

Plenary Lecture

EIN MATHEMATISCHES MODELL DER KONSTITUTIONELLEN CHEMIE
UND DARAUf BERUHENDE COMPUTERPROGRAMME

I. Ugi (Vortragender), J. Bauer, J. Brandt, J. Friedrich,
J. Gasteiger, C. Jochum, W. Schubert

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

und

J. Dugundji

Department of Mathematics, University of Southern
California, Los Angeles, 9007, USA

1. Artificial Intelligence Methoden und Chemie

Im letzten Jahrzehnt sind auf vielen Gebieten Versuche unternommen worden, menschliche Intelligenzleistungen durch Computer zu simulieren. Derlei Bemühungen werden mit dem vielverheißenden Terminus artificial intelligence¹ assoziiert. Die ersten Computerprogramme dieser Art wurden für das Schachspiel, zum Beweisen einfacher mathematischer Theoreme und für die Planung vielstufiger chemischer Synthesen entwickelt. Es gibt inzwischen vielerlei Programme, welche auf artificial intelligence Methoden beruhen. Ihre Bedeutung liegt zunächst in erster Linie darin, daß durch sie die grundsätzlichen Grenzen und Möglichkeiten der Anwendung von Computern aufgezeigt worden sind. Aus heutiger

Sicht ist die Chemie das einzige Gebiet, auf dem artificial intelligence Methoden praktische Bedeutung erlangen werden und einen wesentlichen Fortschritt ergeben.

Computer eignen sich vorzüglich zum Speichern, Abrufen und Manipulieren großer Datenmengen, wobei auch Entscheidungsprozesse mit einbezogen werden können. Für die Nachahmung von Assoziationsvorgängen eignen sich die heutigen Computer relativ wenig. Die direkte Simulation menschlicher Kreativität durch Computer ist praktisch unmöglich. Bei der Entwicklung von Computerprogrammen müssen diese Stärken und Schwächen berücksichtigt werden. Dies gilt auch für Programme, die auf artificial intelligence Methoden beruhen.

Die Entwicklung von Computerprogrammen zur Lösung chemischer Probleme begann mit der computerunterstützten Syntheseplanung. Die ersten Syntheseplanungs-Programme² basieren primär auf Retrieval und Manipulation gespeicherter empirischer chemischer Daten. Aufgrund der Strukturmerkmale eines Syntheseziels werden aus einer Reaktionsbibliothek solche Reaktionen ausgesucht, welche sich zur Herstellung des Ziels eignen. Durch ihre "Retroreaktionen" erhält man aus dem Syntheseziel dessen Vorstufen, die wiederum bei analoger Behandlung zu ihren Vorstufen führen, etc., bis verfügbare Ausgangsstoffe erreicht werden. Solchen Syntheseplanungsprogrammen liegt die Absicht zugrunde, das Denken des Synthetikers nachzuahmen. Dies gelingt nur unvollkommen, denn ein guter Chemiker ist nicht nur fähig, retrosynthetisch zu denken, und er beschränkt sich nicht nur auf die Erzeugung von Synthesewegen, welche sequentiellen Kombinationen bekannter Reaktionen entsprechen. Er ist dem Computer, selbst wenn er sich auf retrosynthetisches Vorgehen beschränkt, durch seine Assoziationsfähigkeit und seine Kreativität überlegen, und er ist dem Computer nur unterlegen hinsichtlich der "speicherbaren" Datenmenge und die Begrenztheit der Anzahl von Operationen, die durchführbar sind. Beim Lösen von Problemen, wie der Syntheseplanung, werden die besonderen Fähigkeiten des Computers nicht optimal genützt,

wenn man versucht Programme zu entwickeln, welche menschliche Denkweisen simulieren sollen. Es ist vorteilhafter, das Ergebnis menschlichen Denkens durch Computer zu erzeugen, ohne die menschliche Vorgehensweise direkt nachahmen zu wollen und statt dessen zur Lösung von Problemen Methoden zu wählen, welche dem Computer angemessen sind.

Die fehlende Kreativität des Computers läßt sich in vielen Fällen dadurch ausgleichen, daß man versucht, für ein gegebenes Problem die Gesamtheit aller denkbaren Lösungen zu berücksichtigen. Diejenigen Lösungen sind nämlich auch dabei, welche der Mensch durch seine Kreativität findet, ohne daß er alle erdenklichen in Betracht ziehen muß.

Nur wenn eine umfassende computerorientierte Theorie der Chemie verfügbar ist, welche die erforderlichen Voraussagen ermöglicht, ist es möglich, für bestimmte chemische Probleme alle chemisch sinnvoll erscheinenden Lösungsvorschläge mittels eines Computerprogramms zu erzeugen bzw. zu berücksichtigen.

2. Ein mathematisches Modell der konstitutionellen Chemie

J. Dugundji und I. Ugi entwickelten 1972/73 eine mathematische Theorie der konstitutionellen Chemie³. Diese Theorie genügt den vorangehend diskutierten Anforderungen und dient nun als Grundlage etlicher Computerprogramme zur deduktiven Lösung chemischer Probleme⁴. Das Wort deduktiv besagt hier, daß die individuellen Probleme in erster Linie anhand allgemeiner Prinzipien gelöst werden.

Diese Theorie entspricht einem mathematischen Modell der logischen Struktur der konstitutionellen Chemie. Sie ergibt sich aus folgenden Aussagen, die bei der Entwicklung der Theorie im Sinne von Axiomen verwendet werden:

- 1) Moleküle und Ensembles von Molekülen (EM) bestehen aus Atomrumpfen (Atomkerne nebst den Elektronen der inneren Schalen), welche durch Valenzelektronen (Elektronen der äußeren Schalen) zusammengehalten werden.
- 2) Die Valenzelektronen können in chemisch stabilen EM als freie Valenzelektronen einzelnen Atomrumpfen zugeordnet sein, oder aber mehreren Atomrumpfen gleichzeitig angehören und bindend wirken. Ein Paar Valenzelektronen, das ein Paar Atomrumpfe verbindet, stellt eine kovalente Bindung der formalen Ordnung eins dar.
- 3) Eine chemische Reaktion ist die Umwandlung eines EM in ein isomeres EM. Sie verläuft unter Erhaltung der Atomrumpfe und der Gesamtzahl der Valenzelektronen.

Diese Invarianz legt nahe, die Mengen aller molekularen Systeme zu betrachten, welche jeweils die gleichen Atomrumpfe und die gleiche Anzahl von Valenzelektronen enthalten. Die Familien isomerer EM (FIEM) sind solche Mengen.

Es sei $\underline{A} = \{A_1, \dots, A_n\}$ ein Sortiment von Atomen. Die Bruttoformel von \underline{A} gibt die Art und Anzahl der Atome in A an; sie ist gleichzeitig die Bruttoformel aller isomeren EM, welche aus \underline{A} erzeugt werden können. Somit ist ein $EM(\underline{A})$ ein Molekül oder ein EM, welches jedes Atom aus \underline{A} genau einmal enthält. Die Konstitutionsformel eines $EM(\underline{A})$ ist die Vereinigung der Konstitutionsformeln aller Moleküle, die zum betrachteten $EM(\underline{A})$ gehören. Da die $FIEM(\underline{A})$ als die Familie aller $EM(\underline{A})$ die ganze Chemie des Atomsortiments \underline{A} darstellt, und die Atomsortimente beliebig wählbar sind, umfaßt eine Theorie der $FIEM(\underline{A})$ gleichzeitig die gesamte Chemie.

Bei festgelegter Bezifferung der Atomrumpfe kann die Konstitution der $EM(A)$ durch Bindungs- und Elektronen-Matrizen (BE-Matrizen) dargestellt werden. Die BE-Matrix $B = (b_{ij})$ eines $EM(\underline{A})$ mit n Atomen ist eine

symmetrische $n \times n$ Matrix. Ihre Zeilen und Spalten sind den Atomrümpfen zugeordnet, und ihre Eintragungen geben die Verteilung der Valenzelektronen an. Die Eintragung b_{ij} ($i \neq j$) in der i -ten Spalte und j -ten Zeile ist die formale Ordnung der covalenten Bindung zwischen den Atomrümpfen A_i und A_j . Es gilt $b_{ij} = b_{ji}$, da die covalente Bindung als Relation symmetrisch ist. Die i -te Diagonaleintragung b_{ii} ist die Anzahl der Valenzelektronen, die zum i -ten Atomrumpf A_i gehört.

Daraus folgt unmittelbar, daß die i -te Zeilen/Spaltensumme

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

angibt, wie viele Valenzelektronen die Formalladung an A_i bestimmen, und die Kreuzsumme

$$\hat{s}_i = 2s_i - b_{ii} \quad (2)$$

ist die Gesamtzahl von Valenzelektronen in den Orbitalen von A_i .

Die Summe über alle Eintragungen

$$S = \sum_{i,j=1}^n b_{ij} \quad (3)$$

entspricht der Gesamtanzahl von Valenzelektronen.

In existenzfähigen EM kommen für jegliche A_i , je nach chemischem Element, nur bestimmte Kombinationen von s_i , \hat{s}_i und Bindungsordnungen b_{ij} vor. Somit ist es leicht festzustellen, ob eine gegebene BE-Matrix ein existenzfähiges EM(\underline{A}) darstellt.

Geht eine BE-Matrix B' aus einer gegebenen B durch eine Zeilen-/Spalten-Permutation gemäß

$$B' = PBP^t \quad (4)$$

hervor, so repräsentieren B' und B das gleiche $EM(A)$. Die Permutation der Zeilen und Spalten von B , welche die Multiplikation mit der Permutations-Matrix P und ihrer Transponierten bewirkt, entspricht einer Permutation der Bezifferung der Atomrümpfe im $EM(\underline{A})$.

Wir nennen Atome innerhalb von $EM(\underline{A})$ konstitutionell gleichwertig, wenn sie dem gleichen chemischen Element angehören und bezüglich aller durch kovalente Bindungen festgelegten Nachbarschaftsverhältnisse übereinstimmen.^{8,9} Den konstitutionell äquivalenten Atomen entspricht eine konstitutionelle Symmetrie der zugehörigen Moleküle oder $EM(A)$. Die Atome eines $EM(A)$ können auf $n!$ verschiedene Weisen beziffert und den Zeilen/Spalten einer BE-Matrix zugeordnet werden. Enthält das EM keine konstitutionell gleichwertigen Atome (s.u.), so gibt es $n!$ verschiedene BE-Matrizen, welche dem $EM(\underline{A})$ als äquivalente Darstellungen zugeordnet werden können. Liegt konstitutionelle Symmetrie vor, so ergeben die $n!$ Zeilen-/Spalten-Permutationen weniger als $n!$ verschiedene BE-Matrizen, denn gegenüber bestimmten Permutationen von Zeilen-/Spalten sind die BE-Matrizen konstitutionell symmetrischer $EM(A)$ invariant.

Bei der Lösung chemischer Probleme mittels Computerprogrammen auf der Grundlage unserer Theorie ist es wichtig, jedem $EM(A)$ eine ganz bestimmte BE-Matrix zuzuordnen. Dem entspricht eine kanonische Bezifferung von Atomen in Molekülen und EM unter Berücksichtigung konstitutioneller Symmetrien.

Für Dokumentationszwecke wurde bereits 1965 der u.a. von CAS verwendete mehrstufige Morgan-Algorithmus⁵ zur kanonischen Bezifferung der Atome in Molekülen entwickelt, allerdings ohne direkte Erkennung und Berücksichtigung konstitutioneller Symmetrien.

An unserem Institut wurden seit 1972 die Atombezifferungs-Algorithmen ASI⁶, NOON⁷ und CANON⁸⁻¹⁰ entwickelt. Unter diesen zeichnen sich ASI und CANON dadurch aus, daß die zugehörigen

Computerprogramme für die Atome innerhalb einer chemischen Konstitution unter direkter Berücksichtigung konstitutioneller Symmetrien eine bis auf die Umbezeichnung konstitutionell äquivalenter Atome eindeutige Bezeichnung liefern. Die wesentliche Eindeutigkeit der Atombezeichnung durch CANON ist kürzlich bewiesen worden¹⁰. CANON erzeugt gleichzeitig für die stereochemische Nomenklatur eine sequentielle Ordnung der Liganden, welche von wenigen Ausnahmen abgesehen, der Ligandensequenz nach CIP¹¹ entspricht⁹. Diese Ausnahmen rühren daher, daß CANON Koordinationszahlen, CIP hingegen Bindungsordnungen als primäre Größe für die Anzahl der Nachbarn eines Atoms verwendet.

Eine chemische Reaktion



ist darstellbar durch die Gleichung

$$B + R = E \quad (5)$$

Darin sind $B = (b_{ij})$ und $E = (e_{ij})$ die BE-Matrizen der $EM(\underline{A})$ zu Beginn und am Ende der Reaktion, und $R = (r_{ij})$ ist eine symmetrische $n \times n$ sogenannte Reaktions-Matrix (R-Matrix), welche die Verschiebung der Valenzelektronen im Zuge der Reaktion repräsentiert. Die Eintragungen $r_{ij} (i \neq j)$ bzw. r_{ii} von R zeigen an, wie sich die Bindungsordnung zwischen A_i und A_j und die Anzahl der freien Valenzelektronen an A_i ändert. Die Matrix R entspricht einem Schema von gebogenen Pfeilen ("Elektronenschieber-Pfeilen"), durch welches die Chemiker den Elektronenfluß bei der Reaktion wiederzugeben pflegen¹².

Beispielsweise könnte man die Reaktion



beschreiben durch

$$\begin{array}{ccccccc}
 & \text{H} & \text{O} & \text{H} & \text{Cl} & & \\
 \text{H} & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 6 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 6 \end{pmatrix} & + & \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -2 & +1 & 0 \\ 0 & +1 & 0 & -1 \\ 0 & 0 & -1 & +2 \end{pmatrix} & = & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix} & \begin{array}{l} \text{H} \\ \text{O} \\ \text{H} \\ \text{Cl} \end{array} \\
 \text{O} & & & & & & \\
 \text{H} & & & & & & \\
 \text{Cl} & & & & & & \\
 \text{B} & & + & \text{R} & = & \text{E} & (6)
 \end{array}$$

Die Gesamtzahl der Valenzelektronen ändert sich bei chemischen Reaktionen nicht.

Es gilt daher:

$$\sum_{i,j} e_{ij} = \sum_{i,j} (b_{ij} + r_{ij}) = \sum_{i,j} b_{ij} \quad (7)$$

Daraus folgt:

$$\sum_{i,j} r_{ij} = \sum_{i,j} (e_{ij} - b_{ij}) = 0 \quad (8)$$

Die R-Matrizen bilden eine freie abelsche Gruppe $R(n)$ vom Rang $\frac{(n+2)(n-1)}{2}$ mit der Basis

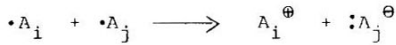
$$\{v_{ij} \mid 1 \leq i < j \leq n\} \cup \{u_{in} \mid i = 1, \dots, n-1\}, \quad (9)$$

$$\text{mit } u_{ij} = p_{ii} - p_{jj} \quad (10)$$

$$\text{und } v_{ij} = p_{ij} + p_{ji} - p_{ii} - p_{jj} \quad (11)$$

Darin ist P_{ij} eine $n \times n$ Matrix mit der Eintragung 1 in der (i, j) Position und Nullen an allen anderen Stellen.

Das Basiselement U_{ij} entspricht der elementaren Redoxreaktion



und das Basiselement V_{ij} repräsentiert die elementare Homolysis



während $-V_{ij}$ den umgekehrten Vorgang, eine Homolyse, wiedergibt. Mit unserer Basis hat eine beliebige R-Matrix $R=(r_{ij})$ die eindeutige Darstellung

$$R = \sum_{i < j} r_{ij} V_{ij} + \sum_{i=1}^{n-1} r_i U_{in} \quad (12)$$

$$\text{mit } r_i = \sum_j r_{ij} \quad (13)$$

Die R-Matrizen der spingepaarten Chemie gehören zur Gruppe $CR(n) \subset R(n)$ vom Rang $\frac{(n+2)(n-1)}{2}$ mit der Basis

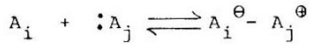
$$\{V_{ij} + U_{ij} \mid 1 \leq i < j \leq n\} \cup \{2U_{in} \mid i = 1, \dots, n-1\} \quad (14)$$

Die R-Matrizen $R \in CR(n)$ sind eindeutig darstellbar als

$$R = \sum_{i < j} r_{ij} (V_{ij} + U_{ij}) + \sum_{i=1}^n \lambda_i 2U_{in} \quad (15)$$

$$\text{mit } \lambda_i = \sum_{k \neq i} r_{ki} + \frac{1}{2} r_{ii} \quad (16)$$

Die Basiselemente ($U_{ij} + V_{ij}$) repräsentieren die heterooptischen und heterolytischen Elementarprozesse:



Den Basiselementen ($2U_{in}$) entsprechen zweielektronige elementare Redoxprozesse. Daraus geht auch hervor, daß chemische Reaktionen aus Redox- und Homoapsis-Anteilen bestehen, und daß diese die sog. Basis- oder Elementar-Reaktionen der Chemie sind.

Die Gleichung $B + R = E$ stellt nur dann eine chemische Reaktion dar, wenn B und E definitionsgemäße BE-Matrizen sind, d.h. wenn keine von ihnen negative Eintragungen enthält.

Daraus folgt, daß R mathematisch nur zu B paßt, wenn keine der negativen Eintragungen $r_{ij} < 0$ von R die zugehörigen Eintragungen b_{ij} von B betragslich übersteigt.

Aufgrund dieser mathematischen Anpaßbedingungen und der durch

$$r_{ij} = r_{ji} \quad \text{und} \quad \sum_{i,j} r_{ij} = 0$$

festgelegten Struktur von R-Matrizen ist es leicht, alle zu einer gegebenen BE-Matrix B passenden R-Matrizen zu konstruieren. Anhand der valenzchemischen Existenzbedingungen können diejenigen BE-Matrizen $E = B + R$ eliminiert werden, welche keine stabilen EM(A) darstellen. Somit ist es möglich, von B bzw. E aus alle chemisch relevanten Lösungen der Grundgleichung $B + R = E$ des Modells zu finden.

Dies ermöglicht bereits die deduktive Lösung vieler chemischer Probleme mittels Computern¹³. Als deduktives Vorgehen bezeichnen wir die Lösung individueller Probleme anhand allgemeingültiger Prinzipien.

3. Die deduktive Lösung chemischer Probleme

Faßt man ein Syntheseziel und seine Coprodukte als das Ergebnis $EM_E(A)$ chemischer Reaktionen auf, so erhält man aus E über die mathematisch und valenzchemisch zu E passenden R-Matrizen -R die $B = E - R$ als Repräsentanten der denkbaren Vorstufen von $EM_E(A)$. Ein solches Vorgehen liegt dem Syntheseplanungsprogramm CICLOPS^{6,13} und dem daraus entwickelten EROS^{4,14} (Erzeugung von Reaktionen für Organische Synthesen) zugrunde.

Auch das kürzlich erstellte Programm ASSOR^{4,9,15} (Allgemeines System zur Simulation Organischer Reaktionen) beruht darauf, daß chemisch relevante Lösungen B der Gleichung $B + R = E$ von E aus gesucht werden, wobei die chemischen Reaktionen aus ihren Elementarschritten sukzessive aufgebaut werden, indem BE-Matrizen durch mathematisch und valenzchemisch passende Basiselemente von $CR(n)$ transformiert werden. Auf diese Weise gelingt es bei der deduktiven Lösung chemischer Probleme, den mechanistischen Aspekt zu berücksichtigen.

Die Programme EROS und ASSOR sind nicht nur für die "retrosynthetische" Planung von Synthesen verwendbar. Mit ihrer Hilfe kann man auch die denkbaren "Vorwärtsreaktionen" eines $EM(A)$ finden.

Wipke et al.¹⁶ erklärten kürzlich die elementaren Heteroapsis-Heterolyse zu "ab initio transforms", um das Syntheseplanungsprogramm SECS "kreativ" zu machen. Er versäumte dabei allerdings, die transforms elementarer Redoxreaktionen hinzuzufügen, so daß dem "ab initio"-Teil von SECS ein Satz von "ab initio transforms" zugrunde liegt, der mathematisch unvollständig ist. Um derlei "ab initio transforms" als R-Matrizen-Äquivalenz überhaupt sinnvoll benützen zu können, ist es zudem notwendig, die Syntheseziele bzw. ihre "subgoals" durch Coprodukte zu isomerisierbaren EM zu ergänzen, was bei konsequenter Durchführung ein ASSOR im wesentlichen entsprechendes Programm ergäbe.

Sind der Beginn $EM_B(A)$ und das Ende $EM_E(A)$ einer Reaktion oder einer Folge von Reaktionen bekannt, so liegt es nahe, die chemische Gesamtveränderung $EM_B(\tilde{A}) \rightarrow EM_E(\tilde{A})$ durch eine R-Matrix

$$R = E - B$$

darzustellen, um durch die Zerlegung von R in die Komponenten R_k

$$R = R_1 + R_2 + \dots + R_r \quad (18)$$

und unterschiedliche sequentielle Anwendung der Komponenten von R_k Netzwerke von Reaktionswegen zu erhalten, z.B. gemäß

$$\begin{aligned} B + R_1 &= B_1 \\ B_1 + R_2 &= B_2 \\ &\vdots \\ &\vdots \\ &\vdots \\ B + R_r &= E \end{aligned} \quad (19)$$

Ein solches Vorgehen wäre bei der Ermittlung von Reaktionsmechanismen nützlich, oder bei einer Syntheseplanung bei gleichzeitiger Kenntnis von Ausgangsstoffen und Synthesziel nebst Coprodukten.

Es setzt jedoch voraus, daß man weiß, wie die einzelnen Atome im $EM_B(\tilde{A})$ den Atomen im $EM_E(\tilde{A})$ zuzuordnen sind. Da eine solche Korrelation meist nicht von vornherein möglich ist, entwickelten wir jüngst ein Computerprogramm (MCD)¹⁷, welches die Korrelation der Atome in $EM_B(\tilde{A})$ und $EM_E(\tilde{A})$ durchzuführen gestattet. Es beruht auf dem heuristischen Postulat der minimalen chemischen Distanz.

Die chemische Distanz³ zwischen isomeren EM mit bezifferten Atomen $EM_B(\underline{A})$ und $EM_E(\underline{A})$ ist laut Definition

$$D(E,B) = \sum_{i,j} |e_{ij} - b_{ij}| = \sum_{i,j} |r_{ij}| \quad (20)$$

Die chemische Distanz entspricht der L_1 -Metrik auf den BE-Matrizen einer FIEM(\underline{A}). Sie ist das Doppelte der Anzahl von Valenzelektronen, die bei gegebener Bezifferung der Atome in $EM_B(\underline{A})$ und $EM_E(\underline{A})$ im Zug der Reaktion $EM_B(\underline{A}) \rightarrow EM_E(\underline{A})$ verschoben wird.

Die chemische Metrik ist geometrisch interpretierbar. Eine beliebige $n \times n$ Matrix B entspricht einem Punkt im n^2 dimensionalen euklidischen Raum \mathbb{R}^{n^2} , oder einem Vektor, falls \mathbb{R}^{n^2} als Vektorraum angesehen wird.

Die Abbildung $B \rightarrow (b_{11}, \dots, b_{1n}, \dots, b_{nn})$ ist eine Einbettung aller $n \times n$ Matrizen im \mathbb{R}^{n^2} . Der euklidische Abstand zwischen den B und E entsprechenden Punkten ist

$$d(B,E) = \sqrt{\sum_{i,j} (e_{ij} - b_{ij})^2}$$

Da $d(B,E) \leq D(B,E) \leq n^2 d(B,E)$ gilt, ist die der chemischen Distanz entsprechende L_1 -Metrik der euklidischen L_2 -Metrik im \mathbb{R}^{n^2} wesentlich äquivalent.

Geht man von einer festgelegten Bezifferung der Atome in $EM_B(\underline{A})$ und $EM_E(\underline{A})$ aus, so ist die chemische Distanz zwischen $EM_B(\underline{A})$ und $EM_E(\underline{A})$, bei Permutation der Indices der Atome in $EM_B(\underline{A})$ eine Funktion F(P) der Permutationsmatrix $P = (p_{ij})$, und ihre Transponierte P^t , durch welche die Zeilen/Spalten von B permutiert werden.

$$D(E, PBP^t) = \sum_{i,j=1}^n |e_{ij} - b_{P(i), P(j)}| = \sum_{i,j=1}^n e_{ij} - \sum_{k,s=1}^n p_{ik} b_{ks} p_{js} = F(P)$$

Unter den Permutations-Matrizen P gibt es mindestens eine, für welche die Funktion $F(P)$ einen Minimalwert annimmt. Die minimale chemische Distanz

$$D(E, P \cdot B \cdot P^t) = F(P) = \min$$

ist ein Maß der chemischen Ähnlichkeit von $EM_B(A)$ und $EM_E(A)$.

Aus der topologischen Äquivalenz von $d(E,B)$ und $D(E,B)$ folgt u.a., daß $F(P) = \min$ und die zugehörigen wechselseitige Zuordnung der Atome in $EM_B(A)$ und $EM_E(A)$ auch durch Maximierung des Skalarproduktes

$$G(P) = E \cdot (P \cdot B \cdot P^t) = \sum_{i,j} e_{ij} \cdot b_{P(i)P(j)} = \max$$

ermittelt werden kann¹⁸. Dies ist von Bedeutung, weil $F(P)$ wegen fehlender Verfahren zur gezielten Lösung solcher Probleme allenfalls durch Ausprobieren von bis zu $n!$ Permutationsmatrizen P minimiert werden kann, während es im Operations Research bereits Algorithmen zur Lösung quadratischer Zuordnungsprobleme gibt, welche sich in modifizierter Form für die Maximierung von $G(P)$ vorzüglich eignen. Mittels eines kürzlich erstellten Computerprogramms (MCD), welches auf einer Modifikation der Ungarischen Methode¹⁸ unter Benützung der Burkhardtschen Störmethode beruht, lassen sich die Bezifferungen der Atome in Paaren relativ großer EM dem Minimum ihrer chemischen Distanz entsprechend korrelieren.

In Übereinstimmung mit dem Rice-Teller-Prinzip²⁰ postulieren wir, daß die chemische Umwandlung von $EM_B(A)$ in ein isomeres $EM_E(A)$ vorzugsweise unter Verschiebung einer minimalen Anzahl von Valenzelektronen stattfindet (s.a. Lit. 3).

Dieses heuristische Postulat der minimalen chemischen Distanz (PMCD) läßt sich bei der Lösung vieler chemischer Probleme als heuristische Regel einsetzen¹⁷. Allerdings erfordert die

Anwendung des PMCD ein Computerprogramm wie MCD.

Die paarweise Korrelation der Atome in den Ausgangsstoffen und Produkten chemischer Reaktionen hängt von deren Mechanismus ab. Umgekehrt sollte es auch möglich sein, auf die denkbaren Reaktionsmechanismen zu schließen, wenn bekannt ist, wie die einzelnen Atome in der EM zu Beginn und am Ende der Reaktion einander paarweise zugeordnet sind. Die Korrelation der Atome in $EM_B(A)$ und $EM_E(A)$ durch das PMCD ergibt die BE-Matrizen B und E, und die R-Matrix $R = E - B$, welche die Gesamtverschiebung der Valenzelektronen im Verlaufe der Reaktion wiedergibt. Aus dem Basiselement von R kann mittels eines Programms vom Typ ASSOR das Netzwerk der denkbaren mechanistischen Wege erzeugt werden, welche von $EM_B(A)$ nach $EM_E(A)$ führen.

Kennt man den Beginn $EM_B(A)$ und das Ende $EM_E(A)$ von Synthesen, so kann man durch Minimieren der chemischen Distanz die Atome in $EM_B(A)$ und $EM_E(A)$ korrelieren und feststellen, welche Bindungen bei der Umwandlung der Ausgangsstoffe $EM_B(A)$ in das Synthesziel und seine Coprodukte $EM_E(A)$ mindestens gebrochen und neu gemacht werden müssen. Dies entspricht einem vollständigen "look ahead" für die Syntheseplanung. Beispielsweise läßt sich mittels des MCD-Programms in 2.3 sec (CYBER 76) feststellen, daß die Strychninsynthese von Woodward et al.²¹ einem "kürzesten Weg" entspricht¹⁷.

Wir nehmen an, daß das PMCD die Entwicklung künftiger Syntheseplanungsprogramme nachhaltig beeinflussen wird. Wir beabsichtigen ein dreistufiges Syntheseplanungsprogramm zu erstellen, in dem das MCD-Programm eine wichtige Rolle spielen wird. Ein auf unserer Theorie beruhendes Programm zur Substruktur-Analyse befindet sich in einem fortgeschrittenen Entwicklungsstadium⁴. Es ist vorgesehen, mit dessen Hilfe aus einer Liste verfügbarer Ausgangsstoffe solche $EM_B(A)$ zusammenzustellen, welche wegen ihrer Substrukturen für die Synthese

gegebener Ziele geeignet erscheinen. Durch Korrelation der Atome in den Ausgangsstoffen und dem Synthesziel nebst seinen Coprodukten mittels des MCD Programms werden die zu brechenden und die zu machenden Bindungen ermittelt.

Davon ausgehend, wird das Netzwerk der denkbaren Synthesewege aus den Komponenten von $R = E - B$ durch eines unserer Syntheseplanungs-Programme⁴ erstellt.

Dieses Vorgehen ist dem retrosynthetischen allein schon dadurch überlegen, daß die kombinatorische Vielfalt der Möglichkeiten vermieden wird, die berücksichtigt werden müssen, wenn man vom Ziel der Synthese aus einen "Baum" von Synthesewegen erzeugt, ohne das jeweilige andere Ende der Synthesewege zu kennen.

Die Grundgleichung $B + R = E$ unserer Theorie läßt sich auch von R aus lösen. Wir entwickelten bereits ein Computerprogramm⁴, welches aus gegebenen R-Matrizen mathematisch und valenzchemisch dazu passende Paare von BE-Matrizen, (B,E), produziert. Dieses Programm soll der systematischen Erzeugung von Repräsentanten chemischer Reaktionen dienen und präzedenzlose Reaktionen zur experimentellen Entdeckung vorschlagen.

Wir sind sicher, daß einige der heute konzipierten oder schon in Entwicklung befindlichen Computerprogramme zur deduktiven Lösung chemischer Probleme bereits in absehbarer Zukunft zum Standard-Rüstzeug des Chemikers gehören werden. Solche Programme werden die Leistungsfähigkeit des Chemikers erhöhen; sie werden ihn aber nie ersetzen.

Für die Förderung unserer Arbeiten danken wir der Deutschen Forschungsgemeinschaft, der Stiftung Volkswagenwerk e.V., den Europäischen Gemeinschaften, dem Bundesministerium für Forschung und Technologie und dem Fonds der Chemischen Industrie.

Literaturverzeichnis

- 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) I. Ugi und G. Kaufhold (unveröff. 1967 - 68); beschrieben in einem Sammelref.: I. Ugi, *Rec. Chem. Progr.*, 30, 289, (1968)
E.J. Corey und W.T. Wipke, *Science*, 166, 178 (1969)
E.J. Corey, W.T. Wipke, R.D. Cramer und W.J. Howe, *J. Amer. Chem. Soc.*, 94, 421, 431 (1972)
H. Gelernter, N.S. Sridharan, A.J. Hart, S.C. Yen, F.W. Fowler und H.J. Shue, *Top. Curr. Chem.*, 41, 113 (1973)
- 3) J. Dugundji, I. Ugi, *Top. Curr. Chem.*, 39, 19 (1973)
- 4) I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, W. Schubert; *Angew. Chem.* (im Druck)
- 5) H.L. Morgan, *J. Chem. Doc.* 5, 107 (1965)
- 6) J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie und I. Ugi in: W.T. Wipke, S. Heller, R. Feldman und E. Hyde "Computer Representation of Chemical Information", Wiley New York, 1974, S. 129
- 7) C. Jochum und J. Gasteiger, *J. Chem. Inf. Comp. Sci.*, 17, 119 (1977)
- 8) W. Schubert und I. Ugi, *J. Amer. Chem. Soc.*, 100, 37 (1978)
- 9) W. Schubert und I. Ugi, *Chimia* (in Vorbereitung)
- 10) J. Dugundji, W. Schubert und I. Ugi (in Vorbereitung)
- 11) R.S. Cahn, C. Ingold und V. Prelog, *Angew. Chem.* 78, 413 (1966); *Angew. Chem. Internat. Edit. (Engl.)* 5, 385 (1966)
- 12) S.u.a.: J. March, "Advanced Organic Chemistry, Reactions, Mechanisms and Structure"; McGraw-Hill, Kogakusha Ltd., London 1977
- 13) I. Ugi, *IBM Nachrichten*, 24, 180 (1974)
- 14) J. Gasteiger und C. Jochum, *Top. Curr. Chem.*, 74, 93 (1978)
- 15) W. Schubert, *Diss. Techn. Univ. München* 1978

- 16) W.T. Wipke, H. Braun, G. Smith, F. Choplin und W. Sieber
in: "Computer Assisted Organic Synthesis", Herausg.:
W.T. Wipke und W.J. Howe; A.C.S. Washington D.C. 1977,
S. 104
- 17) C. Jochum, J. Gasteiger und I. Ugi, Angew. Chem. (in Vorbe-
reitung), idem J. Amer. Chem. Soc. (in Vorbereitung)
- 18) C. Jochum, Diss. Techn. Univ. München 1978
- 19) K.H. Stratmann, "Numerische Untersuchungen über quadratische
Zuordnungsprobleme", Diplomarbeit, Köln 1976
R.E. Burkard und K.H. Stratmann, Naval Research Logistics
Quarterly, im Druck
R.E. Burkard, "Methoden der Ganzzahligen Optimierung",
Springer Verlag, Wien 1972
E.L. Lawler: "The Quadratic Assignment Problem", Man Sci. 9,
586 (1963)
G. Galhus: "Heuristische Verfahren zur Lösung allgemeiner
ganzzahliger linearer Optimierungsprobleme", Z. f. Operat.
Res. 20, 89 (1976)
J.F. Pierce, W.B. Crowston: "Tree-Search Algorithms for
Quadratic Assignment Problems", Nav. Res. Log. Quat. 18,
1 (1971)
R.E. Burkard, K.-H. Stratmann: "Numerische Untersuchungen
über quadratische Zuordnungsprobleme", Nav. Res. Log. Quat.
(im Druck)
R.E. Burkard: "A Perturbation Method for Solving Quadratic
Assignment and Bottleneck Problems", Mathematisches
Institut der Universität Köln, Report 76-2, 1976
- 20) F.O. Rice and E. Teller, J. Chem. Phys. 6, 489 (1938);
s.a.: J. Hine, Adv. Phys. Org. Chem. 15, 1, (1977)