

## Neue Anwendungsgebiete für Computer in der Chemie

Von Ivar Ugi, Johannes Bauer, Josef Brandt, Josef Friedrich, Johann Gasteiger, Clemens Jochum und Wolfgang Schubert<sup>[\*]</sup>

Das hier beschriebene mathematische Modell der konstitutionellen Chemie beruht auf einer Erweiterung des Begriffs der Isomerie auf Ensembles von Molekülen. Damit ist eine chemische Reaktion die Umwandlung eines Ensembles von Molekülen in ein isomeres Ensemble. Ein Ensemble von Molekülen kann repräsentiert werden durch einen Atomvektor und eine zugeordnete Bindungs/Elektronen-(BE-)Matrix, eine Reaktion durch eine Reaktions-(R-)Matrix. Die Algebra der BE- und R-Matrizen dient als Grundlage von Computer-Programmen zur deduktiven Lösung chemischer Probleme. Diskutiert werden Algorithmen und Computerprogramme auf der Grundlage von BE- und R-Matrizen. Sie ermöglichen eine Klassifikation und Dokumentation von Strukturen, Substrukturen und Reaktionen, die Voraussage von Folgeprodukten, die Planung von Synthesen, die Konstruktion von Netzwerken für Reaktionsmechanismen und Synthesewege sowie die Voraussage chemischer Reaktionen.

### 1. Wege zur deduktiven Lösung chemischer Probleme

Der Gebrauch von Computern ist in der Chemie seit langem üblich. Trotzdem wird bei den heutigen Anwendungen in der Chemie erst ein bescheidener Teil der Möglichkeiten moderner Computer genutzt. Entweder werden numerische Probleme bearbeitet wie quantenmechanische Rechnungen und die Erfassung und Auswertung von Meßdaten, oder es werden größere Datenmengen gespeichert und wieder abgerufen. Die Herausforderung, chemische Probleme durch Algorithmen zu lösen, welche menschliche Intelligenzleistungen im Sinne von Entscheidungsprozessen und deduktivem Denken simulieren, wurde frühzeitig erkannt und führte zu einer Arbeitsrichtung, die man heute mit der Bezeichnung „artificial intelligence“ assoziiert.

Diese Entwicklung begann vor mehr als zehn Jahren mit Computerprogrammen zur Strukturermittlung von Molekülen aus physikalisch-chemischen Meßdaten<sup>[1]</sup> (Massenspektren) und auch mit retrievalorientierten Computerprogram-

men zur Syntheseplanung<sup>[2]</sup>. Bei diesen Syntheseplanungsprogrammen erzeugt man aus einem eingegebenen Syntheseziel Vorstufen, welche sich durch eingespeicherte chemische Reaktionen in das Syntheseziel überführen lassen. Hierbei wird anhand der Strukturmerkmale des Syntheseziels untersucht, ob eine der in der Reaktionsbibliothek gespeicherten bekannten Reaktionen zum Syntheseziel führen kann und welches gegebenenfalls die zugehörigen Vorstufen sind.

Um Syntheseplanungsprogramme zu erstellen, welche auch Synthesewege mit präzedenzlosen Reaktionsschritten vorschlagen können, benötigt man eine Theorie der konstitutionellen Chemie, welche es ermöglicht – ohne detaillierte Reaktionskenntnis – von einem gegebenen molekularen System aus andere Systeme zu finden, aus denen es durch chemische Reaktionen entstehen kann oder in welche es sich umwandeln läßt.

Eine solche Theorie der konstitutionellen Chemie eignet sich nicht nur zur Erstellung von Syntheseplanungsprogrammen, sondern sie ermöglicht auch die Lösung einer Vielfalt anderer chemischer Probleme.

Deduktive Computerprogramme auf der Grundlage logischer Strukturmodelle und mathematischer Darstellungen werden künftig in der Chemie eine wesentliche Rolle spielen,

[\*] Prof. Dr. I. Ugi, Dipl.-Inform. J. Bauer, Dr. J. Brandt, Dipl.-Inform. J. Friedrich, Dr. J. Gasteiger, Dr. C. Jochum, Dr. W. Schubert  
Organisch-chemisches Institut der Technischen Universität München  
Lichtenbergstraße 4, D-8046 Garching

weil man sich nicht auf Retrieval und Manipulation eingegebener empirischer Information zu beschränken braucht, sondern stattdessen die *Gesamtheit der denkbaren Lösungen* eines chemischen Problems *deduktiv* anhand einer geeigneten Theorie erfassen kann. Hierbei muß man keineswegs auf die Berücksichtigung relevanter empirischer Daten verzichten, sondern es ist möglich und sinnvoll, sie bei der Bewertung und Sichtung von Zwischenergebnissen und Endresultaten in Form allgemeiner Auswahlregeln heranzuziehen.

In diesem Aufsatz wird ein mathematisches Modell der konstitutionellen Chemie beschrieben und sein Einsatz als theoretische Grundlage deduktiver chemischer Computerprogramme an einigen Beispielen gezeigt. Dieses *mathematische Modell der konstitutionellen Chemie*<sup>[3]</sup> beruht auf einer konsequenten Anwendung des Äquivalenzbegriffs bei der Darstellung und Klassifizierung molekularer Systeme.

Die Grundlagen für die zusätzliche Berücksichtigung der Stereochemie sind an anderer Stelle beschrieben worden<sup>[4]</sup>.

## 2. Ein mathematisches Modell der konstitutionellen Chemie

Um die logische Struktur der konstitutionellen Chemie durch ein mathematisches Modell darzustellen<sup>[3]</sup>, gehen wir von folgendem aus: Ein *Ensemble von Molekülen (EM)*<sup>[5, 6]</sup> besteht aus Molekülen, die gleichartig oder verschieden sein können. Ebenso wie ein Molekül hat auch ein Ensemble von Molekülen eine Bruttoformel; sie ergibt sich als Vereinigung der Bruttoformeln der in dem Ensemble enthaltenen Moleküle.

### 2.1. Der erweiterte Isomeriebegriff

Diese Festlegung führt unmittelbar zu einer Erweiterung des Isomeriebegriffs auf Ensembles von Molekülen:

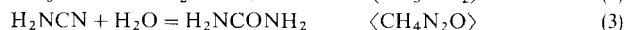
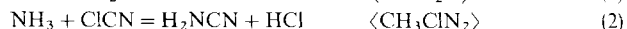
Ensembles von Molekülen (*EM*) sind isomer, wenn sie die gleiche Bruttoformel haben. Die Bruttoformel beschreibt das Sortiment **A** der Atome, aus denen das *EM* zusammengesetzt ist. Alle *EM*, die aus **A** gebildet werden können, haben die gleiche Ensemblebruttoformel  $\langle \mathbf{A} \rangle$ .

Dementsprechend besteht ein *EM(A)* aus einem oder mehreren Molekülen, die man aus **A** herstellen kann, wenn man jedes in **A** enthaltene Atom genau einmal verwendet.

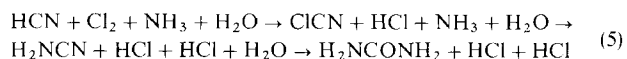
Eine *FIEM(A)*, die *Familie der isomeren Ensembles von Molekülen* des Atomsortiments **A**, ist die Menge aller *EM(A)*. Eine *FIEM* läßt sich einfach durch die Ensemblebruttoformel  $\langle \mathbf{A} \rangle$  des zugrundeliegenden Atomsortiments **A** festlegen.

Eine chemische Reaktion oder eine Folge von chemischen Reaktionen ist eine Isomerisierung, d.h. die Umwandlung eines *EM* in ein dazu isomeres *EM*. Eine *FIEM* enthält also alle von der Stöchiometrie her ineinander umwandelbaren *EM* und somit die gesamte Chemie des zugrundeliegenden Atomsortiments **A**. Da Atomsortimente beliebig gewählt werden können, umfaßt eine Theorie der *FIEM* die gesamte Chemie.

Die linke Seite und die rechte Seite der chemischen „Reaktionsgleichungen“ (1) bis (4) sind jeweils isomere *EM* mit den in Klammern angegebenen (Ensemble-)Bruttoformeln. Addiert man die Reaktionsgleichungen (1) bis (3), so resultiert Gleichung (4), welche die Gesamtreaktion repräsentiert.



Die an den Reaktionen (1) bis (4) beteiligten *EM* lassen sich in die *FIEM*  $\langle \text{CH}_6\text{Cl}_2\text{N}_2\text{O} \rangle$  einbetten, wenn man die Ensembles der reagierenden Moleküle jeweils durch die nicht unmittelbar beteiligten Moleküle ergänzt [Gl. (5)].



### 2.2. BE-Matrizen

Aus der Sicht des Chemikers bestehen die Atome aus einem Rumpf (Atomkern und Elektronen der inneren Schale) und Valenzelektronen (Elektronen der äußeren Schalen). In Molekülverbänden werden die Atomrümpfe durch Valenzelektronen zusammengehalten. Die chemische Konstitution eines molekularen Systems wird durch die Paare *kovalent* gebundener Atomrümpfe beschrieben. Auch die Verteilung der an kovalenten Bindungen nicht beteiligten „freien“ Valenzelektronen kann in die Konstitution mit einbezogen werden. Üblicherweise wird die chemische Konstitution durch Konstitutionsformeln wiedergegeben, in denen die Atomrümpfe durch Elementsymbole, kovalente Bindungen durch Verbindungsstriche zwischen den Elementsymbolen und (freie) Valenzelektronen durch Punkte an den Elementsymbolen dargestellt werden.

In unserem Modell wird die chemische Konstitution eines *EM* durch dessen Atomvektor, der aus den Elementen von **A** besteht, und durch die Bindungs/Elektronen-(BE-)Matrix wiedergegeben, deren Zeilen und Spalten den einzelnen Atomrümpfen zugeordnet sind: Die außerdiagonalen Eintragungen geben die formalen Bindungsordnungen der kovalenten Bindungen und die diagonalen Eintragungen die Zahl der freien Elektronen an<sup>[3-13]</sup>.

In der Vergangenheit wurden schon andere Matrizendarstellungen der chemischen Konstitution eingeführt: die Adjazenz-Matrix **J**, der Graphentheorie entlehnt, gibt lediglich die Nachbarschaftsverhältnisse der Atome wieder. *Spialter*<sup>[14]</sup> definierte die Konnektivitäts-Matrix **C**, deren außerdiagonale Eintragungen die Bindungsordnungen der kovalenten Bindungen und deren diagonale Eintragungen den Atomvektor enthalten. *E. Meyer*<sup>[15]</sup> setzte für die chemische Dokumentation topologische Matrizen ein. *Yoneda*<sup>[16]</sup> verwendet eine Matrizendarstellung zur Behandlung der heterogenen Katalyse, bei der die reaktiven Zentren der Reaktionsteilnehmer und Katalysatoren auf der Diagonalen erfaßt sind.

Die Adjazenz-Matrix **J**(HCN) (6) und die Konnektivitäts-Matrix **C**(HCN) (7) für Cyanwasserstoff mögen zur Erläuterung dienen.

$$\begin{aligned} &\text{H}-\text{C}\equiv\text{N}: \\ \mathbf{J}(\text{HCN}) &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{aligned} \quad (6)$$

$$\mathbf{C}(\text{HCN}) = \begin{pmatrix} \text{H} & 1 & 0 \\ 1 & \text{C} & 3 \\ 0 & 3 & \text{N} \end{pmatrix} \quad (7)$$

Die BE-Matrix **B** eines  $EM(A)$ , dessen Atomsortiment  $A = \{A_1, \dots, A_n\}$  aus  $n$  beliebig numerierten Atomen besteht, ist eine  $n \times n$ -Matrix, deren  $i$ -te Zeile und Spalte dem  $i$ -ten Atomrumpf  $A_i$  zugeordnet sind. Die außerdiagonale Eintragung  $b_{ij}$  der  $i$ -ten Zeile und  $j$ -ten Spalte ist die formale Bindungsordnung der kovalenten Bindung vom Atom  $A_i$  zum Atom  $A_j$ . Da dem auch eine Bindung von  $A_j$  nach  $A_i$  entspricht, haben wir  $b_{ij} = b_{ji}$ . Daraus folgt, daß BE-Matrizen symmetrisch sein müssen. Die  $i$ -te Diagonaleintragung  $b_{ii}$  gibt an, wieviele freie Valenzelektronen das Atom  $A_i$  trägt. Bei  $EM$  mit geschlossenen Valenzschalen und spingepaarten Elektronen sind alle Diagonaleintragungen geradzahlig.

Den mit Atomindizes versehenen Formeln des Cyanwasserstoffs und des Isocyanwasserstoffs entsprechen die BE-Matrizen **B**(HCN) und **B**(HNC).

$$\begin{array}{c} \text{H}^1 - \text{C}^2 \equiv \text{N}^3 \\ \text{H}^1 \text{C}^2 \text{N}^3 \end{array} \quad \text{H}^1 - \overset{\oplus}{\text{N}}^3 \equiv \overset{\ominus}{\text{C}}^2 \\ \text{H}^1 \text{C}^2 \text{N}^3$$

$$\mathbf{B}(\text{HCN}) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 3 \\ 0 & 3 & 2 \end{pmatrix} \begin{matrix} \text{H}^1 \\ \text{C}^2 \\ \text{N}^3 \end{matrix} \quad (8) \quad \mathbf{B}(\text{HNC}) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 1 & 3 & 0 \end{pmatrix} \begin{matrix} \text{H}^1 \\ \text{C}^2 \\ \text{N}^3 \end{matrix} \quad (9)$$

In **B**(HCN) entspricht  $b_{12} = b_{21} = 1$  der H—C-Bindung,  $b_{23} = b_{32} = 3$  der C≡N-Dreifachbindung und  $b_{33} = 2$  dem freien Elektronenpaar am N-Atom.

Wie das Beispiel zeigt, ist die BE-Matrix eine Darstellung einer Grenzstruktur im Sinne der Valence-Bond-Theorie.

Die Anzahl der Valenzelektronen  $s_i$ , die dem Atom  $A_i$  angehören, ist die Summe über die Eintragungen der  $i$ -ten Zeile/Spalte einer BE-Matrix:

$$s_i = \sum_{j=1}^n b_{ij} = \sum_{j=1}^n b_{ji} \quad (10)$$

Die formale elektrische Ladung des Atoms  $A_i$  errechnet sich aus der Differenz von  $s_i$  und der Rumpfladung. Für die zweite Zeile/Spalte von **B**(HNC) haben wir  $s_2 = 0 + 2 + 3 = 5$ . Daraus ergibt sich, daß das C-Atom von HNC bei einer Rumpfladung von +4 eine Formalladung von -1 hat. Die Kreuzsumme

$$\xi_i = 2s_i - b_{ii}$$

über eine Diagonaleintragung  $b_{ii}$  umfaßt alle Eintragungen der  $i$ -ten Zeile und Spalte, d.h. alle Eintragungen  $b_{ij}$  und  $b_{ji}$ , deren Indices  $i$  enthalten. Dies entspricht der doppelten Zeilen/Spalten-Summe  $s_i$ , abzüglich der Diagonaleintragung  $b_{ii}$ . Die Kreuzsumme  $\xi_i$  ist die Zahl der Elektronen, die sich in den Valenzorbitalen des Atoms  $A_i$  befinden. Für das C-Atom von HNC finden wir

$$\begin{aligned} \xi_2 &= b_{21} + b_{22} + b_{23} + b_{12} + b_{23} \\ &= 0 + 2 + 3 + 0 + 3 = 2s_2 - b_{22} \\ &= 2 \cdot 5 - 2 = 8 \end{aligned} \quad (11)$$

d.h. dieses C-Atom erfüllt die Oktettregel.

Die Zeile einer BE-Matrix repräsentiert die Elektronenverteilung am zugehörigen Atomrumpf. Für einen gegebenen Atomrumpf ist nur eine begrenzte Anzahl solcher Verteilungen (Valenzschemata) erlaubt, die sich z.B. als Liste darstellen lassen<sup>[13]</sup>.

Die Summe

$$S = \sum_i s_i = \sum_{ij} b_{ij} \quad (12)$$

über alle Eintragungen einer BE-Matrix ist die Anzahl aller Valenzelektronen, welche das  $EM$  enthält. Für alle  $EM$ , die zur gleichen  $FIEM$  gehören, hat  $S$  den gleichen Wert.

Die Konstitution von chemischen Verbindungen, die man durch lokalisierte Bindungen nicht angemessen wiedergeben kann, z.B. resonanzfähige Systeme oder Systeme mit Mehrzentrenbindungen, läßt sich durch die Gesamtheit der BE-Matrizen aller beteiligten Grenzstrukturen oder auch durch BE-Matrizen mit gebrochenen formalen Bindungsordnungen darstellen.

In einem  $n$ -atomigen  $EM$  können die Atome auf bis zu  $n!$  verschiedene Weisen numeriert werden, so daß es dementsprechend bis zu  $n!$  unterscheidbare, aber äquivalente BE-Matrizen gibt. Durch geeignete Regeln kann eine von diesen zur kanonischen Numerierung erklärt werden<sup>[17]</sup>. Diese Numerierung spielt bei der Behandlung stereochemischer Aspekte eine wichtige Rolle.

## 2.3. R-Transformationen

Eine chemische Reaktion ist die Umwandlung eines  $EM$  in ein isomeres  $EM$  durch Verschieben von Valenzelektronen. Hierbei gelten folgende Invarianzen, die in den Erhaltungssätzen von Ladung und Materie begründet sind:

1. Die Atomrümpfe eines  $EM$  bleiben erhalten.
2. Die Gesamtzahl  $S$  der Valenzelektronen eines  $EM$  bleibt erhalten.

Aus 2 folgt, daß die Umwandlung der Ausgangsstoffe  $EM(\mathbf{B})$  in die Endprodukte  $EM(\mathbf{E})$  durch eine chemische Reaktion



durch solche BE-Matrizen Transformationen  $\mathbf{B} \rightarrow \mathbf{E}$  darstellbar ist, bei denen die Summe über alle Eintragungen ( $S$ ) [siehe Gl. (12)] konstant bleibt.

### 2.3.1. R-Matrizen

Wir definieren die *R-Matrix* (Reaktions-Matrix) **R** durch die Transformation

$$\mathbf{B} + \mathbf{R} = \mathbf{E} \quad (14)$$

Die Addition einer R-Matrix zu einer BE-Matrix kann als Wirkung eines Operators  $\mathbf{R}_{Op}$  auf **B** gemäß

$$\mathbf{R}_{Op}(\mathbf{B}) = \mathbf{B} + \mathbf{R} = \mathbf{E} \quad (15)$$

aufgefaßt werden. Da **B** und **E** die gleiche Eintragungssumme haben:

$$S = \sum_{i,j} b_{ij} = \sum_{i,j} e_{ij} = \sum_{i,j} b_{ij} + \sum_{i,j} r_{ij} \quad (16)$$

ist die Summe der Eintragungen  $r_{ij} = e_{ij} - b_{ij}$  der Matrix  $\mathbf{R} = \mathbf{E} - \mathbf{B}$  Null:

$$\sum_{i,j} r_{ij} = 0 \quad (17)$$

Die Matrix muß symmetrisch sein, weil  $\mathbf{E} = \mathbf{B} + \mathbf{R}$  als BE-Matrix definitionsgemäß symmetrisch sein muß, d. h. aus  $e_{ij} = e_{ji}$  und  $b_{ij} = b_{ji}$  folgt  $r_{ij} = r_{ji}$ .

Die außerdiagonalen negativen Eintragungen  $r_{ij} = r_{ji} = -1$  geben das Lösen jeweils einer kovalenten Bindung  $A_i - A_j$  wieder; eine negative Diagonaleintragung  $r_{ii}$  gibt an, wieviele freie Elektronen das Atom  $A_i$  durch eine Reaktion verliert. Entsprechend zeigen positive, außerdiagonale Eintragungen an, wieviele Bindungen zwischen  $A_i$  und  $A_j$  geknüpft werden, und eine positive Diagonaleintragung läßt erkennen, wieviele freie Valenzelektronen das Atom  $A_i$  durch eine Reaktion gewinnt.

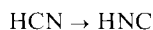
Repräsentiert  $\mathbf{R}$  die Reaktion  $EM(\mathbf{B}) \rightarrow EM(\mathbf{E})$  gemäß  $\mathbf{B} + \mathbf{R} = \mathbf{E}$ , so gilt ihre Inverse  $\bar{\mathbf{R}}$  für die *Retroreaktion*  $EM(\mathbf{E}) \rightarrow EM(\mathbf{B})$  gemäß

$$\mathbf{E} + \bar{\mathbf{R}} = \mathbf{E} - \mathbf{R} = \mathbf{B} \quad (18)$$

Die Eintragungen von  $\bar{\mathbf{R}}$  sind damit gegeben durch

$$\bar{r}_{ij} = -r_{ij} \quad (19)$$

Zum Beispiel wird die Reaktion



dargestellt durch:

$$\mathbf{B}(\text{HCN}) + \mathbf{R} \rightarrow \mathbf{B}(\text{HNC})$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 3 \\ 0 & 3 & 2 \end{pmatrix} + \begin{pmatrix} 0 & -1 & +1 \\ -1 & +2 & 0 \\ +1 & 0 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 1 & 3 & 0 \end{pmatrix} \quad (20)$$

Die inverse Matrix  $\bar{\mathbf{R}} = -\mathbf{R} = \mathbf{B}(\text{HCN}) - \mathbf{B}(\text{HNC})$  entspricht der Reaktion  $\text{HNC} \rightarrow \text{HCN}$ .

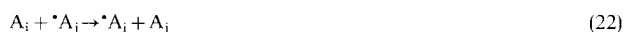
Die R-Matrizen für Reaktionen aus der Chemie von Valenzschalen mit gepaarten Elektronen enthalten in der Diagonalen außer Nulleintragungen nur gerade Zahlen. Die R-Matrizen für Reaktionen aus der Oktett-Chemie von Molekülen ohne Formalladungen haben in der Diagonalen keine von Null verschiedenen Eintragungen. Solche R-Matrizen gibt es erst ab  $n \geq 4$ . Daraus folgt unter anderem, daß man aus einem beliebigen Sortiment von drei Atomen höchstens *ein* Molekül oder *EM* erzeugen kann, das zur Oktettchemie gehört, und daß alle Reaktionen dieses *EM* aus der Oktettchemie herausführen müssen<sup>[31]</sup>.

### 2.3.2. Die Basiselemente der R-Matrizen

Eine R-Matrix des Typs

$$\mathbf{U}^{ij} = \begin{pmatrix} & i & j \\ \cdots & +1 & \cdots & 0 & \cdots \\ & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & \cdots & -1 & \cdots \\ & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (21)$$

in der außer  $u_{ii}^{ii} = +1$  und  $u_{jj}^{jj} = -1$  alle Eintragungen gleich 0 sind, entspricht der elementaren *Redoxreaktion*



durch die ein Elektron von  $A_j$  auf  $A_i$  übertragen wird.

Die elementare *Homoapsis* gemäß



wird durch die R-Matrix

$$\mathbf{V}^{ij} = \begin{pmatrix} & i & j \\ \cdots & -1 & \cdots & +1 & \cdots \\ & \vdots & \vdots & \vdots & \vdots \\ \cdots & +1 & \cdots & -1 & \cdots \\ & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (24)$$

wiedergegeben, in der nur die Eintragungen  $v_{ij}^{ij} = v_{ji}^{ji} = +1$  und  $v_{ii}^{ii} = v_{jj}^{jj} = -1$  von Null verschieden sind.

Die R-Matrix  $-\mathbf{V}^{ij}$  gilt für den umgekehrten Vorgang, die elementare *Homolyse*



Eine R-Matrix mit den Eintragungen  $r_{ij}$  und der  $i$ -ten Zeilen-summe

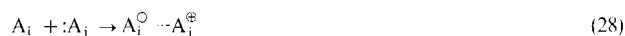
$$r_i = \sum_j r_{ij} \quad (26)$$

läßt sich als Summe von R-Matrizen  $\mathbf{U}^{in}$  und  $\mathbf{V}^{ij}$  (bis auf die Reihenfolge) eindeutig darstellen:

$$\mathbf{R} = \sum_{i < j} r_{ij} \mathbf{V}^{ij} + \sum_{i \leq n} r_i \mathbf{U}^{in} \quad (27)$$

Die Menge der  $\mathbf{U}^{in}$  und  $\mathbf{V}^{ij}$  bildet eine *Basis* zur Darstellung aller R-Matrizen. Da die R-Matrizen als Vektoren des  $\mathbb{R}^{n^2}$  aufgefaßt werden können, werden die  $\mathbf{U}^{in}$  und  $\mathbf{V}^{ij}$  als *Basisvektoren* bezeichnet.

Die Vektorbasis der R-Matrizen für Reaktionen aus der Chemie von Valenzschalen mit gepaarten Elektronen (mit  $r_{ij} = 0, \pm 2, \dots$ ) besteht aus R-Matrizen des Typs  $(\mathbf{U}^{ij} + \mathbf{V}^{ij})$  und  $2\mathbf{U}^{in}$ . Der Basisvektor  $(\mathbf{U}^{ij} + \mathbf{V}^{ij})$  repräsentiert die Reaktionstypen (28) und (29):



Anhand der R-Matrix aus Gleichung (20) sei die Zerlegung von R-Matrizen in Basisvektoren erläutert:

$$\mathbf{R}(\text{HCN} \rightarrow \text{HNC}) = -(\mathbf{U}^{12} + \mathbf{V}^{12}) + (\mathbf{U}^{13} + \mathbf{V}^{13}) \quad (30)$$

$$\mathbf{R}(\text{HCN}/\text{HNC}) =$$

$$\begin{pmatrix} 0 & -1 & +1 \\ -1 & +2 & 0 \\ +1 & 0 & -2 \end{pmatrix} = -\begin{pmatrix} 0 & +1 & 0 \\ +1 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & +1 \\ 0 & 0 & 0 \\ +1 & 0 & -2 \end{pmatrix} \quad (31)$$

### 2.3.3. Einschränkungen für R-Matrizen

Die Transformation einer BE-Matrix  $\mathbf{B}$  durch Addition einer R-Matrix  $\mathbf{R}$  gemäß  $\mathbf{B} + \mathbf{R} = \mathbf{E}$  gibt nur dann eine chemische Reaktion wieder, wenn für alle Eintragungen gilt:

$$e_{ij} = b_{ij} + r_{ij} \geq 0 \quad (32)$$

denn eine BE-Matrix kann per definitionem keine negativen

Eintragungen enthalten. Somit sind die negativen Eintragungen der R-Matrix so zu wählen, daß gilt:

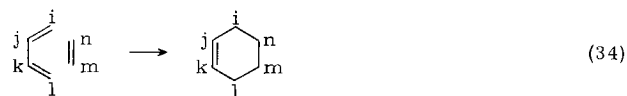
$$|r_{ij}| \leq b_{ij} \text{ für } r_{ij} < 0 \quad (33)$$

Ferner müssen die Eintragungen  $e_{ij}$  der resultierenden Matrix **E** Werte haben, welche für die zugehörigen chemischen Elemente valenzchemisch akzeptabel sind.

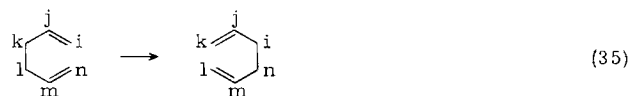
Man kann also bei gegebener BE-Matrix **B** anhand von deren positiven Eintragungen  $b_{ij}$  die negativen Eintragungen  $r_{ij}$  von *mathematisch passenden* R-Matrizen festlegen. Positive Eintragungen werden dann so gewählt, daß passende R-Matrizen mit  $\sum_{ij} r_{ij} = 0$  resultieren. Unter diesen wählt man aufgrund valenzchemischer Begrenzungen in **E** schließlich die erlaubten R-Matrizen aus. Somit ist es möglich, für ein gegebenes *EM* sämtliche denkbaren Reaktionen und deren Produkte aufgrund der Transformations-Eigenschaften von BE-Matrizen und der bekannten valenzchemischen Eigenschaften der chemischen Elemente vorauszusagen. Durch Anwendung aller passenden R-Matrizen erhält man aus einer einzigen BE-Matrix die BE-Matrizen der gesamten Familie der isomeren Ensembles von Molekülen (*FIEM*).

Untersucht man z. B. die Reaktionsmöglichkeiten von  $H-C \equiv N$ : aufgrund der Transformations-Eigenschaften der BE-Matrix **B**(HCN), so findet man, daß es zu **B**(HCN) u. a. eine passende R-Matrix gibt mit  $r_{12} = r_{21} = -1$  und  $r_{33} = -2$ . Aufgrund der zuvor definierten Struktur der R-Matrizen ergibt sich unter Berücksichtigung der valenzchemischen Eigenschaften der Elemente H, C und N, daß  $r_{13} = r_{31} = +1$  und  $r_{22} = +2$  als komplementäre positive Eintragungen in Frage kommen. Daraus resultiert die in Gleichung (20) gezeigte R-Matrix.

Da eine R-Matrix im allgemeinen nicht nur zu einer einzigen, sondern zu mehreren BE-Matrizen mathematisch und valenzchemisch paßt, repräsentiert sie eine Kategorie chemischer Reaktionen, denen das durch die R-Matrix angegebene Elektronenverschiebungsschema und Merkmale der beteiligten Bindungssysteme gemeinsam sind.



Das Reaktionsschema (34) beschreibt eine Diensynthese nach *Diels* und *Alder*, das Schema (35) eine Cope-Umlagerung.



Die Addition einer R-Matrix der Form

$$\mathbf{R} = \begin{bmatrix} & i & j & k & l & m & n \\ \cdot & -1 & \cdot & \cdot & \cdot & \cdot & +1 \\ -1 & \cdot & +1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & +1 & \cdot & -1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & -1 & \cdot & +1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & +1 & \cdot & -1 & \cdot \\ +1 & \cdot & \cdot & \cdot & \cdot & -1 & \cdot \end{bmatrix} \quad (36)$$

zu einer BE-Matrix kann, je nach Konstitution des *EM*, das diese BE-Matrix repräsentiert [vgl. Reaktion (34) und (35)],

u. a. eine Diensynthese oder eine Cope-Umlagerung wiedergeben.

Zwei R-Matrizen **R** und **R'** gehören zur gleichen R-Kategorie und repräsentieren gleichartige Reaktionen, falls es eine R-Matrix **R''** gibt, die durch eine Zeilen/Spalten-Permutation in **R** und durch Einfügen oder Entfernen von Nullzeilen und -spalten in **R'** transformierbar ist. Zwei chemische Umsetzungen gehören zur gleichen R-Kategorie, falls sie durch BE-Matrizen-Transformationen der gleichen R-Kategorie dargestellt werden.

Bei den für organisch-chemische Synthesen wichtigen Reaktionen findet im allgemeinen eine Elektronenverschiebung statt, an der bis zu sechs Atome beteiligt sind, und durch die bis zu je drei kovalente Bindungen gelöst bzw. neu geknüpft werden, wobei sich gleichzeitig an einem der Atome dessen Formalladung durch Beteiligung von freien Valenzelektronenpaaren um eine Einheit erhöhen kann und an einem anderen um eine Einheit erniedrigen kann. Solchen Reaktionen entsprechen R-Matrizen mit bis zu je drei negativen und positiven Paaren außerdiagonaler Eintragungen  $r_{ij} = r_{ji} = -1$  bzw.  $r_{kl} = r_{lk} = +1$ . Die von Null verschiedenen Diagonaleintragungen  $r_{ii} = \pm 2$  liegen so, daß alle Zeilen/Spalten-Summen der R-Matrix gleich Null sind, bis auf ein Zeilen/Spalten-Paar mit  $r_i = \sum_j r_{ij} = \pm 1$ .

## 2.4. Chemische Metrik

Eine  $n \times n$ -BE-Matrix **B** kann auch als ein Vektor **b** mit  $n^2$  Komponenten dargestellt werden:

$$\mathbf{b} = (b_{11}, \dots, b_{1n}; b_{21}, \dots, \dots, b_{n1}, \dots, b_{nn}) \quad (37)$$

Dies ergibt eine Einbettung der BE-Matrizen **B** einer *FIEM* in  $\mathbb{R}^{n^2}$ , einem  $n^2$ -dimensionalen metrischen Raum über dem Körper der reellen Zahlen. Die Eintragungen  $b_{ij}$  von **B** können gleichermaßen als kartesische Koordinaten eines Punktes **P**(**B**) in  $\mathbb{R}^{n^2}$  oder als Komponenten eines Vektors in  $\mathbb{R}^{n^2}$  aufgefaßt werden. Wir nennen **P**(**B**) den *BE-Punkt* der BE-Matrix **B**.

Da BE-Matrizen symmetrisch sind, kann überdies jeder BE-Matrix **B** ein BE-Punkt **b** in einem Raum mit weniger Dimensionen, dem  $\mathbb{R}^{n(n+1)/2}$ , zugeordnet werden, indem das obere Dreieck von **B** als Vektor geschrieben wird. Die Summe der Valenzelektronen muß bei allen elektronischen Grenzstrukturen von **B** gleich sein. Daher wird vorher das untere Dreieck zum oberen addiert:

$$\mathbf{b} = (b_{11}, 2b_{12}, \dots, 2b_{1n}, b_{22}, \dots, 2b_{n-1,n}, b_{nn}) \quad (37a)$$

Alle elektronischen Grenzstrukturen des zweiatomigen Fluorwasserstoffs lassen sich demnach bereits im dreidimensionalen Raum  $\mathbb{R}^3$  darstellen (Abb. 1).

Ebenso entspricht eine R-Matrix **R** einem Vektor **r** in  $\mathbb{R}^{n^2}$ . Eine durch die BE-Matrizentransformation  $\mathbf{B} + \mathbf{R} = \mathbf{E}$  wiedergegebene chemische Reaktion wird durch den Vektor **r** repräsentiert, der vom Punkt **P**(**B**) nach **P**(**E**) führt. Die Summe

$$D(\mathbf{B}, \mathbf{E}) = \sum_{i,j} |b_{ij} - e_{ij}| = \sum_{i,j} |r_{ij}| \quad (38)$$

über die Absolutbeträge der Eintragungen von **R** ist gleich der doppelten Anzahl von Valenzelektronen, die an der Reak-

tion beteiligt sind, denn sie werden sowohl durch die negativen als auch durch die positiven Eintragungen von  $\mathbf{R}$  berücksichtigt.

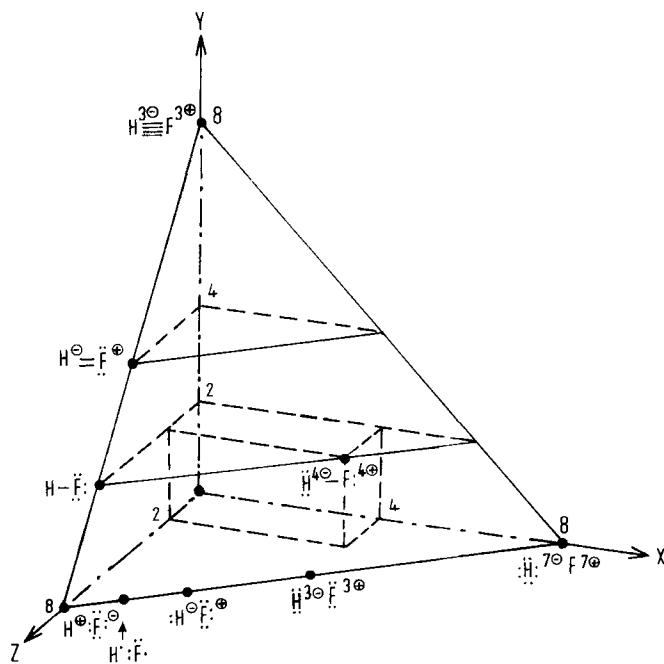


Abb. 1. Die elektronischen Grenzstrukturen des Fluorwasserstoffs liegen auf einer (Hyper-)Fläche, die die Koordinatenachsen jeweils im Punkt 8 (= Zahl aller Valenzelektronen des Moleküls) schneidet.

Wir nennen  $D(\mathbf{B}, \mathbf{E})$  die *chemische Distanz*<sup>[3]</sup> zwischen  $\mathbf{B}$  und  $\mathbf{E}$  sowie  $EM(\mathbf{B})$  und  $EM(\mathbf{E})$ . Durch  $D(\mathbf{B}, \mathbf{E})$  wird im Raum  $\mathbb{R}^n$  eine zur euklidischen Metrik äquivalente Metrik definiert. Die chemische Distanz ist deshalb eine echte Abstandsfunktion der BE-Matrizen einer FIEM. Der Ursprung des Koordinatensystems im  $\mathbb{R}^n$  entspricht der  $n \times n$ -Nullmatrix  $\mathbf{0}$ . Da die Eintragungssumme  $S = \sum_{i,j} b_{ij}$  als Gesamtzahl der Valenzelektronen für alle BE-Matrizen einer FIEM gleich ist, kann eine FIEM auf ein Gitter von Punkten ganzzahliger positiver Koordinaten im  $\mathbb{R}^n$  abgebildet werden, dessen „Radius“  $S = D(\mathbf{B}, \mathbf{0})$  ist. Man beachte, daß der chemische Abstand  $D(\mathbf{B}, \mathbf{E})$  sich unmittelbar auf den Darstellungsraum  $\mathbb{R}^n$  bezieht.

### 3. Algorithmen und Computerprogramme auf der Grundlage von BE- und R-Matrizen

Eine BE-Matrix enthält die gesamte konstitutionsbezogene Information über ein molekulares System, einschließlich der Verteilung der freien Valenzelektronen; eine R-Matrix gibt das Elektronenverschiebungsschema an, welches einer chemischen Reaktion zugrundeliegt.

**Dokumentation von Strukturen und Reaktionen:** BE-Matrizen eignen sich vorzüglich als Grundlage zur Dokumentation von molekularen Strukturen und Substrukturen (siehe Abschnitt 3.1.1).

Die R-Matrizen eröffnen erstmals die Möglichkeit, chemische Reaktionen hierarchisch zu ordnen und zu dokumentieren, wobei die Reaktion direkt beschrieben wird und nicht, wie bisher gebräuchlich, durch die Reaktionsteilnehmer (siehe Abschnitt 3.1.2).

Die Transformation der BE-Matrizen durch R-Matrizen führt zu neuartigen deduktiven Methoden zur Lösung chemischer

Probleme mit Hilfe von Computern. Das deduktive Vorgehen, bei dem man von der allgemeinen Grundlage ausgehend zu speziellen Aussagen kommt, setzt ein universelles Modell der logischen Struktur der Chemie – wie das voranstehend beschriebene – voraus. Unser Modell läßt folgende Ansätze zur Behandlung der dynamischen Aspekte in der Chemie zu:

**Eine BE-Matrix als Ausgangspunkt:** Es ist möglich, von einem gegebenen Ensemble von Molekülen ( $EM$ ) mit der BE-Matrix  $\mathbf{B}$ ,  $EM(\mathbf{B})$ , auszugehen und die hierzu mathematisch und valenzchemisch passenden R-Matrizen zu erzeugen. Aus diesem Ansatz lassen sich Syntheseprogramme und Programme zur Erzeugung von Folgeprodukten vorgegebener Ausgangsstoffe erstellen (siehe Abschnitt 3.2).

**Zwei BE-Matrizen als Ausgangspunkte:** Sind zwei BE-Matrizen  $\mathbf{B}$  und  $\mathbf{E}$  gegeben, welche zur gleichen Familie isomerer Ensembles von Molekülen ( $FIEM$ ) gehören, so ermöglicht die Zerlegung der R-Matrix  $\mathbf{R} = \mathbf{E} - \mathbf{B}$  in Komponenten  $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_r$  gemäß  $\mathbf{R} = \mathbf{R}_1 + \mathbf{R}_2 + \dots + \mathbf{R}_r$  Aussagen über die chemischen Wege, welche von  $EM(\mathbf{B})$  nach  $EM(\mathbf{E})$  führen. Programme auf dieser Grundlage erzeugen Netzwerke von Reaktions- oder Synthesewegen (siehe Abschnitt 3.3).

**Eine R-Matrix als Ausgangspunkt:** Die Ermittlung aller Paare  $(\mathbf{B}, \mathbf{E})$  von BE-Matrizen, welche im Sinne von  $\mathbf{B} + \mathbf{R} = \mathbf{E}$  mathematisch und valenzchemisch zu einer vorgegebenen R-Matrix  $\mathbf{R}$  passen, eröffnet die Möglichkeit, in systematischer Weise chemische Reaktionen vorauszusagen (siehe Abschnitt 3.4).

### 3.1. Klassifikation und Dokumentation

#### 3.1.1. Strukturen und Substrukturen

In Systemen zur chemischen Dokumentation ist die Strukturbeschreibung eines Moleküls zunächst nur einer von mehreren Schlüsseln für den Zugriff zu weiteren Informationen, wie Referaten, Patenten, Aussagen über Stoffeigenschaften. Typisch für Systeme zur chemischen Dokumentation ist, daß die Schlüsselinformation in charakteristischer Weise strukturiert ist.

Einerseits unterliegt die Struktur einer chemischen Verbindung einer Reihe von chemischen Gesetzmäßigkeiten, insbesondere Valenzregeln, andererseits enthält sie stets auch eigenständige Substrukturen (funktionelle Gruppen, Ringsysteme usw.), die wiederum als Zugriffsschlüssel dienen können. Dies gilt bei Anwendungsgebieten wie

- Auffinden von solchen Molekülen aus einer Moleküldatei, die die Anfrage-Substruktur enthalten (reines Retrieval)
- Korrelation von Substrukturen mit Moleküleigenschaften (Substruktur/Wirkungs-Korrelation)
- Analogierecherchen bei unvollständiger Strukturinformation (Aufsuchen „ähnlicher“ Strukturen)
- Erkennen von gemeinsamen Substrukturen von Zielmolekülen und Ausgangsverbindungen in der Syntheseplanung
- Dokumentation und Auswertung von spektroskopischen Daten.

Ein Dokumentationssystem, das Leistungen dieser Art erbringt, muß die Information über die chemische Struktur vollständig speichern und wiedergeben können. Es muß aber auch spezielle Manipulationen durchführen wie Fragmentierung, valenzchemische Plausibilitätsprüfungen, Substrukturzugriff.

Voraussetzung für die Realisierung solcher Operationen ist die Darstellung der konstitutionellen Sachverhalte auf der Grundlage eines mathematischen Modells (siehe Abschnitt 2). Die dadurch mögliche Behandlung chemischer Reaktionen innerhalb einer geschlossenen Theorie wird in den folgenden Abschnitten näher erläutert; zunächst sei auf die Manipulation chemischer Daten, aufbauend auf den im mathematischen Modell definierten BE-Matrizen, eingegangen.

Matrizendarstellungen der chemischen Konstitution, z. B. Adjazenz- und Konnektivitäts-Matrizen<sup>[14, 15]</sup>, sind verschiedentlich in Dokumentationssystemen verwendet worden. Solche Matrizen ermöglichen zwar die Grunddienste eines Dokumentationssystems, also Abspeicherung und Wiederauffinden der Information, jedoch waren Manipulationen der Daten, die chemischen Operationen entsprechen, und valenzchemische Untersuchungen im allgemeinen nicht möglich.

Für eine BE-Matrix gilt jedoch, daß die Eintragungen einer ihrer Zeilen mindestens einem der für den zugehörigen Atomrumpf erlaubten Elektronenverteilungsschemata (Valenzschemata) genügen müssen. Die für einen gegebenen Atomrumpf erlaubten Valenzschemata lassen sich z. B. als Liste darstellen und grenzen den in der jeweiligen Anwendung zugelassenen Bereich der Chemie eines Elementes ab. Valenzschemata ermöglichen wirkungsvolle Plausibilitätsprüfungen bei der Dateneingabe oder beim Umcodieren aus fremden Datenbeständen und insbesondere die Ergänzung und Konsistenzprüfung älterer Datenbestände, die eine erhebliche Investition bilden. In solchen Datenbeständen ist wegen der jeweils verwendeten Codierungsform meist ein Teil der Strukturinformation unterdrückt und muß gegebenenfalls rekonstruiert werden. Häufig fehlen Angaben über freie Elektronen oder auch Ladungen. Ein Vergleich mit den erlaubten Valenzschemata ergibt für die erforderlichen Ergänzungen einen oder wenige Vorschläge, die dann anhand anderer Information, in krassen Fällen auch manuell, selektiert werden müssen. In jedem Fall ist durch diese Bilanz der Valenzelektronen sichergestellt, daß nur erlaubte und vollständige Angaben über die chemische Konstitution in die Datenbestände eingehen. Eine solche vollständige Darstellung durch BE-Matrizen bietet für Dokumentationssysteme den einzigartigen Vorteil, daß bei der Erstellung des Systems noch nicht die Fragestellungen bekannt sein müssen, die in Zukunft interessieren könnten.

Bisher sind Dokumentationssysteme in der Regel im Hinblick auf die jeweils geforderten Fragestellungen entworfen worden, wobei häufig ein Teil der Konstitutionsinformation weggelassen wurde. Dies gilt insbesondere für die vielfach verwendeten Fragmentcodes.

Auf der Grundlage einer solchen vollständigen Darstellung der chemischen Konstitution konnten wir ein interaktives Speicherungs- und Zugriffssystem für chemische Strukturen und Substrukturen aufbauen. Die hierbei verwendeten Techniken der Plausibilitätsprüfung, der Fragmentierung und des Substrukturzugriffs wurden in einer FORTRAN Programmstudie realisiert. Ein Folgesystem, das erweiterte Leistungsmerkmale beim Substruktur-Retrieval und bei der Substruktur/Wirkungs-Korrelation bietet, wird gegenwärtig in PL/1 implementiert.

Um die konkurrierenden Forderungen nach semantischer Unterstützung (z. B. bei der Syntheseplanung: Erkennen von funktionellen Gruppen und gemeinsamen Substrukturen), minimalem Speicherbedarf und minimaler Zugriffszeit zu erfüllen,

zerlegten wir das Substruktur-Retrieval in die rechenintensive Fragmentierungsphase und schnelle Retrievalphasen.

Unter Fragmentierung (1. Phase) verstehen wir die Zerlegung der Struktur von Molekülen in Partialstrukturen (Fragmente), die durch Brechen einer einzigen Bindung oder Erniedrigung einer Bindungsordnung entstehen. Jede so unmittelbar erzeugte Partikel erhält einen Verweis auf das Molekül (*Vaterverweis*) und umgekehrt: Das Molekül erhält Verweise auf die unmittelbar erzeugten Partialstrukturen (*Sohnverweis*). Dieser Prozeß der unmittelbaren Erzeugung wird auf die Partialstrukturen rekursiv so lange angewendet, bis nur noch einzelne Atome als Fragmente auftreten. Man erhält somit ein Netzwerk, in dem die Moleküle und Fragmente die Knoten und die Verweise die Kanten bilden.

Bei der anschließenden Fragmentierung des nächsten Moleküls treten meist nur noch große neue Substrukturen auf, wohingegen kleinere Substrukturen als schon vorhanden erkannt werden und nicht mehr weiter zerlegt werden müssen. Auf diese Weise tritt bei der Generierung der Fragmente eine Sättigung des Substrukturbestandes auf (Abb. 2).

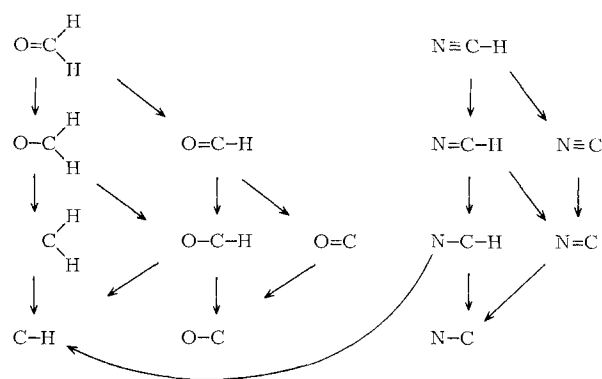


Abb. 2. Fragmentierung von Formaldehyd und Cyanwasserstoff mit Sohnverweisen.

Eine typische Substruktur-Retrieval-Anfrage wird demnach in der Retrievalphase (2. Phase) wie folgt behandelt: Das Anfragefragment (z. B. C—H) wird mit einem Kanonisierungsverfahren in eine eindeutige Darstellung gebracht, aufgrund derer eine Speicheradresse berechnet wird. Über diese steigt man in das Verweisnetz ein und verfolgt alle Vaterverweise bis zu den Molekülen. Die so erreichten Moleküle enthalten gemäß der Konstruktion des Verweisnetzes aus der 1. Phase das Anfragefragment als Partialstruktur.

Bei der Untersuchung von Substruktur/Wirkungs-Beziehungen interessiert man sich in der Retrievalphase für Substrukturen, die möglichst nur den Molekülen gemeinsam sind, die eine bestimmte Wirkung (z. B. oxidationshemmend, fungizid, bakteriostatisch) haben. Das Programm verfolgt dabei Sohnverweise, die von den Molekülen mit der gewünschten Wirkung ausgehen, solange über mehrere Generationen von Söhnen, bis Substrukturen gefunden werden, die statistisch signifikant häufiger bei den Ausgangsmolekülen als bei den restlichen Molekülen der Moleküldatei vorkommen<sup>[18]</sup>.

### 3.1.2. Reaktionen

Die Systematik und Nomenklatur chemischer Reaktionen ist bei weitem noch nicht so ausgearbeitet wie diejenige chemischer Strukturen. Es ist symptomatisch, daß es immer noch

weitgehend üblich ist, Reaktionen mit Trivialnamen zu bezeichnen.

Bisher wurden chemische Reaktionen meist durch ihre Ausgangsstoffe und Endprodukte – ganz im Sinne einer statisch orientierten strukturbezogenen Auffassung – beschrieben. Die bisherigen Dokumentationssysteme bauten auf dieser Grundlage auf.

Mit Begriffen des mathematischen Modells der Chemie ausgedrückt heißt dies, daß man zwei Punkte im Vektorraum einer Familie isomerer Ensembles von Molekülen als Beschreibung einer Reaktion ansieht. Es fehlte indessen eine Beschreibung des Vektors zwischen diesen Punkten.

Bei der Beschreibung einer chemischen Reaktion gehen wir vom zugrundeliegenden Elektronenverschiebungsschema aus. Dieses wird in einer R-Matrix erfaßt. Erst die Erkenntnis, daß sich eine chemische Reaktion durch Matrizen einer ganz bestimmten Art (R-Matrizen) darstellen läßt und daß diese Darstellung in sich wieder strukturiert ist, eröffnet die Möglichkeit der Generalisierung. Die Stufen der Generalisierung sind (am Beispiel chemischer Reaktionen): Irreduzible R-Matrix (*Reaktionskern*)  $\supset$  Reaktionskern und zugehöriger Atomvektor  $\supset$  BE-Matrix des Reaktionskerns  $\supset$  Atomvektoren und BE-Matrizen der folgenden Nachbarsphären des Reaktionskerns.

Ein gegebenes Elektronenverschiebungsschema kann eine Vielzahl von Reaktionen beschreiben, die der gleichen Reaktionskategorie angehören (siehe Abschnitt 2.3.2). Eine unter diesem Gesichtspunkt gegliederte Aufzählung von Reaktionen wurde kürzlich veröffentlicht<sup>[19]</sup> (siehe dazu Abb. 3).

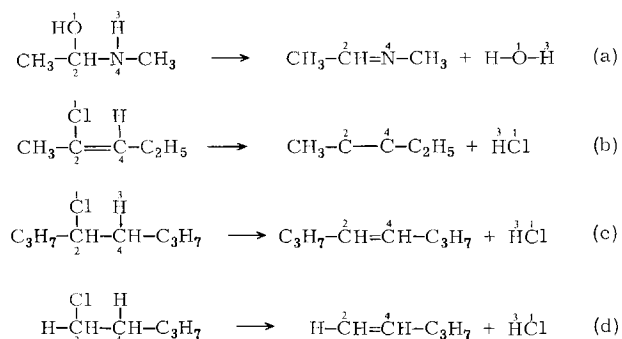


Abb. 3. Vier Reaktionen, die zur gleichen R-Kategorie gehören. (a) unterscheidet sich von (b), (c) und (d) durch die abweichende Besetzung des Atomvektors am Reaktionskern: O, C, H, N bei (a) und Cl, C, H, C bei (b), (c) und (d) (gleicher „RA-Typ“); (b) unterscheidet sich von (c) und (d) durch die abweichende BE-Matrix im Reaktionskern:  $b_{24}=2$  bei (b),  $b_{24}=1$  bei (c) und (d) (gleicher „RB-Typ“); (c) und (d) unterscheiden sich nur durch die abweichende Belegung der ersten Nachbarsphäre des Reaktionskerns. Die zu dieser R-Kategorie gehörende R-Matrix lautet:

$$\mathbf{R} = \begin{pmatrix} 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{pmatrix}$$

Aus dem mathematischen Modell folgt, daß – innerhalb gegebener Schranken – nur eine begrenzte Anzahl irreduzibler R-Matrizen möglich ist. Jede dieser irreduziblen R-Matrizen bestimmt eine R-Kategorie, und jede Reaktion ist in eine dieser R-Kategorien einzuordnen. Einer R-Kategorie kann eine mehr oder weniger große, aber begrenzte Anzahl von Valenzschemata zugeordnet werden, durch welche alle zu einer R-Matrix passenden Atomvektoren bestimmt sind. Hierbei werden nur die Atome betrachtet, die dem Reaktionskern angehören, d. h. denen in der R-Matrix keine Nullzeilen/Null-

spalten entsprechen. Die unterschiedlichen Atomvektoren untergliedern eine R-Kategorie in Reaktionstypen, die wir RA-Typen nennen. Diese sind weiter untergliedert durch Unterschiede in den Bindungen zwischen den Atomen des Reaktionskerns, also durch die unterschiedlichen Eintragungen in dem Block der BE-Matrix, der dem (Atomvektor für den) Reaktionskern entspricht. Die Berücksichtigung weiterer Nachbarsphären des Reaktionskerns kann dann zu weiteren, spezifischen Unterteilungen herangezogen werden. Reaktionen können mit Hilfe dieses Formalismus in genereller Form codiert werden. Damit eröffnet sich die Möglichkeit zu Abfragen von Reaktionen, bei denen Bestimmungsstücke offen bleiben.

### 3.2. Voraussage von Folgeprodukten und Syntheseplanung

Bei der Syntheseplanung analysiert man das Synthesziel nach Teilstrukturen, welche mit bekannten Reaktionen herstellbar sind, und verwendet die Umkehrungen solcher Reaktionen, um die Vorstufen zu konstruieren. Falls man dieses Vorgehen in einem Programm verwirklichen wollte, müßte man eine Reaktionssammlung erstellen. Eine solche Datei wäre eine Anhäufung von Einzelreaktionen, die alle getrennt eingespeichert werden müßten. Die fortschreitende Verbesserung eines solchen Syntheseplanungsprogramms würde eine ständige Erweiterung und Pflege der Reaktionsbibliothek erfordern – und doch könnte ein solches Programm immer nur bekannte Chemie erzeugen.

Das mathematische Modell ermöglicht ein völlig andersartiges Vorgehen bei der Syntheseplanung<sup>[8, 20–22]</sup>. Durch die Anwendung von R-Matrizen auf die BE-Matrix vorgegebener Moleküle und Ensembles von Molekülen lassen sich die BE-Matrizen der Reaktionsprodukte erhalten. Man kann aber die R-Matrizen nicht nur als Beschreibung des Übergangs von Reaktanden zu Produkten betrachten, sondern sie auch als Formulierung von Reaktionen in ihrer Umkehrrichtung, also von den Produkten zurückschauend auf die Ausgangsstoffe, ansehen. Ausgehend vom Synthesziel kann man nun auf sämtliche Vorstufen schließen, die zum Synthesziel führen.

#### 3.2.1. Operationen mit R-Matrizen

Das Programmsystem EROS (Erzeugung von Reaktionen für die Organische Synthese)<sup>[20]</sup> erzeugt aus der BE-Matrix eines Ensembles von Molekülen (*EM*) durch Anwendung von R-Matrizen BE-Matrizen weiterer *EM*. Je nachdem ob die Reaktionen in Vorwärts- oder Rückwärtsrichtung betrachtet werden – dies hängt von der Problemformulierung ab – kann EROS demnach sowohl zur Erzeugung der zu erwartenden Reaktionsprodukte von vorgegebenen Verbindungen als auch zur Syntheseplanung eingesetzt werden.

Besteht das *EM* nur aus einem Molekül, so erzeugt das Programm Umlagerungs- und Spaltprodukte. Sollen auch andere Reaktionstypen erzeugt werden, sind weitere Moleküle hinzuzufügen; deren Auswahl ergibt sich aus der Problemstellung.

Will man die zu erwartenden Reaktionsprodukte eines Moleküls untersuchen, so sind noch potentielle Reaktionspartner hinzuzufügen: Beim Studium der zu erwartenden Folgeprodukte einer Chemikalie (z. B. Pflanzenschutzmittel) in der Umwelt sind dies H<sub>2</sub>O, O<sub>2</sub>, O<sub>3</sub>, CO<sub>2</sub> usw.; bei der Suche nach Verwendungsmöglichkeiten eines industriellen Abfallproduktes kommen leicht zugängliche Chemikalien in Frage.



Bei der Syntheseplanung ist die Frage nach den Molekülen, mit denen das Zielmolekül zu ergänzen ist, nicht so einfach zu beantworten. Ein genaueres Studium von synthetisch wichtigen Reaktionen zeigt, daß hier meist kleine Moleküle freigesetzt werden, und man begegnet dabei immer wieder den gleichen energiearmen Stoffen wie  $\text{H}_2\text{O}$ ,  $\text{HCl}$ ,  $\text{NaCl}$ ,  $\text{NH}_3$ ,  $\text{CO}_2$ ,  $\text{N}_2$ . Für die retrosynthetische Betrachtungsweise sind demnach diese Moleküle mit dem gewünschten Synthesziel zu einem EM zusammenzufassen. Dadurch werden Synthesereaktionen berücksichtigt, bei denen diese Moleküle abgespalten werden.

Da aufgrund des mathematischen Modells sämtliche denkbaren Reaktionsmöglichkeiten erzeugt werden können, gilt es, diese unter dem Gesichtspunkt des vorgegebenen Problems sinnvoll zu bewerten. Einen bedeutenden Teil der Bewertung kann das Programm selbst leisten. Dabei werden physikalisch-chemische und heuristische Kriterien herangezogen.

So werden vornehmlich Mehrfachbindungen und Bindungen zu Heteroatomen und deren Nachbarbindungen gebrochen. Aus Parametern über 1,2- und 1,3-Wechselwirkungen, erhalten aus thermochemischen Daten, läßt sich die Reaktionsenthalpie abschätzen<sup>[23]</sup>.

Es werden mehrere Modelle zur Bestimmung von Aktivierungsparametern getestet. Die sterische Beeinflussung einer Reaktion kann durch Absuchen der Bindungsliste der an der Reaktion beteiligten Atome ermittelt werden. Aufgrund von Elektronegativitätsbetrachtungen<sup>[24]</sup> lassen sich gewisse Atomverknüpfungen als ungünstig erkennen.

Da Synthesen fast immer mehrstufig durchgeführt werden, ist das Problem der Syntheseplanung noch nicht dadurch gelöst, daß man mit den R-Matrizen aus einem Ausgangs-EM die synthetischen Vorstufen des 1. Niveaus erzeugt und diese dann bewertet. Vielmehr führt man sie wieder dem Programm zu und erzeugt aus ihnen ein weiteres Niveau synthetischer Vorstufen. Durch Wiederholen dieses Vorgangs erhält man schließlich einen „Synthesebaum“, dessen Wurzel das Synthesziel ist und dessen Knoten Reaktionen oder Moleküle sind (Abb. 4).

Eine Reaktionsfolge kann von mehreren Molekülen als synthetischen Vorstufen ausgehen. Teilbäume des Synthesebaums, die bis zur Wurzel reichen, repräsentieren vollständige Synthesewege. Die Entwicklung eines Syntheseweges ist als

abgeschlossen zu betrachten, wenn alle erforderlichen Verbindungen zugängliche Ausgangsmaterialien sind.

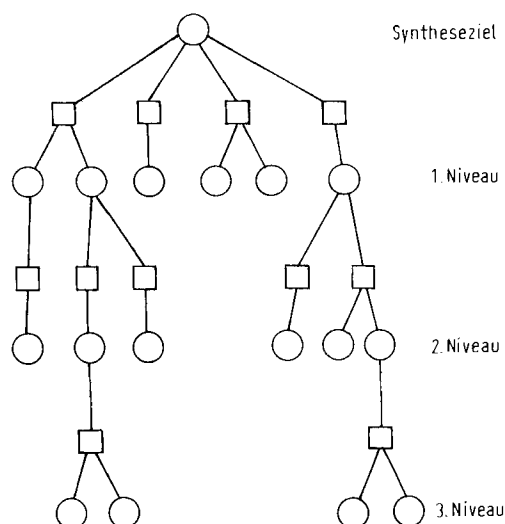


Abb. 4. „Synthesebaum“. ○ Moleküle, ■ Reaktionen.

Bei der Bewertung von Synthesewegen genügt es nicht, nur die einzelnen Schritte – also die individuellen synthetischen Reaktionen – zu bewerten, sondern es müssen ganze Zweige eines Synthesebaums bewertet werden, denn der Erfolg einer Synthese ist abhängig von der Güte sämtlicher Schritte und der Zugänglichkeit aller erforderlichen Verbindungen.

Mit dem Programmsystem EROS wurde die Erzeugung von Folgeprodukten und Syntheseplanung in den Bereichen der organischen Synthese, der industriellen Chemie und der Biochemie an Beispielen studiert. Dabei konnte die Flexibilität und die Vorhersagefähigkeit von EROS unter Beweis gestellt werden.

Als Illustration diene eine Studie zur Synthese von Guanin. Von den erhaltenen Synthesevorschlägen ist eine Auswahl in Abbildung 5 aufgeführt. Im Syntheseweg A werden die beiden letzten Stufen der Biosynthese wiedergegeben. Syntheseweg B entspricht in den ersten beiden Schritten der Traubeischen Harnsäuresynthese; die weitere Entwicklung zu Dicyandiamid und Glycinester ist neuartig. Ob für die Wege C und D Beispiele in der Literatur zu finden sind, ist uns nicht bekannt. Bei einigen der aufgeführten Reaktionsschritte dürfte die

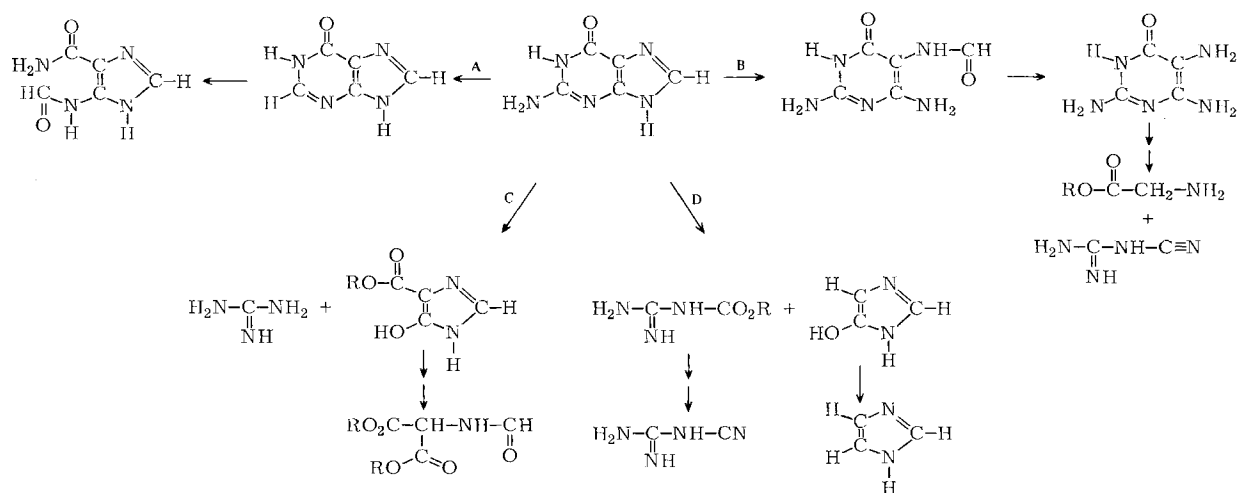


Abb. 5. Eine Auswahl der mit dem Programmsystem EROS erhaltenen Vorschläge zur Guanin-Synthese.

Blockierung weiterer reaktiver Stellen oder eine zusätzliche Aktivierung der reagierenden Gruppen nötig sein, um die Synthesen in die gewünschte Richtung zu lenken.

### 3.2.2. Operationen mit Basiselementen von R-Matrizen

Wir entwickeln zur Zeit ein Programmsystem ASSOR (Allgemeines System zur Simulation organischer Reaktionen)<sup>21)</sup>, das zur Simulation chemischer Reaktionen nicht Linearkombinationen von Basiselementen (siehe Abschnitt 2.3.1), sondern die Basiselemente selbst verwendet. Diese Basiselemente bilden den einfachsten *geschlossenen* Satz von R-Matrizen, mit denen *alle* Reaktionen dargestellt werden können. Da durch die Basiselemente eine Reaktion in Elementarschritten simuliert wird, bietet sich nach jedem Schritt die Möglichkeit, anhand von Plausibilitätsprüfungen zu entscheiden, ob ein bestimmter Reaktionspfad noch weiterverfolgt werden soll. Dies führt zu einem Reaktionsnetzwerk, das aus mechanistischen Elementarschritten besteht und weitgehend frei von mechanistisch nicht realisierbaren Verzweigungen ist. Neben dieser Auswahl bei der Erzeugung von Reaktionswegen ist es notwendig, die Reaktionswege abschließend zu bewerten, um dem Benutzer nicht zu viele Lösungen vorschlagen zu müssen.

Das Ergebnis eines Programmablaufs, bei dem Essigsäureethylester als Zielmolekül und Wasser als Coprodukt vorgegeben werden, wird in Abbildung 6 gezeigt.

### 3.3 Netzwerke von Reaktionsmechanismen und Reaktionsfolgen

Sowohl der Mechanismus einer chemischen Reaktion als auch eine Folge von Reaktionen, z. B. ein Syntheseweg, lassen sich gleichermaßen gut mit dem oben beschriebenen mathematischen Modell der konstitutionellen Chemie wiedergeben. Die Gleichung

$$\mathbf{B} + \mathbf{R} = \mathbf{E} \quad (14)$$

beschreibt bei einem Reaktionsmechanismus die Umsetzung eines Ausgangsstoffs **B** zum Zielprodukt **E**, während sie bei der Beschreibung eines Syntheseweges die Summe aller für die Einzelreaktionen nötigen Ausgangsstoffe (**B**) oder aber alle bei allen Reaktionen des Syntheseweges anfallenden Zielprodukte (**E**) erfaßt. Also gibt **R** im einen Fall die während der Reaktion zu brechenden oder zu knüpfenden Bindungen an und im anderen Fall sämtliche bei allen Reaktionen zu brechenden oder zu knüpfenden Bindungen. Ein großer Teil der mathematischen Behandlung von Reaktionsmechanismen und Synthesewegen kann deshalb gemeinsam durchgeführt werden. Lediglich bei der Aufspaltung der Summenreaktionsmatrix **R** in Matrizen der Teilreaktionen müssen, je nachdem ob Reaktionsmechanismen oder Reaktionsfolgen betrachtet werden, unterschiedliche Gesichtspunkte herangezogen werden (s. u.).

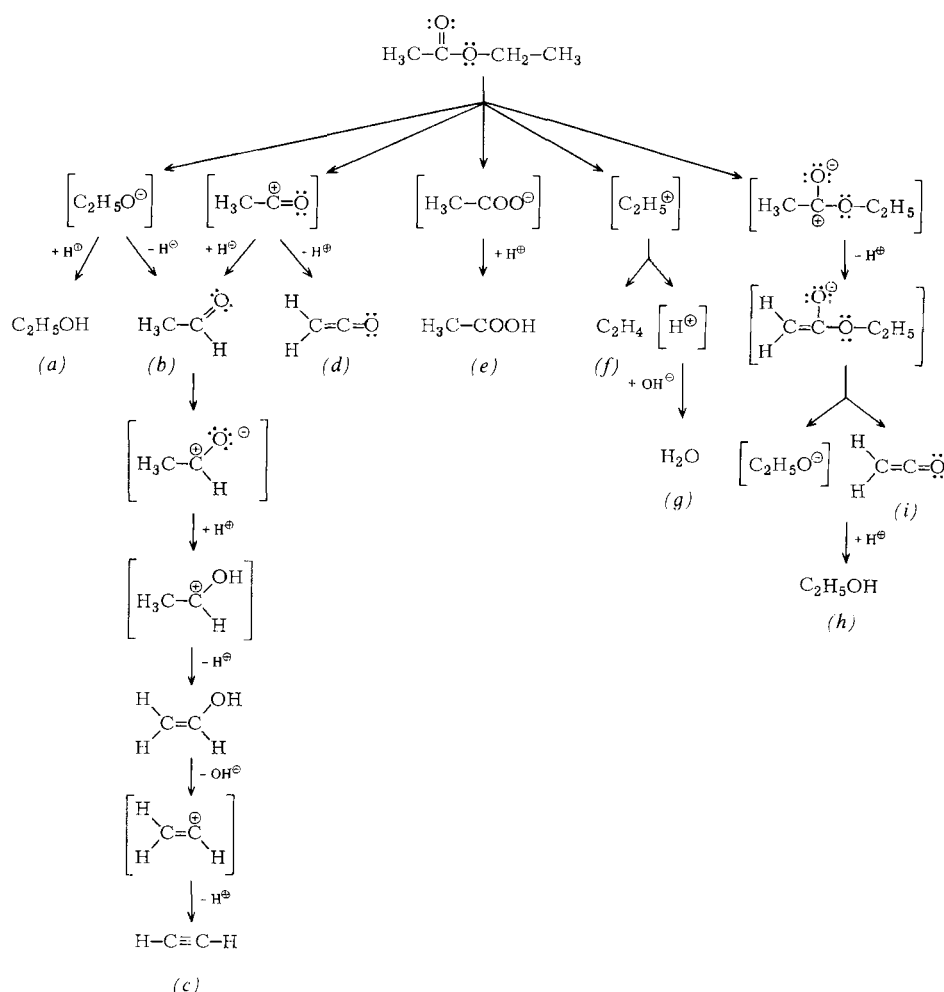


Abb. 6. Mit dem Programmsystem ASSOR erzeugte Wege zur Synthese von Essigsäure-ethylester mit Basiselementen. (a)+(e): Veresterung von Essigsäure mit Ethanol; (b): Dimerisierung von Acetaldehyd (*Tischtschenko*); (c): Wasseranlagerung an Acetylen zu Acetaldehyd; (a)+(d) oder (h)+(i): Anlagerung von Ethanol an Keten (*Wacker*); (f)+(g): Wasseranlagerung an Ethylen zu Ethanol.

### 3.3.1. Kürzeste Reaktionswege

Die chemische Erfahrung zeigt, daß bei einer Reaktion möglichst wenig Elektronen verschoben werden. Wir fordern dies in dem Postulat: Bei einer vorgegebenen Bruttoreaktion (Ausgangsstoffe und Endprodukte sind bekannt) repräsentiert die Gesamtreaktionsmatrix einen Elektronenumordnungsprozeß, der in der Nähe der minimalen chemischen Distanz  $D$  liegt<sup>[25]</sup>.

Sind Ausgangs- und Endverbindungen  $EM(\mathbf{B})$ ,  $EM(\mathbf{E})$  einer chemischen Reaktion willkürlich numeriert, so kann durch wechselseitige Zuordnung von Atomen gleicher Nummer ein komplizierterer Bindungs- und Elektronenumordnungsprozeß vorgetauscht werden (b) als tatsächlich abläuft (a) (Abb. 7).

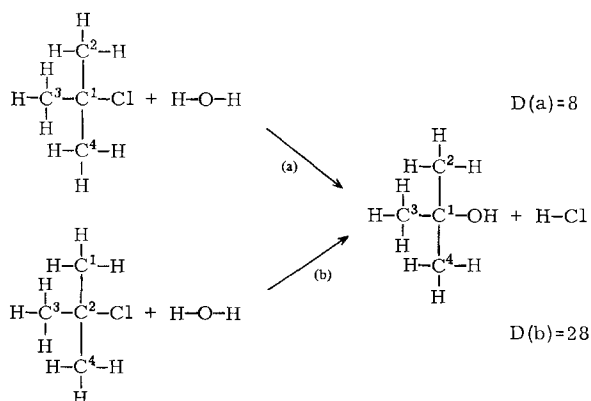


Abb. 7. Zum Postulat der geringstmöglichen Verschiebung von Elektronen.

Aufgabe ist es deshalb, den tatsächlich ablaufenden Bindungs- und Elektronenumordnungsprozeß von solchen formalen Prozessen zu entkleiden: Die Numerierung von Atomen gleicher Ordnungszahl von  $EM(\mathbf{E})$  muß so permutiert werden, daß möglichst wenig Elektronen umgeordnet werden müssen. Dies ist äquivalent mit der Minimierung der chemischen Distanz (siehe Abschnitt 2.4).

Mathematisch formuliert bedeutet dies, daß eine Permutationsmatrix  $\mathbf{P}$  gesucht wird, so daß die Funktion

$$F(\mathbf{P}) = D(\mathbf{B}, \mathbf{P}\mathbf{E}\mathbf{P}^T) \quad (39)$$

minimal wird.

**Lösungsansätze:** Die Durchführung aller Permutationen ist auch mit Hilfe eines Computers selbst bei relativ kleinen Problemen zu aufwendig: Es müßten z. B. bei einer Ausgangs- und Endverbindung mit jeweils zehn C-Atomen mehr als 3 Millionen Permutationen durchgeführt werden.

Wir untersuchen deshalb zur Zeit Methoden aus der ganzzahligen Optimierung, nämlich quadratische Zuordnungsalgorithmen<sup>[26]</sup>, auf ihre Anwendbarkeit auf dieses Problem (Abb. 8). Diese Methoden benötigen jedoch, damit sie nicht zu rechenzeitintensiv werden, bereits eine Permutation  $\mathbf{P}'$ , derart daß  $F(\mathbf{P}')$  bereits in der Nähe des Minimums liegt. Zur Auffindung einer solchen Näherungslösung  $\mathbf{P}'$  haben wir heuristische Methoden entwickelt, die auf dem Vergleich der Umgebungen  $U(A_i, A_j)$  der Atome  $A_i$  aus  $EM(\mathbf{B})$  mit  $A_j$  aus  $EM(\mathbf{E})$  beruhen. Hierbei wird versucht, die Atome aus  $EM(\mathbf{B})$  so auf die Atome auf  $EM(\mathbf{E})$  zu „legen“, daß alle Umgebungen möglichst gut übereinstimmen<sup>[27]</sup>.

Haben wir die minimale chemische Distanz bestimmt, so erhalten wir eine minimale Reaktionsmatrix  $\mathbf{R}_{\text{Min}}$  gemäß der Formel

$$\mathbf{R}_{\text{Min}} = \mathbf{P}\mathbf{E}\mathbf{P}^T - \mathbf{B} \quad (40)$$

Durch geeignete Zerlegung von  $\mathbf{R}$  in Teilreaktionsmatrizen  $\mathbf{R}^i$  mit

$$\mathbf{R}_{\text{Min}} = \sum_i \mathbf{R}^i \quad (41)$$

können dann kürzeste Reaktionswege beschrieben werden.

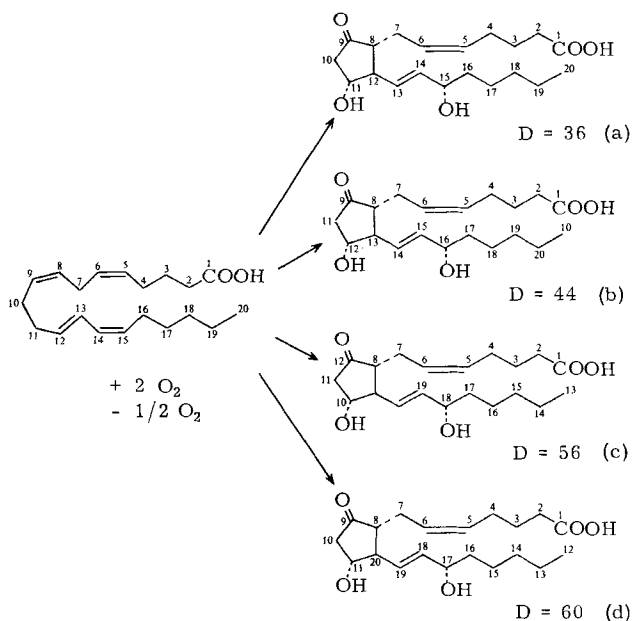


Abb. 8. Verschiedene chemische Distanzen (a) bis (d) bei der Bildung von Prostaglandin  $E_2$  aus 5,8,12,14-Eicosatetraensäure und Sauerstoff. (a) zeigt die Zuordnung der Atome bei der gefundenen minimalen Distanz. (b) bis (d) sind Beispiele der vielen Zuordnungen, die in der Nähe der minimalen chemischen Distanz liegen.

### 3.3.2. Aufbau der Netzwerke

Es seien wieder Ausgangs- und Zielverbindung(en)  $EM(\mathbf{B})$  und  $EM(\mathbf{E})$  vorgegeben. Durch die Bestimmung der minimalen chemischen Distanz  $D(\mathbf{B}, \mathbf{E})$  (siehe oben) erhalten wir eine minimale Reaktionsmatrix  $\mathbf{R}_{\text{Min}}$ , die der Gleichung

$$\mathbf{B} + \mathbf{R}_{\text{Min}} = \mathbf{E} \quad (42)$$

genügt. Um Reaktionsmechanismen oder Reaktionsfolgen zu finden, die  $EM(\mathbf{B})$  in  $EM(\mathbf{E})$  umwandeln, wird die Gesamtreaktionsmatrix in Summen  $\mathbf{S}^k$  von Teilreaktionsmatrizen aufgespalten:

$$\mathbf{S}^k : \mathbf{R}_{\text{Min}} = \sum_i {}^k\mathbf{R}^i \quad (k = 1, \dots, m) \quad (43)$$

Dies entspricht der Aufspaltung einer Summenreaktion in Teilreaktionen. Bei der Ermittlung von Reaktionsmechanismen können diese Teilreaktionen ziemlich klein gewählt werden (z. B. ein Basisvektor oder eine Summe aus wenigen), da hierbei möglichst viele, auch instabile, Zwischenstufen betrachtet werden sollen. Bei der Ermittlung von Synthesewegen hingegen werden die  ${}^k\mathbf{R}^i$  so gewählt, daß stets

stabile Zwischenstufen entstehen. Bei beiden Ansätzen werden auch Teilreaktionsmatrizen zugelassen, deren Eintragungen sich in der Summenreaktionsmatrix zu Null ergänzen, d.h. es werden entlang eines bestimmten Reaktionsweges bestimmte Bindungen gebrochen und wieder geknüpft, die in der Gesamtreaktion als unverändert erscheinen.

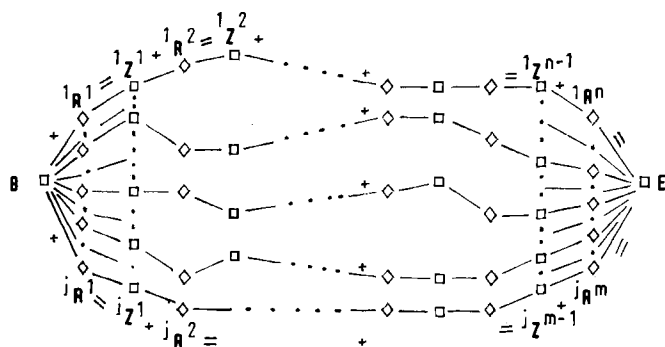


Abb. 9. Aufbau eines Synthesenetzes zwischen Ausgangs- und Zielverbindungen **B** und **E**.

Mit diesen Einzelreaktionsmatrizen wird ein Netz aus Zwischenprodukten  $^kZ^i$  aufgebaut (Abb. 9). Die Teilreaktionsmatrizen müssen jedoch auch so gewählt werden, daß die Zwischenprodukte im Synthese- oder Reaktionsnetz in der Nähe eines Minimalweges liegen, da einerseits Reaktionen in der Natur vermutlich zumindest in der Nähe des Minimalweges verlaufen und andererseits solche Synthesen wegen ihrer Aufwendigkeit verworfen werden müssen, die sich zu weit vom Minimalweg entfernen.

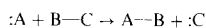
### 3.4. Voraussage chemischer Reaktionen

Während im vorigen Abschnitt die Erstellung aller möglichen R-Matrizen zu gegebenen BE-Matrizen-Paaren behandelt wurde, soll hier der umgekehrte Weg beschrieben werden: die Erzeugung aller zu einem vorgegebenen Elektronenumordnungsschema im chemischen Sinne „passenden“ Reaktionen. Das heißt, daß man zu einer gegebenen R-Matrix „passende“ Paare von BE-Matrizen sucht. Der kreative Aspekt dieses Vorgehens liegt darin, daß in dieser Gesamtheit auch präzedenzlose Reaktionen enthalten sind.

Es ist natürlich chemisch – und übrigens heute auch computertechnisch – nicht sinnvoll, alle kombinatorisch möglichen Reaktionen zu erzeugen. Vielmehr muß man sich durch geeignete Randbedingungen auf einen Bereich beschränken, der dann erschöpfend behandelt werden kann. Für die R-Matrizen kann man dies durch geeignete Begrenzungen erreichen. BE-Matrizen können dann aufgrund von vorgegebenen Valenzschemata erzeugt werden. Solche Valenzschemata sind Abbildungen jeweils einer Zeile von BE-Matrizen. Passende BE-Matrizen werden aus ihnen zeilenweise aufgebaut. Die Addition der zugehörigen Zeile einer R-Matrix ergibt ein neues Valenzschema. Die chemischen Elemente sind jeweils einem oder auch mehreren Valenzschemata zugeordnet. Jede Zeile der Ausgangs- und der Ziel-BE-Matrix ergibt so ein Paar von Valenzschemata. Ist ein solches Paar in der Valenzschemalistes eines Elements enthalten, so kann dieses Element den dieser Zeile zugehörigen Platz im Atomvektor einnehmen. Man erhält so zunächst ein Resultat derart, daß für jeden Platz des Atomvektors eine Liste von allen möglichen Elementbelegun-

gen erscheint (Abb. 10). Die Kombinationen dieser Belegungen werden durch valenzchemische Gesetzmäßigkeiten sowie physikalisch-chemische und heuristische Regeln weiter eingegrenzt.

Vorgegebenes Reaktionsschema:



Äquivalente R-Matrix:

$$R = \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{matrix} A \\ B \\ C \end{matrix}$$

Auszug aus einer Valenzschemalistes:

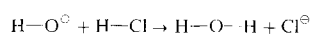
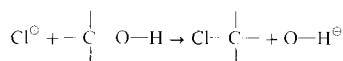
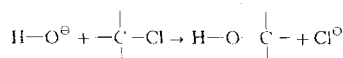
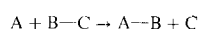
Nr.	Symbol	Valenzschema f, s, d, t	Elemente
v1	$\ddot{X}:$	8, 0, 0, 0,	Cl; Br; I;
v2	$\ddot{X}:$	6, 1, 0, 0,	O; Cl; Br; I;
v3	$-\ddot{X}-$	4, 2, 0, 0,	O;
v4	$\begin{array}{c}   \\ -X \\   \end{array}$	0, 4, 0, 0,	C;
v5	$X$	0, 1, 0, 0,	H;

f = Anzahl der freien Elektronen; s, d, t = Anzahl von Einfach-, Doppel- bzw. Dreifachbindungen

Mögliche Valenzschemapare und Elementbelegungen:

A	v1 → v2	Cl; Br; I;	O; Cl; Br; I;
	v2 → v3	O;	
B	v2 → v2	O; Cl; Br; I;	
	v3 → v3	O;	
	v4 → v4	C;	O; Cl; Br; I; C; H;
	v5 → v5	H;	
C	v2 → v1	Cl; Br; I;	O; Cl; Br; I;
	v3 → v2	O;	

Ausgeschriebene Einzelbeispiele:



etc.

Abb. 10. Vereinfachtes Beispiel der Darstellung der in eine R-Kategorie fallenden Reaktionen: Die für eine (hier stark gekürzte) Valenzschemalistes zu einer R-Kategorie möglichen Elementbelegungen beschreiben die Gesamtheit der denkbaren Reaktionen, zunächst ohne Rücksicht auf die Realisierbarkeit im Einzelfall.

## 4. Perspektiven

Der Fortschritt der Chemie beruht auf Ergebnissen von Experimenten, welche durch heuristisch wertvolle Hypothesen angeregt wurden. Das Bedürfnis nach einer allgemeinen aussagefähigen Theorie konnte bisher nicht befriedigt werden. Auch die Quantenmechanik hat als theoretische Grundlage nicht die Bedeutung erlangt, die man sich aufgrund ihres umfassenden Ansatzes von ihr versprochen hatte. Bisher waren bei der Interpretation und Planung der Experimente induktive Schlüsse und Analogiedenken entscheidend. Bei dieser Vorgehensweise hängt der Fortschritt in hohem Maße von Zufälligkeiten ab. Um systematisch neue Möglichkeiten zu erschließen,

ist es notwendig, die bisherigen Ansätze durch deduktives Vorgehen zu ergänzen. Voraussetzung dafür ist eine Theorie, die es ermöglicht – von allgemeinen Prinzipien ausgehend –, für eine Vielfalt gegebener Probleme die Gesamtheit aller Lösungsmöglichkeiten zu berücksichtigen<sup>[28]</sup>.

Die hier beschriebene Theorie der *Familie isomerer Ensembles von Molekülen* ist ein solcher Ansatz. In der mathematischen Formulierung dieser Theorie werden zur Darstellung der chemischen Konstitution BE-Matrizen verwendet. Die Elektronenverschiebungsschemata, nach denen chemische Reaktionen ablaufen, werden durch R-Matrizen repräsentiert. Eine chemische Reaktion wird dann durch die Addition einer R-Matrix zur vorliegenden BE-Matrix gemäß der Fundamentalgleichung

$$\mathbf{B} + \mathbf{R} = \mathbf{E} \quad (14)$$

ausgedrückt. Darin beschreibt **B** das Ensemble von Molekülen zu Beginn und **E** das Ensemble am Ende der Reaktion.

Für die Anwendung der Fundamentalgleichung ist es wesentlich, daß sie bereits bei Kenntnis eines einzigen Bestimmungsstückes aufgrund der Theorie lösbar ist. Um aus der Gesamtheit aller denkbaren Lösungsmöglichkeiten der Fundamentalgleichung diejenigen auszuwählen, welche für ein Problem relevant sind, ohne daß wertvolle Information verworfen wird, benötigt man Auswahlregeln, welche der chemischen Erfahrung Rechnung tragen. Somit wird das vorhandene chemische Tatsachenwissen zur *Einschränkung* herangezogen, nicht dagegen zur *Erzeugung*.

Die chemische Literatur kann nicht nur verwendet werden, um allgemeine Auswahlregeln für deduktive chemische Computerprogramme zu definieren; sie kann auch in Form von geeignet angelegten Dateien über chemische Verbindungen und Reaktionen unmittelbar herangezogen werden, um festzustellen, ob computererzeugte Problemlösungen bereits bekannt sind. Diese Art von Vorgehen ist für künftige Entwicklungen sehr vielversprechend, denn sie umfaßt sowohl den Zugriff zur vorhandenen dokumentierten Literatur als auch die Vorzüge des mathematischen Modells, welches grundsätzlich die Gesamtheit aller denkbaren Möglichkeiten zu berücksichtigen gestattet. Die Matrizenaddition als Grundoperation ist computermäßig sehr viel rationeller als die Datenmanipulation anhand publizierter Reaktionen.

Die bisher implementierten Programme und die damit erzielten Ergebnisse haben gezeigt, daß in der Zukunft Computerprogramme zu deduktiven Lösungen chemischer Probleme in Kombination mit einer geeignet angelegten Dokumentation den forschenden Chemiker bei der routinemäßigen Informations- und Planungsarbeit entlasten und ihn auf Möglichkeiten aufmerksam machen können, die er zunächst nicht vorausgesehen hätte. Dadurch würde er keineswegs ersetzt werden; mit dem Computer wird ihm jedoch ein Werkzeug in die Hand gegeben, mit dessen Hilfe er seine Aufgaben wirkungsvoller erfüllen kann.

*Unsere Arbeiten wurden ermöglicht durch die dankenswerte finanzielle Förderung durch die Stiftung Volkswagenwerk e. V.*

*(Entwicklung und Erprobung mathematischer Modelle, insbesondere für die Stereochemie), die Deutsche Forschungsgemeinschaft (Syntheseplanung), die Europäische Gemeinschaft (Dokumentation und Umweltschutz) und das Bundesministerium für Forschung und Technologie, Projektträger: Gesellschaft für Information und Dokumentation (Klassifizierung von Reaktionen). Den Professoren J. Dugundji, Los Angeles, und R. E. Burkard, Köln, danken wir für nützliche Diskussionen und Anregungen. Rechenzeit wurde uns freundlicherweise vom Institut für Plasmaphysik (Garching) und vom Leibniz-Rechenzentrum (München) zur Verfügung gestellt.*

Eingegangen am 27. Dezember 1977,  
ergänzt am 9. November 1978 [A 256]

- [1] J. Lederberg, G. L. Sutherland, B. G. Buchanan, E. A. Feigenbaum, A. V. Robertson, A. M. Duffield, C. Djerassi, J. Am. Chem. Soc. 91, 2973 (1969).
- [2] a) I. Ugi, G. Kaufhold (unveröffentlicht 1966–67), beschrieben in [5] und I. Ugi, Rec. Chem. Progr. 30, 289 (1968); b) E. J. Corey, W. T. Wipke, Science 166, 178 (1969); c) E. J. Corey, W. T. Wipke, R. D. Cramer, W. J. Howe, J. Am. Chem. Soc. 94, 421, 431 (1972); d) H. Gelernter, N. S. Sridharan, A. J. Hart, S. C. Yen, F. W. Fowler, H. J. Shue, Top. Curr. Chem. 41, 113 (1973); e) H. Gelernter et al., Science 197, 1041 (1977).
- [3] a) J. Dugundji, I. Ugi, Top. Curr. Chem. 39, 19 (1973); b) I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, Informal Commun. Math. Chem., im Druck.
- [4] J. Blair, J. Gasteiger, C. Gillespie, P. D. Gillespie, I. Ugi, Tetrahedron 30, 1845 (1974).
- [5] I. Ugi, Intra-Sci. Chem. Rep. 5, 229 (1971).
- [6] I. Ugi, P. D. Gillespie, Angew. Chem. 83, 982, 990 (1971); Angew. Chem. Int. Ed. Engl. 10, 914, 915 (1971).
- [7] I. Ugi, P. Gillespie, C. Gillespie, Ann. N. Y. Acad. Sci. 34, 416 (1972).
- [8] J. Blair, J. Gasteiger, C. Gillespie, P. Gillespie, I. Ugi in W. T. Wipke, S. Heller, R. Feldman, E. Hyde: Computer Representation and Manipulation of Chemical Information, Wiley, New York 1974, S. 129.
- [9] I. Ugi, IBM-Nachr. 24, 180 (1974).
- [10] I. Ugi, J. Gasteiger, J. Brandt, J. Brunnert, W. Schubert, IBM-Nachr. 24, 185 (1974).
- [11] J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, I. Ugi in E. V. Ludeña, N. H. Sabelli, A. C. Wahl: Computers in Chemical Education and Research, Plenum, New York 1977.
- [12] J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, I. Ugi: Computer-Assisted Organic Synthesis, ACS Symposium Ser. 61.
- [13] J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert, P. Lemmen, I. Ugi, Pure Appl. Chem. 50, 1303 (1978).
- [14] L. Spialter, J. Chem. Doc. 4, 261, 269 (1964).
- [15] E. Meyer, Angew. Chem. 82, 605 (1970); Angew. Chem. Int. Ed. Engl. 9, 583 (1970).
- [16] a) Y. Yoneda: Saikin no Kagakukogaku (Progress of Chemical Engineering), Maruzen, Tokio 1970, S. 145ff.; b) Y. Yoneda: Kemoguramu 1 – Keisanki Kogyo Kagaku (Chemogram 1 – Computer Industrial Chemistry), Band 1, Maruzen, Tokio 1972, S. 309ff.
- [17] a) W. Schubert, I. Ugi, J. Am. Chem. Soc. 100, 37 (1978); b) C. Jochum, J. Gasteiger, J. Chem. Inf. Comput. Sci. 17, 113 (1977).
- [18] J. Friedrich, I. Ugi, Informal Commun. Math. Chem., im Druck.
- [19] J. C. J. Bart, E. Garagnani, Z. Naturforsch. B31, 1646 (1976); B32, 455, 465, 678 (1977).
- [20] a) J. Gasteiger, C. Jochum, Top. Curr. Chem. 74, 93 (1978); b) J. Gasteiger, C. Jochum, M. Marsili, J. Thoma, Informal Commun. Math. Chem., im Druck.
- [21] a) W. Schubert, I. Ugi, Chimia, im Druck; b) W. Schubert, Informal Commun. Math. Chem., im Druck.
- [22] A. Weise, Z. Chem. 15, 33 (1975).
- [23] J. Gasteiger, Comput. Chem. 2, 185 (1978).
- [24] J. Gasteiger, M. Marsili, Tetrahedron Lett. 1978, 3184.
- [25] C. Jochum, J. Gasteiger, I. Ugi, noch unveröffentlicht.
- [26] a) K. H. Stratmann, Diplomarbeit, Universität Köln 1976; b) R. E. Burkard, K. H. Stratmann, Naval Res. Logistics Quart., im Druck.
- [27] R. E. Burkard: Methoden der Ganzzahligen Optimierung, Springer, Wien 1972.
- [28] I. Ugi, Gießener Universitätsbl. 11, 68 (1978).