

MOLGEN, ein Computeralgebra-System für die Konstruktion molekularer Graphen¹

Dem Andenken an O. E. Polansky gewidmet

R. Grund, A. Kerber, R. Laue

Lehrstuhl II für Mathematik der Univ. Bayreuth

Postfach 101251, 8580 Bayreuth, Deutschland

(received: August 1992)

1 Einführung

Die Existenz des Isomerieproblems der Chemie wurde bereits am Ende des achtzehnten Jahrhunderts behauptet und ein Vierteljahrhundert später bestätigt. Seine Lösung durch Einführung des graphentheoretischen Molekülmodells um die Mitte des neunzehnten Jahrhunderts weckte den Wunsch nach einer Konstruktion chemischer Bindungsgraphen und stimulierte die Entwicklung der Graphentheorie und der kombinatorischen Abzählungstheorie. Das Isomerieproblem hat also großen Einfluß auf die Entwicklung sowohl der Chemie als auch der Mathematik gehabt. Dennoch muß eingeräumt werden, daß diese Aufgabe bisher nur unzureichend gelöst wurde. Dies liegt einmal an der Komplexität der Fragestellung, zum andern aber auch daran, daß erst jetzt Rechenanlagen zur Verfügung stehen, die zu erschwinglichen Preisen solche äußerst umfangreichen und mathematisch sehr anspruchsvollen Rechnungen durchführen können.

Das Problem der Herstellung *aller* Bindungsgraphen zu vorgegebener Summenformel ist vor allem in dem großen amerikanischen Forschungsprojekt DENDRAL angepackt worden (Lederberg, Stanford, seit 1965). Die im Rahmen dieses Projekts entstandenen Programme lassen jedoch noch Wünsche offen, die unter anderem die Reichweite der Systeme betreffen. So ist zum Beispiel der CATALOG der Substrukturen, auf den bei der Konstruktion zurückgegriffen werden kann, viel zu klein. Andere, jüngere Ansätze

¹Das Projekt wird von der Deutschen Forschungsgemeinschaft unter Ke 201 gefördert

benutzen vorwiegend kombinatorische Methoden und laufen deshalb in die Falle der Komplexität, das heißt, ihre Verwendung ist noch wesentlich stärker limitiert als die des mehr algebraisch orientierten Ansatzes von DENDRAL.

Wir haben deshalb eine Eigenentwicklung begonnen mit dem Ziel besserer Portabilität und Redundanzfreiheit, unter Verwendung neuerer Methoden aus Mathematik und Informatik, besser standardisierter Programmiersprachen und eines wesentlich größeren zugrundeliegenden Katalogs von Teilstrukturen.

Hierbei hat uns die Weiterentwicklung der Anwendungen der Graphentheorie in der Chemie stimuliert, die insbesondere auch O.E. Polansky zu verdanken ist. Die von ihm 1975 im Max-Planck-Institut in Mülheim/Ruhr veranstaltete Tagung "Graph Theory in Chemistry" hat bahnbrechend gewirkt. Sie führte u.a. zur Gründung der Zeitschrift MATCH, die seitdem ein Forum für graphentheoretische Überlegungen in der Chemie ist. Der Verlust, der durch den Tod von O.E. Polansky eingetreten ist, trifft uns auch deshalb besonders schwer. Wir widmen diese Arbeit seinem Andenken.

Die finanzielle Unterstützung durch die Stiftung Volkswagenwerk im Rahmen des Projekts "Algebraische und kombinatorische Beschreibung von Molekülen" hat die theoretischen Vorbereitungen ermöglicht, die zu ersten wichtigen Programmen und Manuskripten führte. Danach erhielten wir dankenswerterweise weitere Hilfe durch die Deutsche Forschungsgemeinschaft im Rahmen des Projekts "Molekulare Strukturerkennung" und die Mitarbeit interessierter Studenten, denen wir ebenfalls zu Dank verpflichtet sind, insbesondere F. Bauer, W. Decker, F. Herrmann, R. Hager, D. Moser (der die erste Version von MOLGEN implementierte) und W. Weber.

2 Historische Entwicklung

Bereits 1797 hat Alexander von Humboldt behauptet (vgl. [20], [6]), daß es chemische Substanzen mit verschiedenen Eigenschaften aber derselben atomaren Zusammensetzung geben könne. In seinem Buch "Versuche über die gereizte Muskel- und Nervenfasern, nebst Vermutungen über den chemischen Prozeß des Lebens" schreibt er nämlich ([14], Band 2, Seite 128):

Drei Körper a, b und c können aus gleichen Quantitäten Sauerstoff, Wasserstoff, Kohlenstoff, Stickstoff und Metall zusammengesetzt und in ihrer Natur doch unendlich verschieden sein.

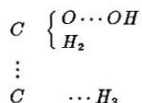
Nachdem dann Gay-Lussac zu Beginn des neunzehnten Jahrhunderts die entsprechenden Analysemethoden entwickelt hatte, wurde diese Vermutung schließlich 1824 durch fast gleichzeitige Veröffentlichungen einerseits von Gay-Lussac und Justus von Liebig, andererseits und unabhängig davon durch F. Wöhler bestätigt. (Alexander von Humboldt war übrigens mit Gay-Lussac eng befreundet und protegierte Justus von Liebig ([2]). Ob die oben erwähnte Behauptung Humboldts direkten Einfluß auf

die Untersuchungen von Gay-Lussac und Liebig gehabt hat, ist uns aber nicht bekannt.) Sie entdeckten tatsächlich Substanzen, die zwar dieselbe Summenformel, aber verschiedene Eigenschaften besitzen. Hier ein entsprechendes Zitat, eine Fußnote von Gay-Lussac zu einer Arbeit von Wöhler ([8]):

... comme ces deux acides sont très différents, il faudrait pour expliquer leur différence admettre entre leurs éléments un mode de combinaison différent.
C'est un objet qui appelle un nouveau examen. ...

Aber erst nachdem dann 1828 Wöhler bemerkte, daß aus Ammoniumcyanat Harnstoff entstehen kann, eine Substanz mit derselben Summenformel, aber anderen Eigenschaften, wurde dieses Phänomen in seiner allgemeinen Bedeutung erkannt, und Berzelius führte 1830 dafür den Begriff der *Isomerie* ein.

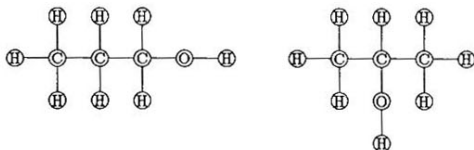
Zur Ermittlung der Ursache dieses Phänomens begann man, Moleküle zu skizzieren. Hier sind einige prominente Vorstellungen von Molekülen (für weitere Einzelheiten vgl. [3]) des Ethanol C_2H_5OH . Zunächst Coupers Version (daß irrtümlich zwei Sauerstoffatome aufgeführt werden liegt daran, daß man damals das Atomgewicht des Sauerstoffs noch nicht kannte):



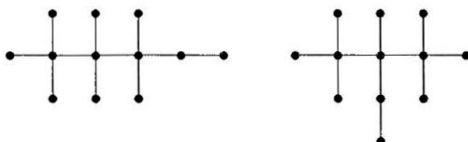
Hier sind Loschmidts und Kekulé's Skizzen für dieses Molekül:



Den Durchbruch zur Lösung brachte die Methode von Alexander Crum Brown, der die Couperschen *Punkte* bzw. die Loschmidtschen und Kekulé'schen *Berührungspunkte*, die atomare Bindungen andeuten, durch die wesentlich deutlicheren *Verbindungskanten* ersetzt. Crum Brown zeigte 1864 unter anderem, daß es *zwei* Möglichkeiten gibt, drei Kohlenstoffatome C mit der Wertigkeit 4 mit acht Wasserstoffatomen H der Wertigkeit 1 und einem Sauerstoffatom O der Wertigkeit 2 so zu kombinieren, daß ein Alkohol entsteht, d.h. eine Hydroxylgruppe OH vorhanden, also ein H -Atom direkter Nachbar des Sauerstoffatoms ist. Beide Möglichkeiten zeigt die folgende Abbildung:



Die zugehörigen Propanole haben z.B. verschiedene Siedepunkte: $97,1^\circ$ bzw. $82,4^\circ$. Läßt man die Bezeichnungen für die Atome dabei weg, dann entstehen aus diesen beiden Skizzen die folgenden *Graphen* (eine genaue Definition dieses Begriffes folgt später).



Crum Browns Lösung des Isomerieproblems lautet deshalb so:

Zu einer Summenformel gibt es unter Umständen mehrere Moleküle, die dieser genügen, aber verschiedene Eigenschaften haben, und verschiedene Bindungsgraphen besitzen.

Damit war die Ursache der Isomerie erkannt und die Graphentheorie geboren. Es blieb die Frage nach einer Bestimmung oder wenigstens einer Abzählung der verschiedenen Isomere zu vorgegebener chemischer Summenformel. Auch bei Mathematikern fiel das Problem auf fruchtbaren Boden, denn etwa zur gleichen Zeit beschäftigte sich A. Cayley mit sogenannten Bäumen, das sind zusammenhängende Graphen, die man leicht gewissen Kohlenwasserstoffen zuordnen kann, was schon die ersten Resultate zum Isomerieproblem lieferte. Auch J. Sylvester interessierte sich für diesen Themenkreis, er schrieb eine Arbeit mit dem Titel "Chemistry and Algebra" ([27]) in der zum erstenmal die Bezeichnung *Graph* auftaucht.

Es begann damit eine fruchtbare Periode, in der viele Isomere graphisch konstruiert oder wenigstens abgezählt wurden, neben der Graphentheorie war also aus dem Isomerieproblem heraus noch die *kombinatorische Abzählungstheorie* entstanden. In einer Arbeit von 1929 haben dann Lunn und Senior (ein Chemiker und ein Mathematiker) gezeigt, daß die *Algebra* mit dem Begriff der Operation von Gruppen auf Mengen und dessen Anwendungen ein wichtiges Werkzeug für die Lösung solcher Abzählprobleme bereithält ([22]).

G. Pólya hat diesen Ansatz dann in seiner berühmten Arbeit [23] systematisch und zudem didaktisch äußerst geschickt systematisiert. Ein erstes Buch über Abzählmethoden in der Graphentheorie haben F. Harary und E. Palmer 1973 veröffentlicht ([13]).

Substanzen derselben Summenformel (z.B. C_3H_8O), deren Atome im Molekül aber verschieden verknüpft sind, heißen *Bindungsisomere*. (Zu einem Bindungsisomer können darüberhinaus noch mehrere *Stereoisomere* gehören, die zwar denselben Bindungszusammenhang besitzen, sich jedoch in der räumlichen Anordnung der Atome unterscheiden. Auf dieses Problem können wir hier aber nicht eingehen, denn dies ist der Gegenstand weiterer Untersuchungen.)

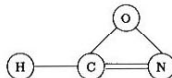
Wichtig ist hier die Bemerkung, daß die Bindungsisomere für eine Summenformel in verschiedene Substanzklassen fallen können. So gibt es z.B. für C_3H_8O noch einen dritten Bindungsgraphen, ihn zeigt die nächste Abbildung. Er repräsentiert jedoch keinen Alkohol, da die Hydroxylgruppe fehlt, der einzige Punkt mit der Wertigkeit 2 ist nämlich mit keinem Punkt der Wertigkeit 1 verbunden:



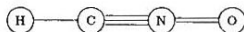
Mit anderen Worten: Zu der Summenformel C_3H_8O gibt es genau drei Bindungsisomere, zu der Strukturformel C_3H_7OH gibt es genau zwei Bindungsisomere. Ähnlich verhält es sich mit der Summenformel $CNOH$. Hierfür gibt es zwei chemische Substanzen, die Cyansäure und die Isocyansäure:



Zu $CNOH$ gibt es sogar noch einen weiteren Bindungsgraphen; eine chemische Verbindung dieser Struktur ist jedoch nicht bekannt:



Andererseits existiert neben Cyansäure und Isocyansäure ein weiteres Isomeres, die Knallsäure, bei dem die Atome in der Reihenfolge $HCNO$ verknüpft sind. Formal kann diese Struktur durch einen Bindungsgraphen mit 5-wertigem Stickstoff beschrieben werden:



Bei der Konstruktion von Bindungsgraphen zu einer gegebenen Summenformel muß also auch die Möglichkeit unterschiedlicher Wertigkeiten einiger Elemente berücksichtigt werden.

Zusammenfassend ist also das Problem der molekularen Strukturerkennung in erster Näherung dieses:

Aus chemischer Information, wie gegebener Summenformel, stofflichen Eigenschaften, spektroskopischen Daten, etc. sollen zunächst die Bindungsgraphen aller Bindungsisomere gewonnen werden, die dieser Information genügen. Diese sollen vorzugsweise nach Substanzklassen getrennt konstruiert werden, d.h. es soll möglich sein, Substrukturen (wie beispielsweise Hydroxylgruppen) vorzuschreiben und/oder zu verbieten.

3 Numerierte Multigraphen

Wollen wir einen molekularen Graphen im Computer speichern, so müssen wir angeben, welche Atome bzw. welche Punkte des Graphen mit welchen Atomen bzw. Punkten verbunden sind, und welche nicht. Dazu bedarf es einer *Numerierung* der Punkte, denn im Computer können keine unnummerierten Strukturen sondern nur nummerierte Strukturen gespeichert werden. Diese Strukturen definieren wir jetzt genauer wie folgt. Ein *numerierter Multigraph* ist ein Paar

$$3.1 \quad (\mathcal{P}, \mathcal{K}),$$

bestehend aus einer Menge \mathcal{P} numerierter *Punkte*, etwa $\mathcal{P} = \{P_1, \dots, P_p\}$, von denen einige durch (ungerichtete) *Kanten* verbunden sind, was durch die Menge der Kanten \mathcal{K} ausgedrückt wird. Diese Kanten können Mehrfachkanten sein, jedoch soll keine der Kanten einen der Punkte mit sich selbst verbinden, d.h. es treten *keine Schleifen* auf. Zur Abkürzung werden wir für die Bezeichnung P_1, \dots, P_p der Punkte meist $1, \dots, p$ schreiben. Unter dem *Kantengrad* bzw. der *Valenz (Wertigkeit)* eines Punktes verstehen wir die Anzahl der Kanten, die von ihm ausgehen.

Die folgende Abbildung zeigt ein Beispiel für die *graphische Darstellung* eines numerierten Multigraphen. Die Menge der Punkte \mathcal{P} ist $\{1, 2, 3, 4\}$, die Menge \mathcal{K} der Kanten besteht aus einer Einfachkante zwischen 1 und 2, einer Einfachkante zwischen 2 und 3, sowie einer Dreifachkante zwischen 3 und 4:



Dieser Graph geht durch geeignetes Bezeichnen der Punkte in den molekularen Graphen der Cyansäure über. Beachten Sie aber, daß dieses Bezeichnen von Punkten mit Atomnamen i.a. nicht eindeutig ist, da es einerseits Atome gibt, die mit verschiedenen Wertigkeiten auftreten können, und es andererseits verschiedene Atome gleicher Wertigkeit gibt.

Ein numerierter Multigraph $(\mathcal{P}, \mathcal{K})$ wird offensichtlich eindeutig durch die zugehörige *Bindungsmatrix* $B(\mathcal{P}, \mathcal{K})$ beschrieben, die wir so definieren:

$$3.2 \quad B(\mathcal{P}, \mathcal{K}) = (b_{ik}), \quad 1 \leq i, k \leq n,$$

wobei b_{ik} die Anzahl der Kanten angibt, die die Punkte i und k verbinden, kurz die *Kantenvielfachheit* des Punktepaars (i, k) . Zum Beispiel ist die Bindungsmatrix für den obigen Multigraphen gleich

$$3.3 \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 3 & 0 \end{pmatrix}.$$

Für eine sparsame Abspeicherung eines Multigraphen $(\mathcal{P}, \mathcal{K})$ mit p Punkten im Rechner müssen aber nicht *alle* p^2 Einträge der Bindungsmatrix abgespeichert werden, für obiges Beispiel also nicht alle $4^2 = 16$ Elemente der Matrix 3.3, vielmehr kommen wir mit wesentlich weniger Speicherbedarf aus. Kanten, die die Punkte i und k verbinden, verbinden nämlich auch die Punkte k und i und es treten nach Voraussetzung keine Schleifen auf: da $B(\mathcal{P}, \mathcal{K})$ symmetrisch ist und die Hauptdiagonale aus lauter Nullen besteht, genügt es, anstelle von $B(\mathcal{P}, \mathcal{K})$ die Folge

$$3.4 \quad \gamma = b_{12}b_{13} \dots b_{1p}b_{23} \dots b_{2p} \dots b_{p-1,p}$$

zu betrachten, die aus denjenigen Elementen von $B(\mathcal{P}, \mathcal{K})$ besteht, welche rechts oberhalb der Hauptdiagonalen stehen; diese Folge nennen wir die *Bindungssequenz* des Graphen. Zu obiger Matrix gehört beispielsweise die Bindungssequenz

$$3.5 \quad 100103.$$

Wenn wir also anstelle der Bindungsmatrix die Bindungssequenz zur Abspeicherung eines Multigraphen mit p Punkten benutzen, müssen wir statt p^2 nur noch $\frac{p^2-p}{2} = \binom{p}{2}$ Zahlen abspeichern, für obiges Beispiel $\binom{4}{2} = 6$, für molekulare Graphen kommen allerdings noch die p Atombezeichnungen hinzu.

Ein klein wenig formaler kann die Bindungssequenz als eine *Abbildung* aufgefaßt werden, und zwar als eine Abbildung von der Menge der *Punktepaare*² $\{i, k\}$ in die Menge $\{0, 1, \dots, m\}$ der *Kantenvielfachheiten*. Die Bindungssequenz aus 3.5 entspricht auf diese Weise der Abbildung mit der folgenden Wertetabelle:

$$3.6 \quad \begin{array}{c} \{i, k\} = \\ b_{ik} = \end{array} \left\| \begin{array}{c|c|c|c|c|c} \{1, 2\} & \{1, 3\} & \{1, 4\} & \{2, 3\} & \{2, 4\} & \{3, 4\} \\ \hline 1 & 0 & 0 & 1 & 0 & 3 \end{array} \right.$$

Wenn wir noch verabreden, die Paare grundsätzlich lexikographisch anzuordnen, dann benötigen wir zur Darstellung dieser Abbildung sogar nur die zweite Zeile dieser Wertetabelle, also nur die Bindungssequenz. Wir wollen eine Abkürzung für solche Mengen von Abbildungen einführen. Dazu benötigen wir zwei Standardmengen der Ordnung p :

$$3.7 \quad \mathbf{p} = \{0, 1, \dots, p-1\}, \text{ und } \underline{\mathbf{p}} := \{1, \dots, p\}.$$

Außerdem kürzen wir die Menge der Paare $\{i, k\}$ so ab:

$$3.8 \quad \mathbf{p}^{[2]} := \{\{i, k\} \mid 1 \leq i < k \leq p\}.$$

Die Bindungssequenz eines numerierten Multigraphen mit p Punkten und maximal m -fachen Kanten ist also eine Abbildung von $\mathbf{p}^{[2]}$ in $m+1 = \{0, 1, \dots, m\}$. Die Bindungssequenz aus 3.5 ist in diesem Sinne eine Abbildung von $\underline{\mathbf{4}}^{[2]}$ in die Menge

²Als Punktepaar verstehen wir hier also eine 2-elementige Teilmenge der Menge aller Punkte.

$3 + 1 = \{0, 1, 2, 3\}$. Da wir, wie oft üblich, mit Y^X die Menge aller Abbildungen von X nach Y abkürzen:

3.9
$$Y^X = \{f : X \rightarrow Y\},$$

können wir mit Hilfe dieser Bezeichnungsweisen jetzt die präzise formale Definition numerierter Multigraphen ohne Schleifen, mit p Punkten und durch m beschränkter Kantenvielfachheit angeben, solche Graphen nennen wir *numerierte m -Graphen mit p Punkten*:

3.10 Definition (Numerierter m -Graph) Ein numerierter m -Graph mit p Punkten ist ein Element $\gamma \in (\mathbf{m} + 1)^{\mathbb{P}^{[2]}}$, also eine Abbildung γ von der Menge

$$\mathbb{P}^{[2]} = \{\{i, k\} \mid 1 \leq i < k \leq p\}$$

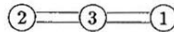
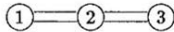
der Punktepaare in die Menge der zugelassenen Kantenvielfachheiten

$$\mathbf{m} + 1 = \{0, 1, \dots, m\}.$$

Dabei bedeutet $\gamma(\{i, k\}) = 0$ bzw. $= j$, daß das Punktepaar $\{i, k\}$ nicht verbunden bzw. durch eine j -fache Kante verbunden ist.

4 Umnummerieren von Multigraphen

Die Tatsache, daß im Computer nur *numerierte Strukturen*, in unserem Beispiel also nur numerierte Multigraphen gespeichert werden können, wirft ein sehr schwerwiegendes Problem auf. Die Numerierung eines Multigraphen ist nämlich im allgemeinen auf sehr viele verschiedene Weisen möglich, so daß wir sehr viele numerierte Multigraphen haben können, die bis auf die Numerierung gleich sind, d.h. denselben chemischen Graphen beschreiben, und es ist sehr schwierig festzustellen, wann dies im konkreten Fall so ist. Ein Beispiel zweier numerierter Multigraphen, die beide CO_2 beschreiben, aber verschieden numeriert sind, ist



Diese Mehrdeutigkeiten müssen wir in den Griff bekommen! Als nächstes wollen wir deshalb die dazu notwendige Mathematik bereitstellen. Betrachten wir ganz allgemein einen numerierten m -Graphen γ mit p Punkten. Die Umnummerierungen der p Punkte von γ bilden eine sogenannte Gruppe, wir müssen also zunächst an diesen Begriff erinnern.

4.1 Definition (Gruppe) Unter einer Gruppe versteht man eine Menge G zusammen mit einer Verknüpfung $*$, also einer Abbildung

$$* : G \times G \rightarrow G : (g, g') \mapsto g * g',$$

die jedem Paar (g, g') von Elementen aus G ein eindeutig bestimmtes Element von G zuordnet, das wir mit $g * g'$ bezeichnen. Es müssen folgende Rechenregeln gelten:

1. Es gilt das Assoziativgesetz, d.h. für alle $g_1, g_2, g_3 \in G$ ist

$$g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3.$$

2. In G existiert genau ein neutrales Element, d.h. ein Element e derart, daß für alle $g \in G$ gilt:

$$g * e = e * g = g.$$

3. Zu jedem $g \in G$ gibt es ein eindeutig bestimmtes Inverses, d.h. ein Element $g' \in G$ mit der Eigenschaft

$$g * g' = g' * g = e.$$

Gruppen mit kommutativer Verknüpfung, d.h. $g * g' = g' * g$, für alle $g, g' \in G$, heißen *abelsche* Gruppen (nach dem Mathematiker N.H. Abel). In diesem Fall schreibt man die Gruppe oft *additiv*, d.h. man nimmt als Verknüpfungssymbol oft $+$ statt $*$, schreibt 0 für das neutrale Element und $-g$ für das Inverse von g . Prominente Beispiele hierfür sind die Zahlbereiche \mathbf{Z} , \mathbf{Q} und \mathbf{C} , die ganzen, die rationalen und die komplexen Zahlen mit der Addition als Verknüpfung. In den Fällen, in denen man offen läßt, ob die Gruppe abelsch ist, oder nicht, schreibt man die Gruppe meist *multiplikativ*, also \cdot anstelle von $*$ als Verknüpfungssymbol oder, wie meist üblich, garnichts. Bei dieser multiplikativen Schreibweise bezeichnet man das neutrale Element mit 1 und das Inverse von g mit g^{-1} . Eine solche Gruppe ist die Gruppe

$$S_{\mathbf{p}},$$

die aus allen *Permutationen*, d.h. Umordnungen, der Menge \mathbf{p} besteht, und deren Verknüpfung die Hintereinanderausführung ist. Diese Gruppe ist für $p > 2$ nicht mehr abelsch, wir schreiben sie deshalb multiplikativ. Ihre Ordnung, d.h. die Anzahl ihrer Elemente, ist $p! = 1 \cdot 2 \cdot \dots \cdot p$.

Benutzen wir kleine griechische Buchstaben π, ρ, σ etc. zur Bezeichnung der Permutationen, und schreiben wir kurzerhand πi für das Bild von $i \in \mathbf{p}$, dann besteht $S_{\mathbf{p}}$ aus den folgenden Elementen:

$$\pi = \begin{pmatrix} 1 & \dots & p \\ \pi 1 & \dots & \pi p \end{pmatrix}.$$

Dabei haben wir einfach das Bild πi unter das Urbild gesetzt. Zu beachten ist bei dieser Schreibweise, daß die Reihenfolge der Spalten $\begin{smallmatrix} i \\ \pi i \end{smallmatrix}$ für die Beschreibung von π offensichtlich keine Rolle spielt. Die Verknüpfung ist, wie bereits erwähnt, die *Hintereinanderausführung*, also bedeutet $\rho \cdot \pi$, was wir wie meist üblich kurz mit $\rho\pi$ bezeichnen, die Permutation "erst π , dann ρ ", d.h. wir lesen Produkte von Permutationen *von rechts nach links*, es gilt also

$$(\rho\pi)i = \rho(\pi i).$$

Das neutrale Element bezeichnen wir wegen der multiplikativen Schreibweise mit 1, es ist die identische Abbildung:

$$1 = \begin{pmatrix} 1 & 2 & \dots & p-1 & p \\ 1 & 2 & \dots & p-1 & p \end{pmatrix},$$

und das Inverse von π (es wird mit π^{-1} bezeichnet) erhält man durch Vertauschen der beiden Zeilen

4.2
$$\pi^{-1} = \begin{pmatrix} \pi 1 & \dots & \pi p \\ 1 & \dots & p \end{pmatrix},$$

also ist zum Beispiel

$$\begin{pmatrix} 123 \\ 231 \end{pmatrix}^{-1} = \begin{pmatrix} 231 \\ 123 \end{pmatrix} = \begin{pmatrix} 123 \\ 312 \end{pmatrix}.$$

Wir wollen die Gruppe $S_{\mathbf{p}}$ jetzt anwenden, um die oben erwähnten Mehrdeutigkeiten beim Numerieren von Multigraphen in den Griff zu bekommen. Dazu betrachten wir einen numerierten m -Graphen γ , also eine Abbildung von $\mathbf{p}^{[2]}$ nach $\mathbf{m} + 1$. Durch Ummunieren der Punkte entstehen aus γ weitere numerierte m -Graphen γ', γ'', \dots , diese wollen wir zu der sogenannten *Isomorphieklasse* von γ zusammenfassen, denn diese durch Ummunierung entstandenen numerierten m -Graphen unterscheiden sich ja nicht wesentlich von γ . Wie dies geschieht wird präzise mit Hilfe des folgenden Begriffs der Operation einer Gruppe G auf einer Menge M beschrieben.

4.3 Definition (Operation einer Gruppe auf einer Menge) *Ist G eine multiplikativ geschriebene Gruppe, M eine Menge, dann wird eine Operation von G auf M durch eine Abbildung*

$$G \times M \rightarrow M: (g, m) \mapsto gm$$

beschrieben (d.h. jedem Paar (g, m) mit $g \in G$ und $m \in M$ ist ein Element von M zugeordnet, dieses wird mit gm bezeichnet) mit folgenden Eigenschaften: Für alle $m \in M$, alle $g, g' \in G$ und für das neutrale Element $1 \in G$ gilt:

$$(gg')m = g(g'm), \quad 1m = m.$$

Ein triviales Beispiel ist die sogenannte *natürliche Operation* von $S_{\mathbf{p}}$ auf \mathbf{p} ,

4.4
$$S_{\mathbf{p}} \times \mathbf{p} \rightarrow \mathbf{p}: (\pi, i) \mapsto \pi i.$$

Diese Operation können wir als *Ummunierung der Punkte eines numerierten m -Graphen* interpretieren. Sie ergibt gleich noch ein weniger triviales Beispiel, das wir als *entsprechende Operation auf den Punktepaaren eines numerierten m -Graphen* interpretieren können:

4.5
$$S_{\mathbf{p}} \times \mathbf{p}^{[2]} \rightarrow \mathbf{p}^{[2]}: (\pi, \{i, k\}) \mapsto \{\pi i, \pi k\}.$$

Schließlich ergibt sich daraus die gewünschte Beschreibung dessen, was bei der Umnummerierung der Punkte mit den nummerierten m -Graphen selbst passiert:

$$4.6 \quad S_{\mathbf{p}} \times (\mathbf{m} + \mathbf{1})^{\{\mathbf{p}^{[2]}\}} \rightarrow (\mathbf{m} + \mathbf{1})^{\{\mathbf{p}^{[2]}\}}: (\pi, \gamma) \mapsto \gamma \circ \pi^{-1}.$$

Die symmetrische Gruppe operiert also in folgender Weise auf der Menge der nummerierten m -Graphen: Dem Paar (π, γ) wird der nummerierte m -Graph $\pi\gamma = \gamma \circ \pi^{-1}$ zugeordnet, das ist die folgende Abbildung γ' :

$$\gamma'(\{i, k\}) = \gamma(\pi^{-1}\{i, k\}) = \gamma(\{\pi^{-1}i, \pi^{-1}k\}).$$

Wir haben damit eine präzise Formulierung dessen, was die *Umnummerierung* nummerierter m -Graphen mit p Punkten vom mathematischen Standpunkt her gesehen bedeutet, es ist die in 4.6 angegebene *Operation der symmetrischen Gruppe $S_{\mathbf{p}}$ auf der Menge der nummerierten m -Graphen mit p Punkten!*

5 Isomorphie numerierter Multigraphen

Wir haben *numerierte* Multigraphen deshalb einführen müssen, weil wir im Rechner Punkte und damit auch Multigraphen nur in numerierter Form speichern können. Wir wollen zwei nummerierte m -Graphen γ, γ' jedoch als im wesentlichen gleich ansehen, wenn sie durch Umnummerieren auseinander hervorgehen, d.h. durch die Operation eines geeigneten $\pi \in S_{\mathbf{p}}$ ineinander übergeführt werden, also wenn es $\pi \in S_{\mathbf{p}}$ gibt mit $\pi\gamma = \gamma'$. Das führt uns auf den folgenden Begriff, der die Zusammenfassung solcher im wesentlichen gleicher Elemente einer Menge beschreibt, auf der eine Gruppe operiert:

5.1 Definition (Bahn einer Gruppe auf einer Menge) *Operiert die Gruppe G auf der Menge M wie oben definiert, und ist $m \in M$, dann heißt die Menge*

$$G(m) = \{gm \mid g \in G\} \subseteq M$$

die Bahn von m unter G .

Die Bahn von m besteht also aus allen $m' \in M$, für die es $g \in G$ gibt mit $m' = gm$. Diese Mengen haben die folgende sehr wichtige Eigenschaft:

5.2 Satz *Je zwei Bahnen $G(m), G(m')$ von G auf M sind entweder identisch oder disjunkt (elementfremd). Die Menge*

$$G \setminus M = \{G(m) \mid m \in M\}$$

der verschiedenen Bahnen bildet also eine vollständige Zerlegung von M in paarweise disjunkte und nicht leere Teilmengen von M .

Solche vollständigen Zerlegungen in disjunkte und nicht leere Teilmengen nennt man auch *Partitionen*. Insbesondere bilden also die Bahnen von $S_{\underline{p}}$ eine Partition von $(m+1)^{\binom{p}{2}}$. Die Elemente dieser Partition sind die Bahnen von $S_{\underline{p}}$. Eine Bahn besteht hier also aus allen numerierten m -Graphen mit den p Punkten $1, \dots, p$, die aus irgendeinem numerierten m -Graphen durch Umnumerieren hervorgehen! Die Operation 4.6 faßt also genau diejenigen numerierten m -Graphen mit p Punkten zusammen, die wir als im wesentlichen gleich ansehen wollen! Oder, anders gewendet: Wir erhalten einen vollständigen Satz wesentlich verschiedener numerierter m -Graphen mit p Punkten, indem wir aus jeder Bahn von $S_{\underline{p}}$ auf $(m+1)^{\binom{p}{2}}$ genau ein Element herausgreifen. Nehmen wir diesen repräsentativen Satz numerierter m -Graphen mit p Punkten her, und lassen wir die Numerierung jeweils weg, dann erhalten wir das, was wir als m -Graphen mit p Punkten, oder, genauer, als unnumerierte m -Graphen mit p Punkten bezeichnen wollen. So besteht beispielsweise die Menge aller 2-Graphen mit 4 Punkten aus 11 Elementen, die Abb. 1 zeigt. Es ist jetzt ganz wichtig zu beachten,

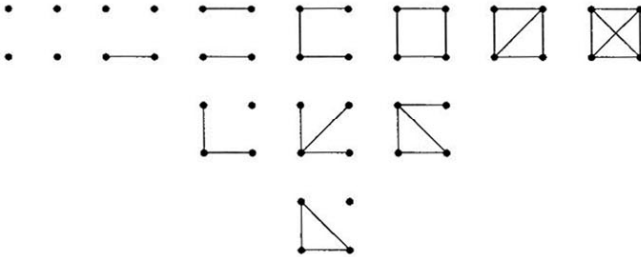


Abb. 1: Die 2-Graphen mit 4 Punkten

daß wir diese repräsentative Menge durch Wahl irgendeines Elements der Bahn und Weglassen der Numerierung erhalten haben, deshalb sind zum Beispiel die beiden unnumerierten Graphen gleich, die sich aus den beiden numerierten Graphen von Abb. 2 ergeben, sie sind in Abb. 3 dargestellt. Das heißt, es gilt tatsächlich die in Abb. 3

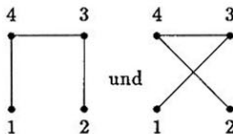


Abb. 2: Zwei isomorphe numerierte 2-Graphen

symbolisch angedeutete Gleichheit!

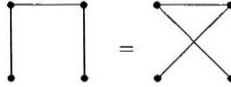
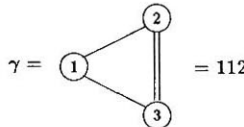
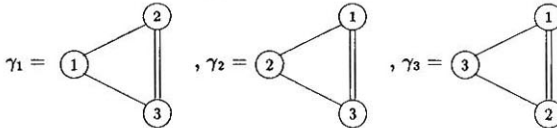


Abb. 3: Die beiden zugehörigen unnummerierten (und gleichen) Graphen

Numerierte m -Graphen, die durch Umnummerieren auseinander hervorgehen, also, mit anderen Worten, *in derselben Bahn liegen*, heißen zueinander *isomorph*. Beispielsweise sind die zu dem im Rechner als Bindungssequenz 112 gespeicherten Graphen



isomorphen numerierten 2-Graphen die folgenden:



Dazu ist übrigens zu bemerken, daß eine Vertauschung der beiden durch eine Zweifachkante verbundenen Punkte die Bindungssequenz nicht ändert!

Wir können jetzt definieren, was wir unter einem m -Graphen verstehen wollen:

5.3 Definition (m -Graph mit p Punkten) *Unter einem m -Graphen mit Punktmenge \underline{p} verstehen wir eine Isomorphieklasse Γ numerierter m -Graphen γ mit Punktmenge \underline{p} , oder, was dasselbe ist, eine Bahn der symmetrischen Gruppe $S_{\underline{p}}$ auf der Menge $(\mathbf{m} + \mathbf{1})^{\{\underline{p}\}}$ aller Abbildungen von der Menge $\underline{p}^{\{2\}}$ der $\binom{p}{2}$ Punktepaare in die Menge $\mathbf{m} + \mathbf{1}$ der Vielfachheiten.*

Die Menge aller m -Graphen mit p Punkten ist also, unter Benutzung der Abkürzungen aus 5.2, gleich

$$S_{\underline{p}} \setminus (\mathbf{m} + \mathbf{1})^{\{\underline{p}\}},$$

und eine Skizze dieser Graphen Γ erhält man durch Herausgreifen eines Systems von Repräsentanten dieser Bahnen und Weglassen der jeweiligen Numerierung. Für $m = 1$ und $p = 4$ gibt Abb. 1 eine solche Skizze.

6 Valenzstrichgraphen

Nachdem wir jetzt eine präzise Definition von m -Graphen mit p Punkten zur Hand haben, können wir unter ihnen diejenigen aussuchen, welche wir einer vorgegebenen chemischen Formel zuordnen wollen. Enthält die Formel p Atome (z.B. $p = 12$ für C_3H_7OH bzw. C_3H_8O), dann ordnen wir die p Valenzen, die wir den Atomen zuschreiben zu einer absteigenden Zahlenfolge, der *Valenzenfolge*

$$V = (v_1, v_2, \dots),$$

wobei v_i die Anzahl der Atome mit Valenz i bezeichnet. Bei C_3H_7OH nehmen wir beispielsweise an, alle drei Kohlenstoffatome C seien vierwertig, das Sauerstoffatom O sei zweiwertig, und die acht Wasserstoffatome H seien einwertig, so daß sich die Valenzenfolge

$$V(C_3H_7OH) = (8, 1, 0, 3)$$

ergibt. Bei den entsprechenden Graphen gibt v_i an, in wievielen Punkten des Graphen genau i Kanten zusammentreffen.

Schließlich bezeichnen wir einen m -Graphen als *zusammenhängend*, wenn wir von einem seiner Punkte alle anderen Punkte erreichen können, indem wir Kanten entlang wandern. Z. B. ist der folgende Graph *nicht* zusammenhängend:



Für die Darstellung der Bindungsstruktur von Molekülen kommen in der Regel nur zusammenhängende Graphen in Betracht. Wir definieren deshalb

6.1 Definition (Valenzstrichgraphen) *Unter den Valenzstrichgraphen zu einer vorgegebenen chemischen Formel und zugeordneter Valenzenfolge $V = (v_1, \dots, v_m)$ verstehen wir die zusammenhängenden m -Graphen Γ , die V als Valenzenfolge haben, d. h. die genau v_i Punkte der Valenz i besitzen. Moleküle, deren Bindungsverhältnisse durch diese Graphen beschrieben werden, heißen auch Valenz- oder Bindungsisomere zu der gegebenen Formel und Valenzenfolge.*

Zur Formel C_3H_7OH und zugeordneter Valenzenfolge $(8, 1, 0, 3)$ gibt es also genau die drei oben angegebenen Valenzstrichgraphen und dementsprechend zumindest drei Valenz- oder Bindungsisomere (denn Valenzisomere zu demselben Bindungsgraphen können sich ja noch in den Bindungsverhältnissen der einzelnen Atomsorten unterscheiden).

7 Bindungsisomere

Leider sind die Valenzstrichgraphen bzw. die Valenzisomere noch nicht ganz genau das, was wir haben wollen. Wenn wir nämlich in der Valenzfolge *verschiedenen Atomen dieselbe Valenz* zugeordnet haben, dann können wir die Punkte desselben Valenzstrichgraphen unter Umständen auf wesentlich verschiedene Weisen aber unter Berücksichtigung der angenommenen Valenzen mit Atomnamen bezeichnen. Betrachten wir beispielsweise den folgenden Graphen:



Nehmen wir einmal an, er beschreibe ein Molekül, von dessen 4 Atomen der Valenz 3 jeweils zwei zur Atomsorte a und zwei zu der von a *verschiedenen* Atomsorte b gehören. Dann können wir die entsprechenden Punkte auf diese zwei wesentlich verschiedenen Weisen beschriften:



Es ist also von Bedeutung, diese wesentlich verschiedenen Beschriftungen zu unterscheiden! Zur präzisen Beschreibung dieses Sachverhaltes gehen wir deshalb zur Definition der Operation einer Gruppe auf einer Menge zurück und benutzen die folgende Definition, für die wir an 4.3 erinnern:

7.1 Definition (Stabilisator eines Punktes) Ist G eine Gruppe und M eine Menge, auf der G gemäß

$$G \times M \rightarrow M: (g, m) \mapsto gm$$

operiert, dann heißt, für $m \in M$, die folgende Untergruppe G_m von G der Stabilisator von m :

$$G_m = \{g \in G \mid gm = m\}.$$

Der Stabilisator des numerierten Graphen $\gamma \in (\mathfrak{m} + 1)^{\mathbb{P}^{(2)}}$ besteht also aus den Umnumerierungen $\pi \in S_{\mathbb{P}}$, die die Bindungssequenz γ fest lassen: $\gamma \circ \pi^{-1} = \gamma$. Dieser Untergruppe geben wir wegen ihrer großen Bedeutung für unsere Problemstellung noch einen speziellen Namen:

7.2 Definition (Automorphismengruppen numerierter m -Graphen) Der Stabilisator eines numerierten m -Graphen $\gamma \in (\mathfrak{m} + 1)^{\mathbb{P}^{(2)}}$ wird mit $A(\gamma)$ bezeichnet und heißt die Automorphismengruppe dieses Graphen:

$$A(\gamma) = \{\pi \in S_{\mathbb{P}} \mid \gamma \circ \pi^{-1} = \gamma\}.$$

Zu jedem Valenzstrichgraphen gehört auf diese Weise eine Automorphismengruppe. Wir erhalten sie, indem wir die Punkte des Valenzstrichgraphen auf alle möglichen Weisen numerieren und die Bindungssequenz mit der ursprünglichen vergleichen. Die Automorphismengruppe $A(\gamma)$ besteht dann aus den $\pi \in S_{\mathbf{p}}$, die γ in sich überführen, das heißt für die $\gamma = \gamma \circ \pi^{-1}$ gilt.

Und nun zur Definition der Beschriftung eines numerierten m -Graphen γ :

7.3 Definition (Beschriftungen numerierter m -Graphen) Als Beschriftungen des numerierten Graphen γ bezeichnen wir nun gewisse Abbildungen

$$\beta \in \mathbf{n}^{\mathbf{p}}.$$

Dabei sei n die Anzahl der verschiedenen Atomsorten, die wir der vorgegebenen Summenformel entnehmen, also z.B. $n = 3$ im Fall der Summenformel C_3H_7OH , und β heißt nur dann Beschriftung, wenn β die vorgegebenen Valenzen respektiert, also wenn für jedes $i \in \mathbf{p}$ gilt, daß $\beta(i)$ die Nummer einer Atomsorte ist, deren Valenz dem Eckengrad des Punktes mit der Nummer i in γ gleicht. Ein numerierter beschrifteter m -Graph ist demnach ein Paar (γ, β) , bestehend aus einem numerierten m -Graphen γ und einer Beschriftung β von diesem. Die Menge aller Beschriftungen von γ bezeichnen wir so:

$$B(\gamma) = \{\beta \mid \beta \text{ Beschriftung von } \gamma\}.$$

Da Umnumerierungen, die γ fest lassen, höchstens Punkte desselben Kantengrades ineinander überführen, permutiert $A(\gamma)$ die Elemente von $B(\gamma)$ untereinander, d.h. $A(\gamma)$ operiert auf $B(\gamma)$ und bildet dementsprechend eine Menge von Bahnen auf $B(\gamma)$:

$$A(\gamma) \backslash B(\gamma).$$

Die wesentlich verschiedenen Beschriftungen von γ erhält man demnach dadurch, daß man aus jeder dieser Bahnen von $A(\gamma)$ auf $B(\gamma)$ je ein Element entnimmt. Eine solche Auswahl heißt auch eine *Transversale* oder ein *Repräsentantensystem* von $A(\gamma) \backslash B(\gamma)$, wir bezeichnen diese(s) mit

$$R(A(\gamma) \backslash B(\gamma)).$$

Damit können wir endlich formulieren, was wir unter den gesuchten *Bindungsisomeren* verstehen wollen, deren Konstruktion wir anstreben:

7.4 Definition (Bindungsisomere) Die Menge aller Bindungsisomere zu einer vorgegebenen chemischen Formel und der ihr zugeordneten Valenzenfolge erhält man, indem man die sämtlichen Valenzstrichgraphen hierzu konstruiert, und jeden von ihnen mit den sämtlichen wesentlich verschiedenen Beschriftungen versieht.

Dabei ergeben sich die wesentlich verschiedenen Beschriftungen eines Valenzstrichgraphen Γ in folgender Weise:

- Durch Numerieren der Punkte von Γ auf irgendeine Weise erhalten wir die Menge p der Punktnummern und den numerierten Graphen γ . Aus diesem ergibt sich die Automorphismengruppe $A(\gamma)$.
- Dann ermittelt man aus der gegebenen chemischen Formel die Menge der vorkommenden Atomsorten und ihre Valenzen v , woraus sich die Menge $B(\gamma)$ der Beschriftungen von γ ergibt.
- Jetzt berechnet man die Menge $A(\gamma) \setminus B(\gamma)$ und ein Repräsentantensystem dieser Bahnen: $R(A(\gamma) \setminus B(\gamma))$. Die Menge aller Bindungsisomere zum Valenzstrichgraphen Γ , der vorgegebenen chemischen Formel und Valenzenfolge der darin auftretenden Atome ist dann die Menge

$$\{(\gamma, \beta) \mid \beta \in R(A(\gamma) \setminus B(\gamma))\}.$$

Betrachten wir als Beispiel die (fiktive) chemische Formel a_2b_2c , der wir die Valenzenfolge $V = (0, 0, 4, 1)$ zuordnen. Wir wollen also alle Bindungsisomere bestimmen, die aus 5 Atomen bestehen, von denen eines die Valenz 4 haben soll, es wird mit c bezeichnet, zwei weitere Atome, die mit a bezeichnet werden, sollen die Wertigkeit 3 haben, die restlichen beiden sollen ebenfalls die Valenz 3 haben, aber zu einer von a verschiedenen Atomsorte b gehören.

Computerberechnungen, deren Organisation und Implementierung weiter unten besprochen werden, ergeben genau 6 zusammenhängende 4-Graphen $\Gamma_1, \dots, \Gamma_6$. Diese wurden numeriert, dies ergab entsprechende numerierte Graphen $\gamma_1, \dots, \gamma_6$, und die Automorphismengruppen dieser numerierten Multigraphen γ_i wurden berechnet. Schließlich lieferte die Berechnung der Bahnen der Automorphismengruppen $A(\gamma_i)$ auf den Mengen $B(\gamma_i)$ aller Beschriftungen die wesentlich verschiedenen Beschriftungen der fünf Punkte von Γ_i mit den Namen a, b, c der verschiedenen Atomsorten. Hier ist zunächst einer von diesen numerierten Graphen γ_1 :



Seine Automorphismengruppe $A(\gamma_1)$ enthält, neben der Identität, die Elemente

$$\begin{pmatrix} 12345 \\ 12435 \end{pmatrix}, \begin{pmatrix} 12345 \\ 13254 \end{pmatrix}, \begin{pmatrix} 12345 \\ 13452 \end{pmatrix}, \begin{pmatrix} 12345 \\ 14253 \end{pmatrix}, \begin{pmatrix} 12345 \\ 14523 \end{pmatrix}, \begin{pmatrix} 12345 \\ 15342 \end{pmatrix}, \begin{pmatrix} 12345 \\ 15432 \end{pmatrix}$$

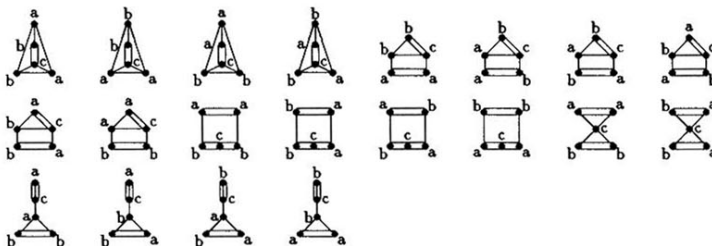
Mit Hilfe dieser Gruppe zeigt man dann, daß es genau die folgenden beiden wesentlich verschiedenen Beschriftungen der Punkte 1,2,3,4,5 gibt:

$$cbbaa, \text{ und } cbaab.$$

Wir erhalten daraus die folgenden beiden *Bindungsisomere* zu der angegebenen chemischen Formel a_2b_2c und der Valenzenfolge $(0, 0, 4, 1)$:



Analog erhält man fünf weitere Valenzstrichgraphen, die jeweils wieder zu verschiedenen Bindungsisomeren führen. Diese haben die folgenden graphischen Darstellungen:



Wir haben also herausbekommen, daß zu der chemischen Formel a_2b_2c und dazu vorgegebener Valenzenfolge $(0, 0, 4, 1)$ genau 6 Valenzstrichgraphen und die oben angegebenen 22 Bindungsisomere gehören!

Wir hoffen, damit eine genaue, aber auch lesbare Beschreibung dessen gegeben zu haben, was unter dem Begriff *Bindungsisomer* zu verstehen ist.

8 Chemische Formeln und Bindungsisomere

Bevor man nun den Strukturgenerator auf eine spezielle chemische Formel anwendet, um alle entsprechenden Bindungsisomere zu bestimmen, kann man sich noch überlegen welche Valenzenfolgen zu chemischen Formeln gehören. Im Rahmen eines systematischen Tests wurden deshalb mit einem Hilfsprogramm Formeln zu vorgegebener Atomanzahl (anz) und maximaler Valenz (max) erzeugt und anschließend wurden jeweils die zugehörigen Bindungsisomere mit dem Generator konstruiert. Wir wollen wenigstens deren Anzahlen angeben.

Wir gehen dabei in zwei Schritten vor, zunächst einmal betrachten wir nur solche Formeln, die zu jeder der auftretenden Valenzen ausschließlich Atome desselben Elements enthält. Jede Formel wird dann nämlich eindeutig

$$V = (v_1, \dots, v_{max})$$

repräsentiert, wobei v_i die Anzahl der Atome mit Valenz i bezeichnet. Es gilt also:

$$\sum_{i=1}^{\max} v_i = anz.$$

Eine beliebige Folge natürlicher Zahlen v_i entspricht — im mathematisch abstrakten Sinne — genau dann einer chemischen Formel, wenn es einen molekularen Graphen hierzu gibt, also eine zusammenhängenden Multigraphen mit dieser Folge als Ecken-gradfolge. Man weiß, daß ein solcher Graph genau dann existiert, wenn die folgenden beiden Bedingungen erfüllt sind: Einmal muß gelten (vgl. etwa [17]):

$$8.1 \quad 2 \cdot \max \leq \sum_i i \cdot v_i,$$

zum andern braucht man, daß die *zyklomatische Zahl* eine natürliche Zahl ist. Diese zyklomatische Zahl ist definiert als

$$8.2 \quad \frac{1}{2} \left(\sum_{i=1}^{\max} (i-2)v_i + 2 \right).$$

(Es ist die Anzahl der Kanten, die man wegnehmen kann ohne daß der Graph in mehrere Zusammenhangskomponenten zerfällt.) In der folgenden Tabelle sind Valenzfolgen angegeben, die diese beiden Bedingungen erfüllen, jede dieser Valenzfolgen stellt also eine chemische Formel dar, im abstrakten mathematischen Sinn, d.h. eine *potentielle chemische Formel*. Zusätzlich ist in eckigen Klammern die zugehörige Anzahl der Bindungsisomere angegeben, die mit Hilfe des nachfolgend beschriebenen Strukturgenerators konstruiert wurden.

Bei C_4H_3NO und $\max = 4$ hat man also:

$$(3, 1, 1, 4) \quad [775].$$

Aus Gründen der Übersichtlichkeit wurde zu jeder Valenz genau *eine* Atomsorte berücksichtigt. Als Beispiel sei die Tabelle mit $anz = 4$ und $\max = 6$ angegeben (vgl. [1]):

(3, 0, 1, 0, 0, 0) [1]	(0, 2, 0, 0, 2, 0) [3]
(2, 2, 0, 0, 0, 0) [1]	(0, 2, 0, 0, 0, 2) [3]
(2, 1, 0, 1, 0, 0) [1]	(0, 1, 2, 1, 0, 0) [4]
(2, 0, 2, 0, 0, 0) [1]	(0, 1, 2, 0, 0, 1) [2]
(2, 0, 1, 0, 1, 0) [1]	(0, 1, 1, 1, 1, 0) [6]
(2, 0, 0, 2, 0, 0) [1]	(0, 1, 1, 0, 1, 1) [6]
(2, 0, 0, 1, 0, 1) [1]	(0, 1, 0, 3, 0, 0) [2]
(2, 0, 0, 0, 2, 0) [1]	(0, 1, 0, 2, 0, 1) [4]
(2, 0, 0, 0, 0, 2) [1]	(0, 1, 0, 1, 2, 0) [4]
(1, 2, 1, 0, 0, 0) [2]	(0, 1, 0, 1, 0, 2) [4]
(1, 2, 0, 0, 1, 0) [1]	(0, 1, 0, 0, 2, 1) [4]
(1, 1, 1, 1, 0, 0) [3]	(0, 1, 0, 0, 0, 3) [2]
(1, 1, 1, 0, 0, 1) [1]	(0, 0, 4, 0, 0, 0) [2]
(1, 1, 0, 1, 1, 0) [3]	(0, 0, 3, 0, 1, 0) [2]
(1, 1, 0, 0, 1, 1) [3]	(0, 0, 2, 2, 0, 0) [5]
(1, 0, 3, 0, 0, 0) [1]	(0, 0, 2, 1, 0, 1) [4]
(1, 0, 2, 0, 1, 0) [2]	(0, 0, 2, 0, 2, 0) [5]
(1, 0, 1, 2, 0, 0) [2]	(0, 0, 2, 0, 0, 2) [5]
(1, 0, 1, 1, 0, 1) [3]	(0, 0, 1, 2, 1, 0) [6]
(1, 0, 1, 0, 2, 0) [2]	(0, 0, 1, 1, 1, 1) [10]
(1, 0, 1, 0, 0, 2) [2]	(0, 0, 1, 0, 3, 0) [3]
(1, 0, 0, 2, 1, 0) [2]	(0, 0, 1, 0, 1, 2) [6]
(1, 0, 0, 1, 1, 1) [3]	(0, 0, 0, 4, 0, 0) [3]
(1, 0, 0, 0, 3, 0) [1]	(0, 0, 0, 3, 0, 1) [3]
(1, 0, 0, 0, 1, 2) [2]	(0, 0, 0, 2, 2, 0) [8]
(0, 4, 0, 0, 0, 0) [1]	(0, 0, 0, 2, 0, 2) [8]
(0, 3, 0, 1, 0, 0) [1]	(0, 0, 0, 1, 2, 1) [9]
(0, 3, 0, 0, 0, 1) [1]	(0, 0, 0, 1, 0, 3) [4]
(0, 2, 2, 0, 0, 0) [3]	(0, 0, 0, 0, 4, 0) [4]
(0, 2, 1, 0, 1, 0) [2]	(0, 0, 0, 0, 2, 2) [11]
(0, 2, 0, 2, 0, 0) [3]	(0, 0, 0, 0, 0, 4) [6]
(0, 2, 0, 1, 0, 1) [2]	gesamt: 63 [203]

Es existieren also beispielsweise 63 Formeln mit 4 Atomen und maximaler Valenz 6 mit insgesamt 203 Bindungsisomeren.

Tabellen dieser Art wurden bei folgenden Atom- und Valenzenanzahlen erstellt: Für die maximale Valenz $max = 4$ ergab sich die Tabelle

Atome	Formeln	Isomere
3	9	9
4	18	37
5	27	146
6	41	772
7	56	4449
8	78	30307
9	101	228605
10	132	1921464

Für $max = 5$:

Atome	Formeln	Isomere
3	14	14
4	36	93
5	59	602
6	107	5847
7	156	66289
8	247	937696

$max = 6$:

Atome	Formeln	Isomere
3	23	23
4	63	203
5	124	2126
6	233	34126
7	392	716141

$max = 7$:

Atome	Formeln	Isomere
3	32	32
4	104	399
5	223	6308
6	468	164965

$max = 8$:

Atome	Formeln	Isomere
3	46	46
4	160	730
5	390	17035

$max = 9$:

Atome	Formeln	Isomere
3	60	60
4	238	1261
5	626	41627

$max = 10 :$

Atome	Formeln	Isomere
3	80	80
4	339	2078
5	983	95165

Insgesamt wurden 5665 Formeln und 4278735 Bindungsisomere erzeugt, wofür ca 1:45:00 Std Rechenzeit benötigt wurden (HP345/68030-Proz.).

Die angegebenen Resultate können auch als Konstruktion von Bindungsisomeren mit vorgegebenen Teilstrukturen interpretiert werden – dies sei an einem Beispiel verdeutlicht:

Die Formel $C_5H_4O_2$ (1821 Isomere) reduziert sich durch die zweifache Vorgabe einer COO Substruktur auf die Formel $CH_4(COO)_2$, wobei COO bei der Konstruktion als ein Atom der Valenz 6 interpretiert werden kann, so daß lediglich die Isomere zur Formel

$$(4, 0, 0, 1, 0, 2)$$

zu berechnen sind. Letzteres macht auch die Isomerenerzeugung bei Formeln sinnvoll, die Valenzen größer als 8 enthalten.

Eine weitere wichtige Verfeinerung ist durch Zulassen mehrerer Atomsorten mit gleicher Valenz leicht möglich. So liefert etwa die Formel $C_5S_2O_3H_{12}Cl_2$ eine Verfeinerung der Valenzenfolge

$$(14, 3, 0, 7)$$

in eine (erweiterte) Folge

$$\underbrace{((12, 2))}_{14}, \underbrace{(3)}_3, \underbrace{(0)}_0, \underbrace{(5, 2)}_7$$

Die Anzahl der Atome gleicher Valenz in einer Valenzenfolge läßt sich also jeweils in verschiedene Atomsorten mit diesen Valenzen *partitionieren*. Zu jeder chemischen Formel kann man auf diese Weise *alle* erweiterten Valenzenfolgen und wiederum die zugehörigen Bindungsisomere berechnen. Uns kommt es dabei allerdings nur auf die *wesentlich* verschiedenen Häufigkeiten der Atome mit gleicher Valenz an, d. h. wir wollen nicht unterscheiden, ob es sich im obigen Beispiel etwa um $H_{12}Cl_2$ oder um $Cl_{12}H_2$ handelt – die Anzahl der Isomere wird dadurch nicht berührt. Die folgenden Tabellen zeigen, daß bei Verfeinerung der Formeln die Zahlen sehr schnell astronomische Dimensionen erreichen:

Für $max = 4$:

Atome	verfeinerte Formeln	Isomere
3	18	18
4	60	179
5	121	1.915
6	302	37.268
7	576	905.266
8	1289	31.233.829
9	2383	1.335.897.151
10	4874	73.741.583.349

Für $maz = 5$:

Atome	verfeinerte Formeln	Isomere
3	25	25
4	108	402
5	223	5.686
6	693	214.937
7	1342	7.524.353
8	3559	603.015.515

9 Der Strukturgenerator MOLGEN

Das Programmsystem, das die Bindungsisomere erzeugt und als Strukturgenerator im Rahmen eines größeren Expertensystems zur molekularen Strukturerkennung dienen soll, heißt MOLGEN. Es ist selbständig lauffähig und kann in entsprechenden Situationen zur redundanzfreien Erzeugung eines vollständigen Satzes von zusammenhängenden Multigraphen mit vorgegebener Eckengradfolge und vorgegebenen sowie verbotenen Teilgraphen dienen.

Das geplante Expertensystem wird ähnlich wie DENDRAL, nach dem Grundsatz

“planen – generieren – testen”

aufgebaut, gemäß dem Schema aus Abb. 4

Dies soll bedeuten, daß dem Strukturgenerator Daten geliefert werden, wie Summenformel, vorgeschriebene Substrukturen, verbotene Teilstrukturen etc. Aus diesen werden dann redundanzfrei und vollständig alle molekularen Graphen generiert, die diesen Bedingungen genügen. Im Anschluß daran werden durch Vergleich der generierten Graphen mit weiteren Daten (z.B. Massenspektren) Strukturkandidaten ausgelesen, evtl. noch versehen mit einer Rangfolge, die dann dem Chemiker zur Entscheidung bzw. Begutachtung vorgelegt werden können.

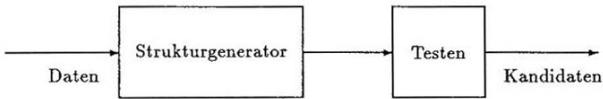


Abb. 4: Schema des Expertensystems

Das vorliegende Teilstück, der Strukturgenerator, ist ohne Zweifel der mathematisch anspruchsvollste Teil, und bei ihm muß größter Wert auf Vermeidung unnötiger Komplexität gelegt werden. Wir haben deshalb für unser System MOLGEN eine möglichst weitgehend *algebraische* Vorgehensweise gewählt im Gegensatz zu der unserer Meinung nach zu kombinatorischen Konstruktion vieler anderer Strukturgeneratoren. Das Programmsystem ist in C geschrieben und kann auf den Betriebssystemen UNIX, VMS und MS-DOS laufen. Es ist für besagte Computeranwendungen in der Chemie geeignet, und zwar in Forschung *und* Lehre. Da input und output graphentheoretisch sind, ist es darüber hinaus für alle entsprechenden Anwendungen in Mathematik, Informatik, Naturwissenschaften und Ökonomie ebenfalls verwendbar. Eine detaillierte Beschreibung des Generators für den Anwender wird weiter unten gegeben.

Die UNIX-Version erlaubt darüberhinaus die Benutzung einer Oberfläche SHOW-MOLE unter X-Windows, welche interaktive Vorgehensweisen ermöglicht und einen Optimierer enthält, der dreidimensionale Realisierungen berechnet, bei denen auch Winkel, Torsionswinkel und Bindungslängen berücksichtigt und interaktiv gemessen und verändert werden können.

Mit der *Abzählung* von Isomeren zu vorgegebener Summenformel ist das Isomerieproblem für einen Chemiker ja keineswegs gelöst, er will vielmehr die entsprechenden Bindungsgraphen *konstruieren* und sichtbar vor Augen haben, damit er entscheiden kann, welcher von ihnen die gerade untersuchte chemische Substanz beschreibt. Diese Aufgabe konnte für komplizierte Fälle natürlich erst mit dem Aufkommen der Computer angepackt werden. Es ist also kein Wunder, daß dieses Problem erst 1965 energisch angegangen wurde, und zwar im Rahmen des DENDRAL-Projekts durch J. Lederberg und seine Arbeitsgruppe. In der Zwischenzeit sind weitere Strukturgeneratoren entstanden, ein interessanter Vergleich von einigen unter ihnen findet sich in [21]. Die meisten von ihnen arbeiten allerdings sehr stark kombinatorisch, so daß ihr Anwendungsbereich wohl wegen der kombinatorischen Explosion auf kleine Beispiele beschränkt bleiben wird.

Wir haben uns demgegenüber der stärker algebraisch orientierten Vorgehensweise von DENDRAL und seinen Ablegern CONGEN und GENOA (vgl. [19], [7]) angeschlossen und eine eigene Version geschrieben, die sich in ganz wesentlichen Teilen von diesen Systemen durch die Verwendung neuer Algorithmen zur Färbung von Bahnen und zur

Berechnung der Automorphismengruppe unterscheidet.

Unser Programmsystem heißt MOLGEN, input und output sind graphentheoretisch. Als input ist zumindest die Summenformel anzugeben, genauer: eine Folge von Valenzen, die den auftretenden Atomen zugeordnet werden. Darüber hinaus können Substrukturen verboten ("BADLIST") bzw. vorgeschrieben ("MAKROATOME" oder "GOODLIST") werden (z.B. Hydroxylgruppe, d.h. eine Substruktur der Form $-OH$ oder $C = C$, Benzolring etc.). Als weiterer input ist die Angabe von Einschränkungen ("RESTRIKTIONEN") möglich (z. B. nur Graphen mit Ringen, Graphen mit einer bestimmten Anzahl zyklischer Teilstrukturen etc.). Als output erhält man eine *vollständige und redundanzfreie* (d.h. doublettenlose) Liste der zusammenhängenden Graphen, die die angebene Valenzenfolge als Eckengradfolge haben, und die vorgegebenen Substrukturen nicht enthalten bzw. enthalten.

Der Konstruktionsprozeß folgt generell der bereits im DENDRAL-System verfolgten Strategie. Dabei sind die einzelnen Schritte mathematisch auf zwei Prinzipien reduzierbar, das *Homomorphieprinzip* und die *Transformation von Färbungsproblemen in Doppelnebenklassenprobleme von Gruppen*. Die mathematische Sicht erleichtert einerseits, das Vorgehen zu gliedern, andererseits können allgemeine Einsichten in diese Prinzipien genutzt werden, um den Prozeß zu verfeinern und zu ergänzen.

Da diese mathematischen Prinzipien von allgemeinerem Interesse sind, beschreiben wir sie zunächst mit einiger Ausführlichkeit und losgelöst von der speziellen Problemstellung der Berechnung aller Bindungsisomere. Hinterher gehen wir detailliert auf die Ermittlung aller Valenzstrichgraphen zu vorgegebener Summenformel ein.

9.1 Das Homomorphieprinzip

Das Homomorphieprinzip ist der Algebra entlehnt. Eine Struktur wird dabei zunächst vergrößert, wobei jedoch wesentliche Merkmale der Ausgangsstruktur unter der vergrößernden Abbildung erhalten bleiben. Das Bild ist meist um Größenordnungen kleiner, so daß es leichter ist, das entsprechend vergrößerte Problem erst dort zu lösen. In der Ausgangsstruktur brauchen dann nur noch gewisse Urbildmengen betrachtet zu werden, um die gesuchten Lösungen des ursprünglichen Problems zu erhalten. In diesem Zusammenhang ist hier die Situation von Bedeutung, daß Repräsentanten aus Bahnen unter einer Gruppe gesucht sind. Ein vollständiges System von Repräsentanten der verschiedenen Bahnen von G auf einer Menge M wollen wir eine *Transversale* der Bahnen nennen und so bezeichnen:

$$T(G \backslash M).$$

Zu beachten ist natürlich, daß es im allgemeinen sehr viele verschiedene solcher Transversalen gibt. Ein weiterer wichtiger Begriff in diesem Zusammenhang ist der Begriff des *Blocks*. Wenn die Gruppe G auf der Menge M operiert, dann heißt eine Teilmenge $B \subseteq M$ Block, wenn sie unter der Operation von jedem $g \in G$ entweder auf sich selbst

oder auf eine zu B disjunkte Menge übergeht:

$$gB \cap B \neq \emptyset \implies gB = B.$$

Bezeichnen wir noch mit G_B die Menge der Gruppenelemente, die ganz B (als Menge, d.h. nicht notwendig elementweise) fest lassen:

$$G_B = \{g \in G \mid gB = B\}.$$

Diese Menge heißt auch der (*mengenweise*) *Stabilisator* oder auch der *Normalisator* von B .

9.1 Satz Operiert G auf M und ist M eine Vereinigung paarweise elementfremder Blöcke $B_i, i \in I$:

$$M = \bigcup_{i \in I} B_i,$$

und bilden die Blöcke $B_j, j \in J \subseteq I$ eine *Transversale der Bahnen von G auf der Menge der Blöcke*,

$$\mathcal{T}(G \backslash \{B_i \mid i \in I\}) = \{B_j \mid j \in J\},$$

dann erhält man wie folgt eine *Transversale der Bahnen von G auf M* :

$$\mathcal{T}(G \backslash M) = \bigcup_{j \in J} \mathcal{T}(G_B \backslash B_j).$$

Die Anwendung dieses Satzes bietet große Vorteile, denn die gesuchte *Transversale von $G \backslash M$* ergibt sich als Vereinigung von *Transversalen von im allgemeinen weit kleineren Gruppen G_B auf weit kleineren Mengen B_j* . Natürlich kann diese Methode nur dann angewendet werden, wenn man eine vollständige Zerlegung von M in Blöcke angeben kann! Einer dieser Glücksfälle wird in folgendem Satz beschrieben:

9.2 Satz Operiert G auf den beiden Mengen M, M' und ist $\varphi: M \rightarrow M'$ eine G -verträgliche Abbildung, d.h. gilt für alle $g \in G$ und alle $m \in M$

$$\varphi(gm) = g\varphi(m),$$

dann ist das Urbild

$$\varphi^{-1}(m')$$

von jedem $m' \in M'$ ein Block, und der *Stabilisator dieses Urbilds ist der Stabilisator von m'* :

$$G_{\varphi^{-1}(m')} = G_{m'}.$$

Zur Ermittlung einer *Transversale $\mathcal{T}(G \backslash M)$* kann man — wenn eine solche Abbildung φ von M auf M' gefunden ist — demnach wie folgt vorgehen:

- Man ermittelt zunächst eine *Transversale $\mathcal{T}(G \backslash M')$* , und zu jedem ihrer Elemente m' ermittelt man den *Stabilisator $G_{m'}$* .

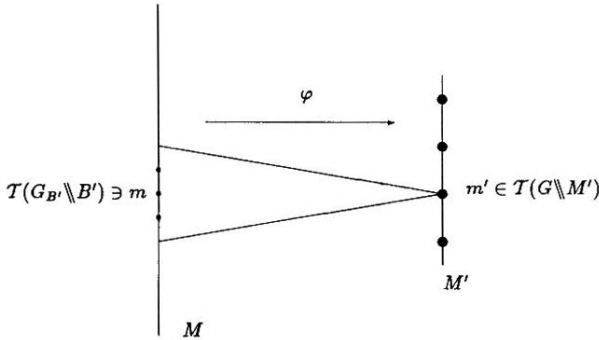


Abb. 5: Das Homomorphieprinzip

- Danach berechnet man die Urbilder, das sind die Blöcke $B' = \varphi^{-1}(m')$, für die $m' \in T(G \setminus M')$.
- Jetzt ermittelt man zu jedem dieser Blöcke eine Transversale der Bahnen des Stabilisators:

$$T(G_{m'} \setminus B').$$

- Die Vereinigung dieser Mengen liefert schließlich, was wir suchen:

$$T(G \setminus M) = \bigcup_{j \in J} T(G_{m'_j} \setminus B'_j).$$

Damit erlaubt eine solche "Vergrößerung" durch die Abbildung φ , zuerst Repräsentanten der Bahnen auf M' zu bestimmen und dann nur in deren Urbildmengen nach Repräsentanten zu suchen. Dabei wird nicht mehr die *gesamte* Gruppe benötigt, sondern nur noch der *Stabilisator* eines Repräsentanten des Bildbereichs. Will man den obigen Satz beweisen, dann hat man folgendes zu zeigen:

9.3 Satz *Unter der G -verträglichen Abbildung $\varphi: M \rightarrow M'$ gehen Elemente m_1, m_2 aus derselben Bahn (von G auf M) in Elemente $\varphi(m_1), \varphi(m_2)$ aus derselben Bahn (von G auf M') über. Ist dabei etwa $g\varphi(m_1) = \varphi(m_2)$, dann gibt es g' im Stabilisator von $\varphi(m_2)$ mit $g'gm_1 = m_2$.*

Für Anwendungen des Homomorphieprinzips auf allgemeine Gruppenoperationen vergleiche [12].

9.2 Doppelnebenklassen

Wir wollen die zweite der zentralen Strategien der Konstruktion erläutern, nämlich die Transformation von Färbungsproblemen in Doppelnebenklassenprobleme von Gruppen. Wir tun dies mit Hilfe von mehreren Beispielen.

Zunächst erinnern wir daran, daß die Menge aller m -Graphen mit p Punkten als Bahnenmenge definiert werden konnte, nämlich als die Bahnenmenge

$$S_{\mathbb{P}} \backslash (m + 1)^{(p^2)},$$

d.h. als ein Spezialfall der Bahnenmenge $G \backslash Y^X$. Wir zeigen deshalb zunächst, daß diese Bahnenmenge $G \backslash Y^X$ mit einer Menge von Doppelnebenklassen identifiziert werden kann. Es sei zuvor noch einmal daran erinnert, was unter Doppelnebenklassen verstanden wird:

9.4 Definition Ist G eine Gruppe mit Untergruppen U und V , dann versteht man unter der Doppelnebenklasse UgV die folgende Teilmenge von G :

$$UgV = \{ugv \mid u \in U, v \in V\}.$$

Die Gesamtheit aller dieser Doppelnebenklassen wird meist so bezeichnet:

$$U \backslash G / V = \{UgV \mid g \in G\}.$$

Die wichtigsten Eigenschaften dieser Doppelnebenklassen fassen wir zusammen in

9.5 Hilfssatz Ist G eine endliche Gruppe mit Untergruppen U und V , dann gilt für die (U, V) -Doppelnebenklassen:

- Je zwei Doppelnebenklassen sind entweder disjunkt oder gleich:

$$UgV \cap Ug'V \neq \emptyset \implies UgV = Ug'V.$$

- Für die Ordnungen von Doppelnebenklassen gilt:

$$|UgV| = \frac{|U||V|}{|U \cap gVg^{-1}|}.$$

Eine ganz wichtige Anwendung finden die Doppelnebenklassen dann, wenn zwei Strukturen "verklebt" werden sollen. Damit meinen wir folgendes: nehmen wir an, es sollen zwei molekulare Graphen zusammengesetzt werden, die beide einen Benzolring enthalten, und diese Zusammensetzung soll auf alle möglichen wesentlich verschiedenen Weisen und dergestalt geschehen, daß ein Benzolring B_1 des einen Graphen mit einem Benzolring B_2 des anderen Graphen "verschmolzen" wird. Es soll ermittelt werden, auf wieviele wesentlich verschiedene Weisen diese Verschmelzung durchgeführt werden kann, darüberhinaus sollen diese Verschmelzungen dann sogar explizit angegeben werden können. Das Problem dabei sind Abbildungen der beiden Benzolringe

B_1, B_2 auf sich selbst, die diese nicht wesentlich ändern, sogenannte *Automorphismen* der Benzolringe. Wir wollen jetzt den entsprechenden mathematischen Satz formulieren, der zeigt, daß die möglichen Verklebungen mit Hilfe von Doppelnebenklassen beschrieben werden können (vgl. [10]). Zunächst eine Vorbemerkung: Eine Verklebung zweier Strukturen S_1 und S_2 , wie etwa die der beiden Benzolringe, ist eine Abbildung $f: S_1 \rightarrow S_2$ zwischen beiden Strukturen, die umkehrbar eindeutig ("bijektiv") ist, und in gewisser Weise einer Umnummerierung entspricht, kurz: f ist ein *Isomorphismus*. Zwei solche Isomorphismen f und f' werden wir als *nicht wesentlich verschieden betrachten*, wenn die eine aus der anderen durch Vorschalten eines Automorphismus π_1 von S_1 und/oder Nachschalten eines Automorphismus π_2 von S_2 hervorgeht

9.6 Satz *Seien S_1, S_2 zwei Strukturen, die verklebt werden sollen, und es sei A_1 eine Gruppe von Automorphismen von S_1 , A_2 eine Gruppe von Automorphismen von S_2 , $f: S_1 \rightarrow S_2$ ein Isomorphismus. Dann ist jeder Isomorphismus $f': S_1 \rightarrow S_2$ von der Form $f' = \pi f$ mit $\pi \in \text{Aut}(S_2)$. Zwei Isomorphismen f', f'' wollen wir als nicht wesentlich verschieden ansehen (kurz: $f' \sim f''$), wenn $\alpha_1 \in A_1$ und $\alpha_2 \in A_2$ existieren mit $f' = \alpha_2 \circ f'' \circ \alpha_1$. Dann gilt:*

$$f' = \pi f, f'' = \pi' f \text{ und } (A_2 \pi_1 (f A_1 f^{-1}) = A_2 \pi_2 (f A_1 f^{-1}) \iff f_1 \sim f_2.$$

Diese Transformation einer Menge von Isomorphismen in eine Gruppe von Automorphismen erlaubt es, gruppentheoretische Algorithmen einzusetzen. Dies geschieht dann für das Problem, Doppelnebenklassenrepräsentanten zu bestimmen.

Ein wichtiger Spezialfall solcher Verklebungen ergibt sich, wenn etwa S_1 die Menge der Gerüstplätze eines Molekülskeletts und S_2 die Menge der auf die Gerüstplätze zu setzenden Liganden ist. Dann operiert auf S_1 die Automorphismengruppe A_1 des Gerüsts und auf S_2 die Gruppe A_2 der Permutationen, die Liganden desselben Typs untereinander permutieren. Die Verklebungen entsprechen dann den Möglichkeiten, die Liganden auf die Gerüstplätze zu verteilen, und die Doppelnebenklassen entsprechen den Klassen nicht unterscheidbarer Ergebnisse. Das wurde für genau diese Situation von Ruch/Hässelbarth/Richter erkannt ([24], [25]). Die mathematische Formulierung dieses Sachverhalts liest sich so:

Sei wieder G eine endliche Gruppe, die auf X und damit auch auf Y^X operiert (vgl. 4.6). \bar{G} bezeichne die Gruppe der Permutationen \bar{g} von X , die zu den Elementen $g \in G$ gehören. Weiterhin bezeichne S_X die symmetrische Gruppe auf X und $(S_X)_f$ den Stabilisator von $f \in Y^X$. Es gilt dann:

9.7 Das Lemma von Ruch/Hässelbarth/Richter *Ist $f \in Y^X$, dann ergibt die Abbildung*

$$\Theta: S_X \rightarrow Y^X : \pi \mapsto f \circ \pi^{-1}$$

eine umkehrbar eindeutige Beziehung zwischen den Doppelnebenklassen und den Bahnen von G auf Y^X , die aus Elementen f' bestehen mit

$$\forall y \in Y: |f^{-1}(y)| = |f'^{-1}(y)|,$$

d.h. die Ordnungen der Urbilder, also die Vielfachheiten, mit denen die Werte $y \in Y$ von f und f' angenommen werden, sind dieselben. Diese Beziehung zwischen Doppelnebenklassen und Bahnen ist die folgende:

$$\Theta(\bar{G}\pi(S_X)_f) = G(f \circ \pi^{-1}).$$

Einen alternativen (und allgemeineren) Zugang zu den Beziehungen zwischen Bahnen und Doppelnebenklassen liefert der folgende Satz ([11]):

9.8 Satz Operiert G transitiv auf M , und ist H eine beliebige Untergruppe von G , m ein beliebiges Element von M und G_m der Stabilisator von m , so besteht eine umkehrbar eindeutige Beziehung zwischen der Menge $H \backslash G / G_m$ der Doppelnebenklassen und der Menge $H \backslash M$ der Bahnen von H auf M . Ist $T \subseteq G$ eine Transversale der Doppelnebenklassen, so ist $T_m := \{gm \mid g \in T\}$ eine Transversale der Bahnen von H auf M .

Man kann also insbesondere aus Repräsentantensystemen von Doppelnebenklassen auch Repräsentantensysteme von Bahnen gewinnen. Auf diese Weise kann man zum Beispiel m -Graphen konstruieren, übrigens auch Permutationsisomere, für die dieses Lemma ursprünglich hergeleitet worden ist. Da im Mittelpunkt von MOLGEN die Bindungsisomere und nicht die Permutationsisomere stehen, wollen wir hier jedoch keine weiteren Einzelheiten zur Konstruktion von Permutationsisomeren diskutieren, dies soll später aber nachgeholt werden. Hier sei nur soviel dazu gesagt: Die Menge X ist bei der Bestimmung der Permutationsisomere die Menge der Gerüstplätze, die Gruppe G ist die Symmetriegruppe des Molekülgerüsts, und die Menge Y ist die Menge der Ligandensorten. Viele nützliche Hinweise auf konkrete Verwendung dieser Methoden finden sich in [9].

10 Die Konstruktion der Bindungsisomere

Wir wollen jetzt die in den vorausgegangenen Paragraphen beschriebenen Techniken auf unser Problem anwenden, Bindungsisomere zu konstruieren. Als Beispiel betrachten wir die empirische Formel:



Wir verlangen dabei, daß jedes der gesuchten Isomere das in Abb. 6 skizzierte Teilmolekül B mindestens dreifach enthält – B wird also zu einem dreifachen "Makroatom". Da B die Summenformel C_6H_5 hat, reduziert sich die Ausgangsformel zu



wobei wir als neue "Atomsorte" B eingeführt haben. Nach dieser ersten Reduktion des Ausgangsproblems bleibt also die Ermittlung aller Bindungsisomere zur Summenformel 10.1, also die Bestimmung aller entsprechend beschrifteten Valenzstrichgraphen hierzu!

Wir lösen dieses Problem deshalb in zwei Schritten:

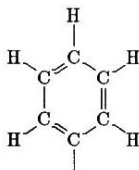


Abb. 6: Teilmolekül B

- In einer ersten Vergrößerung gemäß Homomorphieprinzip werden die zusammenhängenden Multigraphen zu dieser Summenformel bestimmt.
- In einem zweiten Schritt werden dann die Punkte mit Atomnamen "gefärbt".

Bei beiden Problemen werden Doppelnebenklassenalgorithmien benutzt.

10.1 Superatome

Betrachten wir jetzt das erste der beiden Probleme näher, das auf dem Weg zur Berechnung aller Bindungsisomere zu gegebener Summenformel 10.1 gelöst werden muß: die Bestimmung der Isomorphietypen von zusammenhängenden Multigraphen, deren Kantengradfolge sich aus den Wertigkeiten der vorgeschriebenen Atome ergibt. Wir haben die Valenzliste (Kantengradfolge)

$$V = (12, 1, 0, 13)$$

zu betrachten. Der erste Vergrößerungsschritt orientiert sich an der folgenden Vorgehensweise:

Ein Multigraph wird in *zyklische* Strukturen und einen *nichtzyklischen Rest* zerlegt, wobei diese Bestandteile baumartig miteinander verbunden sind. Die zyklischen Strukturen heißen *Superatome*, Punkte vom Grad 1 gehören stets zum nichtzyklischen Rest.

Der Begriff des Superatoms ist deshalb für die nachfolgende Beschreibung des Konstruktionsprozesses von zentraler Bedeutung. Er wurde bereits bei der Konstruktion von DENDRAL verwendet und ist wie folgt definiert:

10.2 Definition Ein Superatom ist ein zusammenhängender Graph, der brückenfrei ist (d. h. es gibt in ihm keine Kante, deren Wegnahme zum Zerfall des Graphen führt), und der eine freie Valenz besitzt (so etwas gibt es, genau genommen, in der Graphentheorie eigentlich nicht, aber es ist dem an den Anwendungen in der Chemie interessierten Leser klar, was gemeint wird). In der Sprechweise der Chemie ist dies

also ein brückenfreies Molekül mit mindestens einer freien Valenz, dessen sämtliche Atome eine Valenz ≥ 2 haben.

Die erste Aufgabe besteht also darin, alle überhaupt möglichen Superatome herauszufinden, die zu der vorgegebenen Summenformel passen. Diese Superatome werden natürlich im allgemeinen nicht alle Punkte und Kanten enthalten, die in der Summenformel bzw. den entsprechenden Graphen vorkommen. Diese überzähligen Punkte und Kanten sind dann später hinzuzufügen, wobei die überzähligen Kanten an die Superatome angefügt werden müssen, so daß sie zu nichtzyklischen Substrukturen, also *nicht* zu Superatomen, sondern zu *baumartigen Teilstrukturen gehören*.

Beim Vergrößern werden also zunächst die Kanten vergessen, welche die baumartige Verbindungsstruktur darstellen. Im späteren Konstruktionsprozeß sind diese wieder hinzuzufügen: jede Kombination geeigneter Superatome mit einem nichtzyklischen Rest kann in der Regel auf *mehrere* Weisen zu baumartigen Verbindungsstrukturen kombiniert werden.

Um nun die zyklischen Bestandteile aus der gegebenen Summenformel $C_{13}H_9OB_3$ zu ermitteln, haben wir als Information nur die Valenzenfolge $V = (12, 1, 0, 13)$. Da bei gegebenen p Punkten genau $p-1$ Kanten für eine baumartige, d. h. zykelfreie Verbindung aller Punkte nötig sind, müssen — wenn mehr als $p-1$ Kanten untergebracht werden sollen — alle übrigen Kanten entweder Zykel schließen, oder Kanten der baumartigen Struktur zu Mehrfachkanten machen. *Da Mehrfachkanten als zyklische Strukturen aufgefaßt werden, müssen alle diese übrigen Kanten also in zyklischen Substrukturen liegen!*

In unserem Beispiel besteht jeder molekulare Graph zu $C_{13}H_9OB_3$ aus 26 Punkten und 33 Kanten. Es sind also 7 Kanten mehr als Punkte vorhanden, insgesamt sind also 8 Kanten "überzählig". Diese Anzahl $U(V)$ heißt, wie bereits oben erwähnt, *zyklomatische Zahl*, und man kann sie wie folgt leicht aus der Valenzenfolge bestimmen: Ist $V = (v_1, \dots, v_n)$ die Valenzenfolge eines Multigraphen (v_i ist also gleich der Anzahl der Punkte vom Grad i), dann errechnet sich die zyklomatische Zahl wie folgt:

$$U(V) := \frac{1}{2} \left(\sum_{i=1}^n (i-2)v_i + 2 \right).$$

Es gilt also beispielsweise

$$U(12, 1, 0, 13) = 8.$$

Es müssen nun alle Kombinationen (Partitionen) in Superatome und einen nichtzyklischen Rest berechnet werden. (In unserem Beispiel werden die Superatome jeweils aus einigen der Kohlenstoffatome, eventuell zusammen mit dem Sauerstoffatom bestehen.) Genauer zerlegt man zunächst die Valenzenfolge in *Superatomfolgen* (diese bestimmen insbesondere die Größe der Superatome) *Restfolgen*. Zum Beispiel

$$(12, 1, 0, 13) = (0, 1, 0, 13) + (12, 0, 0, 0).$$

(Im allgemeinen werden bei solchen Zerlegungen natürlich *mehrere Superatomfolgen* vorhanden sein, aber nur *eine Restfolge*. Die Restfolge kann durchaus auch die Folge aus lauter Nullen sein. Ist sie aber von dieser Nullenfolge verschieden dann kann es sein, daß sie später ebenfalls zerlegt werden wird!) Der erste Summand auf der rechten Seite ist die Superatomfolge, der zweite die Restfolge. Das (in diesem Fall einzige) Superatom enthält alle 13 Kohlenstoffatome und das Sauerstoffatom. Der nichtzyklische Rest besteht hier aus den 9 Wasserstoffatomen samt den 3 "Atomen" der Sorte *B*. Im allgemeinen Fall wird entsprechend *V auf alle möglichen Weisen in eine Summe von Superatomfolgen und Restfolgen zerlegt*. Dabei sind die in Frage kommenden Summanden, also die eventuellen Superatomfolgen bzw. Restfolgen alle Folgen, die der Bedingung 8.1 genügen, und für die der Ausdruck 8.2 eine natürliche Zahl ist. Hinzukommt, daß für die Superatomfolgen *V* gelten muß: $v_1 = 0$ und $\sum_i v_i \geq 2$. Weitere solche Zerlegungen unserer Beispielfolge werden wir gleich angeben. Nach Zerlegung der ursprünglichen Valenzenfolge *V* in eine Summe von Valenzenfolgen V_i , wobei V_1, \dots, V_k Superatomfolgen seien, muß man eine wichtige Entscheidung über die Verwendung der jeweiligen Kanten in der Superatomfolge treffen, denn es muß ja ein Graph mit vorgegebener zyklomatischer Zahl $U(V)$ herauskommen. Sehen wir uns dies erneut an dem Beispiel an.

Die Restfolge ist hier von der Form $(12, 0, 0, 0)$, es handelt sich also um 12 Atome der Valenz 1 (Wasserstoffatome). Das hier einzige Superatom muß demnach genau 12 freie Valenzen haben und die zyklomatische Zahl 8. Im allgemeinen Fall wird $U(V)$ in k (=Anzahl der Superatome) natürliche Zahlen u_1, \dots, u_k zerlegt. Hat man eine solche Zerlegung gefunden, dann bleibt nachzuprüfen, ob es Superatome gibt mit Valenzenfolge V_i und zyklomatischer Zahl u_i . Aus diesen beiden Bedingungen ergibt sich auch die Anzahl der freien Valenzen des jeweiligen Superatoms, und es ist noch zu ermitteln, ob diese freien Valenzen mit der Restfolge oder mit eventuellen Zerlegungen dieser, sowie mit den anderen Superatomen zusammenpassen.

In diesem Partitionierungsprozeß verteilt man die $U(V)$ überzähligen Kanten auf die Superatome, d.h. jedes Superatom enthält mindestens eine dieser zyklischschließenden Kanten. In unserem Beispiel werden also 8 solche Kanten dem einzigen Superatom zugewiesen. Wir müssen also jetzt *alle Superatomfolgen finden, deren Summe $(0, 1, 0, 13)$ ist*.

Da ein Superatom aus mindestens zwei Atomen bestehen muß und darin nur Punkte mit einer Valenz von mindestens zwei vorkommen können, überlegt man sich leicht, daß es im Beispiel höchstens 7 Superatome geben kann.

Wir greifen jetzt einmal 2 Superatompartitionen mit jeweils 5 Superatomfolgen heraus:

	zykl. Zahl	Superatomfolge
Superatome	$u_1 = 3$	$V_1 = (0, 0, 0, 6)$
	$u_2 = 2$	$V_2 = (0, 1, 0, 1)$
	$u_3 = 1$	$V_3 = (0, 0, 0, 2)$
	$u_4 = 1$	$V_4 = (0, 0, 0, 2)$
	$u_5 = 1$	$V_5 = (0, 0, 0, 2)$
Rest	-	$= (12, 0, 0, 0)$
Σ	$U = 8$	$V = (12, 1, 0, 13)$

	zykl. Zahl	Superatomfolge
Superatome	$u_1 = 2$	$V_1 = (0, 0, 0, 2)$
	$u_2 = 2$	$V_2 = (0, 0, 0, 2)$
	$u_3 = 2$	$V_3 = (0, 0, 0, 2)$
	$u_4 = 1$	$V_4 = (0, 0, 0, 2)$
	$u_5 = 1$	$V_5 = (0, 0, 0, 2)$
Rest	-	$= (12, 1, 0, 3)$
Σ	$U = 8$	$V = (12, 1, 0, 13)$

Insgesamt gibt es in 185 wesentlich verschiedene derartige Partitionen mit 5 Superatomen. (Für einen einzigen Topf gibt es 12 verschiedene Partitionen, für 2 Töpfe sind es 210 Partitionen, für 3 Töpfe erhalten wir 567 Partitionen, für 6 Töpfe 25 Partitionen, und schließlich noch 1 einzige Partition, wenn 7 Superatomtöpfe vorhanden sind.) Damit man die gesuchten Superatome erhält, sind jeweils zu jeder Superatomvalenzfolge die zugehörigen Multigraphen zu *konstruieren*. Ein vollständige Liste solcher Graphen heißt *Superatomtopf*.

10.2 Die Superatomtöpfe

Ein Superatomtopf ist also eine komplette Liste von Superatomen zu einer vorgegebenen Superatomfolge V_i . Wir erinnern daran, daß diese Folge V_i zu einer Partition der gegebenen Valenzfolge V in Superatomfolgen V_1, \dots, V_k und eine Restfolge gehört. Mit dieser Partition in Superatomfolgen und Restfolge ist gleichzeitig auch noch eine Partition der zyklomatischen Zahl $U(V)$ in Summanden u_1, \dots, u_k vorgegeben. Wir müssen also zu jedem (u_i, V_i) je einen Superatomtopf SA_i konstruieren, der *aus allen nichtisomorphen zusammenhängenden und brückenfreien Multigraphen besteht mit:*

- V_i als Gradliste,
- $FV_i = 2(U(V_i) - u_i)$ freien Valenzen (interpretiert als Punkte vom Grad 1),
- sowie u_i als Zusammenhangszahl (freie Valenzen eingerechnet).

Die FV_i freien Valenzen werden für den späteren Zusammenbau der zyklischen Bestandteile zu Bäumen benötigt.

Wir wollen uns nun auf die Bestimmung *eines* Superatomtopfs konzentrieren – dieses Teilproblem ist das Herzstück des gesamten Erzeugungsprozesses. Der entscheidende Schritt ist hier, bei vorgegebenem V_i und u_i , die FV_i freien Valenzen auf alle möglichen Weisen auf die $\sum_j j \cdots v_j = p$ Punkte des zukünftigen Superatoms zu verteilen. Dabei können einem Punkt der (späteren) Valenz k natürlich höchstens $k - 2$ freie Valenzen zuteilen. Danach führen wir einen wichtigen Reduktionsprozeß durch. Hier kommen unsere mathematischen Prinzipien voll zum Tragen: wir vergrößern anhand der gegebenen Gradfolge die gesuchten Graphen, indem wir Punkte vom Grad 1 (freie Valenzen der Superatome) und sukzessive auch Punkte vom Grad 2 “vergessen”. Dieses Vergessen ist natürlich genau geregelt und macht einen systematischen Reduktionsprozeß notwendig. Dabei entsteht aus der ursprünglichen Folge V_i eine Sequenz von Folgen $V'_i, V''_i, V'''_i, \dots, V_i^{(m)}$, wobei das Endprodukt $V_i^{(m)}$ nur noch Punkte vom Grad mindestens 3 enthält. Die zugehörigen Multigraphen werden dem sogenannten KATALOG entnommen, d. h. sie sind in vorberechneter Form in einer Datei bereits vorhanden. Diese Katalogstrukturen bilden die “Keime” der Superatome, sie werden anhand der vorher gemachten Reduktionsschritte wieder zu den gesuchten Superatomen “aufgeblasen”. Das Aufblasen geschieht wiederum nach dem Homomorphieprinzip:

Die KATALOG-Graphen werden zunächst unter Berücksichtigung ihrer Automorphismengruppen auf alle möglichen Weisen mit Punkten vom Grad 2 gefärbt (Doppelnebenklassenalgorithm). Praktisch bedeutet dies:

- *Kanten* werden mit Punkten vom Grad 2 belegt,
- *Punkte* werden mit Schleifen versehen und die Schleifen mit Punkten vom Grad 2 belegt.

Dieses Verfahren kann sich, je nach vorheriger Reduktion, auch *mehrfach* wiederholen, und man erhält in allen Zwischenstufen stets *größere* Graphen.

Die neuentstandenen Graphen färbt man, ebenfalls in Abhängigkeit von den vorherigen Reduktionen und ihren Automorphismengruppen, mit den benötigten freien Valenzen, auch dies geschieht mittels Doppelnebenklassenalgorithm.

In einigen Spezialfällen kann auf den KATALOG-Zugriff verzichtet werden – darauf wollen wir aber nicht im Detail eingehen.

10.3 Der Katalog

Da der KATALOG die Leistungsfähigkeit des Programms bestimmt, muß er möglichst viele Fälle vollständig abdecken. Wir haben ihn wiederum durch Vergrößerung erhal-



Abb. 7: Eine KATALOG-Struktur

ten: aus *Multigraphen* werden durch das Vergessen der Kantenvielfachheiten *schlichte* Graphen. Umgekehrt erhält man daher aus den schlichten Graphen durch die Belegung der Kanten mit Vielfachheiten die Multigraphen für den KATALOG. Wiederum ist für die Isomorphieklassifizierung jeweils nur die Automorphismengruppe zu berücksichtigen. Die Belegung mit Vielfachheiten kann auch hier als Doppelnenklassenproblem betrachtet werden (vgl. [18]). Der derzeitige KATALOG-Bestand beträgt ca. 10^6 Strukturen. Er ist zum einen gegliedert nach Punktezahl p und *maximaler* Punktevalenz k einer Struktur. Eine weitere Untergliederung ist dann die jeweilige Gradliste $(0, 0, v_3, \dots, v_k)$, wobei gilt:

$$\sum_{i=3}^k v_i = p.$$

Die erste Untergliederung wird durch folgende Tabelle illustriert:

Punktezahl	2	3	4	5	6	7	8	9	10	12	14
max. Valenz	11	11	11	10	7	6	5	4	4	3	3
Anz. der Graphen	9	72	2646	76338	99919	344891	275927	20966	167893	365	2602

Bemerkung: die höchste in der Chemie bekannte Valenz eines Atoms beträgt 8, die Verwendung noch höherer Valenzen im KATALOG kann aber dennoch sinnvoll sein. Diese können bei Vorgabe komplexer Teilmoleküle in einer Formel auftreten, etwa einer *Makrokomponente* mit 10 freien Valenzen.

Wir illustrieren das eben Gesagte wieder anhand unseres Beispiels und betrachten aus der ersten der beiden oben angegebenen Superatompartitionen die Valenzenfolge

$$(0, 0, 0, 6).$$

Der systematische Reduktionsprozess führt uns zu folgenden reduzierten Gradlisten:

$$(0, 0, 0, 2), \quad (0, 0, 2, 1), \quad (0, 0, 2), \\ (0, 0, 4), \quad (0, 0, 0, 2)$$

Zu jeder Gradliste sind nun dem Katalog alle zugehörigen Strukturen zu entnehmen. Um das Beispiel überschaubar zu halten, beschränken wir uns auf einen einzigen, einfachen Fall: wir greifen die Gradliste $(0, 0, 2)$ heraus – hierzu gibt es genau *eine* KATALOG-Struktur (Abb. 7). Wir wollen uns nun ansehen, wie aus dieser KATALOG-Struktur *ein* Superatom mit 6 Punkten wird. In 3 Schritten werden 4

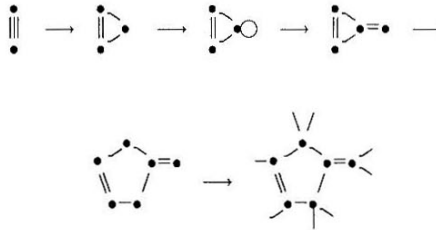


Abb. 8: Die Färbung mit Punkten vom Grad 2 und mit freien Valenzen

Punkte vom Grad 2 angebracht und am Schluß 8 freie Valenzen – diesen “Färbeprozess” zeigt Abb. 8. Damit ist *ein* gesuchtes Superatom konstruiert. Berechnet man zu *allen* KATALOG-Strukturen *jeweils alle* Färbemöglichkeiten mit Punkten vom Grad 1 und 2, so ergeben sich 36 Superatome zur Gradliste $(0, 0, 0, 6)$, also besteht SA_1 aus 36 Repräsentanten.

Die Berechnung von SA_2, \dots, SA_5 ist viel weniger aufwendig – es gibt jeweils *genau einen* Repräsentanten. Damit können wir an dieser Stelle die Kernphase des Erzeugungsprozesses wieder verlassen – zu einer Superatompartition mit 5 Valenzenfolgen haben wir die zugehörigen Superatomtöpfe SA_1, \dots, SA_5 ermittelt.

Im nächsten Verarbeitungsschritt benötigen wir alle *Mengenkombinationen* von Graphen G_i aus den SA_i von der Form: $\{G_1, G_2, G_3, G_4, G_5\}$. Eine solche Kombination zeigt Abb. 9. Nun machen wir erneut einen Vergrößerungsschritt rückgängig: zusammen mit den verbliebenen 12 Atomen aus dem nichtzyklischen Rest wird unsere Superatomkombination *auf alle möglichen* Weisen zu Bäumen verarbeitet – auch hierfür sei in Abb. 10 *ein* Beispiel angegeben. Das “Zusammensetzen” geschieht einmal mehr mit Hilfe von Doppelnebenklassen. Bei der vorliegenden Auswahl von Superatomen gibt es insgesamt bereits 1341(!) verschiedene Möglichkeiten. Ein solcher Baum bildet schließlich das Grundgerüst (Skelett) eines Moleküls und wir sind wieder am Ausgangspunkt angelangt.

Es sei noch erwähnt, daß für den Baum aus Abb. 10 genau 138 verschiedene Belegungsmöglichkeiten mit *Atomnamen* existieren – davon wollen wir noch 2 Möglichkeiten angeben (s. Abb. 11). Einen Gesamtüberblick über unseren Weg durch den “Lösungsbaum” und damit durch die wesentlichen Stationen des Algorithmus gibt uns zum Abschluß noch Abb. 12. Im Anschluß an der Generierungsprozess können alle GOODLIST-Strukturen in einem Molekül wieder eingebettet (*expandiert*) werden, d.h. die künstlichen Atome werden durch die kompletten Teilstrukturen ersetzt. Im allgemeinen ist dieses Einbetten mathematisch auf *mehrere* Weisen möglich.

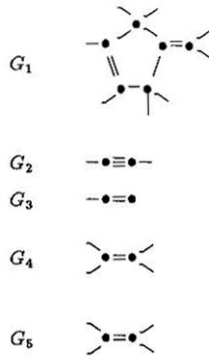


Abb. 9: Eine Kombination von Superatomen

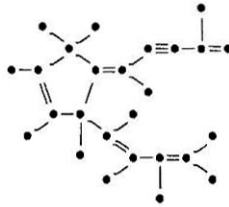


Abb. 10: Eine Zusammensetzung der Superatome

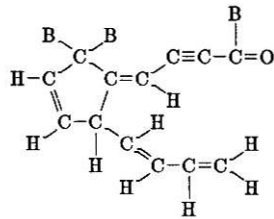
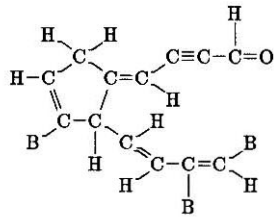


Abb. 11: Belegungen mit Atomnamen

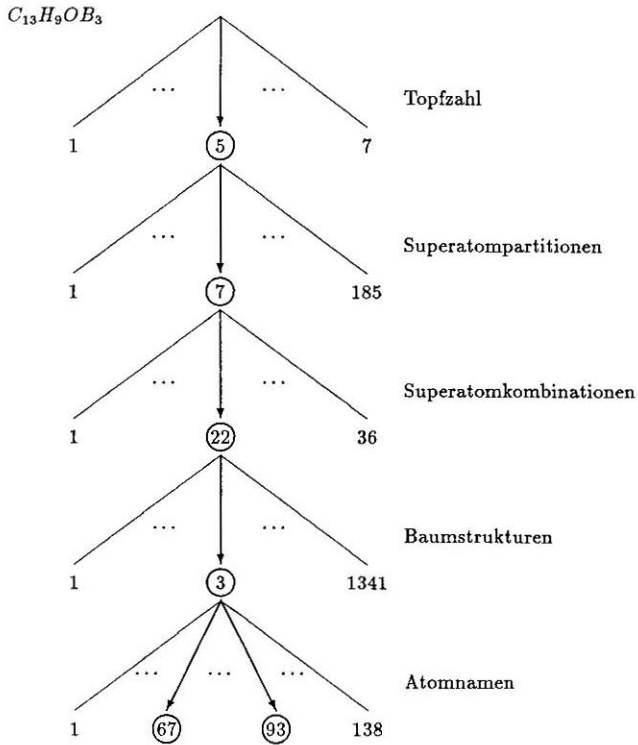


Abb. 12: Lösungsbaum – ein Pfad

In unserem Fall ist das Expandieren einfach: das *B*-Atom hat lediglich die Valenz 1 und kann so nur auf eine einzige Weise eingebettet werden. Die Einbettung in das erste Molekül aus Abb. 11 zeigt schließlich noch Abb. 13 (die H-Atome wurden weggelassen).

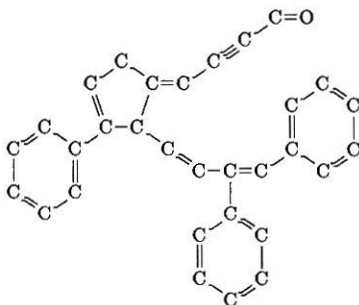


Abb. 13: Einbettung des Teilmoleküls *B*

11 Einige Tabellen

Hier sind einige Ergebnisse von MOLGEN tabelliert. Es sind dies Tabellen von Isomerenanzahlen, dabei ist aber vor allem zu beachten, daß diesen Tabellen nicht Anzahlbestimmungen, sondern *Konstruktionen* der entsprechenden molekularen Graphen zugrunde liegen! Diese Tabellen wurden auf einer Maschine vom Typ DG-UNIX 88000/Risc berechnet. Der Zeilenindex *i* bezeichnet die jeweilige Anzahl von *C*-Atomen, der Spaltenindex *j* die Anzahl von *H*-Atomen.

C_iH_j	0	2	4	6	8	10	12	14	16	18	20	22
0	—	1 0.0	—	—	—	—	—	—	—	—	—	—
1	—	—	1 0.0	—	—	—	—	—	—	—	—	—
2	1 0.0	1 0.0	1 0.0	1 0.0	—	—	—	—	—	—	—	—
3	1 0.0	2 0.0	3 0.0	2 0.0	1 0.0	—	—	—	—	—	—	—
4	3 0.0	7 0.0	11 0.0	9 0.0	5 0.0	2 0.0	—	—	—	—	—	—
5	6 0.0	21 0.0	40 0.0	40 0.1	26 0.0	10 0.0	3 0.0	—	—	—	—	—
6	19 0.0	85 0.1	185 0.2	217 0.2	159 0.2	77 0.0	25 0.0	5 0.0	—	—	—	—
7	50 0.0	356 0.4	920 1.2	1230 1.4	1031 0.9	575 0.4	222 0.1	56 0.0	9 0.0	—	—	—
8	204 0.2	1804 2.2	5308 6.8	7982 8.7	7437 6.5	4679 3.1	2082 1.1	654 0.3	139 0.0	18 0.0	—	—
9	832 0.8	10064 12.6	33860 44.2	56437 1:03.5	57771 50.2	40139 26.4	19983 10.3	7244 3.2	1902 0.7	338 0.1	35 0.0	—
10	4330 4.8	64352 1:24.0	241297 5:35.1	439373 9:01.0	488125 7:45.8	369067 4:20.6	201578 1:46.7	81909 34.4	24938 9.1	5568 1.9	852 0.3	75 0.0

C_iH_jO	0	2	4	6	8	10	12	14	16	18	20	22
0	—	1 0.0	—	—	—	—	—	—	—	—	—	—
1	—	1 0.0	1 0.0	—	—	—	—	—	—	—	—	—
2	1 0.0	3 0.0	3 0.0	2 0.0	—	—	—	—	—	—	—	—
3	2 0.0	9 0.0	13 0.0	9 0.0	3 0.0	—	—	—	—	—	—	—
4	7 0.0	36 0.1	62 0.0	55 0.0	26 0.0	7 0.0	—	—	—	—	—	—
5	21 0.0	151 0.2	318 0.4	337 0.3	205 0.1	74 0.0	14 0.0	—	—	—	—	—
6	85 0.2	738 1.0	1823 1.8	2237 1.6	1623 0.9	747 0.3	211 0.1	32 0.0	—	—	—	—
7	356 0.6	3971 5.4	11332 10.6	15804 10.3	13177 6.5	7166 3.0	2589 0.9	596 0.2	72 0.0	—	—	—
8	1804 3.2	24021 33.5	77431 1:12.7	120427 1:16.7	112484 53.1	69669 26.7	29797 10.1	8796 2.6	1684 0.5	171 0.0	—	—
9	10064 18.2	160114 3:58.8	575884 9:44.3	985744 11:06.7	1013745 8:12.8	697708 4:27.8	338761 1:51.2	118215 35.7	29172 8.4	4745 1.4	405 0.0	—
10	64352 2:00.4	1175685 31:55.7	4654419 1:27:37	8671508 1:48:30	9693195 1:25:29	7288733 49:29.9	3916111 22:13.1	1548361 7:51.5	452458 2:10.6	95312 27.1	13372 3.8	989 0.1

C_iH_jN	1	3	5	7	9	11	13	15	17	19	21
0	—	1 0.0	—	—	—	—	—	—	—	—	—
1	1 0.0	1 0.0	1 0.0	—	—	—	—	—	—	—	—
2	2 0.0	5 0.0	4 0.0	2 0.0	—	—	—	—	—	—	—
3	7 0.0	19 0.0	21 0.0	12 0.0	4 0.0	—	—	—	—	—	—
4	27 0.0	87 0.1	116 0.2	85 0.0	35 0.0	8 0.0	—	—	—	—	—
5	112 0.1	437 0.5	685 0.6	593 0.5	313 0.1	100 0.0	17 0.0	—	—	—	—
6	540 0.6	2447 2.4	4394 3.4	4378 2.7	2732 1.3	1111 0.4	284 0.1	39 0.0	—	—	—
7	2879 2.6	15052 13.3	30478 21.8	34152 19.1	24314 11.0	11673 4.5	3809 1.2	801 0.2	89 0.0	—	—
8	17198 15.1	102012 1:27.9	229260 2:41.3	284065 2:34.8	225296 1:37.8	122819 44.2	47323 15.1	12770 3.6	2258 0.6	211 0.0	—
9	113702 1:41.6	755600 11:05.0	1863935 22:47.4	2521767 23:59.0	2190926 16:20.8	1323028 8:05.4	577485 3:02.8	184124 53.2	41989 11.3	6355 1.7	507 0.0

C_iH_jNO	1	3	5	7	9	11	13	15	17	19
0	1 0.0	1 0.0	—	—	—	—	—	—	—	—
1	3 0.0	5 0.0	3 0.0	—	—	—	—	—	—	—
2	11 0.0	26 0.0	22 0.0	8 0.0	—	—	—	—	—	—
3	46 0.0	136 0.1	154 0.1	84 0.0	21 0.0	—	—	—	—	—
4	216 0.3	775 0.8	1069 0.7	764 0.3	299 0.1	56 0.0	—	—	—	—
5	1111 1.3	4738 3.7	7687 4.4	6637 2.7	3390 1.1	1015 0.3	149 0.0	—	—	—
6	6340 6.7	31325 21.1	58218 28.2	58265 22.3	35759 11.3	13982 3.8	3345 0.8	398 0.0	—	—
7	39727 40.6	223890 2:25.3	467617 3:37.6	528227 3:12.4	372937 1:54.7	174763 47.7	54700 13.6	10777 2.5	1068 0.1	—
8	273106 4:46.0	1729030 19:08.9	3999703 31:45.3	5005355 30:58.0	3955938 20:37.8	2123287 9:49.2	7795607 3:23.9	205672 49.5	34156 7.9	2876 0.3

Literaturverzeichnis

- [1] A. T. Balaban, Chemical Graphs, XVIII, *Revue Roumaine de Chimie* **18** (1973), 635-653.
- [2] H. Beck, Alexander von Humboldt I,II, *Franz Steiner Verlag* 1959/1961.
- [3] N. L. Biggs, K. E. Lloyd, R. J. Wilson, Graph theory 1736-1936, *Clarendon Press*, 1977.
- [4] H. Brown, L. Hjelmeland, L. Masinter, Constructive Graph Labeling Using Double Cosets, *Discrete Math.* **7** (1974), 1-30.
- [5] H. Brown, L. Masinter, An Algorithm for the Construction of the Graphs of Organic Molecules, *Discrete Math.* **8** (1974), 227-244.
- [6] K. Bruhns (Herausgeber), Alexander von Humboldt, Biographie in drei Bänden, *Brockhaus* 1872.
- [7] R. E. Carhart, D. H. Smith, N. A. Gray, J. G. Nourse, C. Djerassi, GENOA: A Computer Program for Structure Elucidation Utilizing Overlapping and Alternative Substructures, *J. Org. Chem.*, **46** (1981) 1708-1718.
- [8] J. L. Gay-Lussac, *Ann. Chim. Phys.*, **27** (1824) 197.
- [9] R. Grund, Symmetrieklassen von Abbildungen und die Konstruktion von diskreten Strukturen, *Bayreuther Math. Schriften*, **31** (1990), 19-54.
- [10] R. Grund, R. Hager, F. Herrmann, A. Kerber, R. Laue, W. Weber, Groups, Graphs and Isomers, in: J. Gmehling (Ed.): *Software Development in Chemistry* 5, Springer-Verlag Berlin Heidelberg 1991.
- [11] W. Hässelbarth, On the interrelation between orbits and double cosets, *Theor. Chim. Acta* **67** (1985), 427-437.
- [12] R. Hager, A. Kerber, R. Laue, D. Moser, W. Weber, Construction of orbit representatives. *Bayreuther Math. Schriften* **35** (1991), 157-169.
- [13] F. Harary, E. Palmer, Graphical enumeration, Academic Press, 1973.
- [14] A. von Humboldt, Versuche über die gereizte Muskel- und Nervenfasern, nebst Vermutungen über den chemischen Prozeß in der Tier- und Pflanzenwelt, *Rottmann, Leipzig* 1797.
- [15] A. Kerber, Algebraic combinatorics via finite group actions, *BI-Wissenschaftsverlag*, 1991.
- [16] A. Kerber, R. Laue, D. Moser, Ein Strukturgenerator für molekulare Graphen, *Analytica Chimica Acta* **235** (1990), 221-228.
- [17] D. König, Theorie der endlichen und unendlichen Graphen, Nachdruck, Chelsea Publishing Company, 1969.

- [18] R. Laue, Computing double coset representatives for the generation of solvable groups, *Lecture Notes in Computer Science* **144**, Springer-Verlag 1982.
- [19] R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, J. Lederberg, Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project, McGraw-Hill, New York, 1980.
- [20] E. O. von Lippmann, Alexander von Humboldt als Vorläufer der Lehre von der Isomerie, *Chemiker-Zeitung* **1909**, Nr. 1, 1-8.
- [21] H.-J. Luinge, EXSPEC a Knowledge-Based System for Structure Analysis of Organic Molecules from Combined Spectral Data, Netherlands, 1989.
- [22] A. C. Lunn, J. K. Senior, Isomerism and Configuration, *Journal of physical chemistry*, **33** (1929) 1027-1079.
- [23] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Sci. Math.* **68** (1937), 145-254.
- [24] E. Ruch, W. Hässelbarth, B. Richter, Doppelnebenklassen als Klassenbegriff und Nomenklaturprinzip für Isomere und ihre Abzählung, *Theor. Chimica Acta* **19** (1970), 288-300.
- [25] E. Ruch, D. J. Klein, Double Cosets in Chemistry and Physics, *Theor. Chimica Acta* **63** (1983), 447-472.
- [26] B. Schmalz, Verwendung von Untergruppenleitern zur Bestimmung von Doppelnebenklassen, *Bayreuther Math. Schriften*, **31** (1990), 109-143.
- [27] J. J. Sylvester, Chemistry and Algebra, *Nature* **17** (1877), 103-104.