

ASSOR - ALLGEMEINES SIMULATIONSSYSTEM  
ORGANISCHER REAKTIONEN

Wolfgang Schubert

Organisch Chemisches Institut, Technische Universität  
München, Lichtenbergstr. 4, D 8046 Garching

Zur Bewältigung des ständig zunehmenden Tatsachenwissens in der Chemie werden schon seit geraumer Zeit Erkenntnisse und Daten maschinell gespeichert und wieder abgerufen. Die dazu entwickelten Informationssysteme ermöglichen es dem Chemiker, sich über bestimmte Gebiete seines Faches schnell und umfassend zu informieren. Ein Charakteristikum dieser Informationssysteme ist es jedoch, daß nur explizit eingespeicherte Daten als Antwort auf eine Frage an das System erhalten werden können.

Es liegt daher nahe zu untersuchen, ob es nicht möglich ist, Systeme zu entwickeln, die dem Chemiker nicht nur Daten und Fakten über Verbindungen liefern, sondern ihn darüberhinaus auch bei der Lösung so vielschichtiger Probleme wie der Planung von Experimenten und Synthesen unterstützen. Solche Systeme gehen weit über den Rahmen einfacher Informationssysteme hinaus, denn sie erfordern neue Strategien und Techniken. Im folgenden wird ein System beschrieben das auf der Grundlage eines mathematischen Modells der konstitutionellen Chemie chemische Reaktionen simuliert. Auf diese Weise ist es möglich, das Reaktionsverhalten von Verbindungen vorherzusagen und somit Experimente und Synthesen sinnvoll zu planen.

### Ein mathematisches Modell der konstitutionellen Chemie

Das automatisierte Lösen chemischer Probleme erfordert eine Repräsentation von Verbindungen und deren Reaktionen. Die Grundlage hierfür bildet ein bereits an anderer Stelle beschriebenes<sup>1</sup> und daher hier nur kurz aufgezeigtes mathematisches Modell der konstitutionellen Chemie.

Eine bestimmte Anzahl  $n$  von Molekülen  $M_i$  nennen wir ein Ensemble von Molekülen (EM)

$$H = \langle M_1, M_2, \dots, M_n \rangle \quad (1)$$

Die Moleküle  $M_i$  können sowohl alle gleich, teilweise gleich oder auch alle verschieden sein. Weiterhin lassen wir EM zu, die nur ein Molekül enthalten.

Genau wie für ein Molekül ist auch für ein EM eine Bruttoformel definiert. Bezeichnen wir mit  $f_i$  die Bruttoformeln der Moleküle, ist die Bruttoformel des EM,  $F$ , durch

$$F = \sum_{i=1}^n f_i \quad (2)$$

gegeben. Konstitutions- und Strukturformeln für EM sind insofern erklärt, als sie sich aus den Konstitutions- und Strukturformeln der in EM enthaltenen Moleküle zusammensetzen. Die Bruttoformel des EM, bestehend z.B. aus Methan und Kohlendioxid

$$H = \langle \text{CH}_4, \text{CO}_2 \rangle \quad (3)$$

lautet demnach

$$F = C_2H_4O_2 \quad , \quad (4)$$

während sich die Konstitutions- und Strukturformeln aus denen von Methan und Kohlendioxid nebeneinander zusammensetzen.

Zwei EM  $H_i$  und  $H_k$  mit der gleichen Bruttoformel bezeichnen wir als isomere EM:

$$H_i \sim H_k \quad . \quad (5)$$

Dabei lassen wir nicht nur jeweils paarweise isomere Moleküle in  $H_i$  und  $H_k$  zu, sondern erweitern den Isomeriebegriff so, daß auch gänzlich verschiedene Moleküle in  $H_i$  und  $H_k$  erlaubt sind. Zum EM aus Methan und Kohlendioxid ist daher das aus Essigsäure bestehende EM isomer.

$$\langle CH_4, CO_2 \rangle \sim \langle CH_3COOH \rangle \quad (6)$$

Alle zueinander isomeren EM fassen wir in der Familie der isomeren EM (FIEM) zusammen.

$$I = \{H : H_i \sim H_k \text{ für alle Paare } i \text{ und } k\} \quad (7)$$

Die FIEM einer bestimmten Bruttozusammensetzung  $F$  enthält also alle Verbindungen, die die durch die Bruttoformel angegebenen Atome überhaupt eingehen können.

Basistransformationen von EM

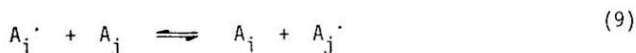
Alle Verbindungen einer bestimmten Bruttozusammensetzung werden durch die FIEM, die zu dieser Bruttozusammensetzung gehört, beschrieben. Alle chemischen Reaktionen werden durch die Transformationen der Elemente der FIEM ineinander dargestellt. Eine Reaktion ist also die Isomerisierung eines EM  $H_i$  in das EM  $H_j$  und durch das geordnete Paar  $(H_i, H_j)$  gegeben. Da wir uns mit Reaktionen selbst beschäftigen wollen, sei durch

$$T_{i \rightarrow j}(H_i) = H_j \quad (8)$$

eine Transformation definiert, die das EM  $H_i$  in das EM  $H_j$  überführt.  $T_{i \rightarrow j}$  ist eine Vorschrift, die besagt, welche Bindungen des EM  $H_i$  gebrochen werden, welche neu hinzukommen und wie freie Elektronen bewegt werden müssen, um von  $H_i$  nach  $H_j$  zu gelangen. Die Transformation  $T_{i \rightarrow j}$  stellt somit den Elektronenumordnungsprozeß dar, der durch die Bruttoreaktion von  $H_i$  nach  $H_j$  durchlaufen wird.

Da chemische Reaktionen, von wenigen Ausnahmen abgesehen, nicht nach konzertierten Mechanismen ablaufen, sondern in einzelne Reaktionsschritte zerlegt werden können, ist es sinnvoll, für diese einzelnen Elementarschritte oder Elementarreaktionen Transformationen zu definieren.

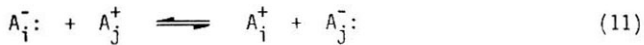
a) Der einfachste Redoxprozeß, bei dem lediglich ein Elektron von einem Atom  $i$  auf ein anderes  $j$  übertragen wird, sei durch die Transformation  $e_{ij}$  repräsentiert.



b) Die Homoapsis und Homolyse einer Bindung zwischen den Atomen  $i$  und  $j$



seien durch die Transformationen  $\tau_{ij}$  bzw.  $-\tau_{ij}$  repräsentiert.  
c) Die Redoxreaktion, bei der zwei Elektronen von einem Atom  $i$  auf ein anderes  $j$  übertragen werden, sei durch  $(\theta_{ij}, \theta_{ij})$  definiert. Wir verstehen darunter zweimaliges Ausführen von  $\theta_{ij}$ .



d) Die Heteroapsis und die Heterolyse einer Bindung



zwischen den Atomen  $i$  und  $j$  seien durch die Transformationen  $(\theta_{ij}, \tau_{ij})$  bzw.  $-(\theta_{ij}, \tau_{ij})$  dargestellt.

Wir gehen davon aus, daß sich durch eine Aufeinanderfolge von Transformationen  $\theta_{ij}$  und  $\tau_{ij}$  sich alle chemischen Reaktionen darstellen lassen. Alle Reaktionen, an denen nur Valenzschalen mit gerader Anzahl von Elektronen beteiligt sind, sind insbesondere durch eine Folge der Transformationen  $(\theta_{ij}, \theta_{ij})$  und  $(\theta_{ij}, \tau_{ij})$  darstellbar. Wir nennen deshalb die Transformationen  $\theta_{ij}$ ,  $\tau_{ij}$  und  $(\theta_{ij}, \theta_{ij})$ ,  $(\theta_{ij}, \tau_{ij})$  Basistransformationen. Daß alle chemischen Reaktionen sich als Linearkombinationen von Basistransformationen darstellen lassen, ist anhand der Matrizendarstellung von Basistransformationen gezeigt worden .

### Die Simulation chemischer Reaktionen

In Gl. (8) ist die Transformation von EM durch zwei Elemente einer FIEM, also zwei isomere EM definiert worden. Es wurde zu einem EM  $H_i$  das dazu isomere EM  $H_j$  vorausgesetzt, um die Transformation  $T_{i \rightarrow j}$  zu erklären. Will man allerdings Reaktionen eines EM simulieren, seine Reaktionsmöglichkeiten aufzeigen, ist das isomere EM  $H_j$  nicht bekannt und damit auch nicht die Transformation  $T_{i \rightarrow j}$  gegeben, die  $H_i$  in  $H_j$  überführt. In der Gleichung

$$T_{i \rightarrow j}(H_i) = H_j \quad (13)$$

stehen folglich zwei Unbekannte, die Transformation  $T_{i \rightarrow j}$  und das EM  $H_j$ . Die Lösungsmenge zu Gl. (13) ist jedoch nicht beliebig groß, denn es gibt nur endlich viele EM  $H_j$ , in die das EM  $H_i$  transformiert werden kann. Diese Beschränkung der Lösungsmenge resultiert aus den zu Gl. (13) existierenden Nebenbedingungen, die erfüllt sein müssen, damit  $H_j$  ein chemisch sinnvolles EM darstellt. Weiterhin gibt es zu zwei isomeren EM  $H_i$  und  $H_j$  zwar nur eine Transformation  $T_{i \rightarrow j}$ , die die Brutto-reaktion repräsentiert, doch diese ist oft nicht mit einer chemischen Reaktion zu identifizieren, sondern nur als eine Folge von Reaktionen interpretierbar, so daß, will man Reaktionen simulieren, ein Arbeiten mit der Transformation  $T_{i \rightarrow j}$  in den meisten Fällen ausscheidet. Lösungsansätze zur Gl. (13) und damit zur Simulation chemischer Reaktionen bestehen folglich in der Auswahl von Transformationen, aus denen  $T_{i \rightarrow j}$  zusammengesetzt werden kann und in der Beschränkung der Lösungsmenge durch Analysieren der EM nach chemischen Kriterien, bevor eine Transformation ausgeführt wird und nachdem sie ausgeführt worden ist.

### Auswahl von Transformationen

Die Auswahl von Sätzen von Transformationen kann nach mehreren Gesichtspunkten geschehen. Es liegt nahe, die Transformationen, die den bekannten chemischen Reaktionen entsprechen, zu einem Satz zusammenzufassen. Dazu muß man die in der Literatur bekannten Reaktionen auf ihre Transformationen zurückführen. Ein solcher Satz von Transformationen ist nicht abgeschlossen, sondern müßte mit fortschreitender Erkenntnis und dem Entdecken neuer Reaktionen ständig erweitert werden. Bei der Anzahl der heute bekannten Reaktionen ist ein solcher Satz von Transformationen bereits so groß, daß, um ihn in Systemen zur Simulation chemischer Reaktionen einsetzen zu können, er nicht vollständig herangezogen werden kann, selbst wenn man berücksichtigt, daß viele Reaktionen durch dieselbe Transformation repräsentiert werden<sup>2</sup>. Durch die Notwendigkeit, nur mit einer Auswahl solcher Transformationen arbeiten zu können, die bekannten Reaktionen entspricht, wird die allgemeine Verwendbarkeit eines Systems zur Simulation chemischer Reaktionen stark eingeschränkt.

Ein anderer Satz von Transformationen kann dadurch gebildet werden, daß nur eine bestimmte Anzahl gebrochener und geknüpfter Bindungen und an der Reaktion beteiligte Elektronen zugelassen werden. Eine solche Annahme ist durchaus vernünftig, denn die Anzahl gebrochener und geknüpfter Bindungen und beteiligter Elektronen an einer Reaktion ist tatsächlich begrenzt. Die Sätze von Transformationen, in denen  $n$  Bindungen gebrochen und jeweils  $m$  Bindungen mit  $n > m \geq 0$  neu geknüpft werden, müssen jedoch nicht vollständig sein in dem Sinne, daß sich alle nur denkbaren Transformationen aus ihnen zusammensetzen lassen. Werden solche Sätze von Transformationen so ergänzt, daß sie vollständig sind, enthalten sie nicht unbedingt nur Transformationen, die auch Reaktionen entsprechen können. Außerdem werden solche Sätze dann wiederum so groß, daß sie sich, bei den

heutigen Möglichkeiten, Systeme zur Simulation chemischer Reaktionen zu realisieren, nicht verwenden lassen.

Bei den erwähnten beiden Möglichkeiten, Sätze von Transformationen für die Simulation chemischer Reaktionen vorzugeben, begibt man sich vor allem der Möglichkeit, neue chemische Reaktionen zu entdecken, denn es können immer nur solche Reaktionen erhalten werden, die schon bekannt sind, weil nur diese in den Sätzen repräsentiert sind. Das hier zugrunde gelegte mathematische Modell läßt jedoch durch die Formulierung von Reaktionen als Isomerisierungen von EM gerade auch die Behandlung präzedenzloser Reaktionen zu.

Die genannten Probleme bei der Wahl von Sätzen von Transformationen werden umgangen, wählt man den kleinsten vollständigen Satz von Transformationen, nämlich die Basistransformationen.

Dabei genügt es, entweder die Transformationen  $\theta_{ij}$ ,  $\tau_{ij}$  bzw.  $(\theta_{ij}, \theta_{ij})$ ,  $(\theta_{ij}, \tau_{ij})$  zu verwenden, je nachdem, ob auch radikalische Reaktionen oder nur solche mit gerader Anzahl von Valenzelektronen betrachtet werden sollen. Durch diese aus den Basistransformationen bestehenden Sätze lassen sich jeweils alle Reaktionen simulieren, und außerdem kommt jeder Basistransformation eine unmittelbare chemische Bedeutung zu.

Ein System zur Simulation chemischer Reaktionen, das auch präzedenzlose Reaktionen behandeln kann und dabei aber immer chemisch sinnvolle Reaktionswege vorschlägt, kann daher nur mit Basistransformationen arbeiten. Durch die präzedenzlosen Reaktionen in der Menge der erhaltenen Reaktionswege erhält ein solches System eine gewisse Vorhersagekraft, so daß es zur systematischen Suche neuer Reaktionen eingesetzt werden kann.



Basistransformationen und Reaktionen

Unabhängig von der tatsächlichen Implementation des beschriebenen Systems zur Simulation chemischer Reaktionen soll hier an ein paar einfachen Beispielen gezeigt werden, wie durch die Anwendung der Basistransformationen auf sehr einfache EM eine Reaktionssequenz erhalten wird.

Die Transformationen  $T_{i \rightarrow j}$ , die die EM  $H_i$  in die isomeren  $H_j$  überführen, sind, abgesehen von einigen sehr einfachen Fällen, eine Folge von Basistransformationen. Abhängig von der Komplexität der Bruttoreaktion  $T_{i \rightarrow j}$  setzt sich die Folge der Basistransformation aus mehr oder weniger vielen Elementen zusammen. An einigen einfachen Reaktionen soll gezeigt werden, wie sie durch Basistransformationen dargestellt werden. Es sei

$$T = (\theta_{12}, \tau_{12}), -(\theta_{23}, \tau_{23}) \quad (14)$$

die Transformation, durch die das EM

$$H_i = \langle A_1^-, A_2 - A_3 \rangle \quad (15)$$

in das EM

$$H_j = \langle A_1 - A_2, :A_3^- \rangle \quad (16)$$

überführt wird. Zuerst wird durch die Transformation  $(\theta_{12}, \tau_{12})$  das Formen der Bindung zwischen den Atomen 1 und 2 simuliert, dann durch  $-(\theta_{23}, \tau_{23})$  das Brechen der Bindung zwischen den Ato-

men 2 und 3. Die Transformation in Gl. (14) simuliert also den Verlauf einer nucleophilen Substitution über den  $S_N2$ -Mechanismus.  $(\theta_{12}, \tau_{12})$  simuliert die Reaktion zum Übergangszustand  $A_1-A_2^-A_3$  hin und  $(\theta_{23}, \tau_{23})$  die Weiterreaktion zu den Produkten. Dies läßt sich auch wie folgt schreiben

$$(\theta_{12}, \tau_{12}) \langle A_1^-, A_2-A_3 \rangle = \langle A_1-A_2^--A_3 \rangle \quad (17)$$

und

$$-(\theta_{23}, \tau_{23}) \langle A_1-A_2^--A_3 \rangle = \langle A_1-A_2, :A_3^- \rangle \quad (18)$$

oder auch

$$-(\theta_{23}, \tau_{23}) ((\theta_{12}, \tau_{12}) H_i) = H_j \quad (19)$$

Werden die beiden Basistransformationen von T in anderer Reihenfolge angewandt

$$-(\theta_{23}, \tau_{23}) \langle A_1^-, A_2-A_3 \rangle = \langle A_1^-, A_2^+, :A_3^- \rangle \quad (20)$$

$$(\theta_{12}, \tau_{12}) \langle A_1^-, A_2^+, :A_3^- \rangle = \langle A_1-A_2, :A_3^- \rangle \quad (21)$$

oder

$$(\theta_{12}, \tau_{12}) (-(\theta_{23}, \tau_{23}) H_i) = H_j \quad (22)$$

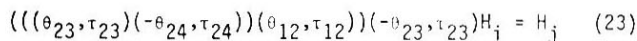
sieht man sofort, daß auf diese Weise die nucleophile Substitution nach dem  $S_N1$ -Mechanismus simuliert wird.

Während  $H_i$  und  $H_j$  in Gln. (19) und (22) jeweils identisch sind, wird durch die Vertauschung der Basistransformationen in beiden Fällen ein anderer Reaktionsweg beschrieben. Daraus folgt, daß die einzelnen Basistransformationen einer Transformation i.allg. nicht vertauschbar sind.

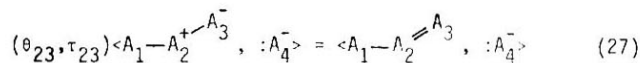
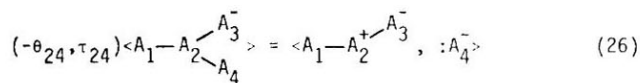
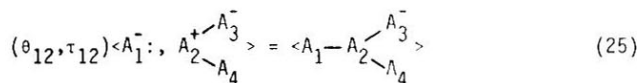
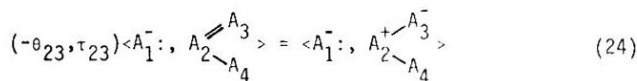
Die Transformation der Gl. (22) repräsentiert nicht nur die nucleophile Substitution nach dem  $S_N1$ -Mechanismus, sondern auch die elektrophile Substitution nach dem  $S_E1$ -Mechanismus. Die elektrophile Substitution nach dem  $S_E2$ -Mechanismus läßt sich durch die Transformation in Gl. (19) darstellen, führt aber über einen Übergangszustand, der sich nicht durch eine

Valence-Bond-Grenzstruktur darstellen läßt. Die elektrophile Substitution am Aromaten besitzt ebenfalls einen Übergangszustand, der sich nur durch mehrere Valence-Bond-Grenzstrukturen beschreiben läßt. Diese Grenzstrukturen werden alle erzeugt, und die Sequenz der Basistransformationen vom Ausgangs-EM  $H_i$  zum stabilen EM  $H_j$  wird entsprechend tief geschachtelt. In Gln. (19) und (22) sind die Transformationen nur bis auf zwei Ebenen geschachtelt.

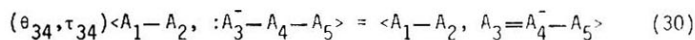
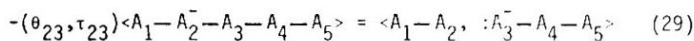
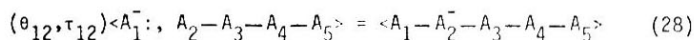
Die nucleophile Substitution an einem ungesättigten System läßt sich durch die Transformation



darstellen, wenn der Mechanismus



zugrundegelegt wird. Ebenfalls durch eine Folge von vier Basistransformationen ist die Elimination



$$-(\theta_{45}, \tau_{45}) \langle A_1 - A_2, A_3 \rightleftharpoons A_4^- - A_5 \rangle = \langle A_1 - A_2, A_3 = A_4, :A_5^- \rangle \quad (31)$$

gegeben.

Den angegebenen Beispielen liegen sehr vereinfachende Annahmen über den Reaktionsablauf zugrunde. Es ist bekannt, daß die meisten Reaktionen durch andere Stoffe katalysiert werden, die an irgendeinem Punkte am Reaktionsgeschehen beteiligt sind. Bezieht man solche Reaktionsschritte mit ein, in denen z.B. ein Katalysator gebunden und wieder abgespalten wird, ändert sich auch die Folge der Basistransformationen entsprechend.

### Analyse und Auswahl von EM

Über die Wahl der Transformationen  $T_{i \rightarrow j}$  in Gl. (13) hinaus ist es auch durch die Analyse der Konstitution des zu transformierenden EM möglich, die Vielzahl der Lösungen einzuschränken. An erster Stelle ist hier die sogenannte fitting-Bedingung <sup>1</sup> zu nennen, die immer erfüllt sein muß, andernfalls negative Bindungsordnungen im EM  $H_j$  entstünden. Da wir immer nur Basistransformationen verwenden, läßt sich diese fitting-Bedingung auch wie folgt formulieren: es darf nicht dort eine Bindung gebrochen werden, wo keine ist. Daraus folgt aber andererseits, daß die fitting-Bedingung bei der Verwendung von Basistransformationen ohnehin eingehalten wird.

Durch konstitutionsbedingte Symmetrieelemente können beim Transformieren der EM Redundanzen entstehen. Dies ist dann der Fall, wenn äquivalente Reaktionszentren zum Ausgangspunkt für Reaktionswege werden. Identische Reaktionswege werden vermieden, wenn äquivalente Reaktionszentren erkannt werden. Reaktionszentren werden als äquivalent angesehen, wenn sie gleiche Arten und Anzahlen von konstitutionell äquivalenten Atomen und Bindungen enthalten. Athan z.B. hat sieben Bindungen, von denen sechs jeweils konstitutionell äquivalent sind. Die Bindungen bilden somit zwei Äquivalenzklassen, aus denen jeweils nur eine Bindung gewählt zu werden braucht, um die reaktiven Zentren des Moleküls vollständig zu beschreiben. Die Anzahl der Reaktionsmöglichkeiten ist auf diese Weise durch Symmetrieeerkennung von sieben auf zwei reduziert worden.

Diese mathematischen Kriterien schränken zwar die Anzahl der Lösungen zu Gl. (13) bereits drastisch ein, doch kann aufgrund heuristischer Regeln die Vielfalt der Lösungen noch weiter reduziert werden. Von allen noch verbleibenden Reaktionszentren eines EM sind nicht alle der Ausgangspunkt für Reakti-

onen, die unter normalen Laborbedingungen auch gleich gut ablaufen. Ausgehend von der Tatsache, daß zwischen Struktur und Reaktivität ein enger Zusammenhang besteht, wird deshalb versucht, aus den Strukturmerkmalen der in den EM vorhandenen Molekülen auf die reaktiven Zentren zu schließen, die höchstwahrscheinlich Ausgangspunkt für eine Reaktion sein werden. Dabei werden keine absoluten Maßstäbe angelegt. Es werden alle Zentren nach den gleichen Kriterien untersucht und dann jene ausgewählt, die in dem betreffenden Molekül am reaktivsten sind. Ein bestimmtes Reaktionszentrum, das in einem Molekül als wenig reaktiv eingestuft worden ist, kann in einem anderen als durchaus reaktiv angesehen werden. Die Einstufung als reaktiv oder nicht reaktiv hängt von den noch vorhandenen Reaktionszentren ab. Dieses Vorgehen stellt insofern keine prinzipielle Einschränkung des allgemeinen Ansatzes des mathematischen Modells dar, als auf diese Weise lediglich unwahrscheinliche Reaktionen gar nicht erst simuliert werden. Die Grenze für die Einstufung eines Zentrums als reaktiv oder nicht reaktiv ist variabel. Sie kann den Erfordernissen bestimmter Probleme angepaßt und gegebenenfalls auch so weit hinausgeschoben werden, daß alle Zentren als reaktiv eingestuft und alle Reaktionsmöglichkeiten simuliert werden.

Ein allgemeiner Ansatz zum Auffinden der reaktivsten Zentren eines Moleküls ist nur möglich, wenn auf das Absuchen von Listen vorgegebener Reaktivitätsabstufungen bestimmter Gruppierungen verzichtet wird. Solche Listen wären so umfangreich, daß sie in derzeit realisierbaren Systemen zur Simulation von Reaktionen nicht verwandt werden könnten, und deshalb nur unvollständige Listen in Betracht kämen.

### Erzeugen des Reaktionsnetzwerkes

Von einem EM  $H_j$  ausgehend werden durch Anwenden der Basis-  
transformationen isomere EM  $H_j$  erzeugt. Diese EM  $H_j$  werden im  
nächsten Schritt wiederum zu EM  $H_k$  transformiert usw. und so  
ein Netzwerk von EM aufgespannt. Die Reaktionswege sind dadurch  
gegeben, daß man die EM, die auseinander hervorgegangen sind,  
entlanggeht.

Die erzeugten EM können jedoch nicht ohne weitere Prüfung  
in das Netzwerk eingefügt werden. Insbesondere werden die Va-  
lenzzustände aller Atome eines EM überprüft. Dadurch wird ver-  
hindert, daß valenzchemisch nicht akzeptable EM zum Ausgangs-  
punkt neuer Reaktionswege werden. Ob der Valenzzustand eines  
Atoms als akzeptabel gilt, wird durch Vergleichen mit einer  
Liste entschieden, die die oberen und unteren Grenzen für die  
Anzahl der Valenzelektronen enthält. Diese Grenzen sind i. allg.  
so weit gefaßt, daß auch mechanistische Zwischenstufen erlaubt  
sind, können ansonsten aber prinzipiell beliebig eng oder weit  
angenommen werden, den Erfordernissen der besonderen Probleme  
entsprechend.

Durch die Prüfung der Valenzzustände wird eine Aussage  
über den Energieinhalt des EM gewonnen, obwohl eine Berechnung  
der Energie nicht vorgenommen wird. Wegen der großen Anzahl  
der zu handhabenden EM sind Berechnungen von Energien meist  
auch gar nicht möglich. Quantenmechanische Berechnungen schei-  
den wegen des zu großen Aufwandes aus, und empirische Methoden  
geben nur selten zuverlässige Werte über die Substanzklassen  
hinaus, aus denen die Parameter zur Berechnung der Energien ab-  
geleitet sind.

Bei der Erzeugung der Reaktionspfade ist es nicht ausge-  
schlossen, daß identische EM über verschiedene Wege entstehen.  
Diese identischen EM wären beim weiteren Aufspannen des Netz-  
werkes der Ausgangspunkt gleicher Reaktionswege, führten daher  
zu Redundanzen. Um die Komplexität des Netzwerkes so gering

wie möglich zu halten, ist es erforderlich, die Identität von EM zu erkennen, um die Pfade, die zu den identischen EM führen, zusammenlaufen zu lassen. Von dem so entstandenen Knoten des Netzwerkes, auf den mehrere Wege zulaufen, werden dann die Reaktionen des EM weiterverfolgt.

Das Aufspannen des Reaktionsnetzwerkes läßt sich als ein algorithmisches Verfahren auffassen. Ein solches Verfahren und der entsprechende Algorithmus sind jedoch nur verwendbar, wenn sie ein Ergebnis nach einer endlichen Anzahl von Schritten liefern. Daraus folgt, daß für das Aufspannen des Reaktionsnetzwerkes Abbruchkriterien definiert werden müssen. Wir fordern, daß alle EM für weitere Transformationen blockiert sein müssen, damit das Aufspannen des Reaktionsnetzwerkes abbricht. EM werden blockiert, wenn sie Atome mit nicht akzeptablen Valenzzuständen enthalten, oder wenn die Moleküle eines EM in einer Liste von Molekülen enthalten sind, die als bekannt und in ihrem Reaktionsverhalten als vollkommen charakterisiert angesehen werden. In dieser Liste könnten z.B. alle käuflichen Chemikalien stehen. Wie das Erkennen der Identität von EM, um das Erzeugen gleicher Reaktionspfade zu verhindern, erfordert das Auffinden von Molekülen in einer Liste eine eindeutige Darstellung der Moleküle.

Um den Aufbau des Reaktionsnetzwerkes an nicht interessierenden Stellen temporär unterdrücken zu können, sollte ferner die Möglichkeit bestehen, bestimmte EM außerhalb der oben genannten Kriterien zu blockieren.



### Repräsentation und Manipulation von EM

In der Vergangenheit sind viele Versuche unternommen worden, die in den zweidimensionalen Konstitutions- und Strukturformeln enthaltenen Aussagen über eine chemische Verbindung in lineare Darstellungen umzusetzen<sup>3</sup>, um sie dadurch einer maschinellen Bearbeitung zugänglich zu machen. Das Prinzip einer solchen Linearisierung ist ähnlich dem der Verbalisierung von Konstitutions- bzw. Strukturformeln. Eine Formel wird dazu in Teile zerlegt, die als allgemein bekannt vorausgesetzt werden, und der Name wird durch Aneinanderfügen der Bezeichnungen für die einzelnen Fragmente gebildet.

Eine Linearisierung von zweidimensionalen Formeln ist für die Simulation chemischer Reaktionen allerdings nicht geeignet. Abgesehen von der Vielfalt der Regeln, die erforderlich wäre, um lineare Darstellungen zu erzeugen, sind immer nur die Eigenschaften einer Verbindung in der linearen Darstellung repräsentiert, die mit einem Fragment assoziiert werden können. Es geht also bei der Erzeugung solcher Linearcodes Information unwiederbringlich verloren.

Eine direktere Darstellung der Konstitution von Molekülen, die auch auf EM und deren Transformationen übertragbar ist, gelingt durch Graphen. In dieser Darstellung müssen die Zerlegung eines EM in seine Moleküle, der Aufbau eines EM aus Molekülen und die Transformationen von EM realisiert sein.

### EM-Graphen

Die Konstitution eines EM kann als Graph dargestellt werden, indem die Atomrümpfe auf die Knoten eines Graphen und die Bindungen auf dessen Kanten abgebildet werden. Weiterhin wer-

den die Ordnungszahlen und die freien Elektronen auf Marken abgebildet, die den Knoten zugeordnet werden. Die Bindungsordnungen und gegebenenfalls auch andere Bindungseigenschaften werden auf Marken abgebildet, die den Kanten des Graphen zugeordnet werden.

Es resultiert ein endlicher, nicht notwendig zusammenhängender, ungerichteter und markierter Graph

$$M = (A, Z, F, B, L, \omega, \phi, \beta, \lambda) \quad , \quad (32)$$

den wir den EM-Graphen oder, falls das EM nur ein Molekül enthält, den Molekülgraphen nennen. Hierin bedeuten

**A** die Menge der Knoten,

**Z** die Menge der Marken, auf die die Ordnungszahlen abgebildet werden,

**F** die Menge der Marken, auf die die freien Elektronen abgebildet werden,

**B** die Menge der Kanten und

**L** die Menge der Marken, auf die die Bindungsordnungen abgebildet werden.

Die Vorschriften  $\omega$ ,  $\phi$ ,  $\beta$  und  $\lambda$  sind definiert gemäß:

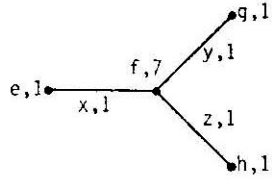
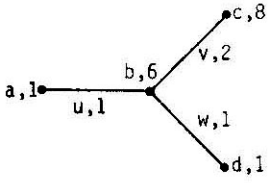
$$\omega(p) = z \quad p \in A, z \in Z \quad , \quad (33)$$

$$\phi(p) = f \quad p \in A, f \in F \quad , \quad (34)$$

$$\beta(e) = (p, q) \quad e \in B, p \in A, q \in A \quad \text{und} \quad (35)$$

$$\lambda(e) = l \quad e \in B, l \in L \quad . \quad (36)$$

Während ein Molekülgraph immer zusammenhängend ist, ist dies für einen EM-Graphen nicht notwendig. Das EM aus Formaldehyd und Ammoniak ist durch den Graphen 1 gegeben.



### Symmetrieeigenschaften von EM

Für ein System zur Simulierung chemischer Reaktionen ist das Erkennen der Symmetrie in Molekülen insofern von entscheidender Bedeutung, als das mehrfache Erzeugen identischer Produkte von äquivalenten Reaktionszentren aus vermieden wird. Da die Darstellungen chemischer Verbindungen innerhalb unseres mathematischen Modells keine Angaben über die Ausdehnung eines Moleküls im dreidimensionalen Raum enthalten, kann nicht die Symmetrie von Molekülen im dreidimensionalen Raum untersucht werden.

Die Symmetrie von Molekülen im dreidimensionalen Raum wird gewöhnlich mittels Punktgruppen beschrieben. Die Punktgruppen basieren auf den orthogonalen Transformationen des dreidimensionalen Raumes, wie Rotationen, Reflexionen und Translationen. Bei der Untersuchung von Symmetrieeigenschaften der Konstitution eines Moleküls braucht hingegen ein Raum, in dem die Symmetrieoperationen erklärt sind, gar nicht zu existieren. Die konstitutionelle Symmetrie ist daher eine allgemeinere Symmetrie als die durch Punktgruppen beschriebene des dreidimensionalen Raumes.

Da die Konstitution eines Moleküls vorteilhaft durch Graphen beschrieben wird, liegt es nahe, die Symmetrieeigenschaften der Konstitution eines Moleküls durch die Untersuchung des der Konstitutionsformel unterliegenden Molekülgraphen zu bestimmen.

Die Symmetrieeigenschaften eines Graphen werden durch dessen Automorphismengruppe<sup>5</sup> beschrieben. Ein Automorphismus ist die Abbildung der Knotenmenge eines Graphen auf sich selbst unter Erhaltung der Nachbarschaften aller Knoten. Wird ein Knoten auf einen anderen abgebildet, so sagen wir, die beiden Knoten sind ähnlich. Knoten, die auf sich selbst abgebildet werden, nennen wir Fixknoten<sup>4,5</sup>.

Für das vorliegende Problem, die Erkennung der konstitutionellen Symmetrie von Molekülen, ist es jedoch nicht erforderlich, die Automorphismengruppe des Graphen explizit zu ken-

nen, sondern nur die untereinander ähnlichen Knoten. Die Eigenschaft der Knoten, ähnlich zu sein, stellt eine Äquivalenzrelation dar und gibt Anlaß zur Bildung von Äquivalenzklassen<sup>6</sup>. Die Teilung der Knotenmenge eines Graphen in Äquivalenzklassen nennen wir die Äquivalenzteilung der Knotenmenge.

Der von uns für die Bestimmung der konstitutionellen Symmetrie von Molekülen entwickelte Algorithmus<sup>7</sup> erzeugt die Äquivalenzteilungen der Knoten- und Kantenmenge der Molekülgraphen. Anhand der Äquivalenzteilungen lassen sich Aussagen über die konstitutionelle Symmetrie eines Moleküls machen. Fallen z.B. alle Knoten eines Molekülgraphen in eine Äquivalenzklasse, ist das Molekül konstitutionell symmetrisch. Bildet dagegen jeder Knoten eine Äquivalenzklasse, ist das Molekül konstitutionell unsymmetrisch.

Konstitutionelle Äquivalenz von Atomen

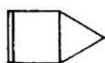
Mit Hilfe der Äquivalenzteilung der Knotenmenge eines Molekülgraphen läßt sich die konstitutionelle Äquivalenz von Atomen wie folgt definieren.

Definition

Sind irgend zwei Knoten eines Molekülgraphen ähnlich unter einem Automorphismus, so sind die durch die Knoten repräsentierten Atome konstitutionell äquivalent.

Aus dieser Definition folgt, daß von der Äquivalenzteilung unmittelbar auf die konstitutionelle Äquivalenz von Atomen geschlossen werden kann, und zwar sind alle jene Atome äquivalent, deren Knoten derselben Äquivalenzklasse angehören.

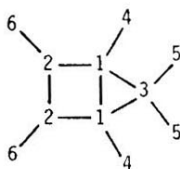
Die Äquivalenzklassen der Atome von z.B. Bicyclo-[2.1.0]pent-2-en (2) und Pentacyclo-[5.1.0.0<sup>2</sup>.6.0<sup>3</sup>.5.0<sup>4</sup>.8]-octan (3) sind in (4) und (5) dargestellt, und zwar sind die mit gleichen Ziffern bezeichneten Atome in (4) und (5) konstitutionell äquivalent.



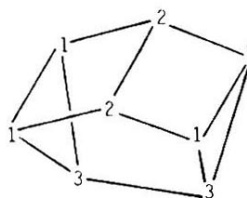
(2)



(3)



(4)



(5)

### Konstitutionelle Äquivalenz von Bindungen

Da die Eigenschaften von Bindungen durch die an der Bindungsbildung beteiligten Atome bestimmt sind, liegt es nahe, auch die konstitutionelle Äquivalenz von Bindungen aus der konstitutionellen Äquivalenz der Atome herzuleiten. So schliesse man z.B. von der Äquivalenz aller C-Atome in Benzol auf die Äquivalenz aller C-C-Bindungen in Benzol. Am Tricyclo[4.2.0.0<sup>2,4</sup>]octan-Gerüst

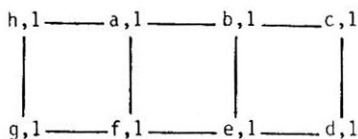


(6)

läßt sich jedoch zeigen, daß dieser Schluß allgemein nicht zulässig ist. Die Atome des mittleren Vierringes sind konstitutionell äquivalent, die Bindungen hingegen sind nur paarweise konstitutionell äquivalent, denn je zwei Bindungen gehören einem bzw. zwei Vierringen an. Um die konstitutionelle Äquivalenz von Bindungen zweifelsfrei zu bestimmen, ist es also erforderlich, einen anderen Weg zu beschreiten.

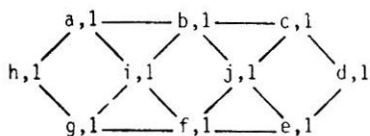
Wir fassen dazu die Kantenmenge eines Molekülgraphen  $M$  als die Knotenmenge eines anderen Graphen  $K(M)$  auf, den wir den Kantengraphen<sup>4</sup> des Molekülgraphen nennen, und in dem die Knoten dann verbunden sind, wenn die Kanten im Molekülgraphen  $M$  benachbart sind. Sowie die Kanten des Molekülgraphen die Knoten des Kantengraphen werden, werden aus den Marken der Kanten im Molekülgraphen unter Einbeziehung der Marken der Knoten, auf die die Kanten jeweils zulaufen, die Marken der Knoten des Kantengraphen. Marken für die Kanten des Kantengraphen brauchen nicht erklärt zu werden.

Aus dem Molekülgraphen



(7)

von (6) ergibt sich dann dessen Kantengraph



(8)

Wird vom Kantengraph (8) die Äquivalenzteilung der Knotenmenge bestimmt, läßt sich wie bei den Atomen eines Moleküls auf die konstitutionelle Äquivalenz von Bindungen schließen.

Wir definieren die konstitutionelle Äquivalenz von Bindungen wie folgt:

Definition

Sind irgend zwei Knoten des Kantengraphen eines Molekülgraphen ähnlich unter einem Automorphismus, so sind die durch sie repräsentierten Bindungen konstitutionell äquivalent.

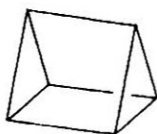
Aus der Äquivalenzteilung



$$A' = \{a,c,e,g\} \cup \{h,d\} \cup \{i,j\} \cup \{b,f\} \quad (37)$$

der Knotenmenge  $V'$  des Kantengraphen  $K(M)$  (8) ist sofort ersichtlich, daß die Kanten  $(a,b)$  und  $(a,f)$  nicht ähnlich sein können, denn die Knoten  $i$  und  $b$  von  $K(M)$ , die die Bindungen  $(a,f)$  bzw.  $(a,b)$  repräsentieren, sind nicht in derselben Äquivalenzklasse von  $A'$  enthalten. Andererseits ist leicht zu sehen, daß der Kantengraph des Molekülgraphen vom Benzolgerüst, genau wie der Molekülgraph selbst, ein regulärer Graph vom Grad zwei mit sechs äquivalenten Knoten ist, und daher alle Bindungen im Benzol äquivalent sind.

Ein weiteres Beispiel, bei dem von der konstitutionellen Äquivalenz der Atome nicht auf die konstitutionelle Äquivalenz der Bindungen geschlossen werden kann, stellt das Tetracyclo-[3.1.0.0<sup>2.4</sup>.0<sup>3.6</sup>]-hexan (Prisman) (9) dar.



(9)

### Reaktive Zentren von EM

In einem System zur Simulierung und Vorhersage chemischer Reaktionen ist das Auffinden der unter bestimmten Bedingungen reaktionsfähigsten Zentren eines Moleküls von entscheidender Bedeutung. Unter den valenzchemisch erlaubten Reaktionsmöglichkeiten der Moleküle eines EM sind es meist nur wenige, die unter üblichen Reaktionsbedingungen auch ablaufen können, denn ein Molekül ist in seinem chemischen Verhalten dadurch charakterisiert, daß es nur ganz bestimmte Reaktionen eingeht. Ein System, das die Anzahl der valenzchemisch erlaubten kombinatorischen Möglichkeiten vorschlägt, präsentiert zwar die Gesamtheit der Transformationen eines EM in seine isomeren EM, es bleibt dann aber die Auswahl der interessanten Lösungen dem Anwender überlassen. Diese nachträgliche Auswahl der für eine experimentelle Überprüfung in Frage kommenden vorgeschlagenen Reaktionen ist selbst bei nicht allzugroßen EM und deren valenzchemisch erlaubten Reaktionen eine praktisch unlösbare Aufgabe. Eine Bestimmung der reaktiven Zentren der Moleküle eines EM und die Reduzierung auf diese reaktiven Zentren verringert die Anzahl der vorgeschlagenen Isomerisierungen eines EM, so daß eine nachträgliche Bewertung mit vertretbarem Aufwand durchzuführen ist.

Es sind verschiedene Ansätze zur Bestimmung reaktiver Zentren möglich. Die einfachste, unspezifischste aber schnellste Methode ist die Markierung reaktiver Bindungen im Molekül aufgrund der Art der gebundenen Atome oder der Bindungsordnung. Auf diese Weise könnten z.B. Kohlenstoff-Heteroatom-Bindungen, Bindungen, an denen Wasserstoff beteiligt ist und Mehrfachbindungen als reaktiv eingestuft werden. Allein die Anzahl der Bindungen, an denen Wasserstoff beteiligt ist, macht im Durchschnitt die Hälfte aller Bindungen eines Moleküls aus. Außerdem wird die chemische Umgebung, in der sich die Bindung befindet, auf diese Weise nicht berücksichtigt. Eine C-H-Bindung in

der Aldehydgruppe wird z.B. einer C-H-Bindung in der Methylgruppe gleichgesetzt. Auch die Aktivierung entfernter Bindungen, wie etwa einer C-H-Bindung in  $\beta$ -Dicarbonylverbindungen, wird auf diese Weise nicht erfaßt. Um von vornherein zu einer größeren Differenzierung bei der Beurteilung der Reaktivität von Bindungen zu gelangen, haben wir einen anderen Weg zur Erkennung reaktiver Zentren beschritten, der von der Ladungsverteilung in einem Molekül und der Polarität von Bindungen ausgeht.

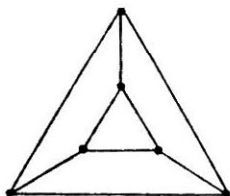
Das Bestimmen der Reaktionsorte eines EM dient der Beschränkung der Anzahl der isomeren EM, die aus einem EM durch Basistransformationen erzeugt werden können. Reaktionsorte eines EM lassen sich jedoch im allgemeinen nicht durch einfaches Berechnen irgendeiner Größe bestimmen, sondern erfordern die Berücksichtigung einer Vielfalt von Fakten, die sich nur in Form heuristischer Regeln in einem System zur Simulation chemischer Reaktionen inkorporieren lassen. Die Bestimmung von Reaktionsorten stellt deshalb nur einen Ansatz dar, auf dem Wege solche heuristischen Regeln zu entwickeln und vor allen Dingen ihre Brauchbarkeit am System selbst zu überprüfen.

Dazu wird die Ladungsverteilung eines Moleküls rekursiv aus korrigierten Gruppenelektronegativitäten<sup>8</sup> berechnet. Von aus der Ladungsverteilung erhaltenen Polaritäten von Bindungen wird auf deren Reaktivität geschlossen. Bindungen werden desto reaktiver eingestuft je polarer sie sind. Dabei werden keine absoluten Maßstäbe angelegt, sondern eine in einem EM als wenig reaktiv klassifizierte Bindung kann in einem anderem EM sehr wohl als reaktiv gelten. Auf diese Weise wird erreicht, daß nur die Reaktionen ausgehend von den reaktivsten Zentren eines EM simuliert werden<sup>9</sup>.

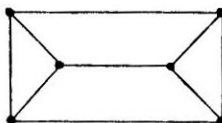
### Eindeutige Darstellung von Molekülen

Zwei Moleküle besitzen die gleiche Konstitution, wenn ihre Molekülgraphen isomorph sind. Diese Tatsache hat weitreichende Konsequenzen, sollen Moleküle anhand ihres Molekülgraphen verglichen bzw. identifiziert werden. Für den Vergleich zweier Moleküle anhand ihrer Molekülgraphen muß also festgestellt werden, ob beide Molekülgraphen isomorph sind.

Zwei Graphen  $M_1$  und  $M_2$  sind isomorph, wenn eine bijektive Abbildung ihrer Knotenmengen aufeinander existiert, so daß je zwei benachbarte Knoten  $p$  und  $q$  von  $M_1$  auch in  $M_2$  benachbart sind<sup>5</sup>. Folgende beiden Graphen sind z.B. isomorph.



(10)



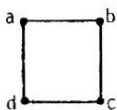
(11)

Aus der Definition der Isomorphie folgt unmittelbar, daß zwei Graphen offensichtlich dann isomorph sind, wenn sie sich zur Deckung bringen lassen. Eine Lösung des Isomorphieproblems besteht dann darin, von einem Knoten ausgehend zu versuchen, alle anderen Knoten und Kanten der beiden Graphen so aufeinanderzulegen, daß sie auch in den ihnen zugeordneten Eigenschaften übereinstimmen.

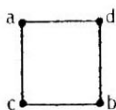
Um die Isomorphie von Graphen festzustellen, könnte man auch eine Reihe von Eigenschaften der Graphen überprüfen, die erfüllt sein müssen, damit Isomorphie besteht. Zu diesen Eigen-

schaften oder auch Invarianten zählen z.B. neben anderen die Anzahl der Knoten, die Anzahl der Kanten und die Anzahl der auf einen Knoten treffenden Kanten. Diesem Verfahren steht aber entgegen, daß kein brauchbarer Satz von Invarianten für Graphen bekannt ist <sup>4</sup>. Es läßt sich anhand von Invarianten oft zeigen, daß zwei Graphen nicht isomorph sind, umgekehrt läßt sich dagegen die Isomorphie - von Sonderfällen abgesehen - zweier Graphen auf diese Weise nicht bestimmen.

Neben der Isomorphie ist für Graphen noch die Gleichheitsrelation Identität definiert <sup>5</sup>. Zwei Graphen  $M_1$  und  $M_2$  sind identisch, wenn jeder Knoten  $p$  in  $M_1$  gleich dem Knoten  $p$  in  $M_2$ , jede Kante  $e$  in  $M_1$  gleich der Kante  $e$  in  $M_2$  ist und jeder Kante  $e$  in  $M_1$  und  $M_2$  die gleichen Knoten zugeordnet werden. Aus der Identität von Molekülgraphen folgt immer deren Isomorphie, da die Identität die feinere Gleichheitsrelation ist. Aus der Isomorphie folgt indessen keineswegs die Identität, wie anhand der Graphen (12) und (13) deutlich wird.



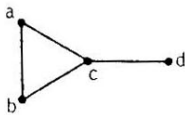
(12)



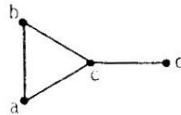
(13)

Die maschinelle Wiedererkennung von Molekülen anhand ihrer Molekülgraphen ist einfacher und effektiver, wenn die Molekülgraphen auf Identität geprüft werden, weil dabei jeder Knoten und jede Kante nur einmal mit einem anderen Knoten bzw. einer anderen Kante verglichen werden. Stimmt nur ein Paar von Knoten oder Kanten nicht überein, sind die Molekülgraphen nicht identisch. Die Bestimmung der Isomorphie von Graphen ist dagegen ein ungleich schwierigeres Problem. Um von der Identität von

Molekülgraphen allerdings auf die Identität der durch sie repräsentierten Moleküle schließen zu können, müssen die Knoten und Kanten isomorpher Molekülgraphen auf immer gleiche Weise geordnet sein; oder, wenn ein Molekülgraph Symmetrieelemente enthält, müssen die Knoten und Kanten so geordnet sein, daß die Ordnungen der Knoten und Kanten sich durch eine isomorphe Abbildung des Molekülgraphen auf sich selbst, also einen Automorphismus, jeweils ineinander überführen lassen. So sind z.B. die Knoten in den Graphen



(14)



(15)

zwar nicht auf die gleiche Weise geordnet, doch läßt sich durch Prüfung der beiden Graphen auf Identität deren Isomorphie zeigen, da die Ordnungen der Knoten durch einen Automorphismus ineinander überführbar sind.

Sollen also durch Molekülgraphen repräsentierte Moleküle anhand der Identität ihrer Molekülgraphen wiedererkannt werden, müssen die Knoten der Molekülgraphen auf bestimmte Weise geordnet sein. Doch das Ordnen der Knoten ist, wenn die Ordnung die beschriebenen Eigenschaften haben soll, gleichwertig mit der Lösung des Isomorphieproblems für alle die Graphen, die dadurch entstehen, daß die Knoten bzw. Kanten eines Molekülgraphen auf bis zu  $n!$  Art und Weisen geordnet werden können. Von den gefundenen isomorphen Ordnungen (gegeben durch die Ordnung der Automorphismengruppe) kann dann jede als repräsentativ ausgewählt werden.

Es ist jedoch nicht sinnvoll, erst einmal alle isomorphen Ordnungen der Knoten und Kanten zu bestimmen, um dann bis auf eine wieder alle zu verwerfen, sondern möglichst nur eine Ord-

nung zu bestimmen und die zu ihr isomorphen gar nicht zu erzeugen.

Der von uns entwickelte Algorithmus CANON 7 erkennt die im Molekülgraphen vorhandenen Symmetrien und vermeidet dadurch die Erzeugung isomorpher Ordnungen von vornherein. Außerdem können aus der so erzeugten kanonischen Bezifferung eines Moleküls mit der R,S-Nomenklatur<sup>10</sup> weitgehend konforme Konfigurationsdeskriptoren hergeleitet werden, ohne mit den der Sequenzregel nach Cahn, Ingold und Prelog<sup>10</sup> anhaftenden Schwierigkeiten konfrontiert zu werden<sup>11</sup>.

### Reaktionsnetzwerke

Werden EM durch Basistransformationen transformiert, stellt eine chemische Reaktion eine Sequenz von Elementarschritten dar. Um aus den Folgen der Elementarschritte die Sequenzen, die von stabilen Ausgangs- zu stabilen Endprodukten führen, auszuwählen, ist es erforderlich, bei jedem Schritt die Ergebnisse der Simulation der Reaktion festzuhalten. Hierzu gehört das Lokalisieren einer Verbindung im Reaktionsnetzwerk ebenso wie das Erkennen von geschlossenen Wegen und Endprodukten im Netzwerk. Unter Umständen ist es sinnvoll, bestimmte Knoten im Reaktionsnetzwerk als Endpunkte zu markieren und somit von einer weiteren Berücksichtigung auszuschließen. Wir sprechen dann von einem Beschneiden des Netzwerkes.

Das durch die Transformation von EM erzeugte Netzwerk wird vorteilhaft in Niveaus eingeteilt. Diese Einteilung ist dadurch gegeben, daß, beginnend mit dem Ausgangs-EM, durch die Basistransformationen das jeweils nächst höhere Niveau erzeugt wird. Dabei wird das nächst höhere Niveau erst erzeugt, wenn alle tiefer liegenden Niveaus vollständig sind. Das Netzwerk wird immer erst in die Breite statt in die Tiefe entwickelt. Der umgekehrte Weg ist prinzipiell auch möglich, doch hat sich gezeigt, daß bei der Erzeugung nur eines Teiles des Netzwerkes der erste Weg vorzuziehen ist, weil das so erzeugte Teilnetzwerk mehr verwertbare Informationen enthält.

Das Netzwerk hat die Eigenschaft, daß es sich mit steigender Anzahl von Niveaus immer mehr verzweigt, und die höheren Niveaus meistens dichter als darunter liegende besetzt sind, wie in Abb. 1 schematisch gezeigt ist.

Bei der Erzeugung der Niveaus des Netzwerkes werden die zu transformierenden EM über ein bestimmtes Niveau lokalisiert und die erhaltenen EM in das nächst höhere Niveau plaziert. Soll allerdings festgestellt werden, ob ein erzeugtes EM sich bereits im Netzwerk befindet, ist es weder sinnvoll noch er-



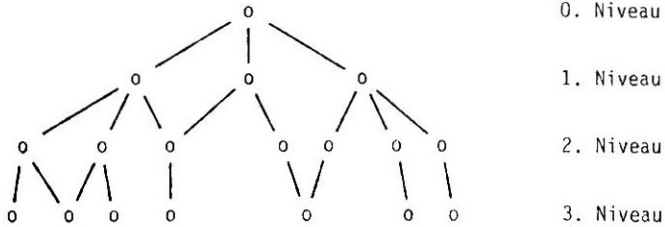


Abb. 1

Schematisches Reaktionsnetzwerk

forderlich, das Netzwerk Niveau für Niveau nach diesem bestimmten EM abzusuchen. Das erzeugte EM hat bestimmte Eigenschaften  $\epsilon$ , die ein anderes EM auch haben muß, damit beide EM gleich sein können. Zu diesen Eigenschaften  $\epsilon$  gehören z.B. die Anzahl der Atome, der Bindungen und der Ladungen. Es genügt daher, nur die EM des Netzwerkes zu vergleichen, die diese Eigenschaften auch besitzen, und das ist meist nur ein kleiner Teil aller im Netzwerk befindlichen EM. Anstatt über die Niveaus ein bestimmtes EM zu suchen, ist es in diesem Fall effektiver, dies über Eigenschaften  $\epsilon$  der EM zu tun.

### Die Simulation des Abbaus von Histidin

Das hier beschriebene Simulationssystem wurde in PL1 programmiert und bildet zusammen mit den notwendigen Steuerrou-tinen das Programmsystem ASSOR. Es wurde auf den Großrechnern (Amdahl 470 und IBM 360/91) des Instituts für Plasmaphysik der Max-Planck-Gesellschaft in 8046 Garching implementiert. Um die Arbeitsweise des Simulationssystems ASSOR zu demonstrieren, sei ein Beispiel für die Simulation von Reaktionen gegeben.

Bei seinem Abbau zu N-Formylisoglutamin reagiert Histidin mit zwei Molekülen Wasser und spaltet Ammoniak ab<sup>12</sup>. Das zur Simulation erforderliche EM besteht folglich anfangs aus Histi-din und zwei Molekülen Wasser, und die Liste der bekannten Ver-bindungen enthält Ammoniak.

Die für die Simulation geltenden Grenzen für die Valenzzu-stände der Atome werden festgelegt. Ein Ausschnitt der Liste der für den Abbau von Histidin verwandten Werte ist in Abb. 2 gezeigt.

VALENCE STATE LIMITS FOR ATOMS										
ATOMIC NUMBERS	1	2	3	4	5	6	7	8	9	10
UPPER LIMIT FOR COORDINATION	2	0	1	2	3	4	5	6	7	0
LOWER LIMIT FOR COORDINATION	0	0	0	2	3	4	5	6	7	0
UPPER LIMIT FOR VALENCE ELEC.	2	0	1	2	3	4	5	6	7	0
LOWER LIMIT FOR VALENCE ELEC.	0	0	0	2	3	4	5	6	7	0
PRINCIPAL VALENCE ELECTRONS	1	0	1	2	3	4	5	6	7	8

Abb. 2

Ausschnitt aus der Liste der erlaubten Valenzzustände von Atomen für den Abbau von Histidin

Ist das System auf diese Weise initialisiert, wird die Si-mulationsphase gestartet. Die Liste der Steuerparameter ist in

Abb. 3 gezeigt.

```

RUN . . . . . 1
MAXIMAL DEGREE OF MOLEC. GRAPHS . . . . . 5
HIGHEST ATOMIC NUMBER . . . . . 53
NUMBER OF ATOMS IN CHEM. TABLE . . . . . 22
LOWEST LEVEL FOR THIS RUN . . . . . 0
HIGHEST LEVEL FOR THIS RUN . . . . . 1
MULTIPLIER FOR REGION SIZE . . . . . 1
NUMBER OF REGIONS . . . . . 19
FORMULA OUTLAY . . . . . 1
MAXIMUM NUMBER OF MOLECULES . . . . . 1023
LIMITING NUMBER FOR REACTIVE BONDS . . . . . 5
FACTOR FOR ELNEG. EQUAL . . . . . 0.8000
OUTPUT CONTROL OPTIONS . . . . .
BASEV BRBD DRAW ELNEG LALIM STORE CHEM TABLE CHARD MARD
ACTION CONTROL OPTIONS . . . . .
CHARD BRBD TERMP CYCLES MARD SICH
```

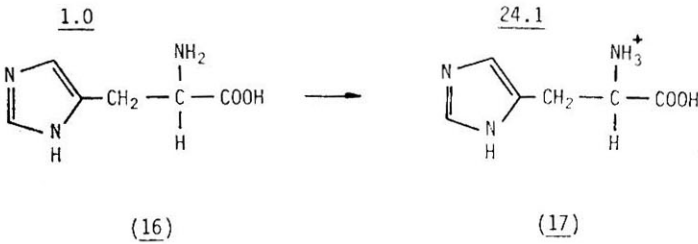
Abb. 3

Liste der Steuerparameter für die Erzeugung des ersten Niveaus des Reaktionsnetzwerkes beim Abbau von Histidin

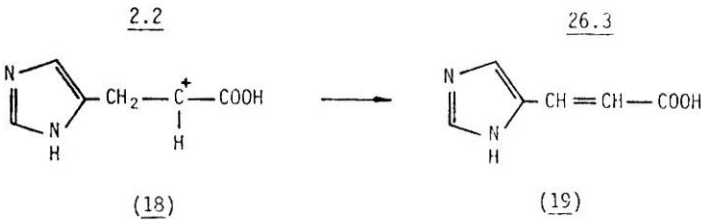
Im folgenden sind die Moleküle im Reaktionsnetzwerk durch die Nummer des Moleküls auf einem Niveau und die Nummer des Niveaus identifiziert, so daß mit n.m das n-te Molekül auf dem m-ten Niveau bezeichnet wird.

Das System schlägt Transformationen des EM <Histidin, 2 Wasser> vor, unter denen sich u.a. die Protonierung der Aminogruppe und die Deprotonierung der Carboxylgruppe befinden. Beide Reaktionen führen schließlich vom Histidin (16) zur zwitterionischen Form des Histidins. Es werden lediglich die Abstraktion und Addition eines Protons in verschiedener Reihenfolge durchgeführt. Alternative Reaktionswege sind in diesem

Beispiel der Übersichtlichkeit halber allerdings nicht gezeigt, so daß nur das an der Aminogruppe protonierte Histidin (17) weiter verfolgt wird.



Aus (17) wird durch Brechen einer Bindung Ammoniak abgespalten und dann durch Ladungsverschiebung unter gleichzeitigem Abspalten eines Protons die Urocainsäure (4-Imidazolacrylsäure) (19) gebildet.



Ammoniak wird als bekannte Verbindung erkannt und so markiert, daß es an den folgenden Transformationen nicht mehr teilnimmt, siehe Abb. 4.

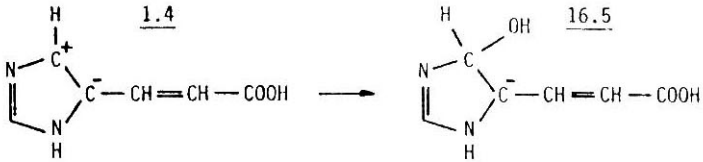
MOLECULE . . . . . 1      LEVEL . . . . . 2

THIS COMPOUND IS A TERMINAL POINT IN THE NETWORK - LOOK FOR:      EMM NIA

Abb. 4

Nachricht des Systems ASSOR über das Erkennen eines Endpunktes im Reaktionsnetzwerk

Die Urocainsäure (19) reagiert mit Wasser zu 4-Imidazolion-5-propionsäure (28).



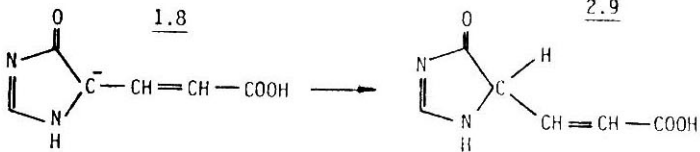
(20)

(21)



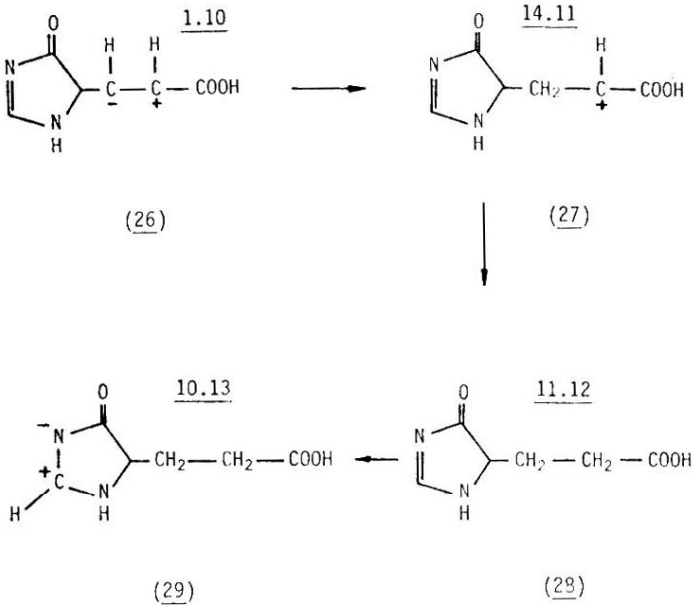
(23)

(22)



(24)

(25)



Zunächst wird eine Hydroxylgruppe an das Atom vier des Imidazolrestes angelagert (21) und anschließend durch das System das Abspalten eines Hydridions vorgeschlagen (22). Bei Zwischenstufen wie (20) wird vom System auch immer die Rekombination der Ladungen vorgeschlagen, so daß in diesem Fall das 10. Molekül auf dem 5. Niveau mit (19) identisch ist. Das System erkennt diesen geschlossenen Reaktionsweg und gibt eine Nachricht, vgl. Abb. 5.

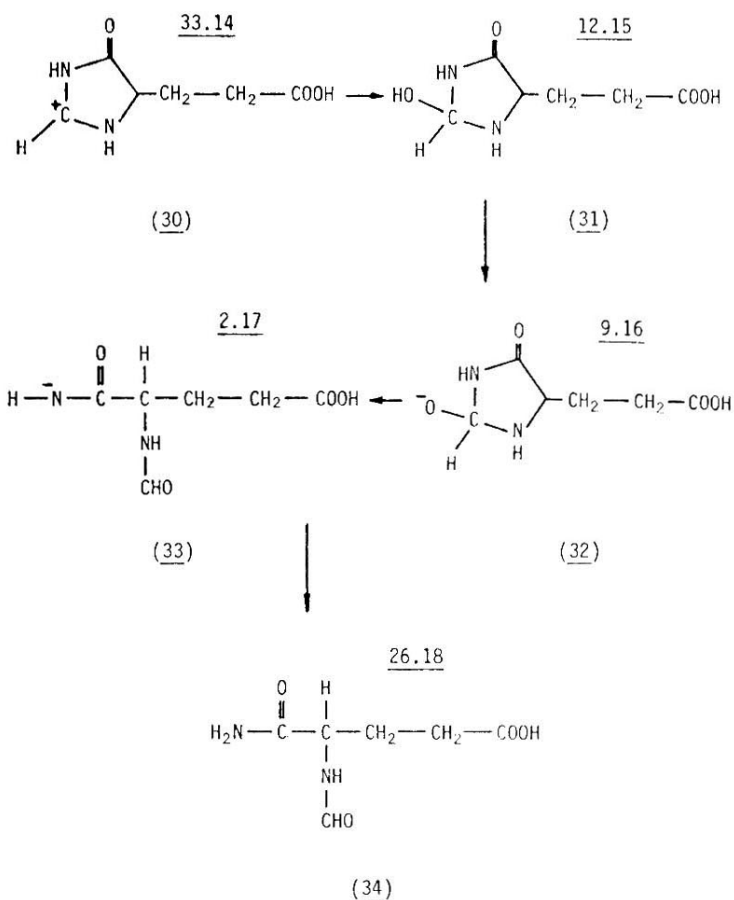
```

MOLECULE . . . . . 10   LEVEL . . . . . 5
THIS COMPOUND CLOSSES A CYCLE IN THE NETWORK - LOOK FOR:
MOLECULE . . . . . 26   LEVEL . . . . . 3
    
```

Abb. 5

Nachricht des Systems ASSOR über das Erkennen eines geschlossenen Weges im Reaktionsnetzwerk

4-Hydroxyimidazol-5- acrylsäure (22) tautomerisiert und wird durch Anlagern eines weiteren Protons und des von (21) abgespaltenen Hydridions in 4-Imidazol-5-propionsäure (28) überführt. Durch Ringöffnung und Anlagerung von Wasser entsteht schließlich N-Formylisoglutamin (34).



Die hier gezeigten Reaktionen des Histidins werden vom System neben anderen Reaktionen gefunden. Auf jedem Niveau ist jeweils ein EM weiter transformiert worden, das die im Hinblick auf den Abbau des Histidins weiterführende Verbindung enthält. Dieses Intervenieren während des Simulationsprozesses ist beim derzeitigen Entwicklungsstand noch erforderlich, um über die hier gezeigte Anzahl von 18 Niveaus mit vertretbarem Aufwand ein solches Ergebnis zu erzielen. Entscheidend ist, daß die tatsächlich abbauenden Reaktionen auch unter den vom System vorgeschlagenen Reaktionen sind, wobei die Zahl der vorgeschlagenen Reaktionen einige wenige nicht übersteigen sollte, d.h. deutlich unter der Anzahl der rein formal möglichen Transformationen liegen sollte.

Ein wichtiger Gesichtspunkt bei der Beurteilung der Effektivität des Simulationssystems ist die für die Simulation erforderliche Zeit. Beim hier angeführten Beispiel des Abbaus von Histidin wurden die 518 Verbindungen des Reaktionsnetzwerkes in etwa 300 sec erzeugt. Der Zeitbedarf wird natürlich steigen, wenn Entscheidungsprozesse, die jetzt noch vom Benutzer vorgenommen, schließlich dem System überlassen werden. Der Zuwachs des Zeitbedarfs dürfte sich allerdings in Grenzen halten, denn mit der besseren Auswahl der Reaktionsmöglichkeiten nimmt die Anzahl der je Niveau zu erzeugenden isomeren EM ab.



### Ausblick

Die Simulation chemischer Reaktionen stellt einen ganz allgemeinen Ansatz zur deduktiven Lösung chemischer Probleme dar. Denn von allgemein gültigen Prinzipien ausgehend, nämlich dem mathematischen Modell und den Nebenbedingungen in Form heuristischer Regeln, werden Lösungen für ganz individuelle Fragestellungen gefunden. Hierzu gehören die Vorhersage von Folgeprodukten einer Verbindung, die systematische Suche nach neuen Reaktionen, das Aufstellen von Reaktionsmechanismen und die Planung von Synthesen<sup>13</sup>. Für das Planen von Synthesen komplizierter Verbindungen wird es im Hinblick auf die zielgerechte Führung des Planungsprozesses wichtig sein, eine Vorschau auf bekannte oder vorhandene Verbindungen vornehmen zu können<sup>14</sup>.

Die Voraussage von Folgeprodukten einer Verbindung mit dem Simulationssystem könnte eine entscheidende Rolle bei der Beurteilung des Umweltrisikos einer Chemikalie spielen. Durch Aufspannen des Reaktionsnetzwerkes des EM, bestehend aus der fraglichen Chemikalie und den in der Umwelt vorherrschenden Substanzen, sollte es möglich sein, die Folgeprodukte aufzuzeigen.

Die hier nur angedeuteten Anwendungsmöglichkeiten der Simulation von Reaktionen erheben keinen Anspruch auf Vollständigkeit. Es ist aber vorauszusehen, daß derartige Systeme immer mehr bei der deduktiven Lösung chemischer Probleme eingesetzt werden, und somit dem Chemiker ein neues, wirkungsvolles Werkzeug in die Hand gegeben wird.

Literaturverzeichnis

- (1) J. Dugundji and I. Ugi, Topics Curr. Chem., **39**, 19-64 (1973)
- (2) (a) J. C. J. Bart und E. Garagnani, Zeitschr. f. Naturforsch., **31b**, 1646-1653 (1976)  
(b) *ibid.*, **32b**, 455-464 (1977)  
(c) *ibid.*, **32b**, 465-468 (1977)  
(d) *ibid.*, **32b**, 678-683 (1977)
- (3) L. Spialter, J. Chem. Doc., **4**, 261-274 (1964)
- (4) F. Harary, "Graph Theory", Addison-Wesley Corp., Reading, Mass., 1969
- (5) K. Wagner, "Graphentheorie", B. I. Wissenschaftsverlag, Mannheim, 1970
- (6) H. Meschkowski, "Einführung in die moderne Mathematik", B. I. Wissenschaftsverlag, Mannheim, 1971
- (7) W. Schubert and I. Ugi, J. Am. Chem. Soc., **100**, 37-41 (1978)
- (8) (a) R. P. Iczkowski and J. L. Margrave, J. Am. Chem. Soc., **83**, 3547-3551 (1961)  
(b) J. Hinze and H. H. Jaffé, J. Am. Chem. Soc., **84**, 540-546 (1962)  
(c) J. Hinze, M. A. Whitehead, and H. H. Jaffé, J. Am. Chem. Soc., **85**, 148-154 (1963)  
(d) J. Hinze and H. H. Jaffé, J. Phys. Chem., **67**, 1501-1506 (1963)  
(e) J. E. Huheey, J. Phys. Chem., **69**, 3284-3291 (1965)  
(f) J. E. Huheey, J. Org. Chem., **31**, 2365-2368 (1966)
- (9) W. Schubert, I. Ugi, in Vorbereitung
- (10) (a) R. S. Cahn and C. K. Ingold, J. Chem. Soc. (London), 612-622 (1951)  
(b) R. S. Cahn, C. K. Ingold, and V. Prelog, Experientia, **12**, 81-124 (1956)  
(c) R. S. Cahn, C. Ingold und V. Prelog, Angew. Chem., **78**, 413-447 (1966)

- (11) W. Schubert, I. Ugi, *Chimia*, eingereicht
- (12) H. R. Mahler and E. H. Cordes, "Biological Chemistry", Harper & Row, New York, 1971
- (13) (a) I. Ugi, *IBM Nachrichten*, **24**, 180-184 (1974)  
(b) I. Ugi, J. Gasteiger, J. Brunnert und W. Schubert, *IBM Nachrichten*, **24**, 185-189 (1974)  
(c) J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, and I. Ugi, in E. V. Ludeña, N. H. Sabelli, A. C. Wahl, "Computers in Chemical Education and Research", Plenum Publishing Corp., New York, 1977, pp. 337-355  
(d) J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, and I. Ugi, in W. T. Wipke and W. J. Howe, "Computer-Assisted Organic Synthesis", *Am. Chem. Soc. Symposium Series No. 61*, 1977, pp. 33-59  
(e) J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, P. Lemmen, W. Schubert, and I. Ugi, "The Deductive Solution of Chemical Problems on the Basis of a Mathematical Model of Chemistry", 26th International Congress of Pure and Applied Chemistry, Tokio, 1977  
(f) I. Ugi, H. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum und W. Schubert, *Angew. Chem.*, im Druck
- (14) W. Schubert und I. Ugi, in Vorbereitung