mit <math>N \subseteq M</math>.</p>	
Nebenklasse	Für ein Element $a \in G$ wird die Menge $aU$ als Linksnebenklasse von $U$ bezeichnet und umfaßt alle Elemente $au$ mit $u \in U$ . Erfolgt die Verknüpfung mit $U$ von rechts, so ergeben sich Rechtsnebenklassen $Ua$ . Ist die Reihenfolge der Verknüpfung aus dem Kontext erschließbar oder ohne Bedeutung, wird schlicht von einer Nebenklasse gesprochen. Die Menge der (Links-/Rechts-)Nebenklassen der Untergruppe $U$ in der Gruppe $G$ bildet einen (Links-/Rechts-)Nebenklassenraum. Ein Nebenklassenraum von $U$ in $G$ stellt eine Partition von $G$ dar. Jeder Nebenklassenraum ist ein Äquivalenzklassenraum.	
Transversale	Eine minimale Menge $T$ mit der Eigenschaft $TU = G$ wird als vollständige Linkstransversale von $U$ in $G$ bezeichnet; analog als vollständige Rechtstransversale, falls $UT = G$ . Geht aus dem Kontext hervor, ob es sich um eine Links- oder eine Rechtstransversale handelt, spricht man von einer Transversale. Wenn $U$ nicht trivial ist – d.h. $U$ enthält nicht nur das neutrale Element – ist die Transversale nicht eindeutig. Die vollständige Transversale von $U$ in $G$ enthält von jeder Nebenklasse von $U$ in $G$ genau einen Repräsentanten. Jedes Element einer Nebenklasse kann als Repräsentant dieser Nebenklasse dienen. Statt von Repräsentanten spricht man auch von Nebenklassenführern, statt von Transversalen von Repräsentantensystemen.	
Erzeugende	Um eine Gruppe $G$ zu definieren, reicht es aus, die Verknüpfung und einige wenige Elemente anzugeben, die zusammen mit ihren Inversen zu $G$ verknüpft werden können. Diese Elemente bezeichnet man als Erzeugende, die aus ihnen bestehende Menge als Erzeugendensystem der Gruppe $G$ . Sind die Elemente einer Gruppe Permutationen, so spricht man auch von einer Permutationsgruppe. Die Menge $P$ aller Permutationen auf einer Menge $L$ von $n$ Symbolen bildet zusammen mit der für Permutationen definierten Hintereinanderausführung $\cdot$ eine Gruppe $Sym(L) = (P, \cdot)$ mit $n!$ Permutationen, die man als symmetrische Gruppe vom Grad $n$ bezeichnet. Da es häufig nicht wesentlich ist, daß Elemente von $L$ permutiert werden, sondern lediglich, wie viele Elemente für Permutationen zur Verfügung stehen, schreibt man auch $S_{ L }$ anstatt $Sym(L)$ . Gruppen finden in der Theorie der CIG unmittelbare Anwendung: Jede Familie von permutierten Modellen $P(E)$ mit Ligandensatz $L$ ist isomorph zur Symmetriegruppe $Sym(L)$ , da ja zwei verschiedene Permutationen aus $Sym(L)$ , angewendet auf das Referenzmodell $E$ , zu zwei verschiedenen Modellen in $P(E)$ führen und jedes Modell aus $P(E)$ durch Anwendung einer Permutation aus $Sym(L)$ auf das Referenzmodell entsteht.	
Konjugierte	Für $\lambda \in G$ bezeichnet man das Produkt $\lambda U \lambda^{-1}$ als ( $\lambda$ -)Konjugierte von $U$ in $G$ . Die Konjugierte einer Gruppe ist wiederum eine Gruppe. Für ein $\mu \in G$ bezeichnet man $\{\lambda \mu \lambda^{-1}: \lambda \in U\}$ als Wignersche Unterklasse von $U$ in $G$ . Die Menge $\{\{\lambda \mu \lambda^{-1}: \lambda \in U\}: \mu \in G\}$ der Wignerschen Unterklassen bildet einen Äquivalenzklassenraum von $G$ .	

- [1] D. Seebach, *Angew. Chem.* **1990**, 102, 1363; *Angew. Chem. Int. Ed. Engl.* **1990**, 29, 1320.
- [2] a) E. L. Eliel, N. L. Allinger, S. J. Angyel, G. A. Morrison, *Conformational Analysis*, Interscience, New York, **1965**; b) M. Hanack, *Conformational Theory*, Academic Press, New York, **1965**; c) G. Chiurdoglu, *Conformation Theory*, Academic Press, New York, **1971**.
- [3] H. Sachse, *Ber. Dtsch. Chem. Ges.* **1890**, 23, 1323; *Z. Phys. Chem.* **1892**, 10, 203.
- [4] E. Mohr, *J. Prakt. Chem.* **1918**, 98, 315.
- [5] H. Meerwein, K. van Emster, *Ber. Dtsch. Chem. Ges.* **1922**, 55, 2500.
- [6] a) W. Reppe, O. Schlichting, K. Klager, T. Toepel, Justus Liebig's, *Ann. Chem.* **1948**, 560, 1; für einen Rückblick siehe [6b]; b) G. Schröder, *Cyclooctatetraen*, Verlag Chemie, Weinheim, **1965**.
- [7] a) O. Roelen, *Angew. Chem.* **1948**, 60, 213; für Übersichten siehe [7b]; b) J. Falbe, *Carbon Monoxide in Organic Synthesis*, Springer, Heidelberg, **1970**; W. A. Herrmann, Kontakte (Darmstadt) **1991** (1), 22.
- [8] J. M. Bijvoet, A. F. Peerdeman, A. J. Bommel, *Nature* **1951**, 168, 271; J. M. Bijvoet, *Endeavour* **1955**, 14, 7.
- [9] G. A. Olah, G. K. S. Prakash, J. Sommer, *Superacids*, Wiley, New York, **1985**.
- [10] G. Domagk, *Klin. Wochenschr.* **1937**, 16, 1412; F. Mietsch, *Ber. Dtsch. Chem. Ges. B* **1938**, 71, 15; *Chem. Zentralbl.* **1938**, 120.
- [11] K. H. Büchel, *Bayer. Landwirtsch. Jahrb.* **1981**, 58, 27; K. H. Büchel, M. Plempel, *Chron. Drug Discovery* **1983**, 2, 235.
- [12] K. Zuse, *Der Computer, mein Lebenswerk*, Springer, Heidelberg, **1984**.
- [13] *Computer Applications in Chemical Research and Education* (Hrsg.: J. Brandt, I. Ugi) Hüthig, Heidelberg, **1989**.
- [14] *Computational Methods in Chemistry* (Hrsg.: J. Bargon), Plenum, New York, **1980**.
- [15] M. D. Johnson, Jr., *Computational Chemistry*, Elsevier, Amsterdam, **1988**; D. M. Hirst, *A Computational Approach to Chemistry*, Blackwell, Oxford, **1990**; *Modern Techniques in Computational Chemistry: MO-TECC-90* (Hrsg.: E. Clementi), ESCOM, Leiden, **1990**.

- [16] *Computer Aids to Chemistry* (Hrsg.: G. Vermin, M. Chanon), Ellis-Horwood, Chichester, **1986**; B. Adler, *Computerchemie – eine Einführung*, VEB Leipzig, **1986**.
- [17] M. A. Ott, J. H. Noordik, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 239.
- [18] J. Zupan, *Algorithms for Chemists*, Wiley, New York, **1985**.
- [19] *PC for Chemists*: (Hrsg.: J. Zupan, D. Hadzi), Elsevier, Amsterdam, **1990**.
- [20] I. Ugi, J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, M. Knauer, K. Reitsam, N. Stein, *Labor 2000* (Sonderheft von *Labor Praxis*) **1992**, 170.
- [21] I. Ugi, E. Fontain, J. Bauer, *Anal. Chim. Acta* **1990**, *235*, 155.
- [22] J. Lederberg, *Proc. Natl. Acad. Sci. USA* **1990**, *53*, 134; *How DENDRAL was conceived and born* (ACM Symposium on the History of Medical Informatics 1987), National Library of Medicine, S. 270.
- [23] 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**.
- [24] N. A. B. Gray, *Computer-assisted Structure Elucidation 2*, Wiley-Interscience, New York, **1986**.
- [25] G. Vleduts, *Inf. Storage Retr.* **1963**, *1*, 117.
- [26] A. Dengler, E. Fontain, M. Knauer, N. Stein, I. Ugi, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 262.
- [27] G. Vleduts, K. A. Finn, *Proc. Dept. of Mechanization and Automation of Information Work*, Acad. Sci. USSR, Moskau, **1960**, S. 66.
- [28] E. J. Corey, *Pure Appl. Chem.* **1967**, *14*, 19; siehe auch *Angew. Chem.* **1991**, *103*, 469; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 455.
- [29] E. J. Corey, W. T. Wipke, *Science* **1991**, *166*, 178; E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, **1989**.
- [30] E. J. Corey, A. K. Long, S. D. Rubenstein, *Science* **1985**, *228*, 4089; siehe auch E. J. Corey, A. K. Long, G. I. Lotto, S. D. Rubenstein, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 304.
- [31] W. T. Wipke, D. Rogers, *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 71.
- [32] H. Gelernter, N. S. Sridharan, A. J. Hart, S. C. Yen, F. W. Fowler, H. J. Shue, *Top. Curr. Chem.* **1973**, *41*, 113; H. Gelernter, J. R. Rose, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492.
- [33] Das von G. Kaufhold und I. Ugi 1965–1967 bei der Bayer AG in Leverkusen entwickelte Computerprogramm zur Planung von Peptidsynthesen ist in [34] beschrieben.
- [34] I. Ugi, *Rec. Chem. Prog.* **1969**, *30*, 389; *Intra-Sci. Chem. Rep.* **1971**, *5*, 229.
- [35] M. Bersohn, *Computer Assisted Organic Synthesis* (ACS Symp. Ser. **1977**, *61*); S. Sasaki, K. Funatsu, *Tetrahedron Comput. Methodol.* **1988**, *1*, 39; R. Barone, M. Arbelot, M. Chanon, *ibid.* **1988**, *1*, 3; E. S. Blurock, *ibid.* **1989**, *2*, 207; A. Weise, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 228; I. Dogane, T. Takabatake, M. Bersohn, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 291; L. Baumer, I. Compagnari, G. Sala, G. Sello, *ibid.* **1992**, *111*, 297; P. Hamm, P. Jauffret, G. Kaufmann, *ibid.* **1992**, *111*, 317.
- [36] E. Hjelt, *Geschichte der Organischen Chemie*, Vieweg, Braunschweig, **1916**.
- [37] a) R. Robinson et al., *J. Chem. Soc.* **1916**, 109, 1029, 1039; *ibid.* **1917**, *111*, 958; *ibid.* **1918**, *113*, 639; *ibid.* **1919**, *115*, 943; *ibid.* **1921**, *120*, 545; b) Sir R. Robinson, *The Structural Relations of Natural Products*, Clarendon Press, Oxford, **1955**.
- [38] S. G. Warren, *Organic Synthesis: The Disconnection Approach*, Wiley, New York, **1982**.
- [39] P. H. Winston, *Artificial Intelligence*, Addison-Wesley, Reading, MA, USA, **1984**; B. A. Hohne, T. H. Pierce, *Expert System Applications in Chemistry* (ACS Symp. Ser. **1989**, *408*).
- [40] H. W. Brown, *Chem. Ind. London* **1988** (5), 43.
- [41] G. M. Downs, V. J. Gillet, J. D. Holliday, M. F. Lynch, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 172.
- [42] E. J. Corey, G. Petersson, *J. Am. Chem. Soc.* **1972**, *94*, 460.
- [43] E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, G. Peterson, *J. Am. Chem. Soc.* **1975**, *97*, 6116.
- [44] M. F. Frèrejaques, *Bull. Soc. Chim. Fr. Mem.* **1939**, *5*, 1008.
- [45] R. B. Woodward, *Angew. Chem.* **1956**, *68*, 13.
- [46] M. Hicks, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 352.
- [47] M. G. Bures, W. L. Jorgensen, *J. Org. Chem.* **1988**, *53*, 2504.
- [48] W. Schubert, *MATCH* **1979**, *6*, 213.
- [49] G. Grethe, T. E. Moock, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 505.
- [50] J. B. Hendrickson, *Top. Curr. Chem.* **1976**, *62*, 49.
- [51] G. Moreau, *Nouv. J. Chim.* **1978**, *2*, 187.
- [52] J. B. Hendrickson, A. G. Toezko, *Pure Appl. Chem.* **1988**, *60*, 1583; J. B. Hendrickson, *Anal. Chim. Acta* **1990**, *235*, 103; *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 323.
- [53] J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie, I. Ugi in *Computer Representation and Manipulation of Chemical Information* (Hrsg.: W. T. Wipke, S. R. Heller, R. J. Feldman, E. Hyde), Wiley, New York, **1974**, S. 129.
- [54] J. Gasteiger, C. Jochum, *Top. Curr. Chem.* **1978**, *74*, 93.
- [55] J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki, *Top. Curr. Chem.* **1987**, *137*, 19; J. Gasteiger, W. D.-Ihlenfeldt, P. R. Röse, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 270.
- [56] R. Doenges, B. T. Groebel, H. Nickelsen, J. Sander, *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 425.
- [57] D. A. Evans, *unveröffentlichtes Manuskript*, **1972**.
- [58] A. Lapworth, *J. Chem. Soc.* **1921**, *120*, 543.
- [59] D. Seebach, *Angew. Chem.* **1979**, *91*, 259; *Angew. Chem. Int. Ed. Engl.* **1979**, *18*, 239.
- [60] E. Fontain, J. Bauer, I. Ugi, *Chem. Lett.* **1987**, 37.
- [61] E. Fontain, *Tetrahedron Comput. Methodol.*, **1990**, *3*, 469.
- [62] E. Fontain, *Dissertation*, Technische Universität München, **1987**.
- [63] E. Fontain, J. Bauer, I. Ugi, *Z. Naturforsch. B* **1987**, *42*, 297.
- [64] D. Hilbert, P. Bernays, *Grundlagen der Mathematik*, Teubner, Berlin, **1934**.
- [65] A. Cayley, *Ber. Dtsch. Chem. Ges.* **1875**, *8*, 172; F. Herrmann, *ibid.* **1880**, *13*, 792.
- [66] *Chemical Applications of Graph Theory* (Hrsg.: A. T. Balaban), Academic Press, London, **1976**.
- [67] *Chemical Applications of Topology and Graph Theory* (Hrsg.: R. B. King), Elsevier, Amsterdam, **1983**.
- [68] R. E. Merrifield, H. E. Simmons, *Topological Methods in Chemistry*, Wiley, New York, **1989**.
- [69] H. R. Henze, C. M. Blair, *J. Am. Chem. Soc.* **1931**, *53*, 3077.
- [70] A. Polyá, *Acta Math.* **1937**, *68*, 145.
- [71] *The Permutation Group in Physics and Chemistry* (Hrsg.: J. Hinz), Springer, Heidelberg, **1979**.
- [72] A. Kerber, K.-J. Thürlings, *Symmetrieklassen von Funktionen und ihre Abzähltheorie* (Bayreuther Math. Schriften), Bayreuth, **1983**.
- [73] J. Dugundji, R. Kopp, D. Marquarding, I. Ugi, *Top. Curr. Chem.* **1978**, *75*, 165.
- [74] E. Ruch, W. Hässelbarth, B. Richter, *Theor. Chim. Acta* **1970**, *19*, 288; W. Hässelbarth, E. Ruch, *ibid.* **1973**, *29*, 259; W. Hässelbarth, E. Ruch, D. J. Klein, T. H. Seligman in *Group Theoretical Methods in Physics* (Hrsg.: R. T. Sharp, B. Kolman), Academic Press, New York, **1977**, S. 617.
- [75] J. Dugundji, J. Showell, R. Kopp, D. Marquarding, I. Ugi, *Isr. J. Chem.* **1980**, *20*, 20.
- [76] I. Ugi, D. Marquarding, H. Klusacek, G. Gokel, P. D. Gillespie, *Angew. Chem.* **1970**, *82*, 741; *Angew. Chem. Int. Ed. Engl.* **1970**, *9*, 703.
- [77] I. Ugi, B. Gruber, N. Stein, A. Demharter, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 485.
- [78] J. Dugundji, I. Ugi, *Top. Curr. Chem.* **1973**, *39*, 19.
- [79] J. Brandt, J. Bauer, R. M. Frank, A. von Scholley, *Chem. Scr.* **1981**, *18*, 53; J. Brandt, A. von Scholley, *Comput. Chem.* **1983**, *7*, 51.
- [80] J. Koča, M. Kratochvíl, V. Kvasnička, L. Matyska, J. Pospichal, *Lect. Notes Chem.* **1989**, *51*; E. Hladka, J. Koča, M. Kratochvíl, V. Kvasnička, L. Matyska, J. Pospichal, V. Potůček, *Top. Curr. Chem.*, im Druck.
- [81] V. Kvasnička, M. Kratochvíl, J. Koča, *Coll. Czech. Chem. Commun.* **1983**, *48*, 2284; J. Koča, *ibid.* **1988**, *53*, 1007, 3108, 3119; *J. Math. Chem.* **1989**, *3*, 73, 91; V. Kvasnička, J. Pospichal, *ibid.* **1989**, *3*, 161; *Int. J. Quantum Chem.* **1990**, *38*, 253; V. Kvasnička, J. Pospichal, *J. Math. Chem.* **1990**, *5*, 309.
- [82] I. Ugi, J. Dugundji, R. Kopp, D. Marquarding, *Lect. Notes Chem.* **1984**, *36*.
- [83] T. F. Brownscombe, R. V. Stevens, persönliche Mitteilung an I. Ugi; T. F. Brownscombe, *Dissertation*, Rice University, Houston, TX, **1972**.
- [84] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum in [14], S. 275.
- [85] I. Ugi, J. Bauer, K. Bley, A. Dengler, E. Fontain, M. Knauer, S. Lohberger, *J. Mol. Struct. (Theochem)* **1991**, *230*, 73.
- [86] A. P. Johnson, C. Marshall, P. N. Judson, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 310; *J. Chem. Inform. Comput. Chem.*, im Druck; A. P. Johnson, C. Marshall, *ibid.*, im Druck.
- [87] J. B. Hendrickson, C. A. Parks, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 209.
- [88] I. Ugi, J. Bauer, R. Baumgartner, E. Fontain, D. Forstmeier, S. Lohberger, *Pure Appl. Chem.* **1988**, *60*, 1573.
- [89] I. Ugi, E. Fontain, M. Knauer, N. Stein, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 262.
- [90] N. S. Zefirov, E. V. Godeeva, *Usp. Khim.* **1954**, *54*, 1753; N. S. Zefirov, *Acc. Chem. Res.* **1987**, *20*, 237; S. S. Tratch, I. I. Baskin, N. S. Zefirov, *Zh. Org. Khim.* **1988**, *24*, 1121; N. S. Zefirov, S. S. Tratch, *Anal. Chim. Acta* **1990**, *235*, 115.
- [91] L. Matyska, J. Koča, *J. Chem. Inf. Comput. Sci.*, im Druck.
- [92] L. Matyska, J. Koča, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 380.
- [93] Z. Hippe in [13], S. 165.
- [94] P. Y. Johnson, I. Burnstein, J. Cray, M. Evens, T. Wang, *Expert Systems Application in Chemistry* (ACS Symp. Ser. **1988**, *408*).
- [95] I. Ugi, N. Stein, M. Knauer, B. Gruber, K. Bley, R. Weidinger, *Top. Curr. Chem.*, **1993**, *166*, 199; R. Weidinger, *Diplomarbeit*, Universität Passau, **1991**.
- [96] J. Gasteiger, P. D. Gillespie, D. Marquarding, I. Ugi, *Top. Curr. Chem.* **1974**, *48*, 1.
- [97] E. O. von Lippmann, *Chem.-Ztg.* **1909**, *1*, 1; R. Grund, A. Kerber, R. Laue, *Bayreuther Universitätsblätter*, im Druck.
- [98] I. Ugi, *Chimia* **1986**, *40*, 340.
- [99] L. Spialter, *J. Chem. Doc.* **1964**, *4*, 261.
- [100] I. Ugi, M. Wochner, E. Fontain, J. Bauer, B. Gruber, R. Karl in *Concepts and Applications of Chemical Similarity* (Hrsg.: M. A. Johnson, G. M. Maggiora), Wiley, New York, **1990**, S. 239.
- [101] J. Dugundji, *Topology*, Allyn and Bacon, Boston, **1966**.

- [102] W. Preuss, *Int. J. Quantum Chem.* **1969**, 3, 123, 131.
- [103] M. Wochner, *Dissertation*, Technische Universität München, **1984**.
- [104] M. Wochner, I. Ugi, *J. Mol. Struct. (Theochem)* **1988**, 165, 229.
- [105] C. Jochum, J. Gasteiger, I. Ugi, *Angew. Chem.* **1980**, 92, 503; *Angew. Chem. Int. Ed. Engl.* **1980**, 19, 495; C. Jochum, J. Gasteiger, I. Ugi, J. Dugundji, *Z. Naturforsch. B* **1982**, 37, 376.
- [106] M. Wochner, J. Brandt, A. von Scholley, I. Ugi, *Chimia* **1988**, 42, 217.
- [107] E. Fontain, *Anal. Chim. Acta* **1992**, 265, 227.
- [108] H. Kolbe, J. Liebig, *Ann. Chem. Pharm.* **1850**, 75, 211; *ibid.* **1850**, 76, 1.
- [109] J. Pospichal, V. Kvasnicka, I. Gutman, *Collect. Sci. Pap. Kragujevac*, **1991**, 12, 109–125; V. Baláz, J. Pospichal, V. Kvasnicka, *Discret Appl. Math.* **1992**, 35, 1–19.
- [110] A. P. Johnson, C. Marshall, *J. Chem. Inf. Comput. Sci.*, im Druck.
- [111] E. Fontain, *Dissertation*, Technische Universität München, **1987**.
- [112] I. Ugi, J. Brandt, A. von Scholley, S. Minker, M. Wochner, H. Schönmann, B. Straupe, *Hierarchisch strukturierte Speicherung und Ermittlung von chemischen Reaktionen* (Forschungsbericht, Information und Dokumentation, BMFT-FB-ID85-005), FIZ Karlsruhe, **1985**.
- [113] J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, I. Ugi in *Computer Assisted Organic Synthesis* (ACS Symp. Ser. **1977**, 61).
- [114] J. Bauer, R. Herges, E. Fontain, I. Ugi, *Chimia* **1985**, 39, 43.
- [115] J. Bauer, *Tetrahedron Comput. Methodol.* **1989**, 2, 269.
- [116] J. Bauer, I. Ugi, *J. Chem. Res. Synop.* **1982**, 298; *J. Chem. Res. Miniprint* **1982**, 3101; J. Bauer, *Dissertation*, Technische Universität München, **1981**.
- [117] D. Forstmeier, J. Bauer, E. Fontain, R. Herges, R. Herrmann, I. Ugi, *Angew. Chem.* **1988**, 100, 1618; *Angew. Chem. Int. Ed. Engl.* **1988**, 27, 1558; J. Bauer, E. Fontain, D. Forstmeier, I. Ugi, *Tetrahedron Comput. Methodol.* **1988**, 1, 129.
- [118] R. Herges, *Dissertation*, Technische Universität München, **1984**.
- [119] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, *Angew. Chem.* **1979**, 91, 99; *Angew. Chem. Int. Ed. Engl.* **1979**, 18, 111.
- [120] E. O. Fischer, G. Bürger, *Z. Naturforsch. B* **1961**, 16, 77; G. Wilke, B. Bogdanović, P. Hardt, P. Heimbach, W. Keim, M. Kröner, W. Oberkirch, K. Tanaka, E. Steinrück, D. Walter, H. Zimmermann, *Angew. Chem.* **1966**, 78, 157; *Angew. Chem. Int. Ed. Engl.* **1966**, 5, 159.
- [121] N. Stein, *Dissertation*, Technische Universität München, in Vorbereitung.
- [122] E. J. Corey, W. J. Howe, D. A. Pensak, *J. Am. Chem. Soc.* **1974**, 96, 7724.
- [123] W. T. Wipke, T. M. Dyott, *J. Am. Chem. Soc.* **1974**, 96, 4825, 4834.
- [124] S. Hanessian, J. Franco, G. Gagnon, D. Laramée, B. Laroade, *J. Chem. Inf. Comput. Sci.* **1990**, 30, 413.
- [125] J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie, I. Ugi, *Tetrahedron* **1974**, 30, 1845.
- [126] C. E. Wintner, *Strands of Organic Chemistry*, Holden-Day, San Francisco, **1979**, S. 9.
- [127] H. C. Longuet-Higgins, *Mol. Phys.* **1963**, 6, 445.
- [128] S. McLane, G. Birkhoff, *Algebra*, McMillan, New York, **1967**, S. 14, 91; T. A. Whitelaw, *Introduction to Abstract Algebra*, Blackie, Glasgow, **1988**, S. 124.
- [129] E. P. Wigner, *Proc. R. Soc. A* **1971**, 322, 181.
- [130] B. Gruber, *Dissertation*, Technische Universität München, **1992**; A. Dietz, *Dissertation*, Technische Universität München, **1993**.
- [131] B. Gruber, A. Dietz, I. Ugi, noch unveröffentlicht.
- [132] J. E. Hopcroft, J. D. Ullman, *Einführung in die Automatentheorie, formale Sprachen und Komplexitätstheorie*, 2. Aufl., Addison-Wesley, Bonn, **1990**, S. 8.
- [133] R. S. Berry, *J. Chem. Phys.* **1960**, 32, 933.
- [134] P. Gillespie, P. Hoffmann, H. Klusacek, D. Marquarding, S. Pfohl, F. Ramirez, E. A. Tsoilis, I. Ugi, *Angew. Chem.* **1971**, 83, 493; *Angew. Chem. Int. Ed. Engl.* **1971**, 10, 503.
- [135] P. Lemmen, R. Baumgartner, I. Ugi, F. Ramirez, *Chem. Scr.* **1988**, 28, 451.
- [136] M. Gielen, N. van Laetern, *Bull. Soc. Chim. Belges* **1970**, 79, 679; J. I. Musher, *J. Am. Chem. Soc.* **1972**, 94, 5662.
- [137] J. Brocas, M. Gielen, R. Willem, *The Permutational Approach to Dynamic Stereochemistry*, McGraw-Hill, New York, **1983**.
- [138] J. Dugundji, D. Marquarding, I. Ugi, *Chem. Scr.* **1976**, 9, 74.
- [139] W. Schubert, I. Ugi, *J. Am. Chem. Soc.* **1978**, 100, 37; *Chimia* **1979**, 33, 183.
- [140] Lit. [96], Kap. 8.
- [141] H. L. Morgan, *J. Chem. Doc.* **1965**, 5, 107.
- [142] N. Stein, I. Ugi, noch unveröffentlicht.
- [143] R. S. Cahn, C. K. Ingold, *J. Chem. Soc.* **1951**, 612; R. S. Cahn, C. K. Ingold, V. Prelog, *Experientia* **1966**, 78, 413; V. Prelog, G. Helmchen, *Angew. Chem.* **1982**, 94, 614; *Angew. Chem. Int. Ed. Engl.* **1982**, 21, 567; H. Dodziuk, M. M. Mirowicz, *Tetrahedron Asymmetry* **1990**, 3, 171.
- [144] E. Fontain, noch unveröffentlicht.
- [145] A. Dengler, I. Ugi, *Comput. Chem.* **1991**, 15, 103; *J. Chem. Res. Synop.* **1991**, 162; *J. Chem. Res. Miniprint* **1991**, 1279.
- [146] J. E. Dubois, *J. Chem. Doc.* **1973**, 13, 8; R. Attias, J. E. Dubois, *J. Chem. Inf. Comput. Sci.* **1990**, 30, 2.
- [147] W. E. Cossum, G. M. Dyson, M. F. Lynch, R. N. Wolfe, *Mimeographed Report*, Chemical Abstracts Service, **1963**.
- [148] E. H. Sussenguth, Jr., *J. Chem. Doc.* **1965**, 5, 36.
- [149] G. W. Adamson, J. Cowell, M. F. Lynch, A. H. W. McLure, W. G. Town, A. M. Yapp, *J. Chem. Doc.* **1973**, 13, 153.
- [150] J. Friedrich, *Dissertation*, Technische Universität München, **1979**; J. Friedrich, I. Ugi, *J. Chem. Res. Synop.* **1980**, 70; *J. Chem. Res. Miniprint* **1980**, 1301.
- [151] M. G. Hicks, C. Jochum, *J. Chem. Inf. Comput. Sci.* **1990**, 30, 191.
- [152] A. V. Aho, J. E. Hopcroft, P. Ullman, *The Design and Analysis of Computer Algorithms*, Addison-Wesley, Reading, **1974**.
- [153] W. D. Ihlenfeldt, *Dissertation*, Technische Universität München, **1991**.
- [154] K. Bley, *Dissertation*, Technische Universität München, **1991**; E. Fontain, *Diplomarbeit*, Technische Universität München, **1986**; K. Bley, I. Ugi, noch unveröffentlicht.
- [155] M. Knauer, *Dissertation*, Technische Universität München, **1992**.
- [156] H. Thalhammer, *Diplomarbeit*, Technische Universität München, **1991**.
- [157] B. Rasper, G. Roden, B. Woost, C. Zirz in *Software-Entwicklung in der Chemie 3* (Hrsg.: G. Gauglitz), Springer, Heidelberg, **1989**, S. 33; W. T. Donner, in [13], S. 9.
- [158] E. Meyer, E. Sens, *Anal. Chim. Acta* **1988**, 210, 135.
- [159] P. Bruck in *Chemical Structures – The International Language of Chemistry* (Hrsg.: W. A. Warr), Springer, Heidelberg, **1988**, S. 113.
- [160] M. Hicks, C. Jochum, *Anal. Chim. Acta* **1990**, 235, 87; M. G. Hicks in *Software-Entwicklung in der Chemie 3* (Hrsg.: G. Gauglitz), Springer, Heidelberg, **1989**, S. 9.
- [161] G. Klopman, *J. Am. Chem. Soc.* **1984**, 106, 7315; *Mutat. Res.* **1984**, 126, 139, 227; *J. Comput. Chem.* **1985**, 6, 28.
- [162] G. Klopman, *J. Math. Chem.* **1991**, 7, 187.
- [163] M. Knauer, I. Ugi, noch unveröffentlicht.
- [164] A. Dengler, I. Ugi, *J. Math. Chem.* **1992**, 9, 1.
- [165] I. Ugi, G. Kaufhold, *Justus Liebigs Ann. Chem.* **1967**, 709, 11.
- [166] I. Ugi, *Angew. Chem.* **1962**, 74, 9; *Angew. Chem. Int. Ed. Engl.* **1962**, 1, 8.
- [167] I. Ugi, S. Lohberger, R. Karl in *Comprehensive Organic Synthesis: Selectivity of Synthetic Efficiency 2* (Hrsg.: B. M. Trost, C. H. Heathcock), Pergamon Press, Oxford, **1991**, Kap. 4.6.
- [168] G. A. Blakley, *J. Chem. Educ.* **1982**, 59, 728; C. Behnke, J. Bargon, *J. Chem. Inf. Comput. Sci.* **1990**, 30, 228.
- [169] J. Bauer, E. Fontain, I. Ugi, *Anal. Chim. Acta* **1988**, 210, 123.
- [170] R. E. Burkhard, *Methoden der ganzzahligen Optimierung*, Springer, Wien, **1972**; R. E. Burkhard, K. H. Stratmann, *Nav. Res. Logistics Qu.* **1978**, 25, 129.
- [171] D. E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison-Wesley, Reading, **1989**.
- [172] *Modern Approaches to Chemical Reaction Searching* (Hrsg.: P. Willett), Gower, Aldershot, **1986**; D. Bawden, *J. Chem. Inf. Comput. Sci.* **1991**, 31, 212.
- [173] A. P. Johnson, *Chem. Br.* **1985**, 21, 59; E. Zass, S. Müller, *Chimia* **1986**, 40, 38; J. H. Borkent, F. Oukes, J. H. Noordik, *J. Chem. Inf. Comput. Sci.* **1988**, 28, 148.
- [174] W. T. Wipke, G. Vladutz, *Tetrahedron Comput. Methodol.* **1990**, 3, 83.
- [175] C. Tonnelier, P. Jauffret, T. Hanser, G. Kaufmann, *Tetrahedron Comput. Methodol.* **1990**, 3, 351.
- [176] K. Bley, J. Brandt, A. Dengler, R. Frank, I. Ugi, *J. Chem. Res. Synop.* **1991**, 261; *J. Chem. Res. Miniprint* **1991**, 2601.
- [177] I. Ugi, J. Bauer, E. Fontain in *Personal Computers for Chemists* (Hrsg.: J. Zupan), Elsevier, Amsterdam, **1990**, S. 135.
- [178] E. Fontain, K. Reitsam, *J. Chem. Inf. Comput. Sci.* **1991**, 31, 96.
- [179] Wir verfügen über ein von J. Bauer implementiertes unveröffentlichtes Computerprogramm, das beliebige r-Matrizen aufgrund ihrer definierten mathematischen Eigenschaften generieren kann.
- [180] I. Ugi, *Chem. Scr.* **1986**, 26, 205.
- [181] J. Bauer, E. Fontain, I. Ugi, *MATCH* **1992**, 27, 31.
- [182] I. Ugi, *Proc. Estonian Acad. Sci.* **1989**, 38, 225; *ibid.* **1990**, 39, 193.
- [183] E. Fontain, *J. Chem. Inf. Comput. Sci.* **1992**, 32, 748.
- [184] I. Ugi, J. Bauer, K. Bley, A. Dengler, E. Fontain, M. Knauer, S. Lohberger, *J. Mol. Structure (Theochem)* **1991**, 230, 73.
- [185] S. Lohberger, E. Fontain, I. Ugi, G. Müller, J. Lachmann, *New J. Chem.* **1991**, 15, 913; *Acta Crystallogr. Sect. C* **1991**, 47, 2444.
- [186] R. E. Valdés-Pérez, *Tetrahedron Comput. Methodol.* **1990**, 5, 277.
- [187] R. Huisgen *Angew. Chem.* **1963**, 75, 604, 742; *Angew. Chem. Int. Ed. Engl.* **1963**, 2, 565, 633; *J. Org. Chem.* **1976**, 41, 403.
- [188] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, 22, 1859; L. J. McBride, M. H. Caruthers, *ibid.* **1983**, 24, 245.
- [189] D. Marquarding, F. Ramirez, I. Ugi, P. D. Gillespie, *Angew. Chem.* **1973**, 85, 99; *Angew. Chem. Int. Ed. Engl.* **1973**, 12, 91.
- [190] F. Westheimer, *Acc. Chem. Res.* **1968**, 11, 70.
- [191] F. Ramirez, S. Glaser, P. Stern, P. D. Gillespie, *Angew. Chem.* **1973**, 85, 39; *Angew. Chem. Int. Ed. Engl.* **1973**, 12, 66; F. Ramirez, J. F. Marecek, I. Ugi, *Synthesis* **1975**, 99.
- [192] W. Richter, R. Karl, I. Ugi, *Tetrahedron* **1990**, 46, 3167.
- [193] F. Ramirez, I. Ugi, *Phosphorus Sulfur* **1976**, 1, 231.
- [194] I. Ugi, P. Jacob, B. Landgraf, C. Rupp, P. Lemmen, U. Verfürth, *Nucleosides Nucleotides* **1988**, 7, 605; I. Ugi, N. Bachmeier, R. Herrmann, P. Jacob, R. Karl, P. Lemmen, W. Richter, U. Verfürth, *Phosphorus Sulfur Silicon* **1990**, 51/52, 57.
- [195] P. Jacob, W. Richter, I. Ugi, *Liebigs Ann. Chem.* **1991**, 519; R. Karl, P. Lemmen, W. Richter, I. Ugi, B. Werner, *Synthesis*, eingereicht.

- [196] W. N. Lipscomb, *Boron Hydrides*, Benjamin, New York, **1966**, S. 57.
- [197] B. Brelloch, H. Binder, *Angew. Chem.* **1988**, *100*, 270; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 262.
- [198] R. E. Carhart, D. H. Smith, H. Brown, C. Djerassi, *J. Am. Chem. Soc.* **1975**, *97*, 5755.
- [199] R. E. Carhart, D. H. Smith, N. A. B. Gray, J. G. Nourse, C. Djerassi, *J. Org. Chem.* **1981**, *46*, 1708.
- [200] V. V. Serov, L. A. Gribov, M. E. Elyashberg, *J. Mol. Struct. (Theochem)* **1985**, *129*, 183.
- [201] A. Kerber, R. Laue, D. Moser, *Anal. Chim. Acta* **1990**, *235*, 221.
- [202] H. Huixiao, X. Xinguan, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 116.
- [203] S. Bohanec, J. Zupan, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 531.
- [204] I. Ugi, E. Wischhöfer, *Chem. Ber.* **1962**, *95*, 136.
- [205] D. H. R. Barton, *Aldrichimica Acta* **1990**, *23*, 3.
- [206] R. Herges, *Tetrahedron Comput. Methodol.* **1988**, *1*, 15; K. M. T. Yamada, M. W. Markus, G. Winnewesser, W. Joentgen, R. Kock, E. Vogel, H.-J. Altenbach, *Chem. Phys. Lett.* **1989**, *160*, 113.
- [207] R. Herges, I. Ugi, *Angew. Chem.* **1985**, *97*, 596; *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 594; *Chem. Ber.* **1986**, *119*, 829.
- [208] F. W. Fowler, *Angew. Chem.* **1971**, *83*, 148; *Angew. Chem. Int. Ed. Engl.* **1971**, *10*, 135.
- [209] S. M. McElvain, *Org. React. (NY)*, **1948**, *4*, 256.
- [210] M. Stoll, *Helv. Chim. Acta* **1947**, *30*, 1837.
- [211] R. Herges, *J. Chem. Inf. Comput. Sci.* **1990**, *3*, 377.
- [212] R. Herges, C. Hoock, *Science* **1992**, *255*, 711.
- [213] R. Herges in *Proc. 4th Chem. Congr. North Am.* **1991**, Division of Chemical Information, New York, im Druck.
- [214] R. Herges, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 37.
- [215] J. Arct, U. H. Brinker, *Methoden Org. Chem. (Houben-Weyl)* 4th ed. 1952–Bd. E19b, **1989**, S. 337.
- [216] H. Dürr, E. M. Frühauf, *Methoden Org. Chem. (Houben-Weyl)* 4th ed. 1952–Band E19b, **1989**, S. 775.
- [217] W. von E. Doering, J. F. Coburn, Jr., *Tetrahedron Lett.* **1965**, 991.
- [218] C. Wentrup, R. Harder, *Top. Curr. Chem.* **1976**, *62*, 173.
- [219] G. B. Fisher, J. J. Juárez-Brambila, C. T. Goralski, W. T. Wipke, B. Singaram, *J. Am. Chem. Soc.*, eingereicht.
- [220] G. Augelman, H. Fritz, G. Rihs, J. Streith, *J. Chem. Soc. Chem. Commun.* **1982**, 1112.
- [221] A. Defoin, G. Augelman, H. Fritz, G. Geoffroy, C. Schmidlin, J. Streith, *Helv. Chim. Acta* **1985**, *68*, 1998.
- [222] W. Ried, R. Dietrich, *Angew. Chem.* **1963**, *75*, 476; *Angew. Chem. Int. Ed. Engl.* **1963**, *2*, 495.
- [223] J. Metzger in *Comprehensive Heterocyclic Chemistry* 6 (Hrsg.: A. R. Katritzky, C. W. Rees, K. T. Potts), Pergamon Press, Oxford, **1984**, S. 294.
- [224] A. Baretta, B. Waegell, *React. Intermed. (Plenum)* **1982**, *3*, 527.
- [225] K. Shishido, E. Shitara, K. Fukumoto, T. Kametani, *J. Am. Chem. Soc.* **1985**, *107*, 5810.
- [226] J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1961**, *94*, 217; J. Oró, S. S. Amat, *Nature* **1961**, *190*, 442; R. E. Dickerson, *Sci. Am.* **1978**, *239* (9), 70; *Spektrum Wiss.* **1979** (9), 98; siehe auch S. Drenkard, J. Ferris, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1373.