

Repräsentation von Molekülstrukturen zur computergestützten Behandlung chemischer Reaktionen

Den Naturwissenschaftlichen Fakultäten
der Friedrich-Alexander-Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades

vorgelegt von

Susanne Bauerschmidt

aus München

Als Dissertation genehmigt von den
Naturwissenschaftlichen Fakultäten der Universität Erlangen-Nürnberg

| | |
|--|------------------------|
| Tag der mündlichen Prüfung: | 16. Mai 1997 |
| Vorsitzender der Promotionskommission: | Prof. Dr. D. Kölzow |
| Erstberichterstatter: | Prof. Dr. J. Gasteiger |
| Zweitberichterstatter: | Prof. Dr. P. Otto |

Meinem Doktorvater

Herrn Prof. Dr. Johann Gasteiger

danke ich für die vielen anregenden Diskussionen und die vielfältige Unterstützung, ohne die die vorliegende Arbeit nicht möglich gewesen wäre.

Mein besonderer Dank gilt:

- Robert Höllering und Dr. Klaus-Peter Schulz für die vielen fruchtbaren Diskussionen bei der gemeinsamen Arbeit.
- Dr. Bruno Bienfait für die Hilfestellung bei Fragen zu objektorientiertem Programmieren.
- Den Systemadministratoren Paul Selzer und Andreas Teckentrup für die Bereitstellung einer stabilen Hard- und Software-Umgebung.
- Herrn Dr. Markus Wagener für seine Geduld, Diskussionsbereitschaft und Unterstützung beim Abfassen dieser Arbeit.

Allen weiteren Mitarbeitern des Arbeitskreises möchte ich für die herzliche Arbeitsatmosphäre danken.

Für finanzielle Unterstützung danke ich dem Bundesminister für Bildung, Wissenschaft, Forschung und Technologie.

für Markus, Friedel und Bruno

| | |
|---|-----------|
| 1. Einleitung | 1 |
| 2. Analyse von massenspektroskopischen Zerfallsschemata organischer Verbindungen | 5 |
| 2.1. Zielstellung | 5 |
| 2.2. Modellannahmen | 9 |
| 2.3. Arbeitsweise von FRANZ | 10 |
| 2.3.1. Einlesen der Eingabedateien und Vorselektion | 10 |
| 2.3.2. Aufbau des Reaktionsnetzwerks | 11 |
| 2.3.3. Analyse des Reaktionsnetzwerks | 12 |
| 2.4. Nachteile der in FRANZ verwendeten Strukturrepräsentation auf der Basis des Valence-Bond-Modells | 14 |
| 2.4.1. Fragmentierungsschemata von Ketonen am Beispiel von Butan-2-on | 14 |
| 2.4.2. Orthogonale π -Systeme | 18 |
| 2.4.3. σ -Ionisation | 20 |
| 3. Datenstrukturen zur Repräsentation von Molekülen | 21 |
| 3.1. Bindungslisten | 21 |
| 3.2. BE-Matrizen und ihre Erweiterungen | 23 |
| 3.3. Z-Matrizen | 24 |
| 3.4. Die Linearnotation SMILES | 25 |
| 4. RICOS - Ein allgemeines Modell zur Repräsentation von Molekülen | 27 |
| 4.1. Atome | 29 |
| 4.2. Elektronensysteme | 31 |
| 4.2.1. σ -Elektronensysteme | 34 |
| 4.2.2. π -Elektronensysteme | 37 |
| 4.2.3. Koordinative Bindungen | 42 |
| 4.3. Moleküle | 48 |
| 4.4. Aggregate | 49 |
| 4.5. Wechselwirkungen | 50 |
| 4.6. Ensembles | 51 |
| 4.7. Topologische Gruppen | 52 |
| 5. Physikochemische und topologische Eigenschaften | 55 |
| 5.1. Übersicht über die implementierten Eigenschaften | 55 |
| 5.2. Ladungen | 56 |
| 5.2.1. Äquilibrierte Ladung | 57 |

| | |
|--|------------|
| 5.2.2. Lokalisierte Ladungen in Molekülen | 58 |
| 5.3. Radikalzentren | 60 |
| 5.4. Hashcode | 61 |
| 5.4.1. Bestimmung der Seedparameter | 62 |
| 5.4.2. Atomhashcodes | 64 |
| 5.4.3. Elektronensystemhashcodes | 66 |
| 5.4.4. Molekül-, Aggregat- und Ensemblehashcodes | 67 |
| 5.4.5. Äquivalenzklassen | 68 |
| 5.5. Physikochemische Parameter auf der Basis von Valenz-Bond-Strukturen | 68 |
| 6. Reaktionen von Molekülen und Aggregaten in einem Ensemble | 73 |
| 6.1. Reaktionen von Elektronensystemen | 73 |
| 6.1.1. Elementarschritte zum Erzeugen und Brechen von σ -Elektronensystemen | 74 |
| 6.1.2. Elementarschritte zum Erzeugen und Brechen von π -Elektronensystemen | 77 |
| 6.1.3. Elementarschritte zum Erzeugen und Brechen von koordinativen Bindungen | 78 |
| 6.1.4. Änderung der Elektronenzahl eines Elektronensystems | 81 |
| 6.1.5. Zusammengesetzte Reaktionen | 82 |
| 6.2. Ändern von Wechselwirkungsgruppen und topologischen Gruppen während einer Reaktion | 83 |
| 6.3. Einfügen und Löschen von Molekülen und Aggregaten im Reaktionsensemble | 85 |
| 6.4. Erkennung von Reaktionsprodukten mit ungültigen Valenzen | 87 |
| 7. Konvertierung zwischen RICOS und dem VB-Modell | 91 |
| 7.1. Umwandlung von Molekülen mit VB-Repräsentation in RICOS | 91 |
| 7.2. Umwandlung von Molekülen in RICOS-Repräsentation in VB-Repräsentation | 95 |
| 7.2.1. Allgemeine Vorgehensweise | 95 |
| 7.2.2. Suche nach geeigneten Atomen zur Lokalisierung von Ladungen und freien Elektronen | 97 |
| 7.2.3. Suche nach Doppelbindungen | 101 |
| 8. Realisierung der RICOS-Repräsentation mit einem objektorientierten Modell | 111 |
| 8.1. Grundbegriffe | 111 |
| 8.2. Allgemeine Klassen | 114 |
| 8.3. Klassenhierarchie der Hauptklassen | 115 |
| 8.4. Chemische Klassen | 118 |

| | |
|---|------------|
| 8.4.1. Die Atomklasse | 119 |
| 8.4.2. Die Klassen für Elektronensysteme | 121 |
| 8.4.3. Die Molekülklasse | 121 |
| 8.4.4. Die Aggregatklasse | 121 |
| 8.4.5. Die Ensembleklasse | 121 |
| 8.4.6. Die Gruppenklasse | 122 |
| 8.4.7. Die Nachbarschaftsklasse | 122 |
| 8.5. Basisklassen | 123 |
| 8.5.1. Chemische Objekte | 123 |
| 8.5.2. Chemische Objekte mit Gruppen | 125 |
| 8.6. Repräsentation von Molekülen in VB-Darstellung | 126 |
| 8.7. Reaktionen | 127 |
| 9. EROS 7 | 131 |
| 9.1. Programmaufbau | 131 |
| 9.1.1. Aufbau des Reaktionsnetzwerks | 132 |
| 9.1.2. Reaktionsgenerierung und Reaktionsregeln | 132 |
| 9.2. Anwendungsbeispiele | 133 |
| 9.2.1. Reaktionen in der Massenspektrometrie | 133 |
| 9.2.2. Thermische Cycloadditionsreaktionen | 138 |
| 10. Zusammenfassung | 141 |
| 11. Literaturverzeichnis | 143 |
| 12. Anhang | 151 |
| 13. Publikationsliste | 161 |
| 14. Lebenslauf | 163 |

1. Einleitung

Seit der Einführung des Computers zur Problemlösung in den Naturwissenschaften befaßt sich auch die Chemie mit dessen Einsatz bei der Lösung chemischer Fragestellungen. Ein wichtiges Gebiet liegt in der Modellierung von Reaktionen. Corey und Wipke berichteten schon 1969 über den Einsatz des Computers in der Planung organischer Synthesen [1]. Seither wurde eine Vielzahl von Programmen entwickelt, die sich mit der Modellierung von Reaktionen organischer Verbindungen befassen. Die beiden großen Einsatzgebiete sind die Syntheseplanung und die Reaktionsvorhersage.

Wissensbasierte Syntheseplanungsprogramme arbeiten retrosynthetisch, sie suchen ausgehend vom Synthesziel leicht verfügbare Vorstufen. Sie stützen sich dabei auf eine Bibliothek von Synthesestrategien, anhand derer aus der Zielverbindung Vorstufen generiert werden. Diese können ebenfalls in weitere Vorstufen zerlegt werden, bis man geeignete Ausgangsmaterialien gefunden hat. Zu den wissensbasierten Programmen gehören unter anderem LHASA [2], SECS [3] und SYNCHEM [4]. Reaktionsvorhersageprogramme gehen in der synthetischen Richtung vor. Sie machen Voraussagen über die Produkte, die entstehen, wenn Edukte und Reaktionsbedingungen vorgegeben sind. Dabei werden nicht nur die Produktstrukturen, sondern auch deren Verteilungen vorhergesagt. CAMEO [5], [6] und EROS 6.0 [7] sind Reaktionsvorhersageprogramme, die anhand einer Wissensbasis die Reaktivität von Atomen und Bindungen abschätzen und daraus den Reaktionsablauf prognostizieren.

Alternativ zu den wissensbasierten Programmen entwickelten Ugi und Dugundji ein logisch-mathematisches Modell zur Behandlung der chemischen Konstitution, die Algebra der BE- und R-Matrizen [8]. BE-Matrizen enthalten die Bindungen und freien Elektronen einer chemischen Verbindung in Matrixform. R-Matrizen kodieren die Elektronenumordnungsprozesse einer Reaktion. Die Reaktionsgenerierung erfolgt also auf einer rein formalen Basis, ohne Einbeziehung von chemischem Wissen. Auf der Grundlage dieser Algebra entstanden Programme, die sowohl zur Syntheseplanung als auch zur Reaktionsvorhersage einsetzbar sind. Zu ihnen gehören CICLOPS [9], IGOR [10], RAIN [11] und frühere Versionen von EROS [12].

Ein weiteres Forschungsgebiet, das sich der Modellierung von Reaktionen bedient, ist die Interpretation und Vorhersage von Massenspektren organischer Verbindungen. Programme, die die Interpretation von Massenspektren einer organischen Verbindung unterstützen, ordnen den Signalen eines experimentellen Spektrums automatisch Fragmente zu, die durch Spaltung von Bindungen dieser Verbindung erhalten werden. Stellvertretend für solche Programme seien MS-Break [13], TIMS [14] und FRANZ [15] genannt. MASSIMO [16] ist ein Programm zur Vorhersage von Massenspektren, das die Anfrageverbindung anhand von formalen Reaktionsschemata in Fragmente spaltet und die Reaktionen bewertet. Aus den Massen der entstandenen Fragmente und der Bewertung der Reaktionen wird ein Massenspektrum berechnet.

Eine zentrale Rolle bei der Behandlung von chemischen Strukturen in Programmen, die Reaktionen modellieren, spielt die Strukturrepräsentation, auf der sie aufbauen. Denn sie ist sowohl für deren Leistungsfähigkeit als auch für deren Anwendungsbreite verantwortlich. Im allgemeinen werden intern Bindungslisten, die auf dem Valenz-Bond-Modell (VB-Modell) zur Beschreibung chemischer Strukturen beruhen, verwendet. Bindungslisten haben sich zur Standardform der Repräsentation chemischer Strukturen mit dem Computer entwickelt [17]. Eine Bindungsliste enthält eine Liste mit allen Atomen eines Moleküls und spezifiziert die Bindung zwischen zwei Atomen durch die Bindungsordnung. Für die Behandlung von chemischen Reaktionen ist auch die Angabe freier Elektronen an Atomen erforderlich [18].

Bindungslisten haben den Vorteil, daß sie einfach zu verwenden und zu interpretieren sind. Sie beschreiben jedoch jeweils nur eine einzige Valenz-Bond Struktur. In VB-Strukturen sind Elektronen durch die Bindungsordnungen implizit kodiert. Eine Einfachbindung enthält zwei Elektronen, in Mehrfachbindungen entspricht die Zahl der Elektronen dem Doppelten der Bindungsordnung. Verbindungen, die durch diese starre Verteilung von Elektronen in zweiatomigen Bindungen und freie Elektronen nicht beschrieben werden können, können auch mit einer Bindungsliste nicht beschrieben werden und sind damit den oben genannten Programmen nicht zugänglich. Diese Einschränkung trifft am stärksten auf Verbindungen mit Mehrzentrenbindungen und Komplexbindungen zu. Beispiele hierfür zeigt Abbildung 1-1. In den Dreizentrenbindungen von Diboran (Abbildung 1-1a) sind zwei Elektronen drei Atomen zugeordnet. Auch die koordinativen Bindungen zwischen Ligand und Metallatom in Komplexen wie Ferrocen (Abbildung 1-1b) lassen sich durch Einfach- oder Mehrfachbindungen nicht beschreiben. Desweiteren kann auch Delokalisation von π -Elektronen in organischen Verbindungen nicht adäquat beschrieben werden. In Furan ist eines der beiden freien Elektronenpaaren im aromatischen Ring delokalisiert, das andere steht orthogonal dazu. Sie werden jedoch in der VB-Struktur nicht unterschieden (Abbildung 1-1c). Mit VB-Strukturen können auch verschieden angeregte Zustände der gleichen Verbindung nicht differenziert werden. Carbene existieren in zwei verschiedenen Zuständen, als Singulettcarbene mit zwei gepaarten Elektronen und als Triplettcarbene mit zwei ungepaarten Elektronen. Nur die Zahl der freien Elektronen an einem Atom anzugeben, erlaubt es nicht, beide Zustände voneinander zu unterscheiden (Abbildung 1-1d).

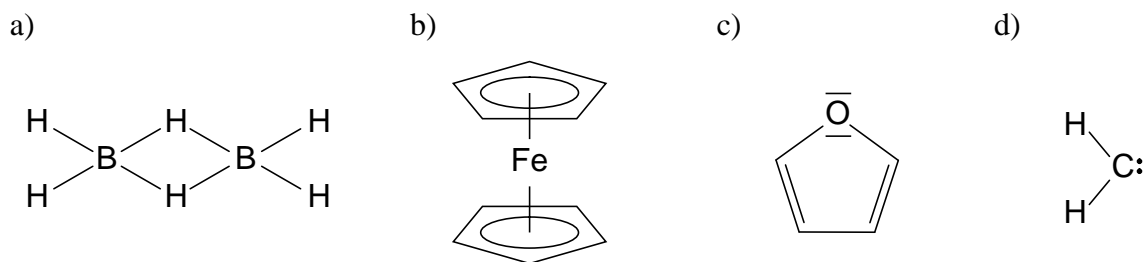


Abb. 1-1: Chemische Verbindungen, die durch eine VB-Struktur nicht eindeutig beschrieben werden können.

In unserer Arbeitsgruppe wurden die beiden Programme FRANZ und MASSIMO zur Analyse und Vorhersage von Massenspektren organischer Verbindungen entwickelt. Bei der Behandlung der dort auftretenden Reaktionen hat sich die Beschreibung von Molekülen in VB-Strukturen als nachteilig erwiesen, da die entstehenden Kationen und Radikalkationen einerseits häufig orthogonale π -Systeme enthalten und andererseits die Ladungen und radikalischen Elektronen in größeren π -Systemen delokalisiert sind. Somit reicht eine einzige VB-Struktur zu ihrer Beschreibung nicht aus. Um den Programmen alle möglichen VB-Strukturen zur Verfügung zu stellen, wurde ein Verfahren entwickelt, das aus einer VB-Struktur alle mesomeren Grenzformen generiert. Es kann jedoch die Orthogonalität von π -Systemen nicht berücksichtigen und erzeugt daher in einigen Fällen auch VB-Strukturen, die keine mesomeren Strukturen der vorgegebenen VB-Struktur sind. Dies führt zu Fehlern bei der Abschätzung der Reaktivität der involvierten Fragmente. Eine weitergehende Diskussion dieser Problematik findet sich in Kapitel 2.

Mit der vorliegenden Arbeit wird ein Modell zur Repräsentation von chemischen Verbindungen vorgestellt, das Konzepte zur Lösung der oben genannten Probleme bietet. Es basiert auf der Trennung der σ - und π -Elektronen und erlaubt Delokalisation über mehr als zwei Atome. Es enthält Konzepte zur Repräsentation von Mehrzentrenbindungen und koordinativen Bindungen in Komplexverbindungen. Desweiteren erlaubt es, Moleküle, die durch Wechselwirkungen, die nicht in Bindungen dargestellt werden können, miteinander verbunden sind, darzustellen. Das Modell integriert außer der Repräsentation der Konstitution auch Verfahren zur Ableitung von Eigenschaften aus der Konstitution und zur Generierung von Reaktionen. Die Anwendung der Strukturrepräsentation wird in einer neuen Version des Reaktionsvorhersageprogramms EROS, das parallel zu diesen Arbeiten von R. Höllering entwickelt wurde, gezeigt.

2. Analyse von massenspektroskopischen Zerfallsschemata organischer Verbindungen

In unserer Arbeitsgruppe wurden die Programme EROS6, FRANZ und MASSIMO entwickelt, die sich mit der Modellierung von Reaktionen in der Reaktionsvorhersage (EROS6, MASSIMO) und der Analyse von Massenspektren (FRANZ) befassen. Alle drei Programme verwenden intern die gleichen Module zur Repräsentation von Strukturen sowie zur Generierung von Reaktionen. Allein die Steuerung, wie das Reaktionsnetzwerk aufgebaut wird, unterscheidet sie voneinander. Stellvertretend für sie soll nun am Beispiel von FRANZ die verwendeten Methoden zur Reaktionsgenerierung beschrieben und die speziellen Probleme bei der Modellierung von Reaktionen in der Massenspektrometrie erläutert werden.

2.1. Zielstellung

Im Rahmen des Projekts zur Simulation von Massenspektren organischer Verbindungen wurde das Program FRANZ (Fragmentation and Rearrangement ANalyZer) von Schulz und Hanebeck entwickelt [15], [19]. Es ist ein Programm, mit dessen Hilfe die Zerfallsprozesse, die während der Aufnahme eines Massenspektrums ablaufen, analysiert werden können. Es verfolgt zwei Ziele.

Abbildung 2-1 zeigt das erste Ziel. Ausgehend vom experimentellen Massenspektrum und der Struktur einer organischen Verbindung wird anhand von formalen Reaktionstypen ein Reaktionsnetzwerk erzeugt, das aus Fragmentierungsreaktionen und Umlagerungen besteht. Die Reaktionstypen werden einer Datei entnommen. Ein für die Analyse von 70 eV Elektronenstoß-Massenspektren häufig verwendeter Satz von Reaktionstypen ist in Abbildung 2-3 dargestellt. Nachdem alle Reaktionen durchgeführt wurden, wird das Massenspektrum rekonstruiert und jeder Reaktion eine Reaktionswahrscheinlichkeit zugeordnet. Im unteren Teil der Abbildung 2-1 ist ein Schema eines Reaktionsnetzwerks gezeigt. Ausgehend vom neutralen Molekül werden zwei Ionisationen durchgeführt und die Ionisationsprodukte weiter in Kationen und Radikalkationen sowie Neutralfragmente und Radikale gespalten. Folgereaktionen werden nur von Kationen und Radikalkationen gebildet. Die ermittelte Reaktionswahrscheinlichkeit ist jeweils neben jeder Reaktion angegeben. Ein so ermitteltes Reaktionsnetzwerk kann bei der Interpretation von Massenspektren wertvolle Hilfe leisten, da größere organische Verbindungen meist in viele verschiedene Fragmente zerfallen und ihre Zuordnung zu den Signalen des experimentellen Spektrum sehr aufwendig sein kann.

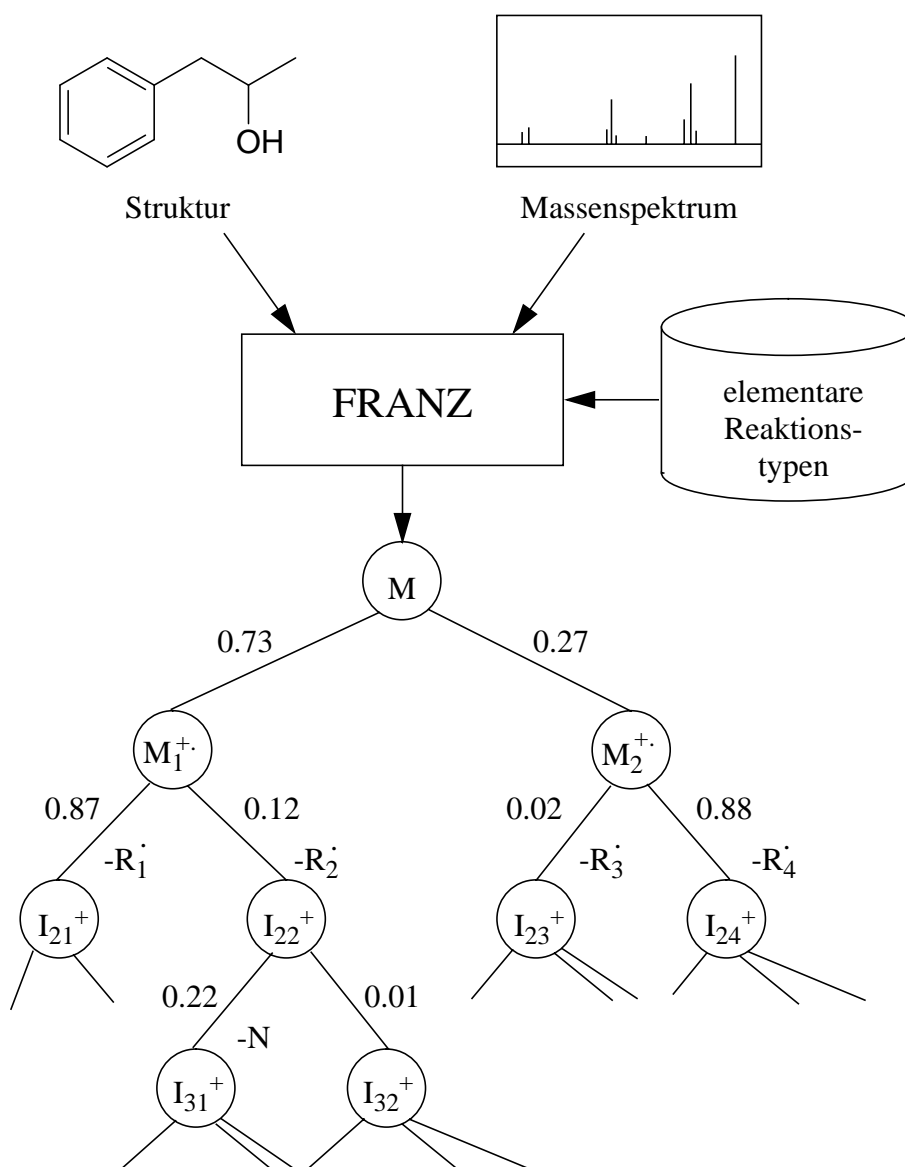


Abb. 2-1: Erzeugung eines Netzwerks von bewerteten Reaktionen mit FRANZ.

Die zweite Aufgabe ist in Abbildung 2-2 dargestellt. Für jeden Reaktionstyp werden alle auftretenden Reaktionen zusammen mit ihrer Reaktionswahrscheinlichkeit und deren Fehlergrenze gesammelt. Dabei wird nicht nur ein Spektrum, sondern ein ganzer Satz von Struktur-Spektren Paaren untersucht, um möglichst viele Daten zu gewinnen. Diese können dann als Grundlage für die Ableitung von allgemeinen Bewertungsfunktionen für einen bestimmten Reaktionstyp mit Hilfe von statistischen Methoden oder neuronalen Netzen verwendet werden.

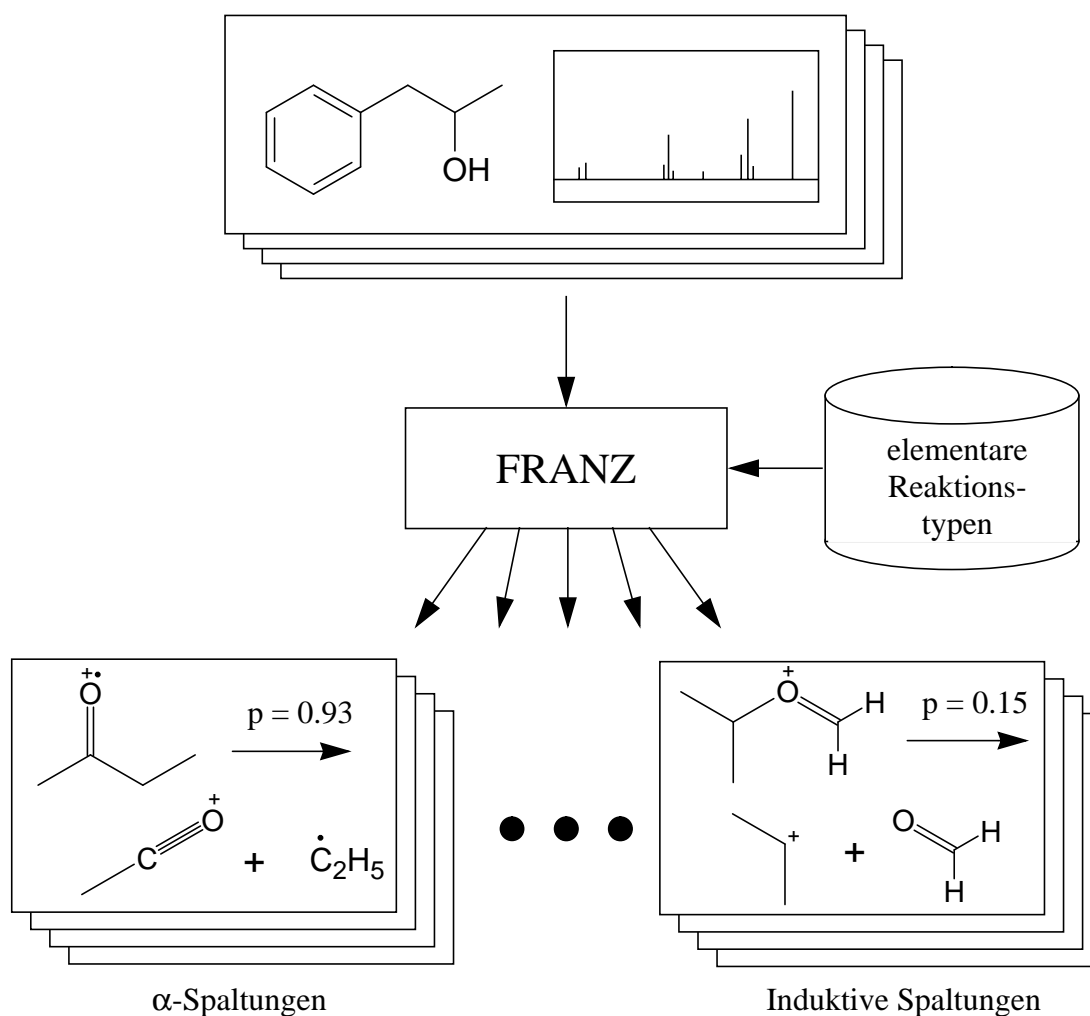


Abb. 2-2: Sammlung von bewerteten Reaktionen eines Reaktionstyps mit FRANZ

Nach McLafferty lassen sich die meisten Reaktionen mit 10 Elementarreaktionen erklären [20]. Aus diesen Elementarreaktionen wurde der in Abbildung 2-3 gezeigte Regelsatz entwickelt [21], der zur Analyse der Massenspektren mit FRANZ verwendet wurde. Die Formulierung der Reaktionsregeln basiert auf der Annahme, daß alle Reaktionen an Atomen ausgelöst werden, die ungepaarte Elektronen oder eine positive Ladung tragen. Deren Positionen stellen die treibende Kraft für die speziellen Reaktionstypen dar, die für jede funktionelle Gruppe charakteristisch sind.

Die verwendete Strukturrepräsentation sowohl der Beschreibung der Reaktionstypen als auch der internen Darstellung von chemischen Strukturen fußt auf dem Valenz-Bond-Modell.

der Ausgangssubstanz zerfallen (fließen) in bestimmte Fragmente. Zuwachs und Zerfall kann nur über Reaktionen des Zerfallsgraphen stattfinden.

- Umlagerungen verlaufen rascher als Fragmentierungen. Diese gehorchen einer Reaktionskinetik erster Ordnung.

2.3. Arbeitsweise von FRANZ

Der Ablauf des Programms gliedert sich in vier Stufen, wie in Abbildung 2-4 gezeigt ist. Die beiden Hauptteile sind der Aufbau des Reaktionsnetzwerks und die Analyse des Zerfallsgraphen, die sich aus mehreren Komponenten zusammensetzt.

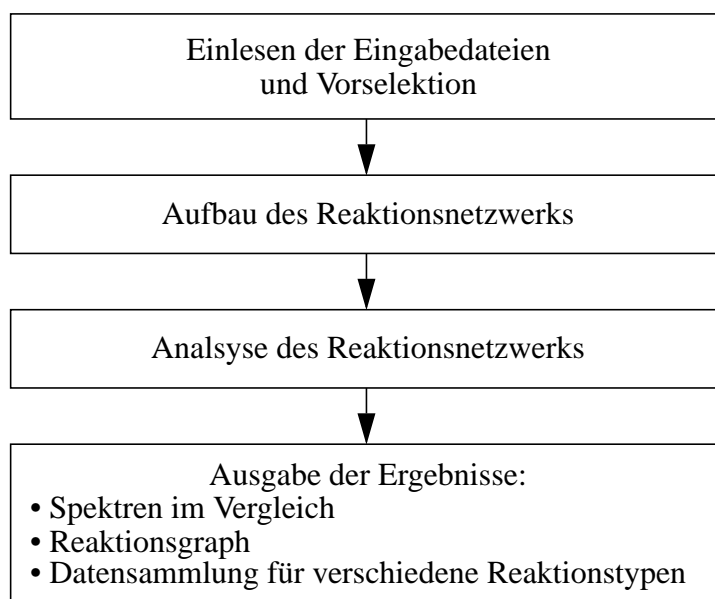


Abb. 2-4: Ablaufschema von FRANZ

2.3.1. Einlesen der Eingabedateien und Vorselektion

Im ersten Schritt werden die drei Eingabedateien, die die Struktur und das experimentelle Massenspektrum, die verwendeten Reaktionsregeln und Konfigurationen enthalten, eingelesen und die Konfigurationsdatei ausgewertet. Die Regeldatei gliedert sich in drei Bereiche. Im ersten werden Bedingungen festgelegt, die im Reaktionszentrum erfüllt sein müssen, damit die Reaktion durchgeführt werden kann. Sie sind meist topologischer Art und bestimmen, welche chemischen Elemente für einen Reaktionstyp erlaubt sind sowie Bindungsordnungen zwischen den beteiligten Atomen. Der zweite Bereich enthält eine Vorschrift, welche Elektronenumordnungen während der Reaktion stattfinden. Der dritte besteht aus einer Bewertung der Reaktion. Er enthält eine Reaktivitätsfunktion, die in Abhängigkeit von physikochemischen Parametern der Atome und Bindungen im Reaktionszentrum Aussagen über die Wahrscheinlichkeit macht, mit der die Reaktion stattfindet. Die Reaktionswahrscheinlichkeit wird jedoch in der Regel in

FRANZ nicht für den Aufbau des Reaktionsnetzwerks herangezogen, und kann daher als Konstante mit dem Wert 1 festgesetzt werden. Eine genauere Beschreibung des Formats der Regeldateien ist in [22] zu finden.

Über die Konfigurationsdatei kann man zwischen verschiedenen Modellen während des Aufbaus und der Analyse des Reaktionsnetzwerks wählen, sowie Eingangskriterien für die Struktur und das experimentelle Massenspektrum festlegen und bestimmen, welche der Ergebnisse auf Dateien ausgegeben werden sollen.

Die Vorselektion dient dazu, ungeeignete Strukturen und qualitativ schlechte Spektren zu verwerfen, die die folgenden Kriterien nicht erfüllen. Jede Spezies muß mindestens ein Kohlenstoffatom enthalten und sie muß aus mindestens drei Atomen bestehen. Im experimentellen Massenspektrum muß die höchste im Spektrum vertretene Masse größer als 32 sein. Zudem muß das Verhältnis von höchster zur niedrigsten Masse mindestens zwei betragen und der Quotient der Signalintensität des Basissignals zur niedrigsten Intensität mindestens 50.

2.3.2. Aufbau des Reaktionsnetzwerks

Im zweiten Schritt wird das Reaktionsnetzwerk gebildet. Auf die eingegebene Struktur werden zunächst die beiden Ionisationstypen angewendet und die Ionisationsprodukte erzeugt. Danach werden für die Molekülonen die Reaktionsmöglichkeiten der anderen Reaktionstypen geprüft. Genügt ein Kation den Bedingungen des jeweiligen Reaktionstyps, so wird die Reaktion durchgeführt und das entstandene Produktkation abgespeichert. Neutralkomplexe und Radikale werden nicht gespeichert, da von ihnen keine weitere Reaktion ausgehen kann. Dieser Prozeß wird nun auch für die entstandenen Kationen solange wiederholt, bis keine Reaktion mehr möglich ist. Dabei werden Produkte, die schon einmal im Reaktionsnetzwerk vorgekommen sind, nicht noch einmal gespeichert, die zusätzlichen Reaktionen werden aber in das Netzwerk eingetragen.

Um einen möglichst vollständigen Reaktionsgraphen zu erzeugen, müssen für jedes Edukt alle mesomeren Grenzformen generiert werden. Denn die Reaktionstypen hängen vom Vorhandensein einer Ladung oder eines Radikalzentrums an den an der Reaktion beteiligten Atomen ab und in einer VB-Strukturrepräsentation ist nur jeweils ein solches Zentrum dargestellt. In delokalisierten Systemen müssen aber alle möglichen Zentren berücksichtigt werden, um keine Reaktionen zu verlieren. Hanebeck hat einen Algorithmus zur Erzeugung aller mesomeren Grenzformen entwickelt [19]. Er beruht auf lokalen Vertauschungsreaktionen von benachbarten Doppelbindungen, Ladungen, Radikalen und freien Elektronenpaaren, die in Abbildung 2-5 dargestellt sind.

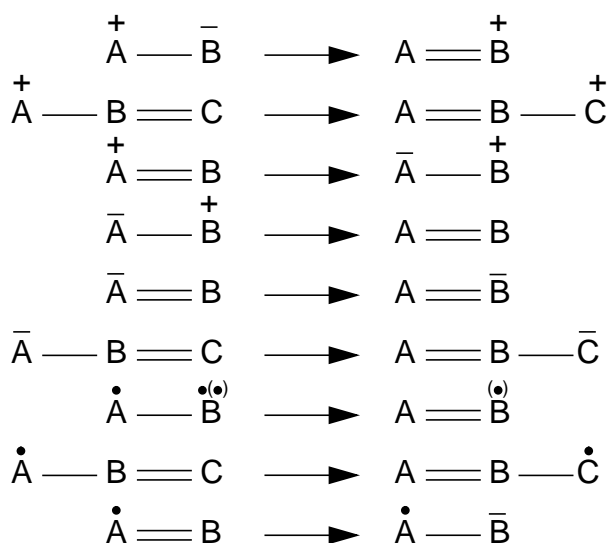


Abb. 2-5: Lokale Vertauschungen von Elektronen bei der Erzeugung von Mesomeren.

Da mit dem Reaktionsnetzwerk das experimentelle Massenspektrum rekonstruiert werden soll, findet während seines Aufbaus ein Vergleich der entstandenen Fragmente mit den zu ihnen passenden Signalen im Spektrum statt. Dabei wird eine Reaktion in zwei Fällen verworfen. Einerseits darf die Masse eines Produktkations nicht kleiner werden, als das Signal mit der niedrigsten Masse im Spektrum. Andererseits wird ein Fragment gelöscht, wenn weder für es selbst noch für keinen seiner direkten Vorgänger ein Signal im Spektrum vorhanden ist. Dabei wird von der Annahme ausgegangen, daß nie alle Sequenzen von zwei aufeinanderfolgenden Fragmentierungen so schnell ablaufen, daß keines der Zwischenprodukte ein Signal im Spektrum verursacht.

Der Vergleich zwischen erzeugten Fragmenten und den Signalen im experimentellen Massenspektrum findet über die Berechnung der theoretischen Isotopenmuster aus der Summenformel jedes Fragments statt.

2.3.3. Analyse des Reaktionsnetzwerks

Nach dem Aufbau des Reaktionsnetzwerks findet dessen Analyse statt, die mit der Berechnung der Stoffmengen aller Fragmente beginnt. Deren theoretische Isotopenmuster werden mit den Signalen im Spektrum so überlagert, daß der größtmögliche Anteil des experimentellen Spektrums erklärt wird. Dabei werden Isotopenmuster, für die es keine Signale im Spektrum gibt, entfernt. Für jedes Isotopenmuster läßt sich eine relative Stoffmenge entsprechend seines Anteils in der Überlagerung berechnen. Die Menge der erklärten Intensität im Verhältnis zur Gesamtintensität ergibt dann die Erklärungsrate, die mit den gebildeten Fragmenten erreicht werden kann.

Als Nächstes findet eine Vereinfachung des Reaktionsnetzwerks statt. Alle Isomere, die durch Umlagerungen ineinander entstehen, werden zusammengefaßt, da der Stoffmengenfluß innerhalb eines Umlagerungssystems nicht bestimmt werden kann. Ausgehend von den Endprodukten wird nun für alle Fragmente die Gesamtstoffmenge, die durch sie geflossen ist, berechnet. Dabei werden jeweils die Stoffmengen der Produktfragmente zu den Stoffmengen der Eduktfragmente einer Reaktion addiert, bis man beim Molekülion angelangt ist. Hierbei wird jedoch nicht jedem Edukt die vollständige Stoffmenge seiner Produkte zugewiesen. Denn falls mehrere Konkurrenzreaktionen zum selben Produkt führen, kann nicht bestimmt werden, welche dieser Reaktionen einen größeren Anteil hat. Daher werden den Edukten in solchen Fällen eine modellierte Gesamtstoffmenge sowie eine Unter- und eine Obergrenze zugeordnet. Die Untergrenze zeigt an, wie hoch die Stoffmenge ist, wenn nur die Konkurrenzreaktion abgelaufen wäre, die Obergrenze, wenn die Konkurrenzreaktion gar nicht stattfindet. Der modellierte Wert kann nach verschiedenen Modellen ermittelt werden, die in [23] erklärt sind. Sie beruhen auf der Wichtung der konkurrierenden Fragmente.

Der Umsatz jeder Reaktion ist nun der Anteil der Gesamtstoffmenge eines Produkts, der von dem Edukt herrührt. Auch für die Umsätze wird außer dem modellierten Wert jeweils die Ober- und Untergrenze aus den Ober- und Untergrenzen der Stoffmengen ermittelt.

Im letzten Schritt werden Reaktionswahrscheinlichkeiten für jede Reaktion berechnet. Es gilt die Annahme, daß alle Zerfallsreaktionen einer Reaktionskinetik erster Ordnung gehorchen. Daher läßt sich die Reaktionswahrscheinlichkeit der Zerfallsreaktionen direkt aus den Umsätzen ableiten.

In die ebenfalls berechneten Unter- und Obergrenzen für die Reaktionswahrscheinlichkeiten gehen die Unter- und Obergrenzen der Umsätze ein. Bei Reaktionen, die keine Konkurrenzreaktionen besitzen, sind die Unter- und Obergrenze gleich dem modellierten Wert. Bei Reaktionen, deren Edukte und Produkte Isomere im Zerfallsgraphen besitzen oder die durch verschiedene Reaktionssequenzen entstehen, ist die Differenz der Ober- und Untergrenze der Reaktionswahrscheinlichkeit sehr groß. Sie ist daher ein Maß für die Zuverlässigkeit der Bewertung einer Reaktion. Reaktionen, für die diese Differenz größer als 0.3 ist, werden als unzuverlässig modelliert angesehen. Sie eignen sich nicht für die Sammlung von Reaktionen zur statistischen Analyse, da hier der mögliche Fehler zu groß ist, um ein zuverlässiges Modell zu entwickeln.

Zuletzt werden die Ergebnisse, die aus den Reaktionsnetzwerken und den ermittelten Reaktionswahrscheinlichkeiten bestehen, in Dateien ausgegeben.

2.4. Nachteile der in FRANZ verwendeten Strukturrepräsentation auf der Basis des Valence-Bond-Modells

Die beiden Aufgaben, die mit FRANZ gelöst werden sollen, stellen unterschiedliche Anforderungen an die erzeugten Reaktionsnetzwerke und damit an die Regelsätze, die zu ihnen führen. Denn für die Analyse und das Verständnis eines Massenspektrums sollte das Reaktionsnetzwerk außer den Reaktionen zu Fragmenten mit charakteristischen Signalen auch Reaktionen zu seltener auftretenden Spezies erklären können und die verschiedenen Konkurrenzreaktionssequenzen aufzeigen. Diese Anwendungsmethode soll jedoch im folgenden nicht weiter diskutiert werden. Für die Sammlung der Daten zur statistischen Analyse ist es jedoch am vordringlichsten, zuverlässige Aussagen über die Reaktionswahrscheinlichkeit von Reaktionen zu erhalten. In flexiblen Netzwerken mit vielen Konkurrenzreaktionen erhalten die Reaktionen zu den Fragmenten mit hohen Signalintensitäten, die meist in der ersten und zweiten Reaktionssebene entstehen, häufig große Intervalle für die Ober- und Untergrenze der Reaktionswahrscheinlichkeit. Solche Reaktionsnetzwerke eignen sich nicht als Grundlage zur Sammlung von Daten für die statistische Analyse. Für diese Zwecke darf der verwendete Regelsatz nicht zu allgemein sein, muß aber dennoch erlauben, mindestens 80% der Gesamtsignalintensität des Spektrums zu erklären, da sonst die Stoffmengen, die in die Reaktionswahrscheinlichkeit eingehen, nicht den Verhältnissen im experimentellen Spektrum entsprechen und ebenfalls zu falschen Bewertungen der Reaktionen führen.

2.4.1. Fragmentierungsschemata von Ketonen am Beispiel von Butan-2-on

Carbonylverbindungen zeigen sehr charakteristische Signale bei der Aufnahme von Massenspektren und eignen sich daher gut, um die Vorzüge und Schwachstellen der Analyse von Massenspektren mit FRANZ aufzuzeigen. Die Fragmente der Hauptsignale lassen sich durch folgende Reaktionstypen aus Abbildung 2-3 erklären [20], [24], [25]: α -Spaltung, Wasserstoffumlagerungen, induktive Spaltungen und Carbonyleliminierung. Der ebenfalls häufig auftretende Reaktionstyp McLafferty-Umlagerung wird nicht eigens formuliert, sondern aus der Wasserstoffumlagerung eines γ -Wasserstoffatoms auf die Carbonylgruppe und anschließender α -Spaltung zusammengesetzt.

Die Abbildung 2-6 zeigt den Vergleich des experimentellen mit dem rekonstruierten Massenspektrum von Butan-2-on, das mit dem Regelsatz aus Abbildung 2-3 analysiert wurde. Die Erklärungsrate beträgt 89.7%. Für sechs der zehn Peaks über 1.5% finden sich passende Fragmente, die mit 24 Reaktionen gebildet wurden. Auf den ersten Blick ist dies ein zufriedenstellendes Ergebnis. Untersucht man jedoch das Reaktionsnetzwerk und die Bewertungen für die einzelnen Reaktionen genauer, so muß man feststellen, daß prinzipielle Probleme auftreten, die eine Übernahme wichtiger Reaktionen dieser Verbindung in eine Datensammlung zur statistischen Analyse wegen ungenauer Bewertungen nicht zulassen.

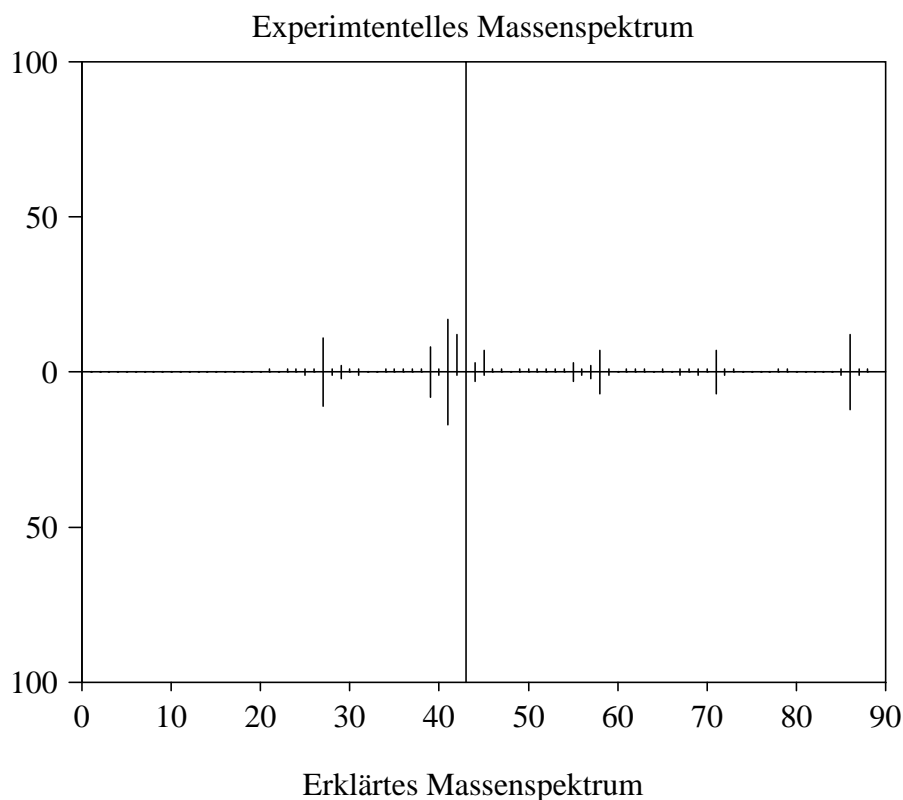


Abb. 2-6: Vergleich des experimentellen und rekonstruierten Massenspektrums von Butan-2-on.

Abbildung 2-7 zeigt einen Ausschnitt des Reaktionsnetzwerks von Butan-2-on. Die Hauptfragmentierungen des Zerfalls von Butan-2-on sind zwei α -Spaltungen, sie führen zu **2** und **3**. Das Produkt **3**, aus dem ein Methylfragment abgespalten wurde, eliminiert anschließend CO (**6**). Beide α -Spaltungsreaktionen wurden als Hauptreaktionen für Butan-2-on nachgewiesen [26], man erwartet eine hohe Reaktionswahrscheinlichkeit mit einer guten Bewertung im Reaktionsschema für sie. Die Reaktionswahrscheinlichkeiten, die jeweils neben dem Reaktionspfeil angegeben sind, entsprechen jedoch nur für die Reaktion zu **3** den Erwartungen. Die Reaktionswahrscheinlichkeit für die Reaktion zu **2** ist niedriger als diejenige zu Fragment **3**, obwohl das korrespondierende Signal im Spektrum der Basepeak ist. Zudem ist die Differenz zwischen Unter- und Obergrenze sehr hoch, die Bewertung ist also als unzuverlässig einzustufen.

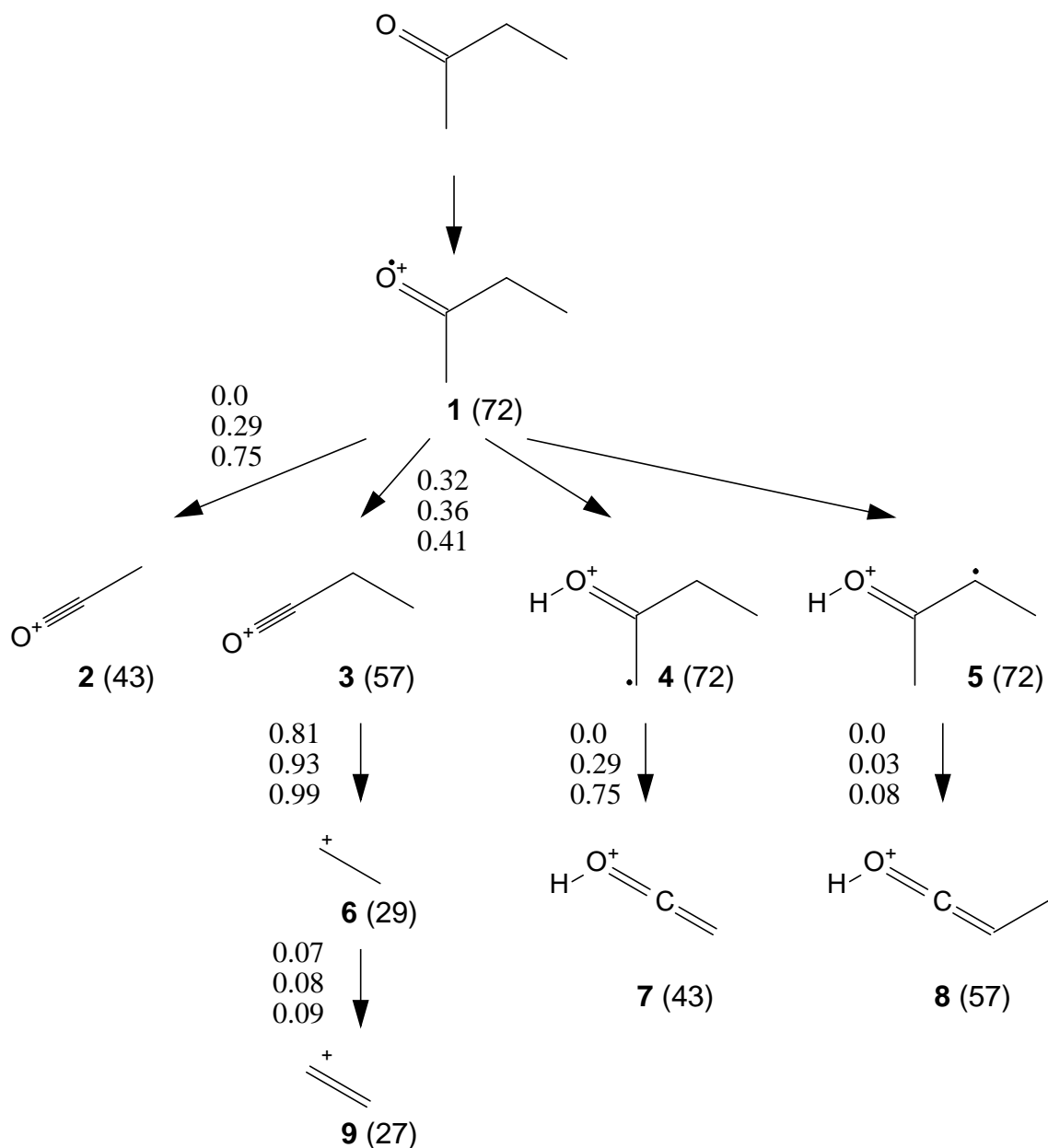


Abb. 2-7: Ausschnitt aus dem Reaktionsnetzwerk von Butan-2-on. Neben den Reaktionspfeilen ist jeweils die ermittelte Reaktionswahrscheinlichkeit mit Unter- und Obergrenze angegeben. Neben den Fragmentnummern ist jeweils die Masse des Hauptsignals (m/z) in Klammern angegeben.

Eine Analyse des restlichen Reaktionsnetzwerks zeigt, daß außer den erwarteten Reaktionen noch zusätzliche Konkurrenzreaktionen zu Fragmenten mit den Massen 43 und 57 gefunden werden, von denen zwei exemplarisch ebenfalls in Abbildung 2-7 dargestellt sind. Die Produkte der Konkurrenzreaktionen sind Isomere von **2** und **3**. Die Gesamtstoffmenge von **3** setzt sich hauptsächlich aus den Stoffmengen seiner Zerfallsprodukte, **6** und **9** zusammen, das zu **3**

korrespondierende Signal selbst hat eine sehr niedrige Intensität. Da **8** aber nicht CO eliminieren kann, muß zwischen beiden Isomeren nur die Stoffmenge des kleinen Signals bei m/z 57 geteilt werden, was zu einer niedrigen Reaktionswahrscheinlichkeit für die Reaktion zu **8** führt. Sie stört bei der Modellierung der α -Spaltung zu **3** nicht.

2 zerfällt nicht weiter, zu seinem CO-Eliminierungsprodukt, dem Methylkation, gibt es kein korrespondierendes kein Signal im experimentellen Spektrum. Daher wird die Bewertung allein auf der Basis der Signalintensität des Basepeaks vorgenommen. Zwischen **2** und seinem Isomeren **7** kann nicht unterschieden werden. Für beide Reaktionen liegt somit die Untergrenze bei Null und die modellierte Reaktionswahrscheinlichkeit ist niedrig, da die Stoffmengen zwischen beiden verteilt werden müssen.

Deuteriummarkierungsexperimente haben jedoch gezeigt [27], daß die Abspaltung eines Ethylradikals aus **4** nicht direkt, sondern erst nach mehreren Wasserstoffumlagerungen, die in **1** resultieren, stattfindet. Betrachtet man die räumliche Lage des an der α -Spaltung beteiligten radikalischen π -Elektrons und der zu spaltenden Einfachbindung (Abbildung 2-8), so ist ersichtlich, daß bei der α -Spaltung ausgehend vom Molekülion **1** beide in der gleichen Ebene liegen und daher leicht miteinander reagieren können, dies aber im Fragment **4** nicht der Fall ist. Dort liegt das Radikal nicht in der Ebene der σ -Bindung, eine α -Spaltung sollte daher nicht stattfinden.

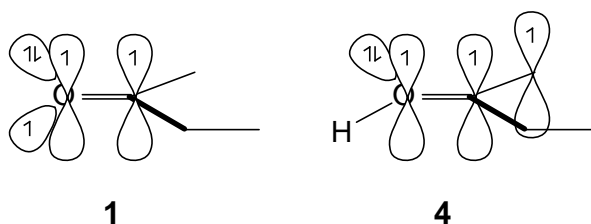


Abb. 2-8: Elektronenkonfigurationen der beiden isomeren Radikalkationen, für die α -Spaltung unter Verlust eines Ethylradikals mit FRANZ gefunden wird.

Die VB-Beschreibung, auf der die Strukturdarstellung in FRANZ beruht, erlaubt jedoch die Unterscheidung von Reaktionszentren aufgrund von Konjugation und räumlicher Gestalt von π -Systemen nicht. Mit den vorhandenen Möglichkeiten zur Formulierung des Reaktionszentrums ist es nicht möglich, diese Konkurrenzreaktion bei der Erzeugung von Reaktionen zu unterdrücken. Sie erlauben ausschließlich, zu spezifizieren, daß eine zu einem Radikalzentrum benachbarte Einfachbindung homolytisch gespalten werden soll. Es kann aber nicht festgestellt werden, ob sich das Radikalzentrum in einem π -System befindet, das in der gleichen Ebene wie die zu brechende Einfachbindung liegt. Diese Konkurrenzreaktion zur α -Spaltung aus dem Molekülion findet sich in allen Reaktionsnetzwerken von Ketonen. Bei der Sammlung der α -Spaltungsreaktionen von Ketonen werden daher fast alle α -Spaltungsreaktionen auf-

grund von hohen Intervallen für die Unter- und Obergrenze der Fragmentierung aus dem Molekülion verworfen. Es ist somit nicht möglich, Datensätze zur statistischen Analyse dieses Reaktionstyps in Carbonylverbindungen zusammenzustellen.

2.4.2. Orthogonale π -Systeme

Die interne Strukturrepräsentation von FRANZ, die auf dem VB-Modell beruht, wirkt sich jedoch nicht nur bei der Modellierung der Fragmentierungsprozesse von Ketonen nachteilig aus, sondern sie bereitet bei allen Strukturklassen, in denen orthogonale π -Systeme auftreten, Probleme. Denn die Beschreibung von Reaktionen in der Massenspektrometrie basiert auf der Lokalisierung von positiven Ladungen und Radikalen auf Atome, die dann einen bestimmten Reaktionstyp auslösen. Eine α -Spaltung kann nur von einem Atom ausgehen, das ein ungepaartes Elektron trägt. Um nun alle Reaktionen eines Radikalkations, die radikalisch induziert sind, erzeugen zu können, werden dem Reaktionsgenerator alle Mesomere zur Verfügung gestellt, die für das gerade behandelte Radikalkation möglich sind.

In Verbindungen mit nur einem π -System verläuft die Mesomerengenerierung unproblematisch. Für das Butadienradikalkation werden alle drei möglichen Mesomere generiert (Abbildung 2-9).

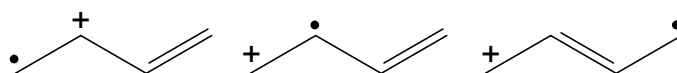


Abb. 2-9: Mesomere Grenzstrukturen des Butadienradikalkations.

Sobald jedoch Verbindungen orthogonale π -Systeme enthalten, werden auch Vertauschungsprozesse von Elektronen durchgeführt, die sich nicht im selben π -System befinden. Ein charakteristisches Beispiel hierfür sind Amide. Die Ionisation kann an den n-Elektronen des Sauerstoffatoms sowie im π -System auftreten. Durch die Elektronenverschiebungsprozesse im Mesomerengenerator werden beide Formen ineinander umgewandelt (Abbildung 2-10). Form A wird durch die Vertauschung einer Doppelbindung und eines freien Elektronenpaares in Form B1 überführt. Dann findet implizit eine Rotation um die Achse der C-O-Bindung statt. Denn mit der Valenz-Bond-Beschreibung kann zwischen B1 und B2 nicht unterschieden werden. Die zweite korrekte Vertauschung findet dann zwischen dem Radikal und der Doppelbindung von B2 zu C statt.

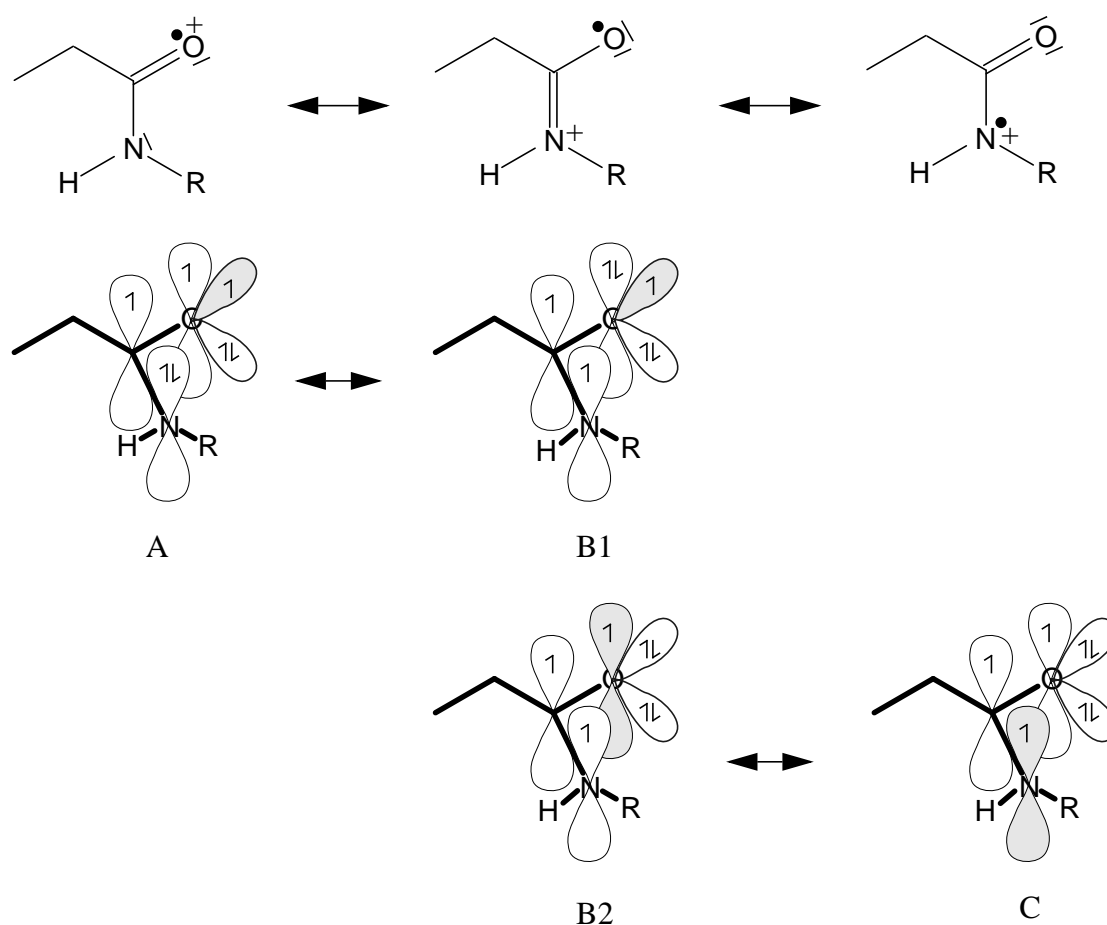


Abb. 2-10: Mesomere Grenzformen eines ionisierten Amids, die vom Mesomerengenerator in FRANZ durch Vertauschung eines freien Elektronenpaares mit einer Doppelbindung (A \longleftrightarrow B1) bzw. eines radikalischen Elektrons mit einer Doppelbindung (B2 \longleftrightarrow C) erzeugt werden. Das radikalische Elektron ist jeweils grau hinterlegt.

Daß beide Ionisationsprodukte als mesomere Grenzformen einer einzigen Struktur betrachtet werden, wirkt sich ebenfalls auf die Bewertung der Reaktionen mit FRANZ aus. Denn der entstehende Zerfallsgraph eines Amids unterscheidet sich deutlich von dem erwarteten. Ausgehend von dem Neutalmolekül sollten zwei Ionisationsprodukte entstehen (Abbildung 2-11a), die dann jeweils fragmentieren. Durch die unkorrekte Formulierung der Mesomerie entsteht jedoch nur ein Ionisationsprodukt (Abbildung 2-11b), von dem ausgehend alle weiteren Reaktionen stattfinden. Bei der Zuordnung der Stoffmengen erhält nun das Molekülion die Gesamtstoffmenge aller Fragmente, die im Reaktionsnetzwerk enthalten sind. Korrekt wäre jedoch eine Aufteilung der Stoffmenge auf beide Ionisationsprodukte. Das Verhältnis zwischen Gesamtstoffmenge des Edukts und Stoffmenge des Produkts einer Reaktion ändert sich dadurch für die Fragmentierungen aus den Molekülionen drastisch. Da jedoch genau dieses

Verhältnis in die Berechnung der Reaktionswahrscheinlichkeiten miteinbezogen wird, erhält man auch in diesen Fällen keine korrekten Werte für die modellierten Reaktionswahrscheinlichkeiten.

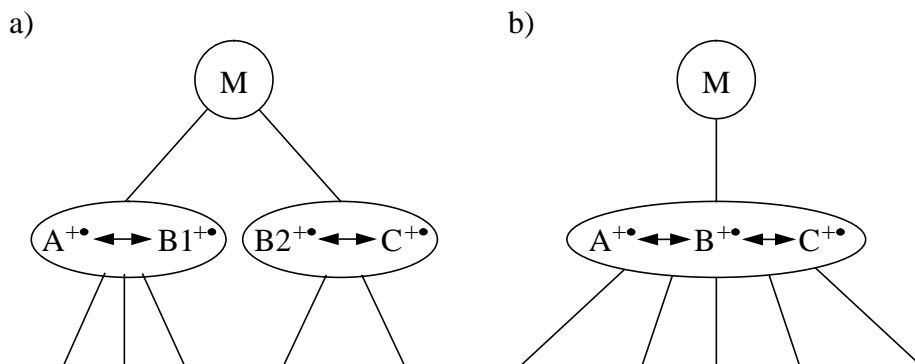


Abb. 2-11: Erwartete und mit FRANZ modellierte Ionisationsprodukte von Amid. A, B und C entsprechen den Grenzformen aus Abbildung 2-10.

2.4.3. σ -Ionisation

Ebenfalls Schwierigkeiten bereitet die korrekte Behandlung von Alkanen, in denen σ -Ionisation auftritt (Abbildung 2-12). Denn deren Ionisationsprodukte enthalten σ -Bindungen mit nur einem Elektron in der Bindung. Solche Bindungen lassen sich mit dem VB-Modell, das implizit jeweils zwei Elektronen in einer Bindung lokalisiert, nicht darstellen.

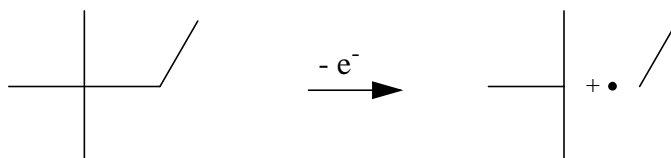


Abb. 2-12: σ -Ionisation von 2,2-Dimethylbutan.

Diese Probleme bei der Reaktionsgenerierung treten natürlich nicht nur in der Analyse von Massenspektren auf, sondern auch bei deren Simulation. Dort wird der umgekehrte Weg beschritten. Mit Hilfe von bewerteten Reaktivitätsfunktionen werden Reaktionswahrscheinlichkeiten für die Fragmentierungen abgeschätzt und daraus Konzentrationen der Fragmente bestimmt. Diese werden dann in Signalintensitäten umgesetzt. Dabei werden die gleichen Reaktionstypen verwendet wie bei der Analyse von Massenspektren, es werden also ebenfalls Reaktionen erzeugt, die nicht im Massenspektrometer auftreten und dann zu falschen Signalintensitäten führen können. Die vorgestellte Problematik machte es erforderlich, die Datenstruktur von Molekülen so zu erweitern, daß Orthogonalität in π -Elektronensystemen berücksichtigt wird.

3. Datenstrukturen zur Repräsentation von Molekülen

Bindungslisten, die auf dem VB-Modell beruhen, sind die am weitesten verbreitete Form, um chemische Strukturen in computerlesbarer Form darzustellen. Sie werden auch häufig als interne Repräsentation in Programmen verwendet, die sich mit der Behandlung von chemischen Problemen befassen. Da Bindungslisten jedoch keine eindeutige Beschreibung chemischer Strukturen zulassen und zudem Mängel bei der Behandlung delokalisierter Teilstrukturen aufweisen, wurden eine Vielzahl von Varianten entwickelt, die je nach Anwendung unterschiedliche Anforderungen erfüllen.

Im folgenden wird ein Überblick über Strukturrepräsentationen gegeben, die entweder in weit verbreiteten Systemen verwendet werden oder die spezielle Lösungen zur Repräsentation von Verbindungen mit delokalisierten π -Systemen anbieten.

3.1. Bindungslisten

In Datenbanksystemen ist es unabdingbar, für jede gespeicherte Verbindung eine eindeutige und kanonische Darstellung ihrer Struktur zu gewährleisten. Auf diese Weise ist es einfach, zu überprüfen, ob eine Verbindung schon in einer Datenbank enthalten ist oder als Neueintrag hinzugefügt werden muß. Somit darf es für jedes Molekül nur eine einzige mögliche Repräsentation geben, unabhängig von der Struktur, in der es eingegeben wurde. In der Datenbank des Chemical Abstracts Service's werden Strukturen in „unique chemistry registration records“ (UCRR) gespeichert [28]. Diese bestehen aus vier Komponenten: der Konstitution in Form einer Bindungsliste, einem Textdeskriptor, Isotopenmarkierungen und einer zusätzlichen Komponente. Da eine sehr große Anzahl von Strukturen gespeichert werden muß, wird die Konstitution einer Struktur nicht jeweils explizit in einer Bindungsliste abgespeichert, sondern diese setzt sich auch aus Bausteinen unterschiedlicher Größe zusammen. So gibt es eine globale Liste mit allen bisher vorgekommenen Ringsystemen. Jedes Ringsystem ist dort nur einmal gespeichert, und erhält eine Identifikationsnummer. Die Bindungsliste eines Eintrags besteht dann aus einer Liste aller Ringidentifizierer, einer Bindungsliste der acyclischen Teile, den Verbindungen zwischen beiden Listen und einer Kreuzreferenz zwischen der speziellen Numerierung des Eintrags und der Numerierung in der Ringliste. Die verwendeten Bindungslisten sind eindeutig. Dies ist notwendig, damit nicht dieselbe Struktur mehrere Bindungslisten und somit auch mehrere UCRRs erhält. Um eindeutige Strukturen zu erzeugen, werden die eingegebenen Strukturen einer Kanonisierung und Normalisierung unterworfen. Die Kanonisierung erlaubt eine eindeutige Numerierung [29], die Normalisierung löst das Problem der Mesomerie und Tautomerie [30].

Bei der Normalisierung werden sich abwechselnde Einfach- und Doppelbindungen in Ringen gesucht und in den speziellen Bindungstyp alternierende Bindungen umgewandelt, bei denen die Bindungsordnung nicht mehr festgelegt ist. Es wird jedoch nicht zwischen aromatischen und nicht aromatischen Verbindungen unterschieden. Die Ringbindungen z. B. in Benzol aber

auch in Cyclooctatetraen werden alle als alternierend eingestuft. Bindungen zwischen Heteroatomen, die eine tautomere Gruppe bilden, wird ebenfalls ein eigener Bindungstyp, die tautomere Bindung zugeordnet, in der die Bindungsordnung nicht festgelegt ist.

In der Beilstein-Faktendatenbank werden Verbindungen getrennt von den Daten gehandhabt [31], [32], die Faktendaten werden in einer Faktendatei, die Strukturen der Verbindungen in einer gesonderten Strukturdatei abgelegt. Letztere enthält für jede Struktur eine „Beilstein Registry Connection Table“ (BRCT). Auch die BRCT genügt den für eine Datenbankanwendung notwendigen Bedingungen: Sie ist eindeutig, jede Bindungsliste kodiert nur eine einzige Struktur und es gibt für jede Struktur nur eine Bindungsliste, d. h. verschiedene Tautomere und mesomere Grenzformen erzeugen den gleichen Eintrag. Um dies zu erreichen, werden zunächst alle möglichen tautomeren Gruppen mit einem Verfahren gesucht, das demjenigen, welches bei CAS verwendet wird, ähnlich ist. Danach wird jede Struktur einem Morgan-Verfahren zur eindeutigen Numerierung unterworfen. Es wird eine Bindungsliste erzeugt, in der für jedes Atom nur die Bindung zum Atom mit der niedrigsten Nummer enthalten ist. Alle weiteren Bindungen eines Atoms sind Ringschlußbindungen, die gesondert gespeichert werden. Dies gewährleistet eine kompakte Darstellung, die den Speicherbedarf gering hält und eine schnelle Übertragung zwischen unterschiedlichen Applikationen ermöglicht. Mesomere Strukturen werden anders behandelt als im CAS Registry System, indem auf ein Verfahren zurückgegriffen wird, das von Gasteiger vorgeschlagen wurde [33],[34], [35]. Darin werden Mehrfachbindungen nicht direkt dargestellt, sondern jedem Atom, das an einem π -System beteiligt ist, wird die Anzahl an Elektronen, die es zum π -System besteuert, zugeordnet. So ergeben die drei mesomeren Grenzformen des Naphthalins nur eine mögliche Darstellung mit je einem π -Elektron an jedem Kohlenstoffatom. Freie Elektronenpaare werden gesondert behandelt.

In den Gmelin-Faktendatenbanken werden Verbindungen hierarchisch gespeichert [36]. Die oberste Ebene der Hierarchie enthält die Information, aus wievielen Komponenten sich eine Substanz zusammensetzt. Jede Komponente kann aus einem oder mehreren Fragmenten bestehen. Jedes Fragment wird klassifiziert, ob sich eine Strukturformel für dieses Fragment ableiten läßt oder nicht, d. h. ob man es zeichnen kann, und entsprechend weiterbehandelt. Fragmente, die sich nicht zeichnen lassen, wie z. B. Legierungen, werden in einer „Inorganic Structure Table“ (IST) abgelegt, die außer der Summenformel weitere Informationen über die Substanz wie den Typ oder ihre Modifikation enthält. Zur Speicherung von Strukturen, die sich zeichnen lassen, werden ebenfalls BRCTs verwendet.

Auch Programme zur Synthesepaltung und Reaktionsvorhersage, wie z. B. LHASA, SECS, CAMEO, verwenden intern Bindungslisten, die auf dem VB-Modell beruhen. Häufig beinhalten die dort verwendeten Datenstrukturen keine besondere Darstellung aromatischer oder delokalisierte Bindungen. Vielmehr kommen Verfahren zum Einsatz, die Hinweise auf die

Aromatizität der Verbindungen aus ihrer Konnektivität ableiten und diese zusätzliche Information bei der Planung von Synthesen oder Reaktionsgenerierung berücksichtigen [5].

Auch bei den Dateiformaten zum Austausch chemischer Information sind verschiedene Formate entwickelt worden. Zu ihnen gehören das MDL-Molfile-Format [37] und das SMD-Format [38], [39]. Das MDL-Molfile-Format ist wohl das bekannteste und am weitesten verbreitete Dateiformat unter ihnen. Es beschreibt chemische Strukturen in Bindungslisten, die auf dem VB-Modell basieren und unterliegt daher ebenfalls deren Restriktionen bei der Darstellung chemischer Strukturen. Für aromatische Bindungen gibt es jedoch einen eigenen Bindungstyp. Das SMD-Format beruht ebenfalls auf der Beschreibung von chemischen Strukturen in Bindungslisten und wurde zum Datenaustausch für organische Verbindungen entwickelt.

Die „concise connection table“ (CCT) [40] wurde entwickelt, um die automatische Übersetzung von IUPAC Namen in eine Moleküldatenstruktur zu erleichtern, die zudem sehr kompakt ist. In ihr werden nicht explizit alle Bindungen zwischen den Atomen eines Moleküls aufgelistet, sondern hierarchisch miteinander verbundene Strukturelemente in tabellarischer Form abgelegt. Zunächst werden Grundgerüste, Ketten und Ringe, aus der Nomenklatur abgeleitet und am Anfang der Tabelle eingetragen. Danach werden Substrukturen aus den Substituenten abgeleitet und hierarchisch unter dem jeweiligen Grundgerüst eingetragen und dem Verknüpfungspunkt zugewiesen.

3.2. BE-Matrizen und ihre Erweiterungen

Die Syntheseprogramme CICLOPS, IGOR und RAIN verwenden Bindungs- und Elektronenmatrizen (BE-Matrizen) [9] zur Darstellung von Molekülen und R-Matrizen zur Repräsentation der Elektronenumordnungsprozesse in Reaktionen. Sowohl BE- als auch R-Matrizen beruhen auf dem Valence-Bond Modell zur Darstellung von Molekülen. Eine BE-Matrix enthält die Bindungen und freien Elektronen eines Reaktionsensembles in einer quadratischen Matrix, deren Dimension der Anzahl der Atome im Ensemble entspricht. Die Diagonalelemente enthalten jeweils die Zahl der freien Elektronen eines Atoms, die Nichtdiagonalelemente die Bindungsordnung zwischen zwei Atomen. In R-Matrizen beschreiben die Diagonalelemente die Änderung der Zahl der freien Elektronen während einer Reaktion und die Nichtdiagonalelemente die Änderung der Bindungsordnungen. Um die Einschränkungen des in BE-Matrizen dargestellten VB-Modells zu überwinden, wurden zwei neue Modelle auf der Basis der BE-Matrizen entwickelt, die die Repräsentation von delokalisierten Bindungen erlauben, sogenannte XBE- [41] und sXBE-Matrizen [42]. Für die Reaktionsgenerierung stehen XR- und sXR-Matrizen zur Verfügung. Diese Erweiterungen der BE-Matrizen wurden entwickelt, um den gewachsenen Anforderungen an die Variationsbreite der darstellbaren chemischen Verbindungen Rechnung zu tragen. XBE-Matrizen enthalten zusätzlich zu einer BE-Matrix weitere Reihen und Spalten. Im Teil der ursprünglichen BE-Matrix werden weiterhin die lokalisierten Bindungen angegeben. Die zusätzlichen Spalten enthalten

delokalisierte Bindungen. Für alle Atome, die zu einer delokalisierten Bindung gehören, sind die Einträge in den entsprechenden zusätzlichen Spalten und Reihen ungleich null, im Diagonalelement ist die Anzahl der darin delokalisierten Elektronen angegeben, die lokalisierten Bindungen werden weiterhin durch die BE repräsentiert. Dies erlaubt eine Darstellung von delokalisierten Elektronen in konjugierten π -Systemen sowie in Elektronenmangelverbindungen. XR-Matrizen enthalten ebenfalls zusätzliche Reihen und Spalten, in denen die Änderung der Zusammensetzung von delokalisierten Bindungen gespeichert wird. *s*XBE-Matrizen unterscheiden sich von XBE-Matrizen durch die Bedeutung der Matrix-Elemente. Sie stellen nun nicht mehr Bindungsordnungen dar, sondern wurden durch symbolische Bindungstypen ersetzt. Es gibt 7 Bindungstypen, 4 für lokalisierte Einfach- und Mehrfachbindungen und drei für delokalisierte Bindungen. Jedem Bindungstyp ist implizit die Zahl der in ihm enthaltenen Elektronen zugeordnet. Sie ist gleich dem Doppelten der Bindungsordnung für die lokalisierten Bindungstypen. Die folgenden delokalisierten Bindungstypen wurden definiert: *pisys* für delokalisierte π -Systeme mit 2 impliziten Elektronen und *edsys* für delokalisierte Sigma-Bindungen sowie *coord* für koordinative Metallkomplexbindungen mit je mit null impliziten Elektronen. Auch die Bedeutung der Einträge einer *s*XR-Matrix hat sich gewandelt. Sie enthalten nun sogenannte Umwandlungsfunktionen, die die verschiedenen Bindungstypen ineinander umwandeln.

Ein alternatives Verfahren zu den XBE- und *s*XBE-Matrizen stammt von Dietz [43], das ebenfalls die Darstellung von Delokalisation, Elektronenmangelbindungen und metallorganischen Verbindungen erlaubt. Die Grundlage seines Modells ist ein molekularer Multigraph, d. h. ein ungerichteter Graph, in dem die Knoten Atome und die Kanten Bindungen repräsentieren. In Multigraphen ist es erlaubt, zwei Knoten durch mehr als eine Kante zu verbinden, es können also mehrere Bindungen zwischen zwei benachbarten Atomen vorhanden sein. Daher ist es nicht möglich, Moleküle in Matrixform darzustellen. Stattdessen werden zwei Mengen gebildet. Die erste enthält alle Atome, wobei jedes Atom wiederum durch ein Tripel aus Ordnungszahl, Zahl der freien Elektronen und einem Index besteht. Die zweite Menge enthält alle Bindungen. Jede Bindung besteht aus einer Menge an Atompaaren und der Zahl der in der Bindung lokalisierten Elektronen. Dabei werden die Atome eines Paares jeweils als Nachbarn betrachtet, woraus sich die Konstitution eines Moleküls ergibt. Da Bindungen mehr als ein Atompaar enthalten können, erlaubt es auch diese Darstellung, delokalisierte Bindungen zu beschreiben.

3.3. Z-Matrizen

Semi-empirische und *ab-initio* Programme optimieren die dreidimensionale Struktur von Verbindungen mit dem Ziel, möglichst nieder-energetische Strukturen zu finden. Z-Matrizen waren bis vor kurzem die einzige Strukturrepräsentation, die als Eingabeformat für diese Programme gedient hat [44]. Sie enthalten die räumliche Position der Atome in Form von internen Koordinaten. Dabei wird zur Beschreibung jedes Atoms eine Zeile verwendet, in der ein

Abstand, ein Winkel und Diederwinkel in Bezug auf schon definierte Atome angegeben sind. Nur die ersten drei Zeilen unterscheiden sich dadurch, daß sie nur einen Winkel und einen Abstand bzw. nur einen Abstand oder, am Anfang der Z-Matrix, überhaupt keine interne Koordinate enthalten. Die Einträge einer Z-Matrix beschreiben nicht unbedingt wirkliche Bindungsverhältnisse, sondern dienen nur dazu, die relativen Positionen der Atome zu definieren.

3.4. Die Linearnotation SMILES

SMILES [45], [46] ist eine flexible, leicht erlernbare Sprache zur Repräsentation chemischer Strukturen in Form einer Linearnotation. Sie wird von Daylight Chemical Information Systems im Daylight Toolkit, aber auch in vielen anderen Systemen zur Eingabe von Strukturen verwendet. Sie hat inzwischen große Bedeutung als Struktureingabeformat für chemische Anwendungen im Internet erlangt, da die kompakte Kodierung als Text kein grafisches Eingabetool verlangt und zudem eine schnelle Übertragung erlaubt. Die Grundlage der Notation ist die Repräsentation eines Moleküls als Graph, wobei Bindungen zwischen Atomen nach dem Valence-Bond Modell beschrieben werden. Außer Einfach- und Mehrfachbindungen wurde auch ein aromatischer Bindungstyp eingeführt. Ein SMILES-String besteht aus einer fortlaufenden Reihe von Atomsymbolen. Nebeneinander stehende Atome sind benachbart. Für jeden Bindungstyp steht ein Zeichen zur Verfügung, mit dem zwei benachbarte Atome verbunden werden können. Verzweigungen werden hinter dem verzweigten Atom in geschachtelten Klammern angegeben. Zur Beschreibung von Ringen, werden Ringschlußbindungen aufgebrochen und beide Atome, deren Bindung nicht direkt angegeben werden kann, mit der selben Ziffer markiert.

4. RICOS - Ein allgemeines Modell zur Repräsentation von Molekülen

In diesem Kapitel wird das RICOS-Modell (*Representation of Inorganic, Coordinative and Organic Structures*) zur Strukturrepräsentation vorgestellt, das Lösungswege für die in den vorigen Kapiteln vorgestellten Probleme bei der Strukturrepräsentation bietet. Die Grundlage dieses Modells ist die Trennung des σ - und des π -Anteils von Bindungen und die Einführung eines Bindungstyps für Bindungen in Metallkomplexen. Außerdem werden Elektronen nicht mehr implizit mit den Bindungen gespeichert, sondern es wird in jeder Bindung angegeben, wieviele Elektronen sie enthält. Außer einer neuartigen Repräsentation von chemischen Bindungen erlaubt das Modell auch die Repräsentation von Wechselwirkungen, die nicht als Bindungen dargestellt werden und eine hierarchische Zuordnung von Molekülen in Ensembles. Die zur Repräsentation von chemischen Verbindungen entwickelten Konzepte werden zunächst kurz eingeführt. Sie werden in den darauffolgenden Kapiteln ausführlich beschrieben.

Elektronensystem. Kovalente Bindungen werden mit Elektronensystemen beschrieben. Es gibt drei Arten von Elektronensystemen, σ -Elektronensysteme, π -Elektronensystem, und koordinative Bindungen. Einfachbindungen werden durch σ -Elektronensysteme beschrieben. Mehrfachbindungen werden in ihre orthogonalen Bindungsanteile aufgetrennt. Sie werden durch je ein σ -Elektronensystem und so viele π -Elektronensysteme, wie sie Mehrfachbindungsanteile enthalten, beschrieben. Für Koordinationsbindungen in metallorganischen Komplexen wurde ein eigener Bindungstyp eingeführt.

Wechselwirkungsgruppe. Nicht-kovalente Wechselwirkungen zwischen Atomen oder Elektronensystemen werden durch sogenannte Wechselwirkungsgruppen dargestellt. Sie enthalten die miteinander wechselwirkenden Atome oder Elektronensysteme und spezifizieren die Art und Stärke der Wechselwirkung. Diese Wechselwirkungen können sowohl intra- als auch intermolekular sein.

Molekül. Ein Molekül ist als chemische Einheit definiert, in der alle Atome mit kovalenten Bindungen gebunden sind. Nicht nur Neutralverbindungen, sondern auch Ionen oder Übergangszustände werden im vorgestellten Modell als Moleküle betrachtet.

Aggregat. Aggregate setzen sich aus mehreren Molekülen zusammen, die durch intermolekulare Wechselwirkungen miteinander verbunden sind.

Ensemble. Ein Ensemble setzt sich aus mehreren Molekülen oder Aggregaten zusammen, die im gleichen Zusammenhang verwendet werden, aber zwischen denen keinerlei Wechselwirkungen bestehen.

Die Bestandteile eines solchen Ensembles zeigt die graphische Darstellung in Abbildung 4-1. Es enthält ein Aggregat und ein Molekül. Das Aggregat setzt sich aus einem Ionenpaar zusammen, das ein Salz bildet. Eine Wechselwirkungsgruppe wird verwendet, um die elektrostatischen Wechselwirkung der beiden Ionen des Aggregats zu spezifizieren. Alle Atome in allen Molekülen sind durch σ - und π -Elektronensysteme miteinander verbunden.

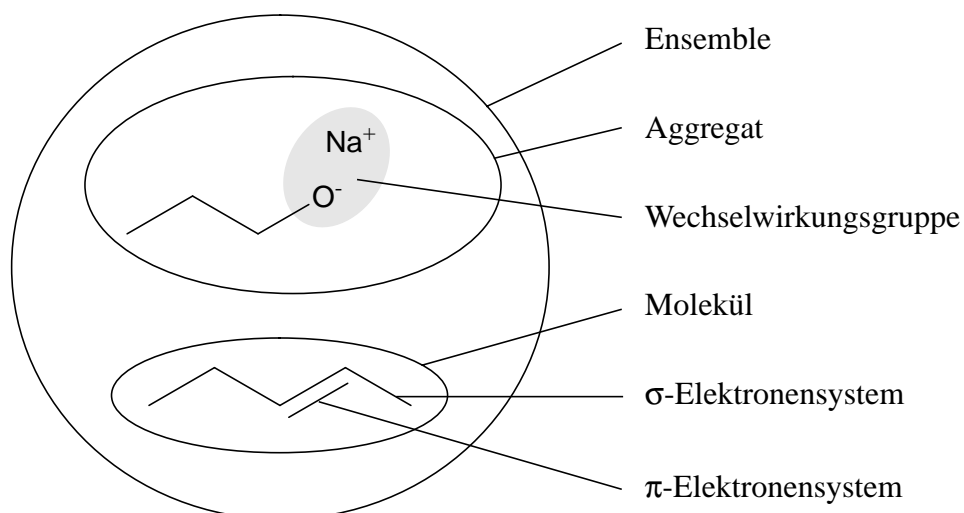


Abb. 4-1: Bestandteile eines Ensembles.

Die hierarchischen Beziehungen innerhalb eines Ensembles sind in Abbildung 4-2 gezeigt. Ein Ensemble kann aus Molekülen oder Aggregaten bestehen, ein Aggregat enthält Moleküle und die Wechselwirkungen, die zur Aggregation führen. Ein Molekül enthält Atome und Elektronensysteme sowie intramolekulare Wechselwirkungen. Die Wechselwirkungsgruppen selbst sind ebenfalls aus Atomen und Elektronensystemen zusammengesetzt.

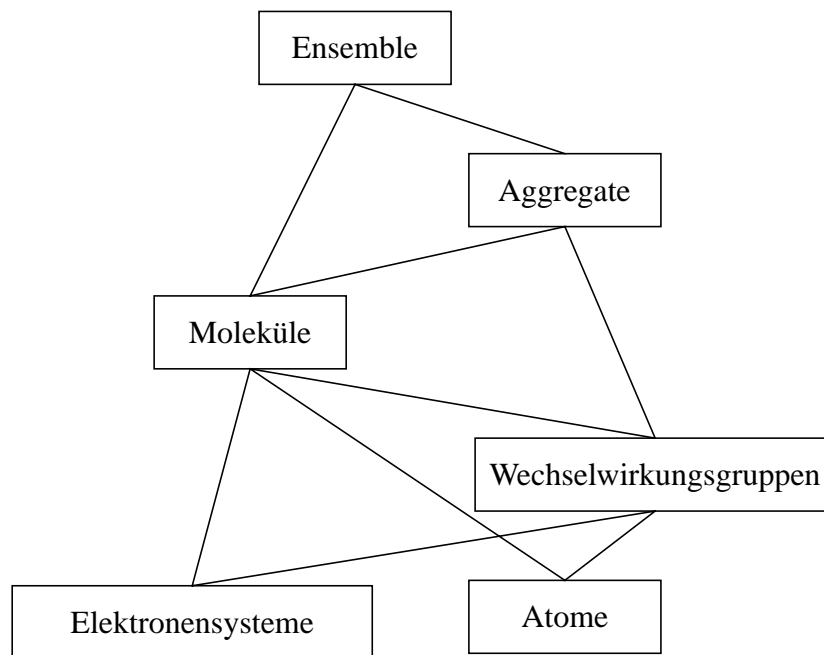


Abb. 4-2: Hierarchie innerhalb eines Ensembles.

4.1. Atome

Atome bestehen aus dem Atomkern und der Elektronenhülle. Da die Elektronen der inneren Schalen an der Bildung von Bindungen nicht beteiligt sind, werden sie im vorgestellten Modell als impliziter Bestandteil eines Atoms betrachtet. Die Valenzelektronen jedoch, die die Reaktivität und Zahl der möglichen Bindungen eines chemischen Elements bestimmen, werden gesondert behandelt. Jedes Orbital eines Atoms kann mit maximal zwei Elektronen besetzt werden. Da aber nicht nur besetzte Orbitale an der Bindungsbildung beteiligt sind, werden für jedes Atom sowohl die besetzten als auch unbesetzte Orbitale der höchsten Schale in das Modell miteinbezogen.

Tabelle 4-1 zeigt für einige Elemente die für isolierte Atome verwendete Zahl der Orbitale und die Verteilung der Valenzelektronen. Dabei werden die einzelnen Orbitale bezüglich ihres Energieniveaus nicht unterschieden. Wasserstoff und Helium sind Elemente der ersten Periode, sie besitzen nur das 1s-Orbital. Daher wird ihnen auch nur ein Orbital zugeordnet. Bei den Elementen der zweiten Periode werden das 2s- und die drei 2p-Orbitale aufgefüllt, sie erhalten so insgesamt vier Orbitale. Für die Hauptgruppenelemente der dritten und höheren Perioden werden auch die d-Orbitale berücksichtigt. Da Bindungen häufig mit Hybridorbitalen beschrieben werden und die Elemente der dritten und vierten Hauptgruppe in der Regel ein Elektron pro Orbital zur Bindungsbildung beisteuern, entspricht die angegebene Verteilung für diese Elemente nicht dem Grundzustand mit zwei Elektronen im s-Orbital. Die Orbitale von Bor und Kohlenstoff und ihrer Homologen werden alle einzeln besetzt.

Tab. 4-1: Besetzung der Valenzschalen am Beispiel der Elemente der ersten drei Perioden, sowie von je drei Elementen der Nebennmetallgruppen und der Lanthanide. n gibt die Zahl der Orbitale des Elements an.

| Element | Besetzung | | | | | | | | | | | | | | |
|---------|-----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|
| | n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| H | 1 | 1 | | | | | | | | | | | | | |
| He | 1 | 2 | | | | | | | | | | | | | |
| Li | 4 | 1 | 0 | 0 | 0 | | | | | | | | | | |
| Be | 4 | 1 | 1 | 0 | 0 | | | | | | | | | | |
| B | 4 | 1 | 1 | 1 | 0 | | | | | | | | | | |
| C | 4 | 1 | 1 | 1 | 1 | | | | | | | | | | |
| N | 4 | 2 | 1 | 1 | 1 | | | | | | | | | | |
| O | 4 | 2 | 2 | 1 | 1 | | | | | | | | | | |
| F | 4 | 2 | 2 | 2 | 1 | | | | | | | | | | |
| Ne | 4 | 2 | 2 | 2 | 2 | | | | | | | | | | |
| Na | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | |
| Mg | 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | |
| Al | 9 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | |
| Si | 9 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | | | | |
| P | 9 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | | | | |
| S | 9 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | | | | |
| Cl | 9 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | | | | | |
| Ar | 9 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | | | | | |
| Sc | 11 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| Ti | 11 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| V | 11 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| Ce | 9 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pr | 9 | 2 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Nd | 9 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Nebengruppenelemente erhalten 11 Orbitale, die von fünf d-, einem s-, drei p- und zwei weiteren d-Orbitalen stammen. Die beiden höchsten d-Orbitale werden benötigt, da sie an Bindungen in High-Spin-Komplexen teilnehmen. Abbildung 4-3 zeigt die Elektronen und Orbitale, die für Eisen in das Modell miteinbezogen werden. Die Elektronenverteilung entspricht dem Grundzustand, wie er in gängigen Lehrbüchern für diese Elemente angegeben wird [47].

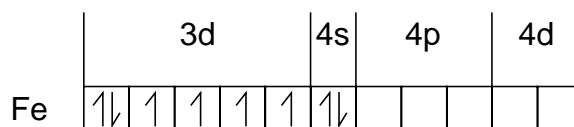


Abb. 4-3: Elektronen und Orbitale des Eisens, die im Grundzustand für das isolierte Atom verwendet werden.

Lanthanide und Actinide erhalten je 13 Orbitale. Sie entsprechen einem s-, fünf d- und sieben f-Orbitalen.

Die vollständige Liste aller verwendeten Elektronenkonfigurationen findet sich im Anhang A.

4.2. Elektronensysteme

Verschiedene Strukturrepräsentationen unterscheiden sich am stärksten voneinander in der Art, in der Bindungen dargestellt werden. Das vorgestellte Verfahren führt Neuerungen bei der Darstellung von Bindungen ein, die mit herkömmlichen, auf dem VB-Modell beruhenden Bindungslisten nicht beschrieben werden können.

Es wurden drei verschiedene Bindungstypen entwickelt. Sie werden im folgenden σ -Elektronensysteme, π -Elektronensysteme und koordinative Bindungen genannt.

σ -Elektronensysteme und π -Elektronensysteme beschreiben die orthogonalen σ - und π -Anteile einer Bindung. Jedes Elektronensystem wird durch die an ihm beteiligten Atome und die Zahl seiner Elektronen definiert. Eine Einfachbindung wird als σ -Elektronensystem beschrieben, eine Doppelbindung als ein σ -Elektronensystem und ein π -Elektronensystem, eine Dreifachbindung als ein σ -Elektronensystem und zwei π -Elektronensysteme, die jeweils aus zwei Atomen und zwei Elektronen bestehen. Freie Elektronen werden als π -Elektronensysteme dargestellt. Die Beschränkung, daß eine Bindung zwischen zwei Atomen lokalisiert sein muß, wurde aufgehoben. Sowohl σ - als auch π -Elektronensysteme können über mehrere Atome delokalisiert sein. Es gibt geschlossene und verbrückte Mehrzentren- σ -Elektronensysteme (siehe unten).

Um den Bindungsverhältnissen in Metallkomplexen, die sich grundlegend von z. B. denen zwischen Nichtmetallen unterscheiden, Rechnung zu tragen, wurde die „koordinative Bindung“ als weiterer Bindungstyp eingeführt. Er wird sowohl für Metallkomplexe mit anorganischen Liganden als auch für Organometallverbindungen verwendet. Die Verwendung der

einzelnen Bindungstypen unterliegt Einschränkungen, die in Tabelle 4-2 zusammengefasst sind.

Tab. 4-2: Typen von Elektronensystemen, die in Bindungen zwischen Nichtmetallelementen (E) und Metallen (M) erlaubt sind.

| an der Bindung beteiligte Elemente | möglicher Elektronensystemtyp |
|------------------------------------|-------------------------------|
| E - E | σ, π |
| M - M | σ, π |
| M - E | coord |

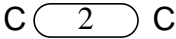
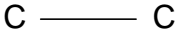
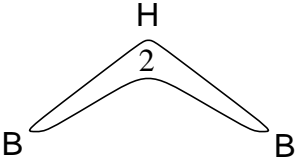
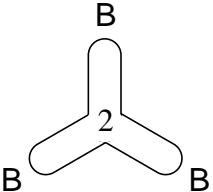
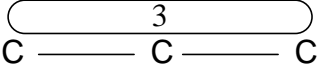

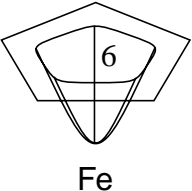
Da die Grenze zwischen Metallen und Nichtmetallen nicht scharf gezogen ist, wurden solche Elemente als Metalle definiert, die metallorganische Verbindungen mit Liganden mit einer Haptizität größer eins eingehen. Tabelle 4-3 gibt an, welche Elemente dies sind.

Tab. 4-3: Elemente, die im vorgestellten Modell als Metalle betrachtet werden.

| Gruppe | Metallelemente |
|-----------------------|----------------|
| 1. und 2. Hauptgruppe | alle |
| 3. Hauptgruppe | Al, Ga, In, Tl |
| 4. Hauptgruppe | Ge, Sn, Pb |
| 5. bis 8. Hauptgruppe | keine |
| Nebengruppen | alle |
| Lanthanide | alle |
| Actinoide | alle |

Tabelle 4-4 gibt einen Überblick über die verschiedenen Darstellungsformen für Elektronensysteme, die in den folgenden Kapiteln Verwendung finden. In graphischen Darstellungen werden die an einem Elektronensystem beteiligten Atome durch ein Oval miteinander verbunden, die Zahl der in diesem System lokalisierten Elektronen wird in diesem Oval angegeben. Zur Vereinfachung können reguläre Zweielektronen-Zweizentren- σ -Elektronensysteme jedoch auch mit der für Einfachbindungen gebräuchlichen Linie dargestellt werden.

Tab. 4-4: Unterschiedliche Darstellungsformen von Elektronensystemen.

| Art des Elektronensystems | graphische Darstellung | Notation |
|--|---|--------------------|
| reguläres σ -Elektronensystem |  | (1 2 s r; 2) |
| reguläres σ -Elektronensystem |  | (1 2 s r; 2) |
| verbrücktes σ -Elektronensystem |  | (1 2 3 s b 2; 2) |
| geschlossenes σ -Elektronensystem |  | (1 2 3 s c; 2) |
| π -Elektronensystem |  | (1 2 3 p; 3) |
| freie Elektronen als π -Elektronensystem |  | (1 p; 2) |
| koordinative Bindung |  | (1 2 3 4 5 6 c; 6) |

Zur Notation von Elektronensystemen wird folgendes Schema verwendet: Ein Paar runder Klammern umfaßt jeweils ein Elektronensystem. Die ersten Ziffern geben die an dem Elektronensystem beteiligten Atome an, danach folgt die Bezeichnung des Typs, wobei s für σ -Elek-

tronensystem, p für π -Elektronensystem und c für eine koordinative Bindung steht. σ -Elektronensysteme werden durch einen zweiten Buchstaben noch genauer spezifiziert: r für reguläre (engl. regular), b für verbrückte (engl. bridging) und c für geschlossene (engl. closed) σ -Elektronensysteme. In verbrückten σ -Elektronensystemen wird zusätzlich das zentrale Atom angegeben. Die Zahl der Elektronen folgt nach dem Semikolon.

4.2.1. σ -Elektronensysteme

Es gibt drei verschiedene Arten von σ -Elektronensystemen: reguläre σ -Elektronensysteme, geschlossene und verbrückte Mehrzentren- σ -Elektronensysteme. Reguläre σ -Elektronensysteme (siehe Abbildung 4-4a) bestehen aus zwei Atomen und ein oder zwei Elektronen, sie entsprechen also Einfachbindungen bzw. ionisierten Einfachbindungen. Verbrückte σ -Elektronensysteme (siehe Abbildung 4-4b) beschreiben Mehrzentrenbindungen mit einem zentralen Atom. Darunter fallen unter anderem BHB-Bindungen in Boranen. Geschlossene σ -Elektronensysteme (siehe Abbildung 4-4c) werden zur Darstellung zentrierter Mehrzentrenbindungen verwendet, in denen sich die bindenden Orbitale im Zentrum aller beteiligten Atome überlappen. Beide Typen können Elektronenüberschuß- oder Elektronenmangelbindungen beschreiben. An jedem σ -Elektronensystem muß jedoch mindestens ein Elektron beteiligt sein, und die Anzahl der Elektronen pro beteiligtem Atom darf zwei nicht überschreiten. In geschlossenen Elektronensystemen sind alle Atome als benachbart definiert, in verbrückten Systemen sind die Eckatome nur jeweils zum zentralen Atom benachbart, nicht jedoch untereinander. So sind alle Boratome eines geschlossenen BBB-Elektronensystems benachbart, in einem BHB-Elektronensystem dagegen besteht Nachbarschaft nur zwischen dem Wasserstoffatom und den Boratomen, nicht jedoch zwischen den beiden Boratomen direkt.

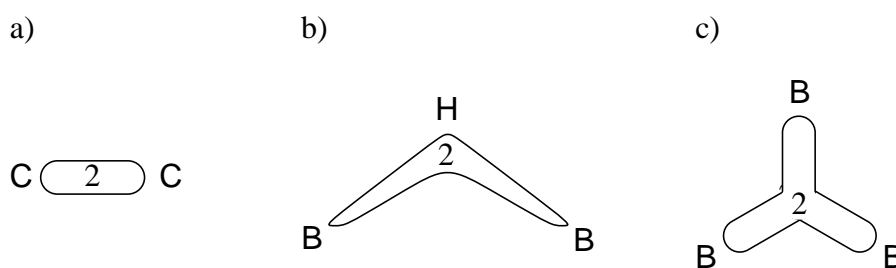


Abb. 4-4: Die drei Typen von σ -Elektronensystemen, a) regulär, b) verbrückt, und c) geschlossen.

Borane wie B_5H_{11} können unter Verwendung dieser drei σ -Elektronensystemtypen beschrieben werden. B_5H_{11} besteht aus drei verbrückten σ -Elektronensystemen, (1 2 8 s b 8; 2), (2 4 10 s b 10; 2) und (4 5 14 s b 14; 2), zwei geschlossenen σ -Elektronensystemen, (1 2 3 s c; 2) und (3 4 5 s c; 2), sowie acht regulären σ -Elektronensystemen (Abbildung 4-5).

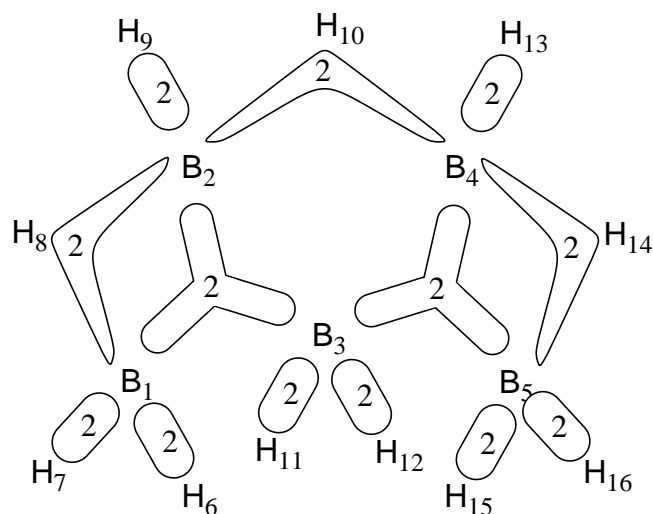


Abb. 4-5: B_5H_{11} als Beispiel einer anorganischen Verbindung mit Elektronenmangelbindungen. Atomlabel sind als Subskripte angegeben.

Die VB-Struktur von SF_4 zeigt Abbildung 4-6a. Das Schwefelatom hat 10 Valenzelektronen, die von 4 Einfachbindungen und dem freien Elektronenpaar stammen. Sie bedingen eine trigonal-bipyramidale Geometrie, wobei das freie Elektronenpaar eine äquatoriale Position einnimmt. Eine genauere Analyse der Energien der an den Bindungen beteiligten Hybridorbitale zeigte jedoch, daß die axialen SF-Bindungen besser durch charge-transfer Konfigurationen oder eine Dreizentren-Vierelektronenbindung beschrieben werden [48]. In der vorgestellten Kodierung werden zwei der Fluoratome durch ein verbrücktes σ -Elektronensystem mit vier Elektronen an Schwefel gebunden, die beiden anderen S-F-Bindungen sind reguläre σ -Elektronensysteme (Abbildung 4-6b). Das Schwefelatom besitzt außerdem vier leere π -Elektronensysteme, die von seinen d-Orbitalen stammen. Jedem Fluoratom sind für die freien Elektronenpaare drei π -Elektronensysteme mit je zwei Elektronen zugeordnet. Folgende Elektronensysteme stellen die vollständige Struktur dar:

(1 2 3 s b 1; 4) (1 p; 2) (1 4 s r; 2) (1 5 s r; 2) (1 p; 0) (1 p; 0) (1 p; 0) (1 p; 0) (2 p; 2)
 (2 p; 2) (2 p; 2) (3 p; 2) (3 p; 2) (3 p; 2) (4 p; 2) (4 p; 2) (4 p; 2) (5 p; 2) (5 p; 2) (5 p; 2).

In Abbildung 4-6 wurden außer dem freien Elektronenpaar am Schwefelatom alle π -Elektronensysteme weggelassen, um eine übersichtlichere Darstellung zu erhalten.

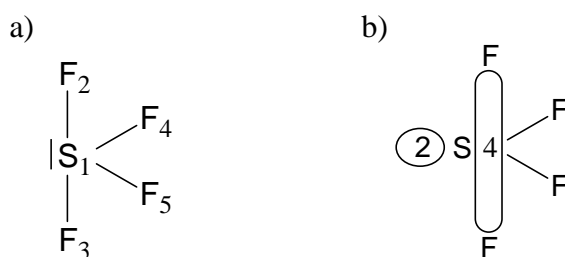


Abb. 4-6: VB-Darstellung (a) und RICOS-Repräsentation (b) von SF_4 . Die leeren π -Elektronensysteme am Schwefel und die freien Elektronenpaare an den Fluoratomen wurden der Übersichtlichkeit halber weggelassen.

In organischen Verbindungen treten hauptsächlich reguläre σ -Elektronensysteme auf. Bei der Behandlung von Alkanen in der Massenspektrometrie werden sie jedoch ionisiert und enthalten bisweilen nur ein Elektron. Abbildung 4-7 zeigt das Ionisationsprodukt von 2,2,3,3-Tetramethylbutan nach Elektronenstoßionisation. Das mittlere C-C σ -Elektronensystem wird bevorzugt ionisiert [20], das resultierende Elektronensystem enthält nur noch ein Elektron.

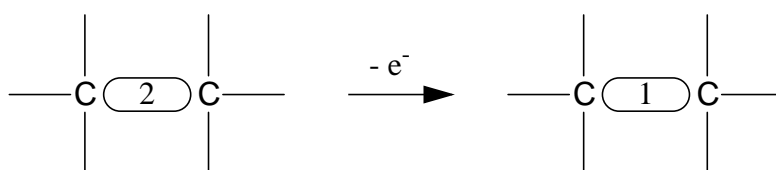


Abb. 4-7: Elektronenstoßionisation von 2,2,3,3-Tetramethylbutan.

Mehrzentren- σ -Elektronensysteme eignen sich auch, um Zwischenstufen oder Übergangszustände bei der Modellierung von Reaktionen zu beschreiben.

Nichtklassische Kationen treten in Substitutionsreaktionen als Zwischenstufen [49] auf. Sie besitzen delokalisierte σ -Bindungen, die als σ -Elektronensysteme dargestellt werden können. Das 2-Norborylkation enthält ein verbrücktes σ -Elektronensystem (Abbildung 4-8) mit drei Zentren und zwei Elektronen. Das Kohlenstoffatom 3 ist dessen zentrales Atom, die Atome 1 und 2 sind endständig. Das gesamte Kohlenstoffgerüst besteht aus den folgenden Elektronensystemen:

(1 2 s r; 2) (1 2 6 s b 6; 2) (1 7 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (4 5 s r; 2) (4 7 s r; 2) (5 6 s r; 2)

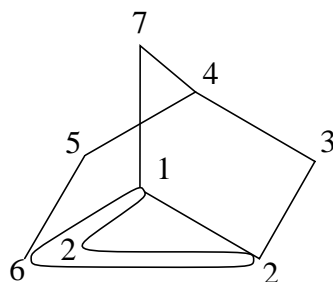


Abb. 4-8: Repräsentation des 2-Norbornylkations mit Hilfe eines Dreizentren-Zweielektronen- σ -Elektronensystems.

Der Übergangszustand von S_N2 Reaktionen läßt sich ebenfalls unter Verwendung einer Mehrzentrenbindung darstellen (Abbildung 4-9). Da aber sowohl die freien Elektronen der angreifenden Gruppe als auch die Bindungselektronen, die zum freien Elektronenpaar der Abgangsgruppe werden, in das Elektronensystem miteinbezogen werden, befinden sich in dem verbrückten σ -Elektronensystem vier Elektronen.

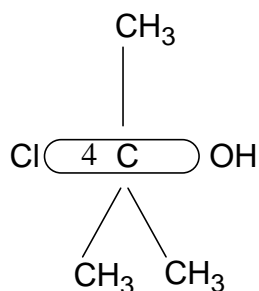


Abb. 4-9: Übergangszustand einer S_N2 -Reaktion.

4.2.2. π -Elektronensysteme

π -Elektronensysteme beschreiben π -Bindungen. Wenn mehrere orthogonale π -Bindungen vorhanden sind, wird für jede dieser Bindungen ein π -Elektronensystem gebildet (Abbildung 4-10a). π -Elektronensysteme können über mehrere Atome, die mit σ -Elektronensystemen verbunden sind, delokalisiert sein (Abbildung 4-10b). Auch freie Elektronen werden als π -Elektronensysteme dargestellt, da sie ebenfalls zu σ -Elektronensystemen orthogonal sind und sich in Reaktionen oft ähnlich wie π -Bindungen verhalten. π -Elektronensysteme, die freie Elektronen repräsentieren, enthalten genau ein Atom und maximal zwei Elektronen. Wenn ein Atom mehr als zwei freie Elektronen besitzt, muß eine entsprechende Anzahl π -Elektronensysteme gebildet werden (Abbildung 4-10c). Auch leere Orbitale werden als π -Elektronensysteme

repräsentiert. Sie enthalten ein Atom und keine Elektronen (Abbildung 4-10d). Kationen werden unter Verwendung leerer π -Elektronensysteme dargestellt.

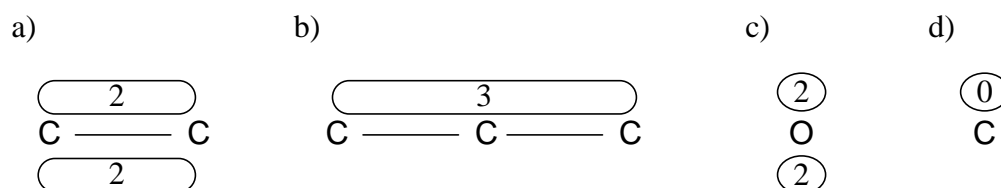


Abb. 4-10: π -Elektronensysteme. a) orthogonal, b) delokalisiert, c) freie Elektronenpaare, d) leere Orbitale

Die in Kapitel 4.1 vorgestellten Elektronenkonfigurationen für isolierte Atome werden mit π -Elektronensystemen realisiert. Elemente der zweiten Periode erhalten vier Elektronensysteme, wie in Abbildung 4-11 gezeigt ist. So wird ein Lithiumatom mit einem ein Elektron enthaltenden und drei leeren π -Elektronensystemen dargestellt. Bei Bor und seinen Homologen sind drei Elektronensysteme mit je einem Elektron besetzt und eines bleibt leer, in der Kohlenstoffgruppe sind alle Elektronensysteme mit je einem Elektron besetzt, bei den Elementen der höheren Gruppen sind schließlich auch paarweise besetzte Elektronensysteme zu finden.

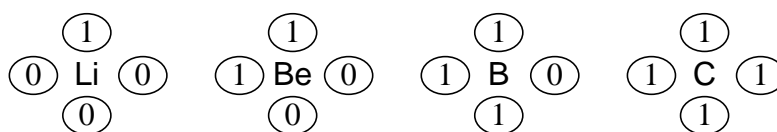


Abb. 4-11: Isolierte Atome der zweiten Periode mit π -Elektronensystemen, die die Valenzelektronen repräsentieren.

Die Kodierung von freien Elektronen als π -Elektronensysteme erlaubt außerdem auch, verschieden angeregte Zustände von Verbindungen zu beschreiben, in denen nicht alle π -Elektronensysteme vollständig besetzt sind. Carbene kommen sowohl im Singulett als auch im Triplett-Zustand vor. Beide Zustände lassen sich beschreiben, da nicht mehr alle freien Elektronen zusammengefaßt sind. Im Singulett-Carben besetzen beide Elektronen das selbe π -Elektronensystem, im Triplettcarben sind sie auf zwei Systeme verteilt (Abbildung 4-12).

Die Elektronensysteme beider Formen lauten:

C: 1 H: 1,2

Singulett-Carben: (1 2 s r; 2) (1 3 s r; 2) (1 p; 2) (1 p; 0)

Triplett-Carben: (1 2 s r; 2) (1 3 s r; 2) (1 p; 1) (1 p; 1)

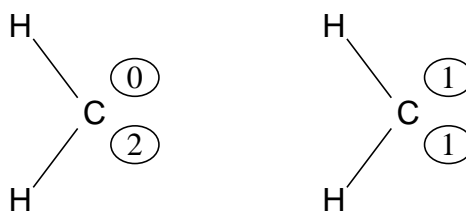


Abb. 4-12: Besetzung der π -Elektronensysteme in Singulett- und Triplett-Carben.

π -Elektronen, die miteinander konjugiert sind, werden in einem π -Elektronensystem zusammengefaßt. Sie können aus konjugierten Doppelbindungen, aromatischen Systemen oder freien Elektronenpaaren, die einer Doppelbindung benachbart sind, stammen. Das Kohlenstoffgerüst in Benzol wird mit sechs regulären σ -Elektronensystemen und einem π -Elektronensystem mit 6 Elektronen beschrieben, das sich über den Ring erstreckt (Abbildung 4-13a): (1 2 s r; 2) (1 6s r; 2) (2 3 s r; 2) (3 4 s r; 2) (4 5 s r; 2) (5 6 s r; 2) (1 2 3 4 5 6 p; 6)

In Buta-1,3-dien sind zwei Doppelbindungen miteinander konjugiert, das Kohlenstoffgerüst besteht aus folgenden Elektronensystemen:

(1 2 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (1 2 3 4 p; 4) (Abbildung 4-13b).

Im Enolatanion ist ein freies Elektronenpaar mit der Doppelbindung konjugiert, die entsprechenden Elektronensysteme lauten:

(1 2 s r; 2) (2 3 s r; 2) (1 2 3 p; 4) (3 p; 2) (3 p; 2) mit C: 1, 2 und O: 3 (Abbildung 4-13c).

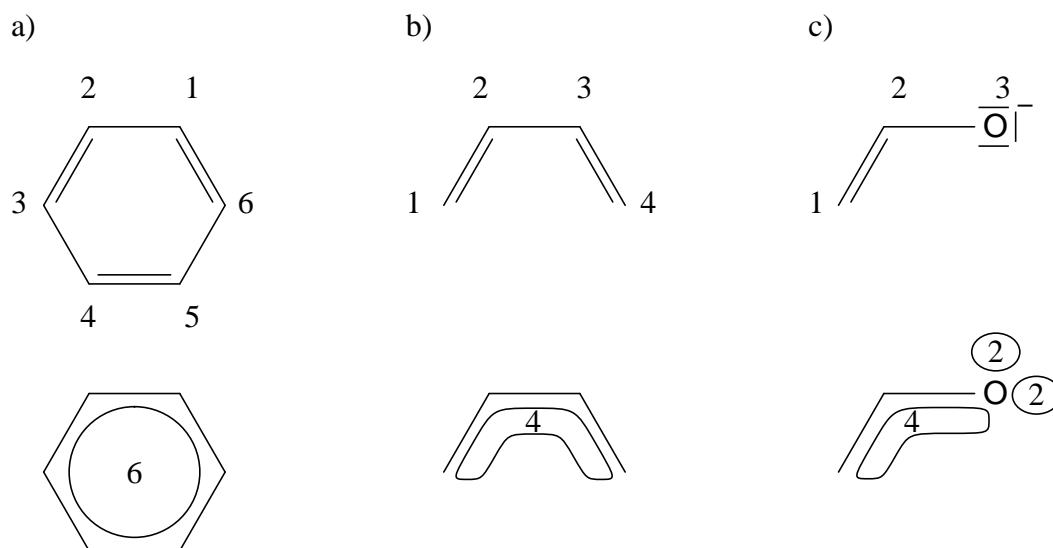


Abb. 4-13: Verbindungen mit delokalisierten π -Systemen in VB- und RICOS-Darstellungsweise.

Die Einführung von π -Elektronensystemen ermöglicht es, Mehrdeutigkeiten, die bei der Modellierung von orthogonalen π -Systemen mit dem VB-Modell auftreten, aufzulösen. Sie treten immer dann auf, wenn sich verschieden große π -Systeme überlappen. Dies sei am Beispiel von Tolan verdeutlicht.

Für Tolan sind zwei verschiedene Geometrien denkbar, planar (Abbildung 4-14a) oder mit beiden Phenylringen orthogonal zueinander (Abbildung 4-14b). Im Kristall liegt die planare Konformation vor [50]. Diese Geometrien resultieren aus unterschiedlicher Konjugation der π -Elektronen. Die planare Form ergibt sich, wenn beide Phenylringe über die Alkinbindung miteinander konjugiert sind, die nicht planare Form entsteht, wenn die Elektronen jedes Phenylrings mit einem anderen Orbital der Dreifachbindung überlappen. Diese Unterschiede sind mit einer Valence-Bond-Kodierung nicht darstellbar, da diese keine Information über Konjugation enthält. Mit π -Elektronensystemen lassen sich beide Formen klar voneinander unterscheiden. Das σ -Elektronensystemgerüst ist für beide Formen gleich. Die planare Form wird jedoch durch zwei verschieden große π -Elektronensysteme beschrieben, eines erstreckt sich über das ganze Molekül, das andere nur über die beiden Kohlenstoffatome der Dreifachbindung. Sie lauten wie folgt:

(1 2 3 4 5 6 7 8 9 10 11 12 13 14 p; 14) (7 8 p; 2).

Die nicht planare Form besteht aus zwei gleich großen π -Elektronensystemen, die jeweils einen Phenylring und die Dreifachbindung miteinander verbinden, (1 2 3 4 5 6 7 8 p; 8)

(7 8 9 10 11 12 13 14 p; 8).

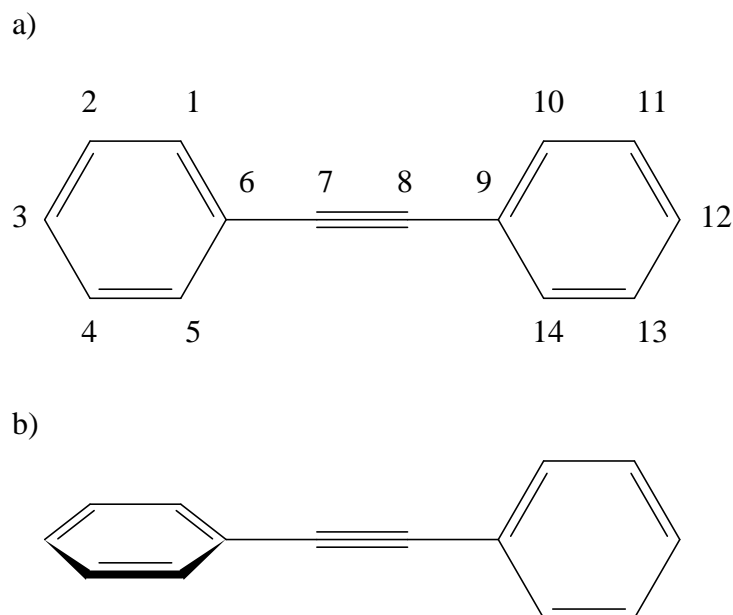


Abb. 4-14: Zwei Geometrien von Tolan, die sich nicht durch eine Valence-Bond-Beschreibung unterscheiden lassen.

Da beide Formen nun unterscheidbar sind, können ihnen auch unterschiedliche Eigenschaften wie ihre räumliche Gestalt zugeordnet werden.

π -Elektronensysteme ermöglichen jedoch nicht nur die Unterscheidung orthogonaler Elektronensysteme, sondern auch die korrekte Beschreibung äquivalenter Atome in konjugierten Systemen. Somit ist es nicht mehr notwendig, mehrere mesomere Grenzstrukturen einer Substanz zu verwenden, um sie adäquat zu beschreiben. Die Darstellung von Nitrogruppen löst unter Verwendung einer VB-Repräsentation immer Schwierigkeiten aus. Es gibt verschiedene weit verbreitete Ansätze, sie darzustellen: Eine valenzchemisch richtige Repräsentation ist nur unter Ladungstrennung möglich. Einerseits kann das Stickstoffatom der Nitrogruppe einfach positiv geladen sein, und unterschiedliche Bindungen zu den Sauerstoffatomen haben. So sind zwei mesomere Formen notwendig, um zu zeigen, daß beide Sauerstoffatome äquivalent sind (Abbildung 4-15a). Andererseits kann der Stickstoff als zweifach positiv geladen betrachtet werden, und jedes Sauerstoffatom als einfach negativ geladen (Abbildung 4-15b). Dies ist aber keine sehr gebräuchliche Form der Darstellung, die auch die Delokalisation der Elektronen über alle drei Atome nicht berücksichtigt. Die dritte, häufig verwendete Kodierung vermeidet zwar Ladungsaufteilung, ordnet dem Stickstoffatom aber 10 Valenzelektronen zu, was für dieses Element verboten ist (Abbildung 4-15c). Es gibt also keine zufriedenstellende Darstellungsweise. Mit der RICOS-Repräsentation wird eine Kodierung möglich, die sowohl der Delokalisation als auch der Valenzchemie Rechnung trägt (Abbildung 4-15d). Zwischen dem Stickstoffatom und den Sauerstoffatomen besteht neben den σ -Elektronensystemen ein π -Elektronensystem, das vier Elektronen enthält. Beide Sauerstoffatome haben zudem noch je vier freie Elektronen, die auf je zwei π -Elektronensysteme verteilt werden. Beide Sauerstoffatome haben gleichartige Elektronensysteme, sie sind also äquivalent.

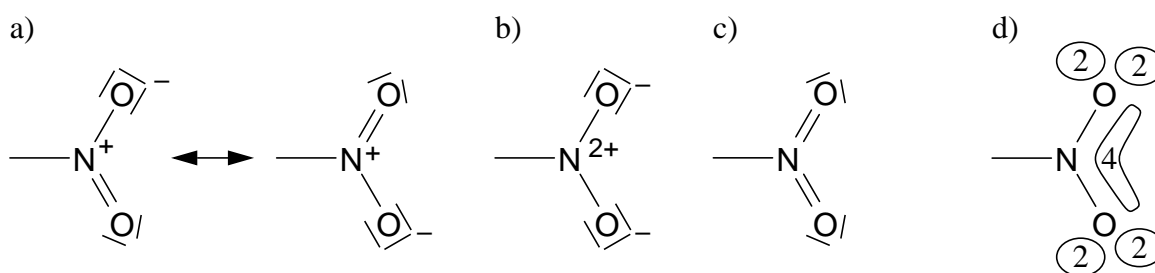


Abb. 4-15: Repräsentation der Nitrogruppe. a) - c) traditionelle Kodierungen, d) RICOS

Unter Einsatz von π -Elektronensystemen können auch Spezies, die in der Massenspektrometrie auftreten, eindeutig beschrieben werden. Im neutralem Furan wird ein Elektronenpaar des Sauerstoffs in den aromatischen Ring doniert, das andere ist dazu orthogonal. Wenn ein freies Elektronenpaar des Furans ionisiert wird, können zwei verschiedene Produkte entstehen. Zum einen wird das π -System des aromatischen Rings ionisiert, zum anderen das isolierte freie Elektronenpaar. Beide Produkte sind mit einer VB-Repräsentation, wie sie in Abbildung 4-16a

gezeigt ist, ununterscheidbar. Die beiden π -Elektronensysteme, die zur Beschreibung der π -Elektronen in Furan dienen, lassen sich unabhängig voneinander ionisieren, man erhält also auch zwei verschiedene Ionisationsprodukte (Abbildung 4-16b).

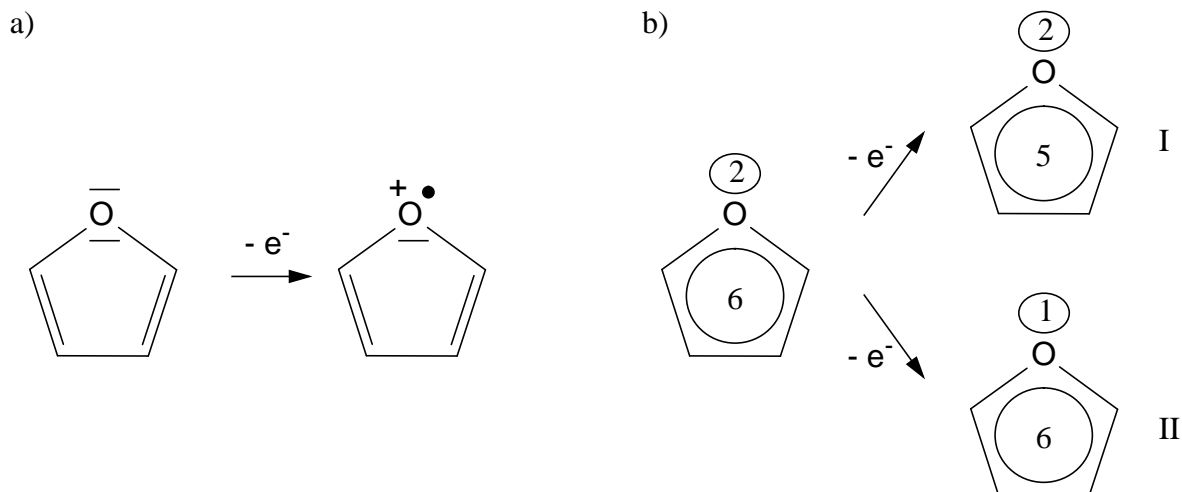


Abb. 4-16: Ionisation von Furan.

Auch das Problem, daß die beiden Ionisationsprodukte in Amiden nicht unterschieden werden können, wird mit dieser Repräsentation gelöst. Die beiden verschiedenen Ionisationsprodukte eines Amids sind in Abbildung 4-17 dargestellt. Einmal ist das delokalisierte π -Elektronensystem ionisiert, das andere Mal ein freies Elektronenpaar.

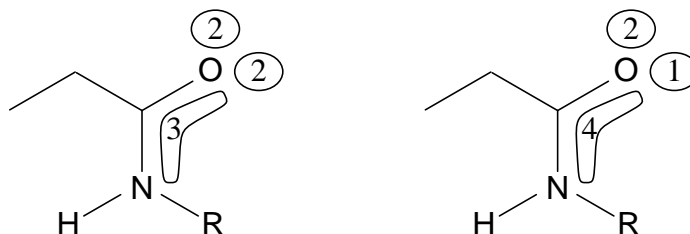


Abb. 4-17: Zwei Ionisationsprodukte eines Amids.

4.2.3. Koordinative Bindungen

In der Komplexchemie gibt es eine Vielfalt unterschiedlicher Bindungen zwischen Metallatomen und ihren Liganden. Aufbauend auf einer Klassifizierung der verschiedenen Bindungsformen wird im folgenden ein allgemeines Modell zu ihrer Beschreibung vorgestellt.

Bindungen zwischen Metallen und organischen Liganden unterscheiden sich in der Zahl der an der Bindung beteiligten Atome. Dabei gibt es zwei Varianten. Einerseits können mehrere Atome eines Liganden an ein Metallatom gebunden sein. Die Anzahl der gebundenen Atome

wird als Haptizität des Liganden bezeichnet und mit dem Buchstaben η gekennzeichnet. Für eine Haptizität von 1 bis 8 sind Verbindungen bekannt. Abbildung 4-18 zeigt je ein Beispiel für η^1 - bis η^6 -Bindungen.

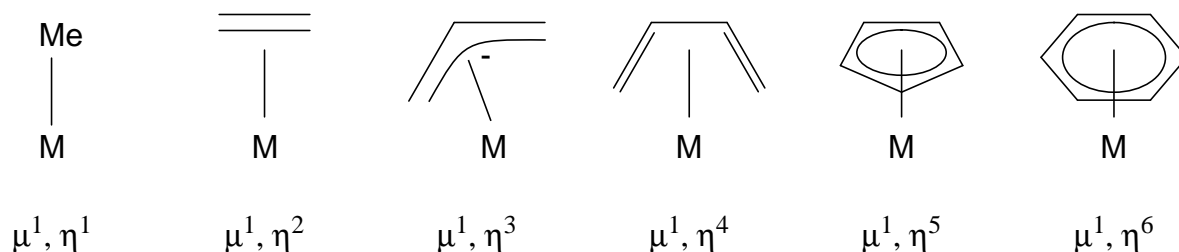


Abb. 4-18: Komplexbindungen mit Liganden unterschiedlicher Haptizität

Andererseits können mehrere Metallatome an der Bindung zum selben Liganden beteiligt sein. Solche Bindungen werden als verbrückt bezeichnet. Die Anzahl der verbrückten Metallatome wird mit dem Buchstaben μ gekennzeichnet. Metallcarbonyle bilden mit Metallen der höheren Nebengruppenelemente häufig verbrückte Komplexe aus. Dabei kann Kohlenmonoxid bis zu drei Metallatome verbrücken. Außerdem kann es zusätzlich mit einer seiner π -Bindungen eine weitere η^2 -Bindung zu einem benachbarten Metallatom eingehen (Abbildung 4-19).

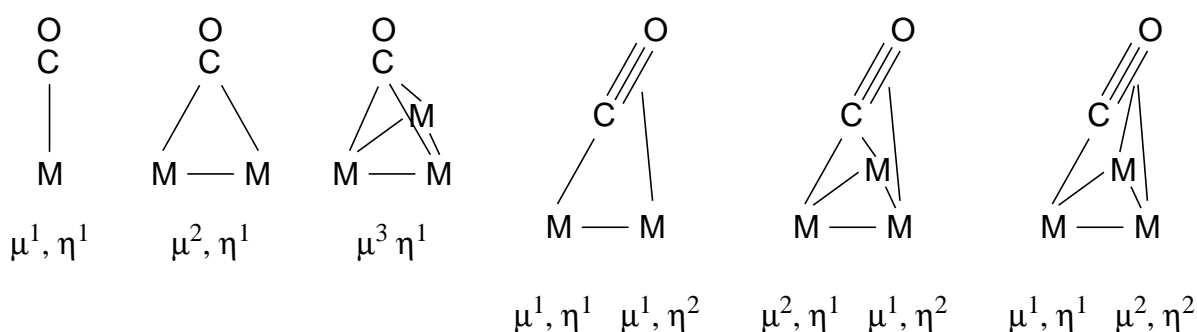


Abb. 4-19: Unterschiedliche Arten von Komplexbindungen zwischen Metallatom und Kohlenmonoxid.

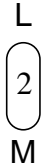
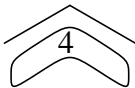
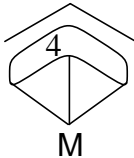
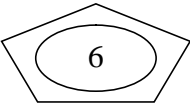
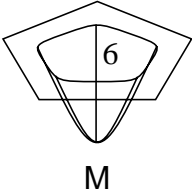
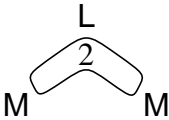
In Komplexen mit anorganischen Liganden, wie z. B. $[\text{Cu}(\text{NH}_3)_4]^{2+}$, ist jeweils ein Atom des anorganischen Neutralmoleküls oder Anions durch ein freies Elektronenpaar an das Zentralatom gebunden. Sie lassen sich also formal in die Klasse der Bindungen einordnen, die in metallorganischen Verbindungen mit μ^1, η^1 bezeichnet werden.

Für alle Arten von Metall-Ligand-Bindungen wurde ein gemeinsamer Bindungstyp eingeführt, die sogenannte koordinative Bindung. Eine koordinative Bindung wird aus einem Metallatom und einem Liganden gebildet, der Elektronen aus einem π -Elektronensystem in die koordinative Bindung beisteuert. Dabei wird nur die Hinbindung gefüllter Orbitale vom Ligand zum

Metall berücksichtigt, nicht jedoch die Rückbindung vom Metall zum Ligand, da es in der Regel nicht möglich ist, diese Elektronen einem Liganden zuzuordnen und in einem Elektronensystem zu lokalisieren.

Zur Bildung einer koordinativen Bindung wird pro Elektronenpaar in der Bindung jeweils ein leeres Orbital des Metallatoms herangezogen. So steuert das Metallatom für einen Liganden mit einem freien Elektronenpaar ein leeres π -Elektronensystem bei, für eine Bindung zu einem Allylanion oder Buta-1,3-dien zwei, und in Cyclopentadienylkomplexen drei leere π -Elektronensysteme. Tabelle 4-5 zeigt je ein Beispiel für die Bildung unverbrückter koordinativer Bindungen mit zwei, vier und sechs Elektronen. Die unterste Zeile illustriert die Bildung einer verbrückten koordinativen Bindung.

Tab. 4-5: Bildung koordinativer Bindungen.

| Metall | Ligand | koordinative Bindung | Verbrückung |
|--------|---|--|-------------|
| M (0) | (2) L |  | |
| M (0) |  |  | unverbrückt |
| M (0) |  |  | unverbrückt |
| M (0) | (2) L |  | verbrückt |

Die Verwendung koordinativer Bindungen in Metallverbindungen wird anhand einiger Beispiele illustriert, stellvertretend für Verbindungen der Hauptgruppenelemente seien die Repräsentationen von Bleitetraethyl und Methylithium näher erläutert.

In Bleitetraethyl (Abbildung 4-20) ist jede Ethylgruppe über eine koordinative Bindung an das Pb^{4+} -Kation gebunden. Dem Bleiatom sind im Grundzustand neben den besetzten s- und p-Orbitalen noch 5 leere d-Orbitale zugeordnet, es besitzt also insgesamt 9 π -Elektronensysteme. Vier dieser π -Elektronensysteme werden für die koordinativen Bindungen verwendet, die anderen 5 bleiben leer. Sie sind in Abbildung 4-20 der Übersichtlichkeit wegen weggelassen. Die vollständige Liste der Elektronensysteme am Bleiatom lautet: (1 2 c; 2) (1 3 c; 2) (1 4 c; 2) (1 5 c; 2) (1 p; 0) (1 p; 0) (1 p; 0) (1 p; 0) (1 p; 0).

Alle C-C und C-H Bindungen der Ethylgruppen sind reguläre σ -Elektronensysteme.

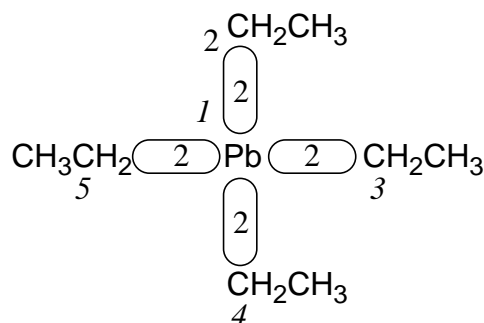


Abb. 4-20: Bleitetraethyl als Hauptgruppenelementorganyl mit vier koordinativen Systemen.

Methylithium liegt in der Gasphase als Tetramer, $\text{Li}_4(\text{CH}_3)_4$, vor [51]. Die vier Lithiumatome bilden einen Tetraeder, jede Seite des Tetraeders wird von einer Methylgruppe überbrückt. Wenn die Bindung zwischen einer Methylgruppe und dreier Lithiumatome als 4-Zentren-2-Elektronenbindung betrachtet wird, wird diese mit einer koordinativen Bindung dargestellt (Abbildung 4-21). Jede dieser koordinativen Bindungen setzt sich aus je einem leeren π -Elektronensystem der Lithiumatome und dem zweifach besetzten π -Elektronensystem der Methylgruppe zusammen. Jedes Lithiumatom hat zusätzlich noch ein leeres Orbital, das von den Ecken des Tetraeders wegzeigt. Auch dieses wird als π -Elektronensystem repräsentiert. Der vollständige Satz der Elektronensysteme zur Beschreibung des $\text{Li}_4(\text{CH}_3)_4$ lautet:

Li: 1- 4, C: 5 - 8, H: 9 - 20

Koordinative Bindungen:

(1 2 3 5 c; 2) (1 2 4 6 c; 2) (1 3 4 7 c; 2) (2 3 4 8 c; 2)

Unbesetzte π -Elektronensysteme der Lithiumatome:

(1 p; 0) (2 p; 0) (3 p; 0) (4 p; 0)

C-H σ -Elektronensysteme der 4 Methylgruppen:

(5 9 s r; 2) (5 10 s r; 2) (5 11 s r; 2) (6 12 s r; 2) (6 13 s r; 2) (6 14 s r; 2)

(7 15 s r; 2) (7 16 s r; 2) (7 17 s r; 2) (8 18 s r; 2) (8 19 s r; 2) (8 20 s r; 2)

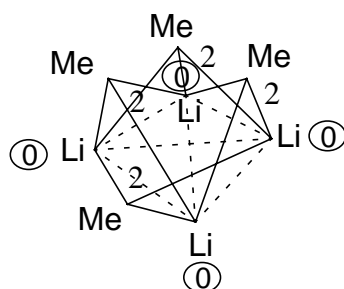


Abb. 4-21: Vier verbrückte koordinative Bindungen in $\text{Li}_4(\text{CH}_3)_4$. Der Lithiumtetraeders ist durch gestrichelte Linien markiert, die keine Bindungen darstellen.

Am Beispiel von 1,1'-Dimethylmanganocen und Re_3Cl_9 soll die Repräsentation von Nebengruppenmetallverbindungen in einkernigen Komplexen und mehrkernigen Clustern gezeigt werden.

1,1'-Dimethylmanganocen ist ein Sandwichkomplex, der verdünnten Lösungen in zwei verschiedenen Zuständen nebeneinander vorliegt, einerseits als Highspin-, andererseits als Lowspinkomplex [52]. Die 5 Valenzelektronen des Mangan(II)kations liegen somit entweder ungepaart oder gepaart vor.

Zur Ausbildung der koordinativen Bindungen zwischen dem Mangandikation und den beiden Methylcyclopentadienanionen werden je drei leere π -Elektronensysteme des Mangans benötigt. Die fünf Valenzelektronen des Mangans werden nun auf zwei verschiedene Arten auf die restlichen fünf π -Elektronensysteme verteilt. Im Highspinkomplex ist jedes einfach besetzt (Abbildung 4-22a), im Lowspinkomplex sind je zwei Elektronen paarweise besetzt, eines einfach und zwei bleiben leer (Abbildung 4-22b). Es ergeben sich damit folgende Elektronensysteme am Mangankation:

Mn: 1, C: 2 - 13

Highspin Form: (1 2 3 4 5 6 c; 6) (1 8 9 10 11 12 c; 6) (1 p; 1) (1 p; 1) (1 p; 1) (1 p; 1) (1 p; 1)

Lowspin Form: (1 2 3 4 5 6 c; 6) (1 8 9 10 11 12 c; 6) (1 p; 2) (1 p; 2) (1 p; 1) (1 p; 0) (1 p; 0)

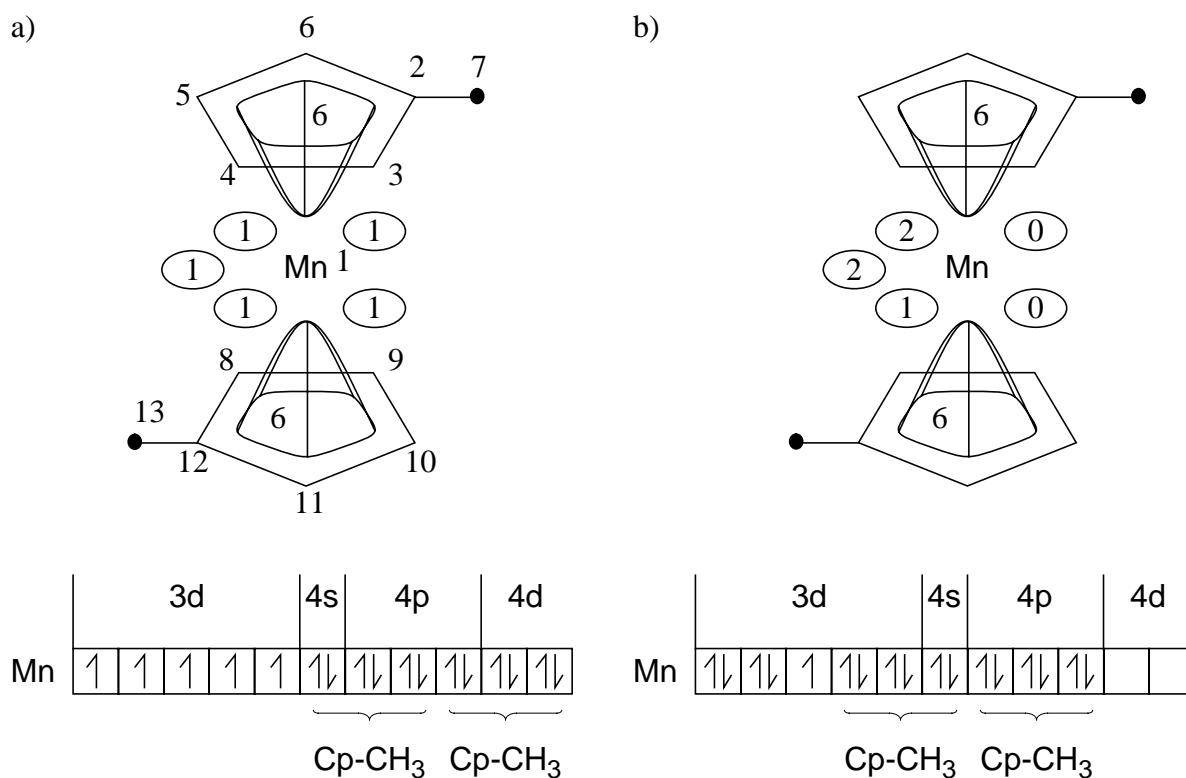


Abb. 4-22: Highspin- (a) und Lowspin-Komplex (b) von 1,1'-Dimethylmanganocen. Im unteren Teil der Abbildung ist jeweils die Elektronenkonfiguration des Mangans gegeben.

Bindungen zwischen Metallatomen in Clusterverbindungen werden als σ - und π -Elektronensysteme dargestellt, da in solchen Bindungen beide Metallatome Elektronen in das Elektronensystem donieren. Die Bindungsordnung der M-M-Bindungen beträgt maximal vier, zwei Metallatome können also mit je einem σ -Elektronensystem und höchstens drei π -Elektronensystemen verbunden sein.

Abbildung 4-23 zeigt die Beteiligung von σ - und π -Elektronensystemen sowie von koordinativen Bindungen in Clustern am Beispiel von Re_3Cl_9 . Die drei Rhenium(III)kationen sind jeweils mit Doppelbindungen aneinander gebunden [53], die mit je einem σ - und einem π -Elektronensystem repräsentiert werden. Zur Ausbildung dieser Bindungen werden die vier Valenzelektronen jedes Rhenium(III)kations (Elektronenkonfiguration: $5d^4$) verwendet. Drei der Chloridionen bilden μ^2 -verbrückte koordinative Bindungen zu benachbarten Rheniumatomen aus, von den anderen sechs Chloridionen sind je zwei über koordinative Bindungen an eines der Rheniumatome gebunden.

Die Atome und Elektronensysteme des Clusters lauten:

Re: 1 - 3, $\text{Cl}_{\text{verbrückt}}$: 4 - 6, $\text{Cl}_{\text{terminal}}$: 7 - 12

σ - und π -Elektronensysteme der Rheniumatome:

(1 2 s r; 2) (1 2 p; 2) (1 3 s r; 2) (1 3 p; 2) (2 3 s r; 2) (2 3 p; 2)
 (1 p; 0) (1 p; 0) (1 p; 0) (2 p; 0) (2 p; 0) (2 p; 0) (3 p; 0) (3 p; 0) (3 p; 0)

Koordinative Bindungen zu den verbrückten Chloratomen:

(1 2 4 c; 2) (1 3 5 c; 2) (2 3 6 c; 2)

Koordinative Bindungen zu den terminalen Chloratomen:

(1 7 c; 2) (1 8 c; 2) (2 9 c; 2) (2 10 c; 2) (3 11 c; 2) (3 12 c; 2)

π -Elektronensysteme der Chloratome:

(4 p; 2) (4 p; 2) (4 p; 2) (4 p; 0) (4 p; 0) (4 p; 0) (4 p; 0) (4 p; 0)
 (5 p; 2) (5 p; 2) (5 p; 2) (5 p; 0) (5 p; 0) (5 p; 0) (5 p; 0) (5 p; 0)
 (6 p; 2) (6 p; 2) (6 p; 2) (6 p; 0) (6 p; 0) (6 p; 0) (6 p; 0) (6 p; 0)
 (7 p; 2) (7 p; 2) (7 p; 2) (7 p; 0) (7 p; 0) (7 p; 0) (7 p; 0) (7 p; 0)
 (8 p; 2) (8 p; 2) (8 p; 2) (8 p; 0) (8 p; 0) (8 p; 0) (8 p; 0) (8 p; 0)
 (9 p; 2) (9 p; 2) (9 p; 2) (9 p; 0) (9 p; 0) (9 p; 0) (9 p; 0) (9 p; 0)
 (10 p; 2) (10 p; 2) (10 p; 2) (10 p; 0) (10 p; 0) (10 p; 0) (10 p; 0) (10 p; 0)
 (11 p; 2) (11 p; 2) (11 p; 2) (11 p; 0) (11 p; 0) (11 p; 0) (11 p; 0) (11 p; 0)
 (12 p; 2) (12 p; 2) (12 p; 2) (12 p; 0) (12 p; 0) (12 p; 0) (12 p; 0) (12 p; 0)

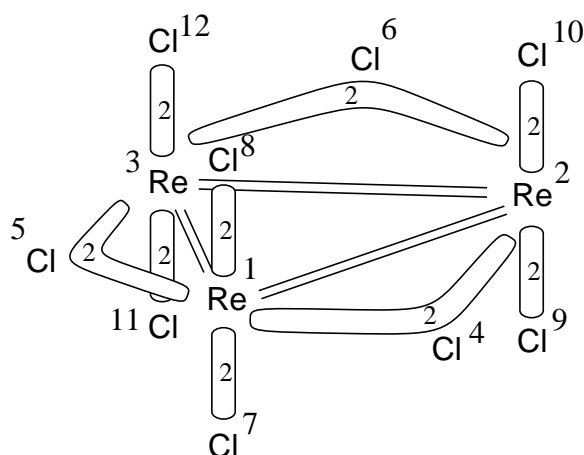


Abb. 4-23: Re_3Cl_9 , ein Metallcluster mit M-M-Mehrfachbindungen und verbrückten koordinativen Bindungen. Der Übersichtlichkeit wegen sind das σ - und π -Elektronensystem jeder Doppelbindung mit zwei parallelen Linien dargestellt, die koordinativen Bindungen als Ovale. Alle π -Elektronensysteme, die freie Elektronen bzw. unbesetzte Orbitale repräsentieren, sind nicht gezeichnet.

4.3. Moleküle

Im vorgestellten Modell werden alle Spezies, in denen jedes Atom durch mindestens einen der drei Bindungstypen mit den anderen Atomen verbunden ist, als Moleküle betrachtet. Dies

müssen aber nicht notwendigerweise neutrale Verbindungen sein, auch Anionen und Kationen fallen unter den Begriff Molekül. Instabile Verbindungen, die als Übergangszustände oder Zwischenstufen in Reaktionen modelliert werden, gelten ebenfalls als Moleküle. Auch isolierte Atome wie Metallkationen oder Halogenanionen werden als Molekül betrachtet. Alle im Kapitel über Elektronensysteme gezeigten Beispielverbindungen sind als Moleküle anzusehen.

4.4. Aggregate

Moleküle wurden als Spezies definiert, die nur Atome enthalten dürfen, die durch kovalente Bindungen miteinander verknüpft sind. Um mehrere Moleküle mit intermolekularen Wechselwirkungen nicht kovalenter Natur als Einheit darzustellen, wurde das Konzept des Aggregats entwickelt. Ein Aggregat besteht aus mehreren Molekülen, die durch mindestens eine intermolekulare Wechselwirkung miteinander verbunden sind. Wechselwirkungen werden in sogenannten Wechselwirkungsgruppen repräsentiert, die in Kapitel 4.5 vorgestellt werden.

Eine wichtige Anwendungsmöglichkeit ist die Repräsentation von Salzen. Natriumacetat ist ein Beispiel für ein solches Aggregat, das aus einem Anion und einem Kation besteht. Salze, die sich aus mehr als zwei Ionen zusammensetzen, enthalten entsprechend viele Wechselwirkungsgruppen. Das Aggregat, das Dilithiummalonat codiert, besteht aus zwei Lithiumkationen, dem Malonatdianion und zwei Wechselwirkungsgruppen zwischen je einer Säuregruppe und einem Lithiumion (Abbildung 4-24). Die an der Wechselwirkung beteiligten Elektronensysteme sind in Abbildung 4-24 grau unterlegt.

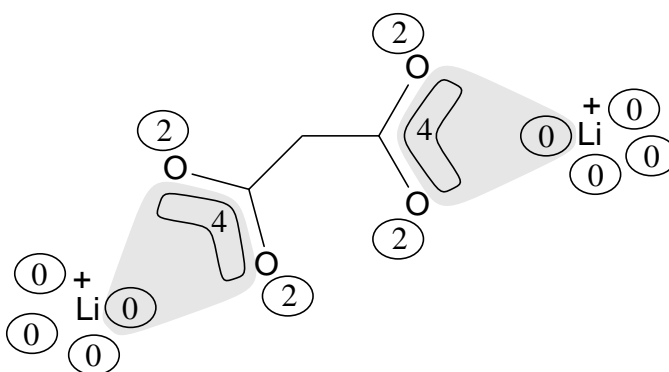


Abb. 4-24: Dilithiummalonat, ein Aggregat mit zwei ionischen Wechselwirkungsgruppen, die grau markiert sind.

Aber auch andere miteinander in Wechselwirkung stehende Moleküle können in Aggregate zusammengefaßt werden. Darunter fallen Verbindungen mit intermolekularen Wasserstoffbrückenbindungen, oder auch Catenane und Rotaxane, in denen die Aggregation durch mechanische Kräfte zustandekommt.

4.5. Wechselwirkungen

Um Wechselwirkungen Rechnung zu tragen, wurde das Konzept der Wechselwirkungsgruppen entwickelt. Eine Wechselwirkungsgruppe ist definiert als eine Menge von Atomen bzw. Elektronensystemen, die aufgrund einer spezifischen Wechselwirkung miteinander in Beziehung stehen. Diese Wechselwirkung kann verschiedener Art sein und kann unter anderem ionische Bindungen, Wasserstoffbrückenbindungen oder die ineinandergreifenden Ringe in Catenanen beschreiben.

In ionischen Bindungen findet eine Wechselwirkung zwischen einem gefüllten π -Elektronensystem des Anions und einem leeren π -Elektronensystem des Kations statt. Die entsprechende Wechselwirkungsgruppe enthält daher in der Regel nur zwei π -Elektronensysteme. In Natriumacetat (Abbildung 4-25) sind das π -Elektronensystem mit vier Elektronen, das sich über beide Sauerstoffatome und den Carboxylkohlenstoff erstreckt, und ein leeres π -Elektronensystem des Natriumkations daran beteiligt. Sie sind in Abbildung 4-25 grau unterlegt. Da Natrium ein Element der dritten Periode ist, werden ihm außer dem 3s-Orbital drei unbesetzte p- und fünf unbesetzte d-Orbitale zugeordnet, die jedoch nicht an der Wechselwirkungsgruppe beteiligt sind. Sie wurden der Übersichtlichkeit wegen nicht gezeichnet.

Die Atome und Elektronensysteme des Aggregats lauten wie folgt:

C: 1-2, O: 3-4, Na: 5

Acetatanion: $\{(1\ 2\ s\ r; 2)\ (2\ 3\ s\ r; 2)\ (2\ 4\ s\ r; 2)\ (2\ 3\ 4\ p; 4)\ (3\ p; 2)\ (4\ p; 2)\}$

Natriumkation: $\{(5\ p; 0)\ (5\ p; 0)\ (5\ p; 0)\ (5\ p; 0)\ (5\ p; 0)\ (5\ p; 0)\ (5\ p; 0)\ (5\ p; 0)\ (5\ p; 0)\}$

Wechselwirkungsgruppe: $[(2\ 3\ 4\ p; 4)\ (5\ p; 0)]$.

Alle Elektronensysteme eines Moleküls sind in geschweiften Klammern zusammengefaßt, die Elektronensysteme der Wechselwirkungsgruppe sind in eckigen Klammern zusammengefaßt.

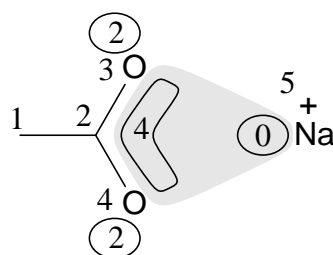


Abb. 4-25: Ionische Bindung von Natriumacetat als Wechselwirkungsgruppe. Die acht leeren π -Elektronensysteme des Natriums, die 3 p- und 5 d-Orbitalen entsprechen und nicht an der ionischen Bindung beteiligt sind, wurden nicht gezeichnet.

Eine Wasserstoffbrückenbindung enthält einerseits das Heteroatom zusammen mit einem freien Elektronenpaar, das als Wasserstoffakzeptor fungiert und andererseits das Wasserstoff-

atom zusammen mit dem σ -Elektronensystem und dem Heteroatom, die den Wasserstoffdonor bilden. Abbildung 4-26a zeigt die intramolekulare Wasserstoffbrücke der Enolatform des 1,3-Diketons Pentan-2,4-dion in herkömmlicher Kodierung, Abbildung 4-26b die Wechselwirkungsgruppe, die im vorgestellten Modell verwendet wird. Die σ -Elektronensysteme sind in Abbildung 4-26b als Linien gezeichnet, das π -Elektronensystem besteht aus der Kette der beiden Sauerstoffatome und der dazwischenliegenden Kohlenstoffatome und enthält sechs Elektronen. Die Atome und Elektronensysteme, die an der Wechselwirkungsgruppe der Wasserstoffbrücke teilnehmen, sind grau unterlegt.

Das Molekül setzt sich folgendermaßen zusammen:

C:1-5, O: 6, 7 H: 8-15

(1 2 s r; 2) (1 10 s r; 2) (1 11 s r; 2) (1 12 s r; 2) (2 3 s r; 2) (2 7 s r; 2) (2 3 4 6 7 p; 6)

(3 4 s r; 2) (3 9 s r; 2) (4 5 s r; 2) (4 6 s r; 2) (5 13 s r; 2) (5 14 s r; 2) (5 15 s r; 2) (7 8 s r; 2)

mit der Wechselwirkungsgruppe: [6 7 8 (6 p; 2) (7 8 s r; 2)]

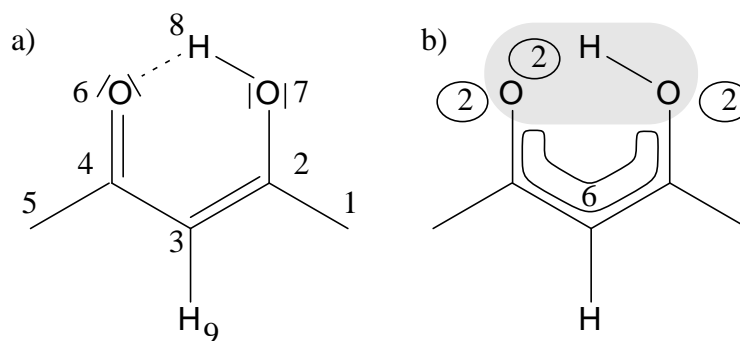


Abb. 4-26: Repräsentation der Wasserstoffbrückenbindungen der Enolatform von 1,3-Diketonen am Beispiel von Pentan-2,4-dion. (a) Valenz-Bond Repräsentation, (b) Repräsentation als Wechselwirkungsgruppe.

4.6. Ensembles

Ensembles sind die größte Einheit, in die Moleküle oder Aggregate zusammengefaßt werden können. Sie bestehen aus isolierten Molekülen oder Aggregaten, die nicht miteinander in Wechselwirkung stehen, aber die im gleichen Zusammenhang verwendet werden. Ein Ensemble kann alle Edukte oder Produkte einer Reaktion umfassen, ein weitere Anwendungsmöglichkeit liegt im Bereich der Synthesepaltung, wo ein Ensemble alle Vorstufen einer Zielverbindung beinhalten kann. Auch im Bereich der Interpretation und Vorhersage von Infrarot- oder Massenspektren lassen sich Ensembles einsetzen. Sie könnten dort alle Strukturvorschläge, die zu einem experimentellen Spektrum gehören, zusammenfassen. Abbildung 4-27 zeigt zwei Ensembles, die die Edukte und Produkte einer S_N2 -Reaktion enthalten. Sie bestehen jeweils aus einem Aggregat und einem Molekül. Das Aggregat Natriumethoxid reagiert mit

n-Butylbromid zu *n*-Butylmethylether und Natriumbromid, das ebenfalls als Aggregat dargestellt wird. Die acht leeren π -Elektronensysteme des Natriumkations, die an keiner Bindung oder Wechselwirkung beteiligt sind, wurden der Übersichtlichkeit wegen nicht gezeichnet.

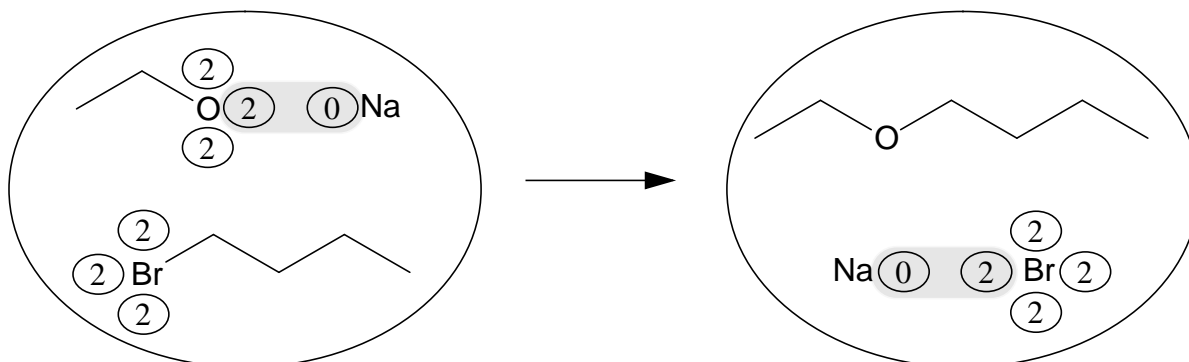


Abb. 4-27: Edukt- und Produktensembel der Reaktion von *n*-Butylbromid mit Natriumethoxid.

4.7. Topologische Gruppen

Bei der Behandlung von chemischen Strukturen ist es günstig, wenn Atome oder Elektronensysteme eines Moleküls, die gemeinsame strukturelle Eigenschaften besitzen, zu einer größeren strukturellen Einheit zusammengefaßt sind, da sie dann gemeinsam behandelt werden können.

Besonders häufig werden die Ringsysteme eines Moleküls verwendet. Gruppen, die die Ringatome eines Ringsystems zusammenfassen, werden daher oft eingesetzt. In Abbildung 4-28a ist die Gruppe der Ringatome in Cyclohexanol grau unterlegt. Die Notation dieser Gruppe lautet: [1 2 3 4 5 6]

Ein zweiter oft benötigter Gruppentyp enthält alle Nachbaratome eines vorgegebenen Atoms. Abbildung 4-28b zeigt die Gruppe aus Nachbaratomen des substituierten Kohlenstoffatoms in Cyclohexanol, die aus folgenden Atomen besteht: [2 6 7 8]

Die an den Gruppen beteiligten Atome sind in Abbildung 4-28 grau unterlegt.

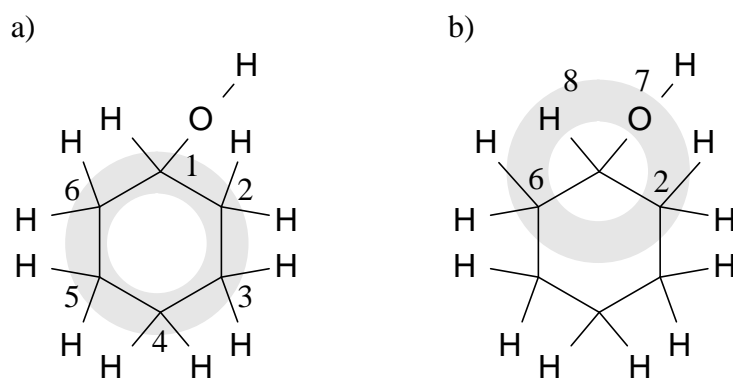


Abb. 4-28: Topologische Gruppen in Cyclohexanol. (a) Ringgruppe, (b) Nachbaratome des substituierten Kohlenstoffatoms. Für die an den Gruppen beteiligten Atome ist jeweils die Numerierung angegeben.

5. Physikochemische und topologische Eigenschaften

Im vorigen Kapitel wurden Konzepte vorgestellt, anhand derer eine allgemeine Repräsentation von Molekülen möglich ist. Viele Anwendungen, mit deren Hilfe Reaktionen oder Zusammenhänge zwischen chemischen Verbindungen und spektroskopischen oder biologischen Daten modelliert werden, benötigen außer der Konstitution eines Moleküls auch Eigenschaften, die sich aus der Konnektivität ableiten lassen. Darunter fallen physikochemische und topologische Eigenschaften sowohl der Atome und Bindungen als auch der Moleküle oder Ensembles in ihrer Gesamtheit. In unserer Arbeitsgruppe existiert bereits eine große Anzahl von Methoden zur Berechnung von physikochemischen bzw. topologischen Eigenschaften von Molekülen, die jedoch häufig nur für neutrale organische Verbindungen mit O, N, S, Cl, F und P als Heteroatomen parametrisiert sind. Die Methoden zur Berechnung dieser Eigenschaften sollen in Zukunft auf eine breitere Basis gestellt werden und dem RICOS-Modell angepaßt werden. Da in diese Methoden jedoch ein großer Schatz an Erfahrung und Zeit investiert wurde, war es im Rahmen dieser Arbeit nicht möglich, sie alle in das neue Modell umzusetzen. Um nun auch für RICOS physikochemische Eigenschaften zur Verfügung stellen zu können, wurden zwei parallele Ansätze verfolgt. Einerseits wurde eine Konvertierung in das bisherige Datenformat, das auf einer Bindungsliste basiert geschaffen, um über diesen Weg, die schon vorhandene Eigenschaftsberechnungen in einer Übergangsphase verwenden zu können. Dies ist natürlich nur für organische Verbindungen möglich, die sich auch mit Bindungslisten darstellen lassen, jedoch nicht für Verbindungen mit Mehrzentrenbindungen oder koordinativen Bindungen. Andererseits wurde begonnen, einige der Verfahren zur Bestimmung topologischer Eigenschaften, die für die Behandlung von Molekülstrukturen in einem Reaktionsvorhersagesystem unerlässlich sind, zu erweitern und auf RICOS zu übertragen.

5.1. Übersicht über die implementierten Eigenschaften

Tabelle 5-1 gibt einen Überblick über alle Eigenschaften, die für RICOS bisher implementiert wurden. Der Zugriff auf diese Eigenschaften erfolgt über die Schlüsselwörter, die in der zweiten Spalte aufgeführt sind. Jedes Schlüsselwort beginnt mit einem Präfix, das angibt, für welchen chemischen Typ die Eigenschaft definiert ist, A_ für Atome, EL_ für Elektronensysteme, M_ für Moleküle, AG_ für Aggregate und E_ für Ensembles.

Tab. 5-1: Eigenschaften, die für RICOS zur Verfügung stehen.

| Beschreibung | Schlüsselwort |
|--------------------------------|----------------|
| Ladungszentrum | A_CHARGECENTER |
| Äquilibrierte Ladung des Atoms | A_EQUILCHARGE |
| Atomhashcode | A_HASH |
| Atomhashgruppe | A_HGROUP |

| Beschreibung | Schlüsselwort |
|--|----------------|
| Flag, das angibt, ob ein Atom als Metallatom definiert ist | A_ISMETAL |
| Isotopenmasse des Atoms | A_ISOTOPE |
| IUPAC Gruppennummer | A_IUPAC_GROUP |
| Flag, das angibt, ob sich ein Atom in einer endständigen Kette eines Moleküls befindet. | A_PERIPH |
| Index des größten π -Elektronensystem, an dem das Atom beteiligt ist | A_PISYS |
| Radikalzentrum | A_RADICCENTER |
| Summe der äquilibrierten Ladung aller Atome des Elektronensystems | EL_EQUILCHARGE |
| Elektronensystemhashcode | EL_HASH |
| Elektronensystemhashgruppe | EL_HGROUP |
| Flag, das angibt, ob sich ein Elektronensystem in einer endständigen Kette eines Moleküls befindet | EL_PERIPH |
| Molekülladung | M_CHARGE |
| Molekülhashcode | M_HASH |
| Molekülhashgruppe | M_HGROUP |
| Molekülname | M_NAME |
| Aggregathashcode | AG_HASH |
| Aggregathashgruppe | AG_HGROUP |
| Ensemblehashcode | E_HASH |
| Identifikator des Ensembles | E_IDENT |
| Flag, das angibt, ob die Hashcode Berechnung mit einem Seed initialisiert werden soll. | E_SEEDHASH |
| Eigenschaft, mit der die Hashcode Berechnung initialisiert werden soll. | E_SEED_PROP |

5.2. Ladungen

Die Repräsentation von Molekülen mit delokalisierten Elektronensystemen bedeutet, daß eine ganzzahlige Verteilung von Formalladungen auf einzelne Atome eines Moleküls nicht direkt aus den Elektronensystemen eines Atoms abgeleitet werden kann. Um jedoch abschätzen zu

können, ob Atome partial positiv oder negativ geladen sind, wurden zwei Methoden zur Berechnung von Ladungen an Atomen entwickelt. Einerseits werden äquilibrierte Ladungen berechnet, in dem die Elektronen gleichmäßig auf alle Atome eines Elektronensystems verteilt werden. Andererseits wird die Ladung denjenigen Atomen eines π -Elektronensystems zugeordnet, die in einer VB-Struktur eine ganzzahlige Formalladung tragen können.

5.2.1. Äquilibrierte Ladung

In die Berechnung der äquilibrierten Ladung eines Atoms gehen die Anzahl der Elektronen im Grundzustand, sowie die Elektronensysteme, an denen es beteiligt ist ein. Für Atome, die nur über σ -Elektronensysteme und π -Elektronensysteme gebunden sind, wird sie wie folgt, berechnet:

$$q_e = nel_G - \sum_{esys} nel_i / natom_i \quad (5-1)$$

Dabei steht q_e für die äquilibrierte Ladung des betrachteten Atoms, nel_G für die Anzahl der Valenzelektronen des Elements, nel_i für die Anzahl der Elektronen eines Elektronensystems und $natom_i$ für die Anzahl der Atome des Elektronensystems.

In koordinativen Bindungen steuert der Ligand häufig sämtliche Elektronen in die Bindung bei. Bei einer Gleichverteilung der Elektronen auf alle Atome eines Elektronensystems würden Metallatome in diesen Fällen eine sehr hohe negative äquilibrierte Ladung erhalten. Daher wurde für Metallatome, die an koordinativen Bindungen beteiligt sind, die Berechnung der äquilibrierte Ladung folgendermaßen definiert:

$$q_e = nel_G - \sum_{sigmasys, pisy} nel_i / natom_i \quad (5-2)$$

Bei der Berechnung der äquilibrierten Ladung werden also nur Elektronen aus σ - und π -Elektronensystemen berücksichtigt, Elektronen koordinativer Bindungen jedoch nicht.

Für Nichtmetallatome, die Bestandteil eines Liganden sind, werden alle Elektronensysteme berücksichtigt (Gleichung 5-3), es wird in koordinativen Bindungen ($coordsys$) jedoch nur die Zahl der Ligandatome ($nligatom$) miteinbezogen.

$$q_e = nel_G - \sum_{sigmasys, pisy} nel_i / natom_i - \sum_{coordsys} nel_i / nligatom_i \quad (5-3)$$

Beispiele für äquilibrierte Ladungen an Atomen einer Verbindung mit Mehrzentren-Elektronensystemen, sowie einer organischen und metallorganischen Verbindung zeigt Abbildung 5-1. In Diboran (Gleichung 5-1) sind die beiden Boratome partial negativ geladen, die verbrückenden Wasserstoffatome partial positiv, und die terminalen Wasserstoffatome sind ungeladen. In Furan (Gleichung 5-1) steuert das Sauerstoffatom zwei Elektronen in das delokalisierte π -Elektronensystem bei. Daher ist es positiv geladen, die Kohlenstoffatome, sind negativ gela-

den, die nicht dargestellten Wasserstoffatome sind ungeladen. In Ferrocen tragen alle Kohlenstoffatome die Ladung -0.2 (Gleichung 5-3), das Eisenatom die Ladung 2.0 (Gleichung 5-2), die nicht dargestellten Wasserstoffatome sind formal ungeladen.

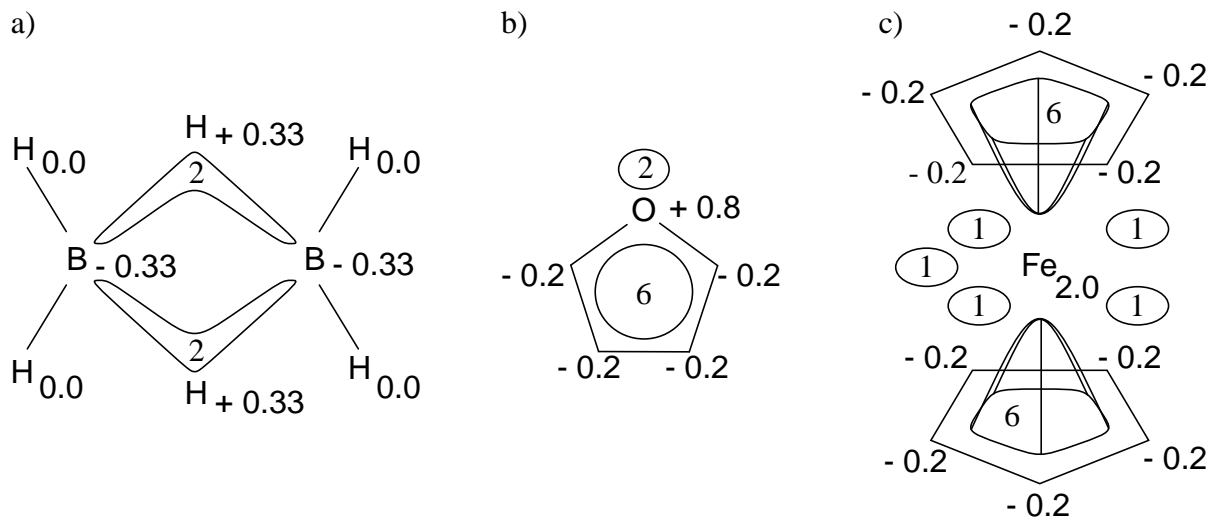


Abb. 5-1: Äquilibrierte Ladungen in Diboran (a), Furan (b) und Ferrocen (c).

Ausgehend von den äquilibrierten Ladungen von Atomen werden auch äquilibrierte Ladungen von Elektronensystemen durch Summation der Ladungen aller Atome eines Elektronensystems berechnet. Die Ladung von Molekülen ergibt sich ebenfalls aus der Summe der äquilibrierten Ladung aller Atome eines Moleküls. Die Ladung der Aggregate und Ensembles setzt sich aus den Molekülladungen zusammen.

5.2.2. Lokalisierte Ladungen in Molekülen

Die Modellierung von Reaktionen in der Massenspektrometrie basiert auf der Modellierung der Reaktivität von Kationen. Einige der dort verwendeten Reaktionsregeln basiert auf heterolytischen Bindungsbrüchen, an denen formal positiv geladene Atome beteiligt sind. Auch bei organischen Reaktionen können geladenen Atome Bestandteile des Reaktionszentrums sein. Um diese Information zur Verfügung zu stellen, wurde ein Verfahren entwickelt, das alle Atome im Molekül sucht, die eine ganzzahlige Ladung tragen können.

Bei geladenen Atomen, die nicht an delokalisierten Elektronensystemen beteiligt sind, ergibt die oben beschriebene Methode zur Bestimmung von äquilibrierten Ladungen (Gleichung 5-1) geradzahlige Werte (Abbildung 5-2).

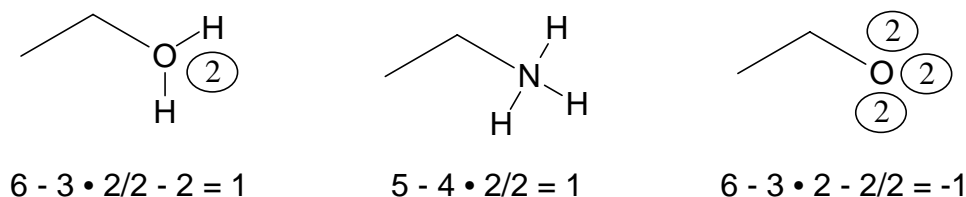


Abb. 5-2: Ladung von Atomen ohne delokalisierte Elektronensysteme.

Für geladene Verbindungen, in denen die Ladung über mehrere Atome in π -Elektronensystemen delokalisiert ist, wurde ein Algorithmus entwickelt, der alle Atome sucht, die die Ladung tragen können.

Bevor die eigentliche Suche stattfindet, werden alle Atome eines π -Elektronensystems vorbehandelt. Bei der Suche nach positiven Ladungen wird für jedes π -Elektronensystem geprüft, ob die Zahl der Atome kleiner als die Zahl der Elektronen ist. Trifft dies zu, enthält das π -Elektronensystem Heteroatome, die formal ungeladen sind, wenn ihnen ein freies Elektronenpaar zugeordnet wird und positiv geladen, wenn ihnen eine Doppelbindung zugeordnet wird. Diese Atome werden gesondert markiert. Wenn die Verbindung negativ geladen ist, wird geprüft, ob das π -Elektronensystem Heteroatome enthält, die keine negative Ladung tragen können und diese ebenfalls markiert.

Auf den restlichen Atomen des π -Elektronensystems wird dann ein Verfahren zur Bestimmung von Atomen angewendet, die die Ladung tragen können und gleichzeitig erlauben, eine gültige VB-Struktur zu erzeugen, in der zwischen allen anderen Atome alternierende Einfach- und Doppelbindungen liegen. Dieses Verfahren wird in der Konvertierung von RICOS zu einer VB-Struktur ebenfalls verwendet und in Kapitel 7.2.2 detailliert beschrieben. Jedem so gefundenen Atom wird dann als Eigenschaftswert der Quotient aus Ladung und Zahl der geladenen Atome zugewiesen. Wenn ein geladenes Atome Teil mehrerer π -Elektronensysteme ist, die keine freien Elektronen repräsentieren, wird die Ladung auf alle Atome der π -Elektronensysteme verteilt. Abbildung 5-3 zeigt Beispiele für Kationen und Anionen. Die π -Elektronensysteme wurden der Übersichtlichkeit halber in VB-Darstellung mit lokalisierter Ladung gezeichnet. Sie erstrecken sich jeweils über das gesamte Gerüst aus Schweratomen. Die ersten drei Beispiele sind einfach positiv geladen, das letzte einfach negativ.

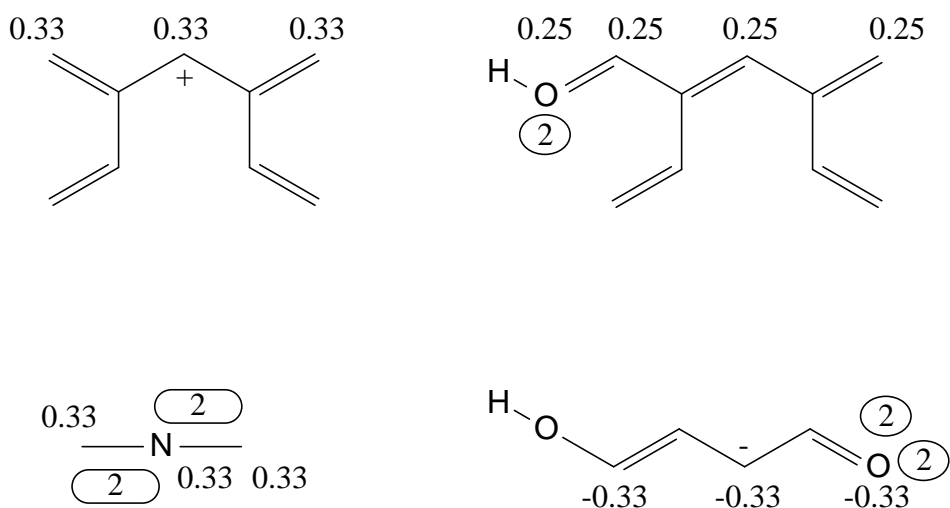


Abb. 5-3: Lokalisierte Ladungen in delokalisierten π -Elektronensystemen. Die delokalisierten π -Elektronensysteme sind der Übersichtlichkeit halber als VB-Grenzform mit lokalisierter Ladung gezeichnet. Sie erstrecken sich jeweils über das gesamte Gerüst aus Schweratomen.

5.3. Radikalzentren

Die Behandlung der radikalinduzierten Reaktionen in der Massenspektrometrie erfordert es, in radikalischen Verbindungen diejenigen Atome zu erkennen, die ein Radikalzentrum sein können. Die Bestimmung der radikalischen Atome verläuft ähnlich wie die der geladenen Atome. In allen π -Elektronensystemen einer Verbindung, die eine ungerade Anzahl von Elektronen tragen, wird das Verfahren zur Bestimmung von Atomen, die freie Elektronen tragen können, angewendet und jedem so gefundenen Atom, die Eigenschaft zugewiesen, ein Radikalzentrum zu sein. Im π -Elektronensystem der Verbindung in Abbildung 5-4 sind die Atome, die das radikalische Elektron tragen können, mit einem Stern markiert. Auch hier ist das π -Elektronensystem der Übersichtlichkeit wegen in einer der möglichen VB-Strukturen dargestellt.

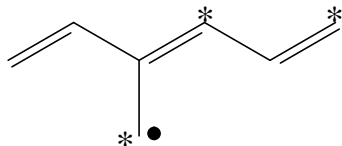


Abb. 5-4: Radikalzentren in verzweigten delokalisierten π -Elektronensystemen. Alle Atome, die das radikalische Elektron tragen können, sind mit einem Stern markiert. Das π -Elektronensystem enthält alle Kohlenstoffatome und ist der Übersichtlichkeit wegen als VB-Struktur dargestellt.

5.4. Hashcode

Die schnelle und nicht redundante Behandlung von Reaktionen in den Reaktionsvorhersageprogrammen EROS und MASSIMO benötigt ein schnelles und zuverlässiges Verfahren zur Erkennung identischer Moleküle. Um zu bestimmen, ob zwei Moleküle identisch sind, wurde in unserer Arbeitsgruppe für das EROS-System ein Hashcoding-Verfahren [54] entwickelt, das sich als sehr zuverlässig erwiesen hat. Ein Hashcode ist eine hochkomprimierte Darstellung einer komplexen Datenstruktur in einem festen Wertebereich und mit konstanter Länge. Die Berechnung eines Hashcodes ist ein nicht umkehrbares Verfahren, bei dem der Hashcode nicht in die ursprüngliche Informationseinheit zurücktransformiert werden kann. Hashcodes werden im allgemeinen als Zugriffsschlüssel oder Indizes in großen Tabellen eingesetzt. Da der Wertebereich eines Hashcoding-Verfahrens begrenzt ist, kann es zu sogenannten Kollisionen kommen, wenn die Zahl der zu transformierenden Daten der Zahl der möglichen Hashcodes nahekommt. Als Kollision wird der Fall bezeichnet, wenn zwei verschiedene Datenstrukturen denselben Hashcode erhalten. Das oben genannte Hashcode-Verfahren erzeugt einen 32-bit Hashcode aus Molekülen, die in Valenz-Bond-Repräsentation kodiert sind. Es ist ausreichend, um Datensätze mit bis zu 10 000 Molekülen kollisionsfrei transformieren zu können. Da aber die Größe verfügbarer Strukturdatenbanken in den letzten Jahren stark angewachsen ist, wurde das Verfahren auf einen 64-bit Code erweitert, mit dem es auch möglich ist, Datenbanken mit bis zu 10 000 000 Einträgen kollisionsfrei zu kodieren [55]. Dieses Verfahren zur Hashcodeerzeugung wurde nun dahingehend geändert, daß es auf Moleküle, die mit RICOS repräsentiert werden, anwendbar ist.

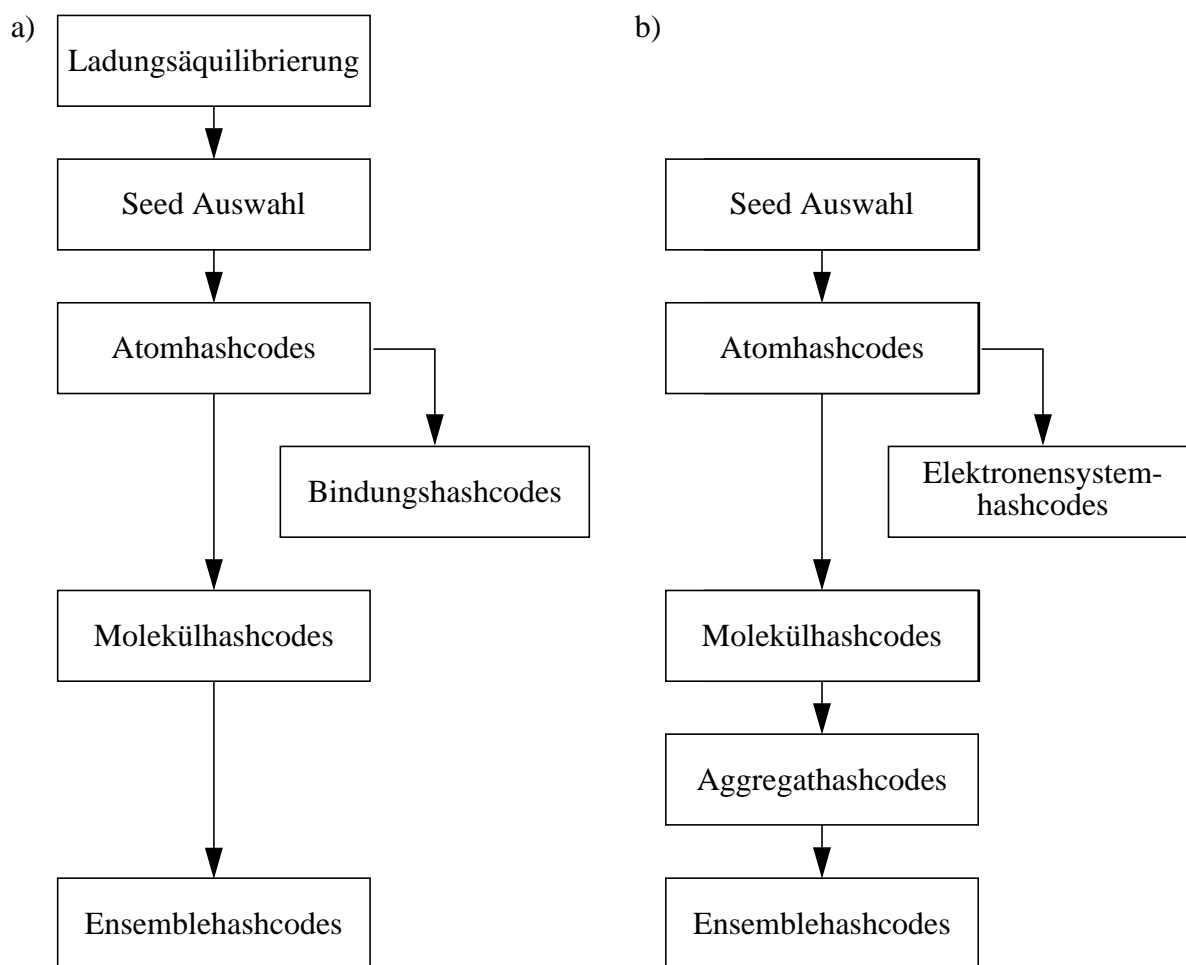


Abb. 5-5: Schematischer Ablauf des Hashcoding-Verfahrens, (a) nach Ihlenfeldt, (b) Modifikation für RICOS.

Abbildung 5-5a zeigt schematisch den schrittweisen Ablauf des ursprünglichen Verfahrens, Abbildung 5-5b die Modifikationen dieses Schemas für RICOS. Beiden Verfahren ist gemein, daß zunächst Atomhashcodes bestimmt werden, aus denen sich Molekül- und Bindungshashcodes bzw. Elektronensystemhashcodes berechnen lassen. Ensemblehashcodes wurden ursprünglich direkt aus den Molekülhashcodes ermittelt. In RICOS wird jedoch noch die Berechnung der Aggregathashcodes eingeschoben. Im Folgenden werden alle Stufen des Verfahrens erläutert.

5.4.1. Bestimmung der Seedparameter

Die erste Stufe des Hashcoding-Verfahrens nach Ihlenfeldt besteht in der Ladungsäquilibrierung. Wie bereits mehrfach erläutert, erweist sich das Auftreten von Resonanzstrukturen bei der Repräsentation von geladenen Verbindungen nach dem Valenz-Bond-Modell als problematisch. Die Berechnung eines Hashcodes darf natürlich nicht von der gewählten Resonanzstruktur abhängig sein. Daher werden zunächst alle π -Systeme eines geladenen Moleküls bestimmt

und die darin lokalisierte Ladung gleichmäßig auf alle Atome verteilt. In RICOS ergibt sich das Problem der Ladungsverteilung nicht, da delokalisierte Ladungen bereits in π -Elektronensysteme verteilt sind. Dieser Schritt entfällt daher.

Der zweite Schritt besteht darin, jedem Atom einen Anfangswert zuzuweisen, der einerseits unterschiedliche Atome von Anfang an differenzieren soll, andererseits für äquivalente Atome gleich sein soll. Um dies zu erreichen, werden Parameter, die die Atome unterscheiden, als Indizes in einer Primzahltable verwendet. Alle so bestimmten Primzahlen werden miteinander zum Startwert, dem Seed, multipliziert. Jeder Parameter hat einen Wertebereich, der mit einem eigenen Sektor in einer Primzahltable korrespondiert, um zufällige Übereinstimmung zu vermeiden. Zur Bestimmung des Seeds werden die in Tabelle 5-2 angegebenen Faktoren herangezogen.

Tab. 5-2: Parameter, aus denen der Seed der Hashcodeberechnung bestimmt wird.

| Parameter | Ihlenfeldt | RICOS |
|--|------------|-------|
| Zahl der Nachbarn (0 . . . 8) | X | X |
| Zahl der H-Nachbarn (0 . . . 6) | X | X |
| Ordnungszahl (1 . . . 113) | X | X |
| Zahl der Atome des Moleküls modulo einer Primzahl (257) | X | X |
| optional: Atom Sterodeskriptor (-1, 0, 1) | X | |
| Äquilibrierte Ladung (-3 . . . 3) | X | |
| Summe der Elektronen aller Elektronensysteme modulo einer Primzahl (29) | | X |
| Summe aller Elektronensysteme, die mindestens ein Elektron enthalten. | | X |
| optional: Atomeigenschaften, normiert auf einen Bereich zwischen 1 und 100 | | X |

Für RICOS wurden nur die ersten vier Parameter zur Bestimmung des Seeds übernommen, die anderen wurden neu formuliert, um die speziellen Eigenschaften der Datenstruktur zu berücksichtigen. Der einzige Parameter, der die Elektronenverteilung innerhalb eines Moleküls berücksichtigt, wird bei Ihlenfeldt aus der äquilibrierten Ladung wie folgt bestimmt:

$$index = mod((int)((q_e + 3.0) \cdot 5), 29) \quad (5-4)$$

Die äquilibrierte Ladung q_e eines Atoms bewegt sich in einem Bereich zwischen -3.0 und +3.0. Sie wird durch die Addition in eine positive Zahl transformiert und dann durch die Multiplikation mit 5 auf einen größeren Wertebereich gespreizt. Da die Modulooperation nur mit ganzen Zahlen möglich ist, wird das reellwertige Ergebnis vorher abgerundet. In größeren π -Systemen mit einer Gesamtladung von Eins führt diese Rundung dazu, daß Unterschiede zwischen geladenen Verbindungen und ihren neutralen Äquivalenten aufgehoben werden. Dies sei am Beispiel von Benzol und seinem Radikalkation erläutert. Die äquilibrierte Ladung der Kohlenstoffatome in Benzol ist Null, es ergibt sich also vor der Modulooperation ein Wert von 15. Die äquilibrierte Ladung der Kohlenstoffatome des Radikalkations ist 0.17, der Wert nach der Spreizung beträgt 15.85 und wird auf 15 abgerundet. Die Kohlenstoffatome in beiden Verbindungen erhalten demnach den gleichen Seedwert und entsprechend identische Hashcodes. Dieses Manko würde sich in der Behandlung von Reaktionen im Massenspektrometer stark auswirken, da hier häufig Verbindungen auftreten, die sich nur durch ihre Ladung, nicht aber durch die anderen Parameter unterscheiden. Zudem kann die äquilibrierte Ladung für RICOS auch nicht verwendet werden, um zwischen Verbindungen zu unterscheiden, die zwar das gleiche σ -Elektronensystemgerüst besitzen, jedoch unterschiedliche Verteilung der Elektronen auf die π -Elektronensysteme, wie dies in Carbenen oder ionisierten Amiden der Fall ist. Um solche Verbindungen ebenfalls unterscheiden zu können, wurden zwei neue Startparameter eingeführt. Zum einen wird für jedes Atom die Summe aller Elektronen aller Elektronensysteme, an denen es beteiligt ist, bestimmt. Dies erlaubt, zwischen geladenen und ungeladenen Verbindungen zu unterscheiden. Der zweite Parameter ist die Summe aller Elektronensysteme eines Atoms, die mindestens mit einem Elektron besetzt sind. Somit werden verschieden angeregte Zustände eines Moleküls leicht unterscheidbar.

Der letzte Parameter ist optional, es können mehrere zusätzliche Atomeigenschaften in den Seed miteinbezogen werden. Diese Möglichkeit war schon in der 32-bit Hashcode Version gegeben, ist aber für den 64-bit Hashcode des CACTVS-Systems nicht implementiert. Alle Werte dieser Atomeigenschaften innerhalb eines Ensembles werden im Bereich von 1 bis 100 skaliert. Als Eigenschaften, die optional in den Hashcode miteinbezogen werden können, sind Stereodeskriptoren oder auch Isotopenmarkierungen geeignet.

5.4.2. Atomhashcodes

Das eigentliche Verfahren zur Berechnung von Atomhashcodes aus den Seedwerten besteht aus Rotations- und Exklusive-Oder-Verknüpfungen (XOR) der Startwerte in 32 Zyklen. Über die Rotationsoperationen wird die Bitverteilung, die zu Beginn hauptsächlich in den unteren Bits dominant ist, auf den gesamten Bitvektor verteilt. Danach werden XOR-Verknüpfungen mit den Hashcodes der benachbarten Atome durchgeführt. Dabei muß jeweils überprüft werden, ob ein Hashcode schon einmal aufgetreten ist. In diesem Fall wird vorher noch eine Rotationsoperation um so viele Bits durchgeführt, wie oft der Hashcode bisher aufgetreten ist. Damit wird vermieden, daß sich eine gerade Anzahl von XOR-Verknüpfungen mit dem selben

Wert gegenseitig auslöschen. Auf diese Weise wirkt sich die Umgebung jeden Atoms auf seinen Hashcode aus. Je größer die Zyklenzahl, desto weitreichender ist der Einfluß der äußeren Sphären eines Atoms.

Abbildung 5-6 zeigt für Furan und Carben, wie sich die unterschiedlichen Startwerte auf die Atomhashcodes auswirken. In neutralem Furan erhalten das Sauerstoffatom und das benachbarte Kohlenstoffatom als Seedparameter den Wert 12, in der Form, in der das aromatische π -Elektronensystem ionisiert ist, je 11, in der Form in der das freie Elektronenpaar ionisiert ist, 11 und 12. Der Einfluß der Hashcodes der Nachbaratome bewirkt jeweils, daß Atome mit dem gleichen Seedwert unterschiedliche Hashcodes erhalten. In den beiden unterschiedlichen Carbenen erhält jedes der beiden Kohlenstoffatome einen anderen Seed für die Zahl der besetzten Elektronensysteme. Auch hier wirkt sich dieser Anfangswert auf die benachbarten Atome aus, die Wasserstoffatome beider Zustände haben verschiedene Hashcodes.

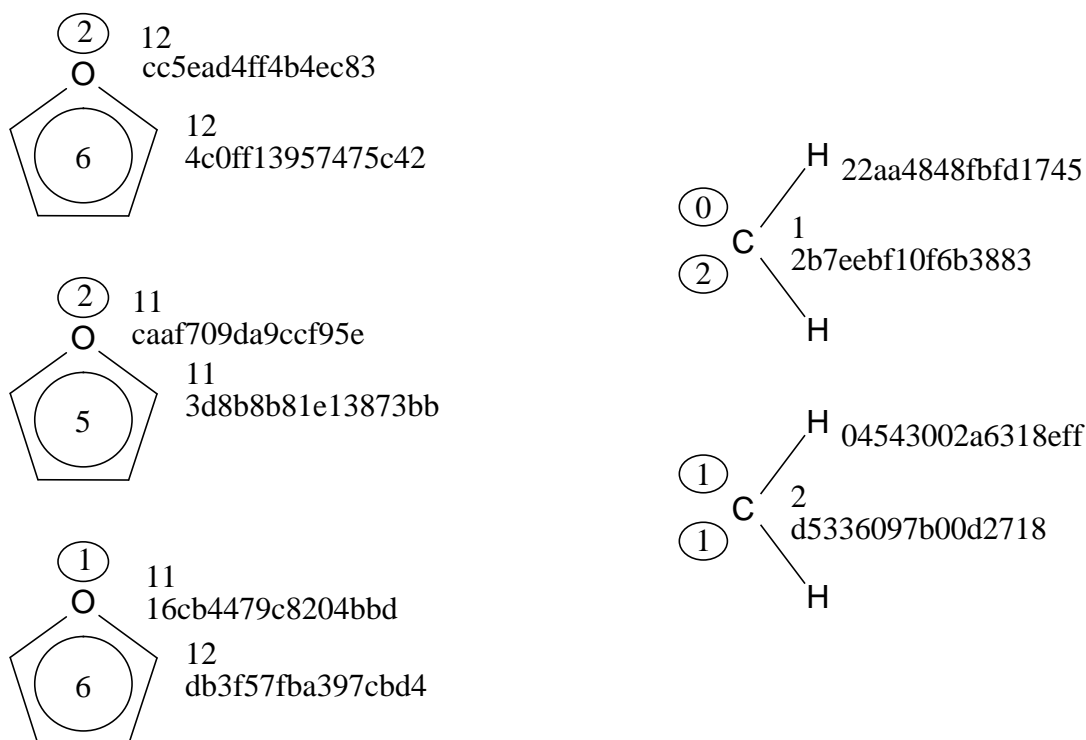


Abb. 5-6: Einfluß des Seeds auf Verbindungen mit dem gleichen Elektronensystemen in unterschiedlicher Besetzung mit Elektronen. Neben den Atomen steht jeweils der Seedwert und der ermittelte Atomhashcode. Die Hashcodes sind als Hexadezimalzahlen angegeben.

Auf der Grundlage der Atomhashcodes können nun sowohl Elektronensystem- als auch Molekülhashcodes berechnet werden.

5.4.3. Elektronensystemhashcodes

Im EROS-System sind Bindungseigenschaften gerichtete Eigenschaften, das heißt, es gibt pro Bindung ein Wertepaar. Diese Werte können sich entweder vollständig oder nur im Vorzeichen unterscheiden. Nur in symmetrischen Bindungen sind beide Werte gleich. Um dieses Verhalten auch in den Bindungshashcodes wiederzuspiegeln, berechnet das Verfahren nach Ihlenfeldt für jede Bindung zwei gerichtete Hashcodes, die nur in symmetrischen Bindungen identisch sind. Da die Atomzahl in Elektronensystemen bei RICOS jedoch nicht auf zwei festgelegt ist, gibt es pro Elektronensystem nicht exakt ein Paar aus benachbarten Atomen. Wieviele Paare benachbarter Atome ein Elektronensystem enthält, hängt von der Art des Elektronensystems und der Anzahl seiner Atome ab. Es ist daher nicht sinnvoll, für jedes Elektronensystem zwei gerichtete Hashcodes zu berechnen, jedem Elektronensystem wird nur ein Hashcode zugeordnet.

Auch bei der Bestimmung von Elektronensystemhashcodes wird jedes Elektronensystem zunächst mit einem Startwert initialisiert, der dann in mehreren Zyklen durch die anderen Elektronensystemhashcodes verfeinert wird. Dabei werden drei Fälle unterschieden. Wenn ein Atom π -Elektronensysteme besitzt, die sowohl freie Elektronen als auch unbesetzte Orbitale repräsentieren, muß zwischen diesen unterschieden werden. Der Elektronensystemhashcode wird aus dem Atomhashcode bestimmt, indem dieser um die Anzahl der Elektronen im Elektronensystem nach links rotiert wird. In Fällen, in denen ein Elektronensystem aus zwei Atomen besteht, wird jeweils der größere der beiden Atomhashcodes um ein Bit nach links rotiert, bevor er mit dem zweiten XOR-verknüpft wird. Wenn ein Elektronensystem aus mehr als zwei Atomen besteht, werden zur Bestimmung des Initialwerts alle Atomhashcodes durch XOR-Operationen miteinander verknüpft. Falls der gleiche Atomhashcode mehrfach vorhanden ist, wird bestimmt, wie oft er bisher aufgetreten ist, und vor der XOR-Operation um diesen Wert nach links rotiert. Dadurch werden Auslöschungen durch eine gerade Anzahl gleicher Hashcodes vermieden.

In Verbindungen ohne Ringsysteme ist damit die Bestimmung des Elektronensystemhashcodes beendet. Für cyclische Verbindungen reicht dieses Verfahren nicht aus, da in Molekülen wie Biphenylen unterschiedliche σ -Elektronensysteme zwischen äquivalenten Atomen auftreten (Abbildung 5-7) und allein auf der Basis von Atomhashcodes nicht differenziert werden können. Sie können jedoch durch eine Pfadsuche unterschieden werden. Dazu werden zuerst alle Bindungen ausgewählt, die sich nicht in einer terminalen Kette des Moleküls befinden und die aus zwei Atomen bestehen. Für diese werden dann alle möglichen Pfade zwischen beiden an der Bindung beteiligten Atomen gesucht, und sich deren Länge und Häufigkeit gemerkt. Sich kreuzende Pfade werden nicht berücksichtigt, d. h. jedes Atom wird nur einmal traversiert. Jeder Elektronensystemhashcode wird nun durch mehrere Rotationen und XOR-Verknüpfungen modifiziert, wobei die Bitverschiebungen von der Länge und Häufigkeit der gefundenen Pfade abhängen. Danach werden wiederum mehrere Zyklen durchgeführt, in denen diese Has-

hcodes auf die Elektronensystemhashcodes benachbarter Elektronensysteme einschließlich der zuvor ausgenommenen Elektronensysteme propagiert werden, und sich so auch auf diese auswirken.

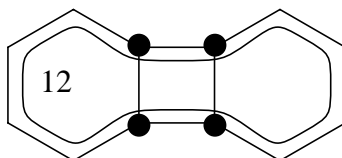


Abb. 5-7: Zwei Paare von σ -Elektronensystemen zwischen den äquivalenten Atomen in Biphenylen sind unterscheidbar, diejenigen, die Teil eines Benzolrings sind und diejenigen, die zwischen beiden Benzolringen liegen.

Durch die Pfadsuche werden σ -Elektronensysteme wie in Biphenylen sicher differenziert. So werden für die σ -Elektronensysteme in den Benzolringen 4 Pfade der Länge 1, 3, 5 und 7 gefunden, für die σ -Elektronensysteme zwischen den Benzolringen 5 Pfade der Länge 1, 3, 7 (2x) und 11.

5.4.4. Molekül-, Aggregat- und Ensemblehashcodes

Ein Molekülhashcode wird durch XOR-Verknüpfungen der Hashcodes aller Atome ermittelt. Wenn ein Molekül nur ein Atom enthält, ist der Molekülhashcode gleich dem Atomhashcode. Um eine Auslöschung durch identische Werte zu vermeiden, wird vor der Hashcodeberechnung für jeden Atomhashcode bestimmt, das wievielte Mal er im Molekül auftritt. Wenn ein Atomhashcode mehrfach vorhanden ist, wird er vor der XOR-Verknüpfung um den zuvor ermittelten Wert nach links rotiert.

Ein Aggregathashcode wird im ersten Schritt nach dem gleichen Verfahren aus den Molekülhashcodes generiert wie diese aus Atomhashcodes. Danach wird zusätzlich noch ein Gruppenhashcode aus allen Atomen der Wechselwirkungsgruppen, die ein Aggregat bilden, erzeugt und dieser ebenfalls mit dem Aggregathashcode durch den XOR-Operator verknüpft. Dadurch werden für Aggregate, die zwar aus den gleichen Molekülen bestehen, aber unterschiedliche Wechselwirkungsgruppen haben, verschiedene Hashcodes erzeugt. Dies gilt sowohl für Wechselwirkungsgruppen verschiedener Art als auch für Wechselwirkungen der gleichen Art, die jedoch zwischen verschiedenen, nicht äquivalenten Atomen oder Elektronensystemen bestehen.

Die Ensemblehashcodes werden auch durch XOR-Verknüpfungen aus den Molekül- und Aggregathashcodes eines Ensembles generiert. Enthält ein Ensemble nur ein Molekül oder Aggregat, erhält es dessen Hashcode.

5.4.5. Äquivalenzklassen

Anhand der Hashcodes werden die Äquivalenzklassen der Atome, Elektronensysteme, Moleküle und Aggregate in einem Ensemble bestimmt und gezählt, wieviele äquivalente Objekte es pro Äquivalenzklasse gibt. Diese Information wird in der Reaktionsgenerierung verwendet, um zu verhindern, daß Reaktionen mit äquivalenten Atomen oder Elektronensystemen mehrfach ausgeführt werden, und um die Symmetriezahl einer Reaktion zu bestimmen.

5.5. Physikochemische Parameter auf der Basis von Valenz-Bond-Strukturen

Im Rahmen dieser Arbeit konnten nur die wichtigsten topologischen Eigenschaften auf das RICOS-Modell übertragen werden. Um aber Reaktionsvorhersage auf der Basis von physikochemischen Eigenschaften durchführen zu können, war es vonnöten, auch in der Übergangsphase den größtmöglichen Satz an bisher entwickelten physikochemischen Eigenschaften zur Verfügung zu haben. Es wurde eine Konvertierung organischer Verbindungen von RICOS in VB-Strukturen entwickelt, die in Kapitel 7.2 detailliert beschrieben wird. Sie wird auch dazu verwendet, um Verbindungen, deren Eigenschaften berechnet werden sollen, in konventionelle Bindungslisten zu übertragen, und auf ihnen die Verfahren zur Berechnung von physikochemischen Eigenschaften anzuwenden. Diese Konvertierung kann natürlich nicht ohne Informationsverlust in Bezug auf die Elektronenverteilung in orthogonalen π -Elektronensystemen stattfinden, zudem wird für jede Struktur nur eine der möglichen Resonanzstrukturen erzeugt. Sie ist aber ausreichend, um alle Eigenschaften, in deren Berechnungsmethode eine Behandlung von Resonanzstrukturen integriert ist, in das neue Modell zu übernehmen. Der Ablauf der Berechnung dieser Eigenschaften ist in Abbildung 5-8 schematisch dargestellt. Ein Ensemble in RICOS-Kodierung wird in das Datenformat von EROS konvertiert. Auf dieser Datenstruktur werden die Eigenschaften berechnet. Anhand einer Übersetzungstabelle werden danach die Eigenschaften den entsprechenden chemischen Objekten in RICOS zugewiesen und die EROS-Datenstruktur wieder verworfen.

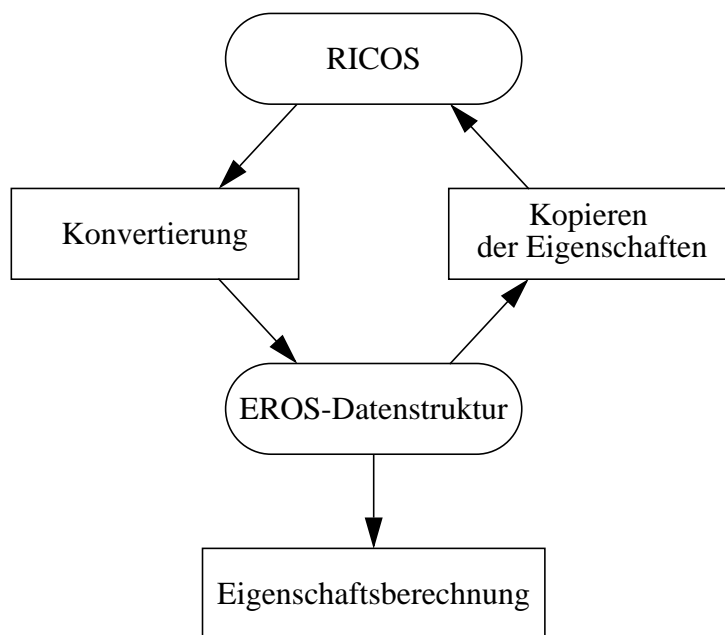


Abb. 5-8: Berechnung physikochemischer Eigenschaften über die Konvertierung in das VB-Modell.

Die Datenstruktur in EROS besteht aus Ensembles, die Moleküle aus Atomen und Bindungen enthalten. In EROS gibt es das Konzept eines Aggregats und von Wechselwirkungsgruppen nicht. Abbildung 5-9 zeigt, welche Konzepte in beiden Datenstrukturen miteinander korrespondieren. Die Linien verbinden die chemischen Einheiten miteinander, deren Eigenschaften aufeinandergepaßt werden. Eigenschaften von Atomen und Molekülen sind direkt übertragbar. Die Eigenschaften von Ensembles sind in der Regel aus allen Moleküleigenschaften zusammengesetzt. Sie sind dann übertragbar, wenn die Wechselwirkungen innerhalb von Aggregaten vernachlässigt werden. Am schwierigsten gestaltet sich die Übertragung von Bindungseigenschaften, da die Definition einer Bindung in beiden Konzepten stark voneinander abweicht. In der EROS-Datenstruktur beinhalten Bindungseigenschaften nicht nur Eigenschaften des σ -Bindungsgerüsts, sondern auch des π -Anteils einer Bindung. Eine direkte Korrespondenz mit σ - oder π -Elektronensystemen ist nicht gegeben. Daher werden die Eigenschaften von Bindungen nicht Elektronensystemen in RICOS zugeordnet, sondern Paaren von benachbarten Atomen. Da Bindungseigenschaften in EROS gerichtet sind, gibt es für jede Bindung ein Wertepaar. Beide Werte werden so kopiert, daß der erste Wert mit der Bindung vom Atom mit einem niedrigeren Index zum Atom mit einem höheren Index korrespondiert und der zweite Wert für die umgekehrte Bindungsrichtung gültig ist.

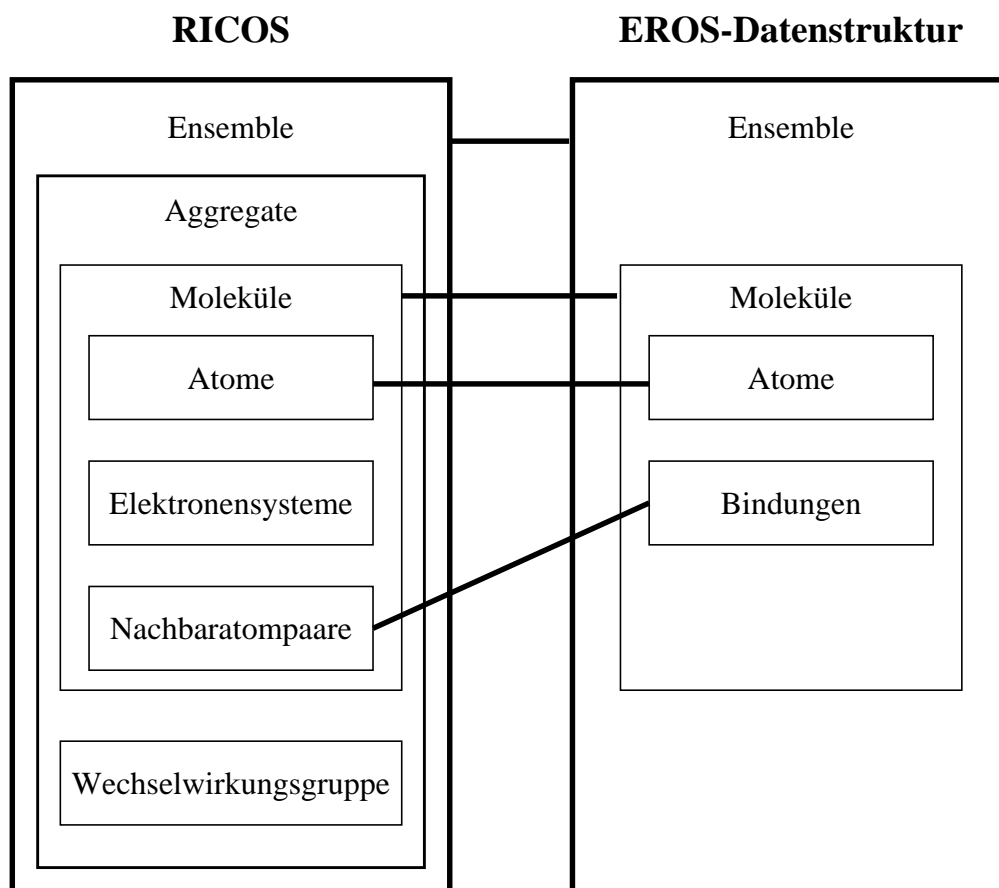


Abb. 5-9: Korrespondenz zwischen RICOS und der EROS-Datenstruktur bei der Übertragung von physikochemischen Eigenschaften.

Tabelle 5-3 enthält alle Eigenschaften, die auf der EROS-Datenstruktur berechnet werden und auf RICOS übertragen werden. Neben der Kurzbeschreibung ist jeweils das Schlüsselwort und die Einheit der Eigenschaft angegeben. Bindungseigenschaften erhalten den Präfix B_.

Tab. 5-3: Physikochemische Parameter, die von der EROS-Datenstruktur übernommen werden.

| Beschreibung | Schlüsselwort | Einheit |
|---|---------------|---------------|
| Aromatizitätsinformation | A_AROMATIC | |
| Atompolarisierbarkeit (α) [56] | A_POLARIZ | [Angström**3] |
| Elektronegativität der freien Elektronen | A_ENLP | [eV] |
| π -Elektronegativität (χ_π) [57] | A_ENPI | [eV] |
| σ -Elektronegativität (χ_σ) [57] | A_ENSIG | [eV] |

| Beschreibung | Schlüsselwort | Einheit |
|---|---------------|---------------|
| Mesomeriestabilisierung durch freie Elektronenpaare | A_LPSTAB | |
| π -Atompartialladung (q_π) [58] | A_QPI | [e] |
| σ -Atompartialladung (q_σ) [59] | A_QSIG | [e] |
| gesamte Atompartialladung | A_QTOT | [e] |
| Standardbildungswärme (ΔH_f°) [60] | M_DELTAHF | [kJ/mol] |
| Molekölpolarisierbarkeit | M_POLARIZ | [Angström**3] |
| aromatische Stabilisierungsenergie | M_STABIL | [kJ/mol] |
| Standardbildungswärme aller Moleküle des Ensembles | E_DELTAHF | [kJ/mol] |
| Polarisierbarkeit aller Moleküle des Ensembles | E_POLARIZ | [Angström**3] |
| aromatische Stabilisierungsenergie aller Moleküle des Ensembles | E_STABIL | [kJ/mol] |
| Differenz der π -Elektronegativität der Bindungsatome je Bindung ($\Delta\chi_\pi$) | B_DENPI | [eV] |
| Differenz der σ -Elektronegativität der Bindungsatome je Bindung ($\Delta\chi_\sigma$) | B_DENSIG | [eV] |
| Differenz der σ -Atompartialladungen der Bindungsatome je Bindung (Δq_σ) | B_DQSIG | [e] |
| Differenz der π -Atompartialladungen der Bindungsatome je Bindung (Δq_π) | B_DQPI | [e] |
| Differenz der (Gesamt-)Atompartialladungen der Bindungsatome je Bindung (Δq) | B_DQTOT | [e] |
| Bindungspolarisierbarkeit | B_POLARIZ | [Angström**3] |
| die durch den PEOE-Algorithmus über die Bindung verschobene Ladungsmenge [59] | B_SQIT | [e] |
| (Gesamt-)Delokalisierungsstabilisierung von Ladungen bei heterolytischen Bindungsbruch (D^\pm) [61] | B_SDELOC | [eV] |
| Delokalisierungsstabilisierung einer negativen Ladung bei heterolytischem Bindungsbruch (D^+) [61] | B_NDELOC | [eV] |

| Beschreibung | Schlüsselwort | Einheit |
|--|---------------|---------|
| Delokalisierungsstabilisierung einer positiven Ladung bei heterolytischem Bindungsbruch (D^-) [61] | B_PDELOC | [eV] |
| Bindungsordnung in der EROS-Datenstruktur | B_BOORD | |

Auf der Basis der auf diese Weise zur Verfügung gestellten physikochemischen Eigenschaften können nun in der Neuentwicklung des EROS-Systems sofort Modelle zur Reaktionsvorhersage organischer Reaktionstypen entwickelt werden. Auch die bisher entwickelten Modelle zur Reaktionsvorhersage können auf das neue System übertragen werden und sind damit auch in Zukunft verfügbar.

6. Reaktionen von Molekülen und Aggregaten in einem Ensemble

Das folgende Kapitel beschäftigt sich mit Methoden zur Generierung von Reaktionen. Sie sind im Vergleich zur Reaktionsgenerierung der Programme EROS6, FRANZ und MASSIMO wesentlich erweitert worden, um eine flexiblere Gestaltung der formalen Regeln für die Reaktionsgenerierung zu erlauben. Die Reaktionsgeneratoren dieser Programme erlauben lediglich Elektronenumordnungsprozesse in Form von Änderungen der Bindungsordnungen und der Anzahl der freien Elektronen an Atomen. Dies ist jedoch für eine flexible Formulierung von Reaktionstypen, wie sie in der Neuentwicklung von EROS verwirklicht wurden, nicht ausreichend. Die im folgenden vorgestellten Konzepte bilden die Schnittstelle zu deren Reaktionsgenerator, der die Formulierung von Reaktionsregeln sowie die Steuerung der Reaktionsgenerierung enthält.

Außer Methoden zur Neubildung und zum Brechen von Elektronensystemen, die die speziellen Eigenschaften der drei Arten von Elektronensystemen berücksichtigen, werden auch Methoden zur Änderung von Wechselwirkungs- und topologischen Gruppen vorgestellt. Zudem gibt es Module, die sowohl das Einschleusen, als auch das Verwerfen von Aggregaten und Molekülen während einer Reaktion erlauben.

Eine zweite Anwendungsmöglichkeit der in den folgenden Abschnitten vorgestellten Konzepte zur Modellierung von Reaktionen liegt in der Entwicklung eines Moleküleditors, der die Reaktionsschritte beim internen Aufbau von Molekülen aus Atomen nutzen könnte. Denn die Methode, die zur Konvertierung von Molekülen in VB-Kodierung nach RICOS entwickelt wurde, kann nur Strukturen konvertieren, für die es auch eine gültige Bindungsliste gibt, nicht jedoch Verbindungen mit Mehrzentrenbindungen oder koordinativen Bindungen. Es wäre daher wünschenswert, einen Editor zur Verfügung zu haben, der direkt die Abspeicherung in RICOS erlaubt.

6.1. Reaktionen von Elektronensystemen

Die Modellierung einer Reaktion erfolgt im VB-Modell durch sequentielles Anwenden von mehreren Elementarschritten auf die Edukte. Es gibt zwei Elementarschritte, aus denen sich sämtliche Reaktionstypen aufbauen lassen. Zum einen kann die Bindungsordnung zwischen zwei Atomen geändert werden. Eine Erhöhung der Bindungsordnung umfaßt dabei auch das Bilden von Einfachbindungen zwischen nicht aneinander gebundenen Atomen, das Erniedrigen der Bindungsordnung beinhaltet auch den Bindungsbruch einer Einfachbindung. Der zweite Elementarschritt ist die Änderung der Zahl der freien Elektronen an einem Atom.

Um nun die Behandlung von Reaktionen zwischen Molekülen, die mit der vorgestellten Repräsentation kodiert werden, zu ermöglichen, wurden ebenfalls einfache Elementarschritte formuliert, aus denen sich komplexere Reaktionsmechanismen zusammensetzen lassen. Jeder Elementarschritt besteht aus der Kombination oder Teilung von Elektronensystemen, wobei

sich die Gesamtzahl der Elektronen der an dem Reaktionsschritt beteiligten Elektronensysteme nicht ändern darf. Weiterhin dürfen Elektronensysteme nur dann zwischen Atomen gebildet werden, wenn die für den speziellen Elementarschritt notwendigen Elektronensysteme an beiden Atomen zur Verfügung stehen.

Die Elementarschritte werden in drei Gruppen eingeteilt, Schritte zum Machen und Brechen von σ -Elektronensystemen, π -Elektronensystemen und koordinativen Bindungen. π -Elektronensysteme sind als Elektronendonoren oder -akzeptoren an allen Schritten beteiligt. Die Elementarschritte werden im folgenden zusammen mit erläuternden Beispielen vorgestellt. In den Schemata der Abbildungen 6-1, 6-4 und 6-6 sind die gebildeten Elektronensysteme jeweils dick umrandet. Wenn ein π -Elektronensystem an der Seite offen gezeichnet ist, bedeutet dies, daß seine Ausdehnung auf dieser Seite nicht festgelegt ist.

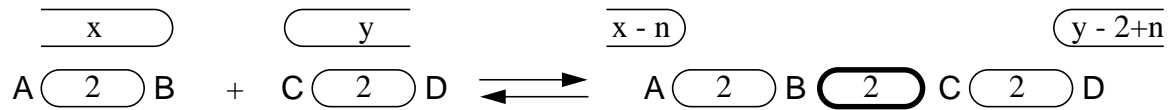
6.1.1. Elementarschritte zum Erzeugen und Brechen von σ -Elektronensystemen

Reguläre σ -Elektronensysteme werden aus zwei nicht benachbarten Atomen gebildet, die π -Elektronensysteme als Elektronendonoren zur Verfügung haben (Abbildung 6-1a). Im Zuge eines solchen Reaktionsschritts werden die beteiligten π -Elektronensysteme um die zur Bildung des regulären σ -Elektronensystems benötigten Atome verkürzt und die Zahl der Elektronen reduziert. Dabei kann sowohl ionische Bindungsbildung ($n = 0, 2$), bei der ein Atom keine Elektronen, das zweite Atom zwei Elektronen zum Elektronensystem beisteuert, als auch eine Bindungsbildung, bei der beide Atome je ein Elektron in das Elektronensystem donieren ($n = 1$), erfolgen. Zwischen den beiden Atomen wird dann das neue σ -Elektronensystem gebildet. Dabei müssen die beteiligten Atome jeweils am Ende eines π -Elektronensystems liegen.

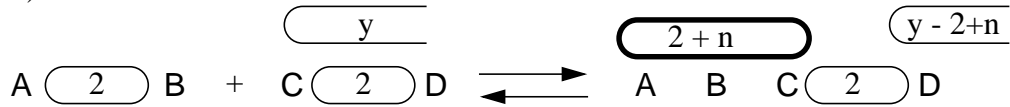
Mehrzentren- σ -Elektronensysteme werden aus je einem schon vorhandenen σ -Elektronensystem und einem π -Elektronensystem gebildet. Abbildung 6-1b zeigt die Bildung eines verbückten σ -Elektronensystems, Abbildung 6-1c die Bildung eines geschlossenen σ -Elektronensystems. Auch hier ist es möglich, entweder größere π -Elektronensysteme um je ein Atom zu verkürzen oder freie Elektronen in einatomigen π -Elektronensystemen zu verwenden.

Die Umkehrreaktionen aller gezeigten Elementarschritte führen zum Bruch des jeweils stark umrandeten σ -Elektronensystems in Abbildung 6-1. Dabei gibt es jeweils ebenfalls zwei Fälle. Einerseits können isolierte π -Elektronensysteme erzeugt werden, andererseits können die freigebliebenen π -Elektronen mit schon vorhandenen π -Elektronensystemen konjugiert werden.

a)



b)



c)

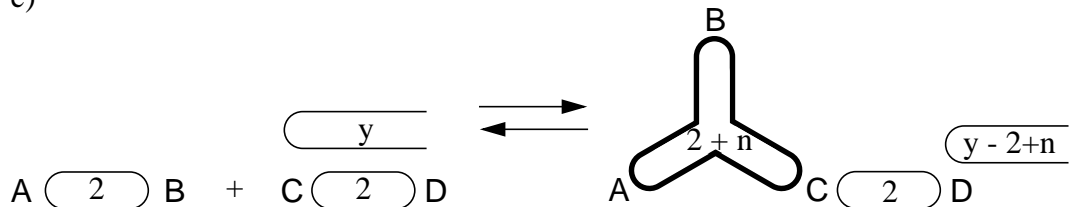


Abb. 6-1: Reaktionsschritte, mit denen σ -Elektronensysteme gemacht oder gebrochen werden. Die Zahlen innerhalb der Ovale, die die Elektronensysteme darstellen, bezeichnen die Anzahl der Elektronen. n bzw. $2+n$ geben an, wieviele Elektronen des π -Elektronensystems zur Bildung des σ -Elektronensystems verwendet werden. x, y ist beliebig.

Die Bildung regulärer σ -Elektronensysteme wird anhand der Abbildung 6-2 veranschaulicht. Abbildung 6-2a zeigt die traditionelle VB-Kodierung der Addition eines Methoxylatanions an Acetaldehyd in der basenkatalysierten Bildung des Halbacetals. Abbildung 6-2b stellt den Elementarschritt, der zur Erzeugung des σ -Elektronensystems führt, dar. Das neu entstandene σ -Elektronensystem ist durch ein Oval gekennzeichnet, alle übrigen σ -Elektronensysteme als Linien zwischen den Atomen. Beim Angriff des Methoxylatanions auf Acetaldehyd während der basenkatalysierten Bildung des Halbacetals wird ein reguläres σ -Elektronensystem zwischen dem Carbonylkohlenstoffatom und dem Sauerstoffatom des Alkohols gebildet. Das π -Elektronensystem der Carbonylgruppe wird um ein Atom verkürzt, beide Elektronen bleiben am Sauerstoffatom. Das zweite π -Elektronensystem, welches ein freies Elektronenpaar des Alkoholsauerstoffatoms bezeichnet, gehört nur zu einem Atom und steuert beide Elektronen in das neue Elektronensystem bei.

Die Elektronensysteme zwischen den Schweratomen der Edukte und des Produkts sehen folgendermaßen aus:

Edukte: C: 1, 2, 5; O: 3, 4

{(1 2 s r; 2) (2 3 s r; 2) (2 3 p; 2) (3 p; 2) (3 p; 2)} {(4 p; 2) (4 p; 2) (4 p; 2) (4 5 s r; 2)}

Produkt: C: 1, 2, 5; O: 3, 4

(1 2 s r; 2) (2 3 s r; 2) (2 4 s r; 2) (3 p; 2) (3 p; 2) (3 p; 2) (4 p; 2) (4 p; 2) (4 5 s r; 2)

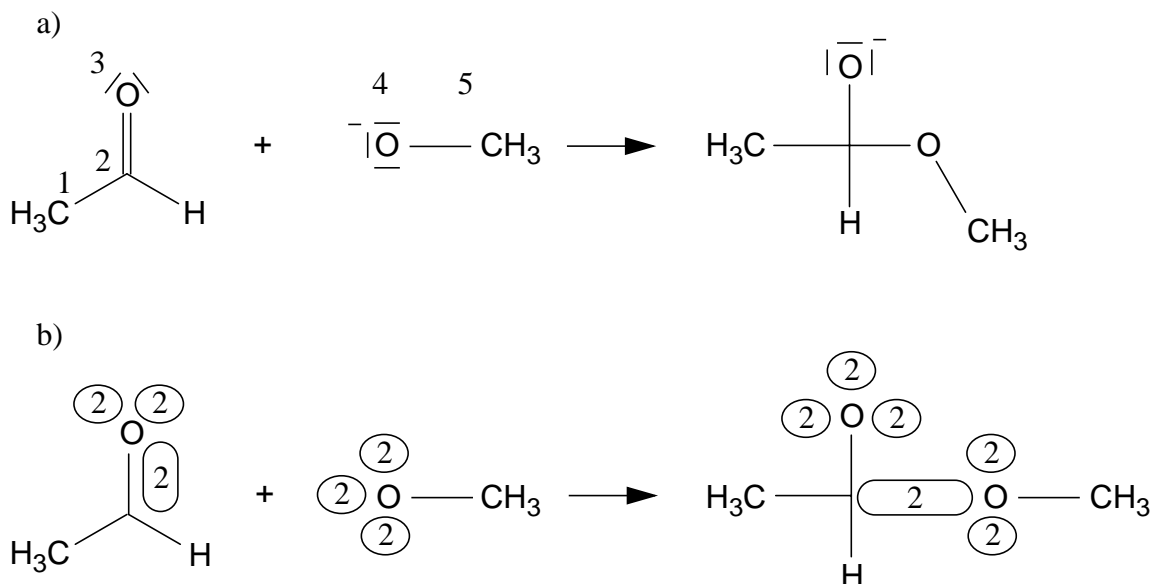


Abb. 6-2: Bildung eines regulären σ -Elektronensystems. Addition des Methoxyanions an Acetaldehyd bei der basenkatalysierten Bildung eines Halbacetals, (a) VB-Kodierung, (b) RICOS-Kodierung.

In Boranen treten geschlossene und verbrückte σ -Elektronensysteme auf. Abbildung 6-3 zeigt, wie aus einem regulären BH σ -Elektronensystem und der freien Valenz von Bor ein verbrücktes BHB σ -Elektronensystem gebildet werden kann. Die dargestellte Reaktion ist der erste Schritt der Dimerisierung von Boran. Das reguläre σ -Elektronensystem zwischen dem Boratom 5 und dem Wasserstoffatom 6 wird um das Boratom 1 erweitert. Das Wasserstoffatom wird als Brückenatom gekennzeichnet. Da das π -Elektronensystem, welches die freie Valenz am Boratom 1 repräsentiert, keine Elektronen enthält, ändert sich die Zahl der Elektronen im neu gebildeten σ -Elektronensystem nicht.

Edukte: B: 1,5; H: 2-4, 6-8

{(1 2 s r; 2) (1 3 s r; 2) (1 4 s r; 2) (1 p; 2)} {(5 6 s r; 2) (5 7 s r; 2) (5 8 s r; 2) (5 p; 2)}

Produkt: B: 1,5; H: 2-4, 6-8

(1 2 s r; 2) (1 3 s r; 2) (1 4 s r; 2) (1 5 6 s b 2; 2) (5 7 s r; 2) (5 8 s r; 2) (5 p; 2)

Die analoge Reaktion zwischen dem regulären σ -Elektronensystem zwischen den Atomen 1 und 4 sowie der freien Valenz am Boratom 5 vervollständigt die Dimerisierung.

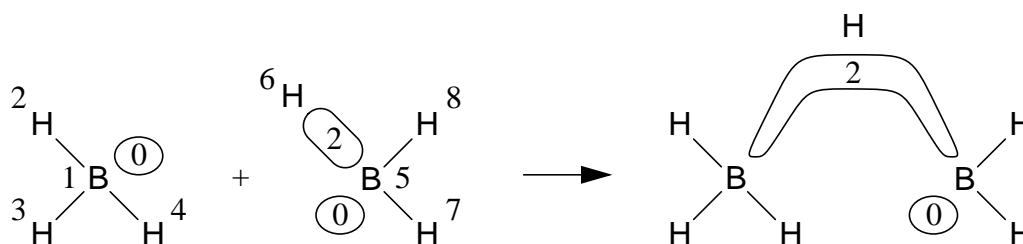


Abb. 6-3: Bildung eines verbrückten σ -Elektronensystems aus einem regulären σ -Elektronensystem und einem Atom mit einem π -Elektronensystem. Der erste Reaktionsschritt in der Dimerisierung von Boran.

6.1.2. Elementarschritte zum Erzeugen und Brechen von π -Elektronensystemen

In den in Abschnitt 6.1.1 vorgestellten Elementarschritten zur Bildung von σ -Elektronensystemen gilt die Bedingung, daß die Atome, die aus einem π -Elektronensystem entfernt werden, immer am Ende eines solchen Elektronensystems lokalisiert sein müssen. In komplexeren Reaktionsmechanismen ist dies aber nicht immer gewährleistet. Daher wurde ein Elementarschritt entwickelt, der es erlaubt, die Konjugation innerhalb von π -Elektronensystemen aufzuheben, bzw. benachbarte π -Elektronensysteme miteinander zu verbinden. Abbildung 6-4 zeigt die Kombination zweier benachbarter π -Elektronensysteme, die an ihren Eckpunkten miteinander verbunden werden. Es ist aber auch möglich, verzweigte π -Elektronensysteme zu bilden, wenn eines der beiden zu verbindenden Atome innerhalb des π -Elektronensystems liegt.



Abb. 6-4: Reaktionsschritt, mit dem π -Elektronensysteme gemacht oder gebrochen werden.

Abbildung 6-5 zeigt die Konjugation in π -Elektronensystemen. In Buta-1,3-dien (Abbildung 6-5a) werden die beiden Doppelbindungen miteinander verbunden. Die Elektronensysteme zwischen den Kohlenstoffatomen vor und nach dem Reaktionsschritt lauten wie folgt:

nicht konjugiert:

(1 2 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (1 2 p; 2) (3 4 p; 2)

konjugiert:

(1 2 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (1 2 3 4 p; 4)

In Abbildung 6-5b wird die Bildung eines verzweigten π -Elektronensystems anhand der Konjugation des aromatischen Rings mit der Doppelbindung von Styrol gezeigt.

Die Elektronensysteme der Kohlenstoffatome der nicht konjugierten und der konjugierten Form lauten wie folgt:

nicht konjugiert:

(1 2 s r; 2) (1 6 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (4 5 s r; 2) (5 6 s r; 2) (6 7 s r; 2) (7 8 s r; 2)
(1 2 3 4 5 6 p; 6) (7 8 p; 2)

konjugiert:

(1 2 s r; 2) (1 6 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (4 5 s r; 2) (5 6 s r; 2) (6 7 s r; 2) (7 8 s r; 2)
(1 2 3 4 5 6 7 8 p; 8)

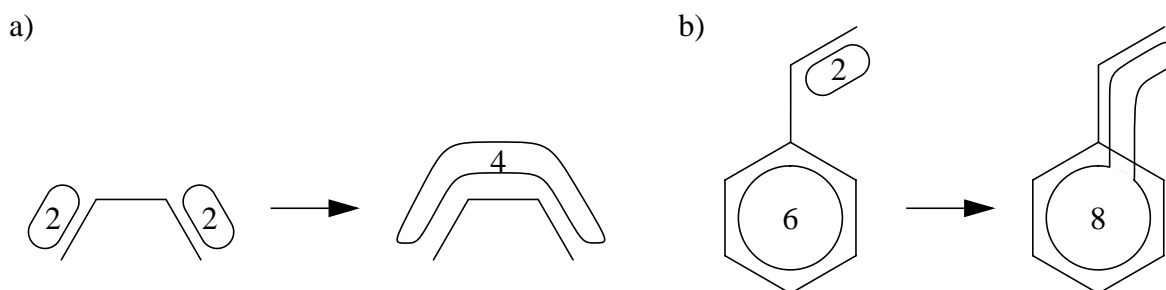


Abb. 6-5: Konjugation von π -Elektronensystemen. a) Konjugation der beiden Doppelbindungen in Buta-1,3-dien, b) Konjugation des aromatischen π -Elektronensystems mit der Doppelbindung in Styrol.

6.1.3. Elementarschritte zum Erzeugen und Brechen von koordinativen Bindungen

Es gibt drei verschiedene Elementarschritte, die die Behandlung von koordinativen Bindungen erlauben (Abbildung 6-6). Abbildung 6-6a zeigt die Bildung einer koordinativen Bindung zwischen einem Metallatom und einem Liganden mit einem π -Elektronensystem. Bei der Bildung einer koordinativen Bindung werden alle Elektronen des Ligand- π -Systems in die Bindung übernommen. Das Metallatom muß jeweils so viele leere π -Elektronensysteme in die Bindung beisteuern, wie sich Elektronenpaare im Liganden befunden haben (siehe auch Abschnitt 4.2.3). Für eine Bindung zu einer Doppelbindung ist also ein π -Elektronensystem am Metall notwendig. Die beiden Elementarschritte aus Abbildung 6-6b und Abbildung 6-6c dienen zur Erweiterung einer schon vorhandenen koordinativen Bindung. Im ersten Fall wird eine verbückte Bindung gebildet, indem ein weiteres Metallatom an die Bindung addiert wird. Das Metallatom muß ebenfalls halb soviel leere π -Elektronensysteme zur Verfügung haben, wie in der Bindung lokalisiert sind. Im zweiten Fall wird ein weiteres Atom des Liganden in die Bindung miteinbezogen. Dieser Elementarschritt wird dann gebraucht, wenn Reaktionen an einem Liganden des Metallkomplexes stattfinden, die zur Änderung seiner Haptizität führen. Alle drei Elementarschritte können in beide Richtungen verwendet werden, zur Bindungsbildung und zum Bindungsbruch.

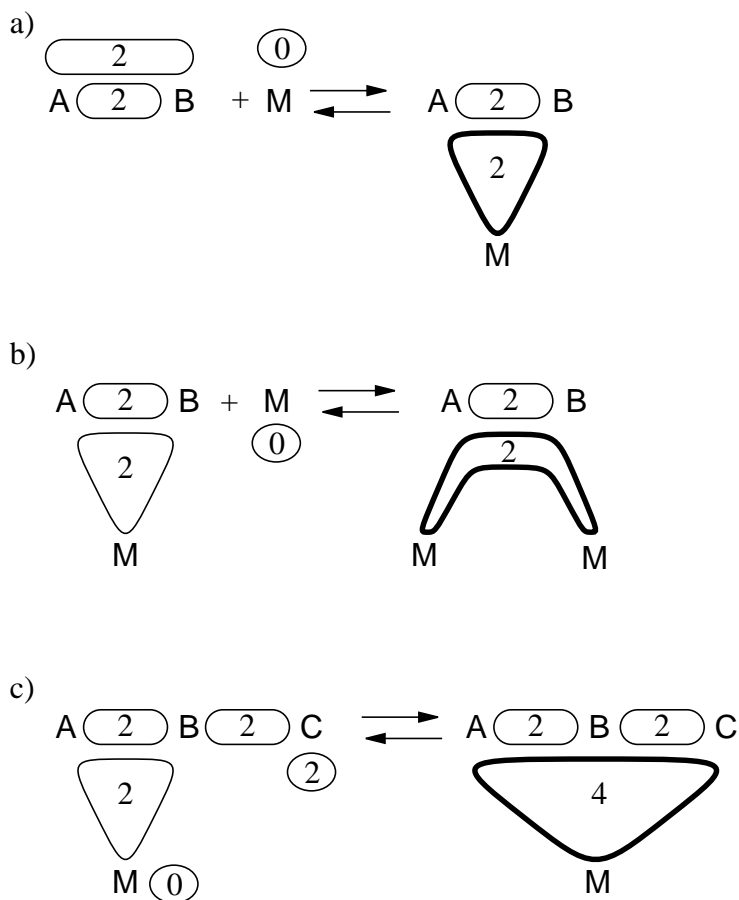


Abb. 6-6: Reaktionsschritte, mit denen koordinative Bindungen gemacht oder gebrochen werden. a) Bildung einer koordinativen Bindung aus einem Metall und einem Liganden, b) Bildung einer verbrückten koordinativen Bindung durch Addieren eines weiteren Metallatoms, c) Erweiterung eines koordinativen Systems um einen weiteren Liganden.

Im folgenden werden drei Beispiele beschrieben, die den Einsatz dieser drei Elementarschritte bei der Modellierung von Reaktionen mit Metallkomplexen demonstrieren.

Abbildung 6-7 zeigt die Bildung einer koordinativen Bindung zwischen dem Mangan(II)kation und Methylcyclopentadienylanion. die Elektronenkonfiguration des Mangankations lautet: $3d^5$. Es stehen also von den elf im Modell als π -Elektronensystemen abgebildeten Orbitalen sechs leere Orbitale zur Ausbildung einer Bindung mit einem Liganden zur Verfügung. Um die sechs Elektronen eines Liganden aufnehmen zu können, steuert das Manganion je drei leere π -Elektronensysteme in die koordinative Bindung bei. Eine Wiederholung dieses Reaktionsschritts mit einem zweiten Methylcyclopentadienylliganden liefert den Sandwichkomplex.

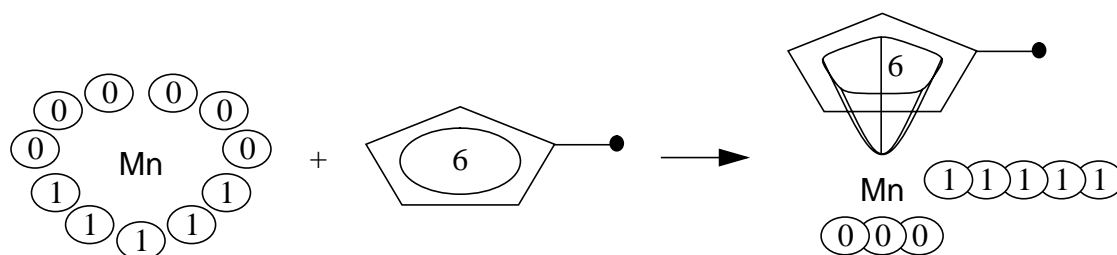


Abb. 6-7: Bildung einer koordinativen Bindung aus dem π -Elektronensystem des Methylcyclopentadienylanions und drei leeren π -Elektronensystemen des Mangan(II)kations (Elektronenkonfiguration: $3d^5$).

$\text{Co}_2(\text{CO})_8$ (Abbildung 6-8) liegt in Lösungen im Gleichgewicht zwischen der offenen Form (I) und der verbrückten Form vor, in der zwei Carbonylliganden über beide Metallatome verbrückt sind [62]. Der erste Schritt der Modellierung dieser Gleichgewichtsreaktion ist in Abbildung 6-8 dargestellt. Jedes Cobaltatom hat in der nicht verbrückten Form fünf π -Elektronensysteme, von denen zwei unbesetzt sind, und daher weitere Bindungen eingehen können. Zwischen beiden Cobaltatomen besteht ein σ -Elektronensystem, die Carbonylgruppen sind, wie in Abbildung 6-8 angegeben, koordinativ an die Cobaltatome gebunden. Die an der Reaktion beteiligten Elektronensysteme sind als Ovale gezeichnet. Eines der leeren π -Elektronensysteme wird zur Bildung der verbrückten koordinativen Bindung verwendet. Der zweite Reaktionsschritt ist eine Wiederholung dieses Schrittes mit je einer koordinativen Carbonyl-Cobaltbindung und einem leeren π -Elektronensystem des anderen Cobaltatoms.

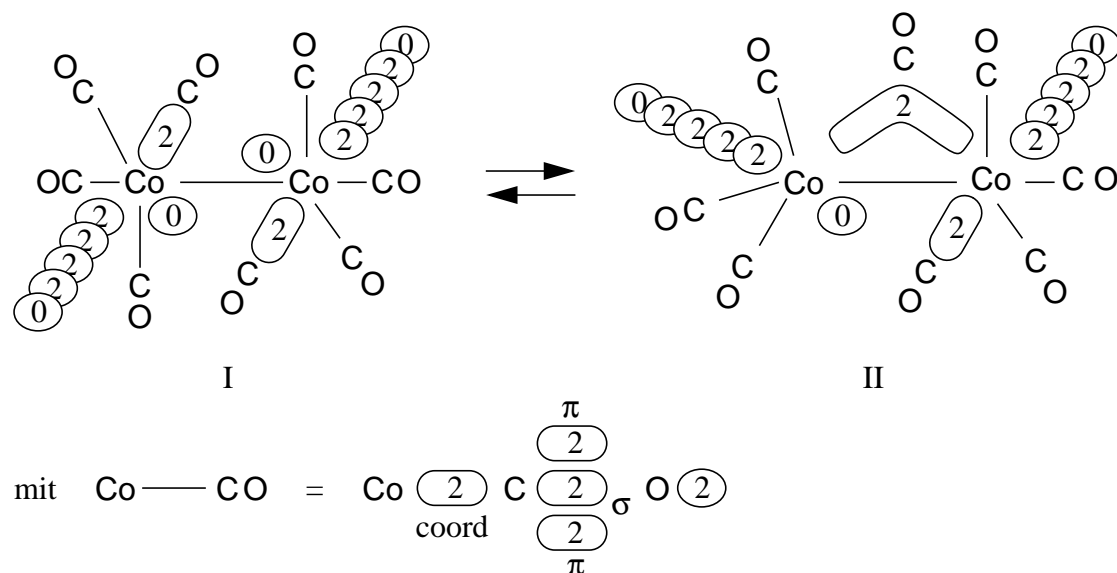


Abb. 6-8: Bildung einer verbrückten koordinativen Bindung. Modellierung des Gleichgewichts zwischen verbrückter und nicht verbrückter Form von $\text{Co}_2(\text{CO})_8$.

Der Elementarschritt, mit dem Reaktionen am Liganden einer koordinativen Bindung eingeleitet werden, kann z. B. in der Modellierung der katalytischen Hydrierung von Alkenen verwendet werden. Ein wichtiger Schritt im Katalysezyklus der Hydrierung mit dem Wilkinson-Katalysator $(\text{Ph}_3\text{P})_3\text{RhCl}$ ist die Insertion des koordinativ gebundenen Alkens in eine Rhodium-Wasserstoffbindung (Abbildung 6-9) [63].

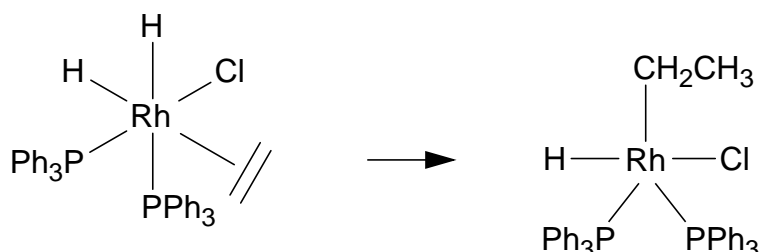


Abb. 6-9: Insertion des koordinativ gebundenen Alkens in die Rh-H-Bindung während der Katalyse der Hydrierung von Alkenen mit dem Wilkinson-Katalysator $(\text{Ph}_3\text{P})_3\text{RhCl}$.

Die Modellierung der Insertion von Ethylen in die Rh-H-Bindung beginnt mit der Erniedrigung der Haptizität der koordinativen Bindung zwischen Ethylen und dem Rhodiumatom so, daß Ethylen nur noch η^1 -gebunden ist und eine freie Valenz erhält (Abbildung 6-10). Diese kann dann in weiteren Reaktionsschritten dazu verwendet werden, ein σ -Elektronensystem zum Wasserstoffatom der Rh-H-Bindung auszubilden.

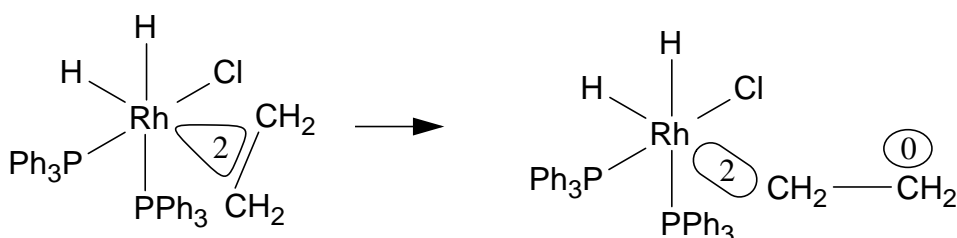


Abb. 6-10: Erniedrigung der Haptizität eines Liganden in einer koordinativen Bindung. Die sechs Valenzelektronen von Rhodium(III), die sich in fünf π -Elektronensystemen befinden, wurden der Übersichtlichkeit wegen weggelassen.

6.1.4. Änderung der Elektronenzahl eines Elektronensystems

Um Reaktionen valenzchemisch richtig zu modellieren, ist es notwendig, darauf zu achten, daß sich die Gesamtelektronenzahl eines Ensembles während einer Reaktion nicht ändert. Dies wird durch die Elementarschritte zum Bilden und Brechen von Elektronensystemen gewährleistet. Für Fälle, in denen sich die Elektronenzahl eines Ensembles durch äußere Einflüsse ändert, ist es jedoch erforderlich, die Zahl der Elektronen von speziellen Elektronensystemen

gezielt zu ändern. Solche Prozesse sind vor allem für die Modellierung der Ionisation in der Elektronenstoßmassenspektrometrie oder in elektrochemischen Reaktionen von Bedeutung.

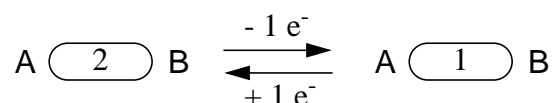


Abb. 6-11: Oxidation und Reduktion eines Elektronensystems durch äußere Einflüsse.

6.1.5. Zusammengesetzte Reaktionen

Komplexere Reaktionsmechanismen lassen sich aus mehreren Elementarschritten zusammensetzen. Die Diels-Alder-Reaktion zwischen Buta-1,3-dien und Acrolein (Abbildung 6-12a) kann mit drei Reaktionsschritten modelliert werden (Abbildung 6-12b). Im ersten Schritt wird das konjugierte π -Elektronensystem von Acrolein gespalten, um die Doppelbindung, die mit dem π -Elektronensystem von Buta-1,3-dien reagiert, zu isolieren. Der zweite und dritte Schritt erzeugen jeweils ein neues σ -Elektronensystem zwischen den Kohlenstoffatomen 1 und 6 sowie 4 und 5. Beide Male wird je ein Elektron aus beiden beteiligten π -Elektronensystemen in das neu gebildete σ -Elektronensystem beigesteuert. Das π -Elektronensystem, das die Doppelbindung in Acrolein repräsentiert, wird dabei vollständig verbraucht, vom π -Elektronensystem des Buta-1,3-diens bleibt ein π -Elektronensystem zwischen den Atomen 2 und 3 zurück, das der Doppelbindung des entstandenen Cyclohexenrings entspricht.

Die Elektronensysteme der Edukte, der beiden Zwischenprodukte und des Produkts sind im folgenden dargestellt. In Ensembles aus mehreren Molekülen sind die Elektronensysteme eines Moleküls in geschweiften Klammern zusammengefaßt.

- I: {(1 2 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (1 2 3 4 p; 4)}
 {(5 6 s r; 2) (6 7 s r; 2) (7 8 s r; 2) (5 6 7 8 p; 4) (8 p; 2) (8 p; 2)}
- II: {(1 2 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (1 2 3 4 p; 4)}
 {(5 6 s r; 2) (6 7 s r; 2) (7 8 s r; 2) (5 6 p; 2) (7 8 p; 2) (8 p; 2) (8 p; 2)}
- III: (1 2 s r; 2) (1 6 s r; 2) (2 3 s r; 2) (3 4 s r; 2) (2 3 4 p; 3) (5 6 s r; 2) (5 p; 1) (6 7 s r; 2)
 (7 8 s r; 2) (7 8 p; 2) (8 p; 2) (8 p; 2)
- IV: (1 2 s r; 2) (1 6 s r; 2) (2 3 s r; 2) (2 3 p; 2) (3 4 s r; 2) (4 5 s r; 2) (5 6 s r; 2) (6 7 s r; 2)
 (7 8 s r; 2) (7 8 p; 2) (8 p; 2) (8 p; 2)

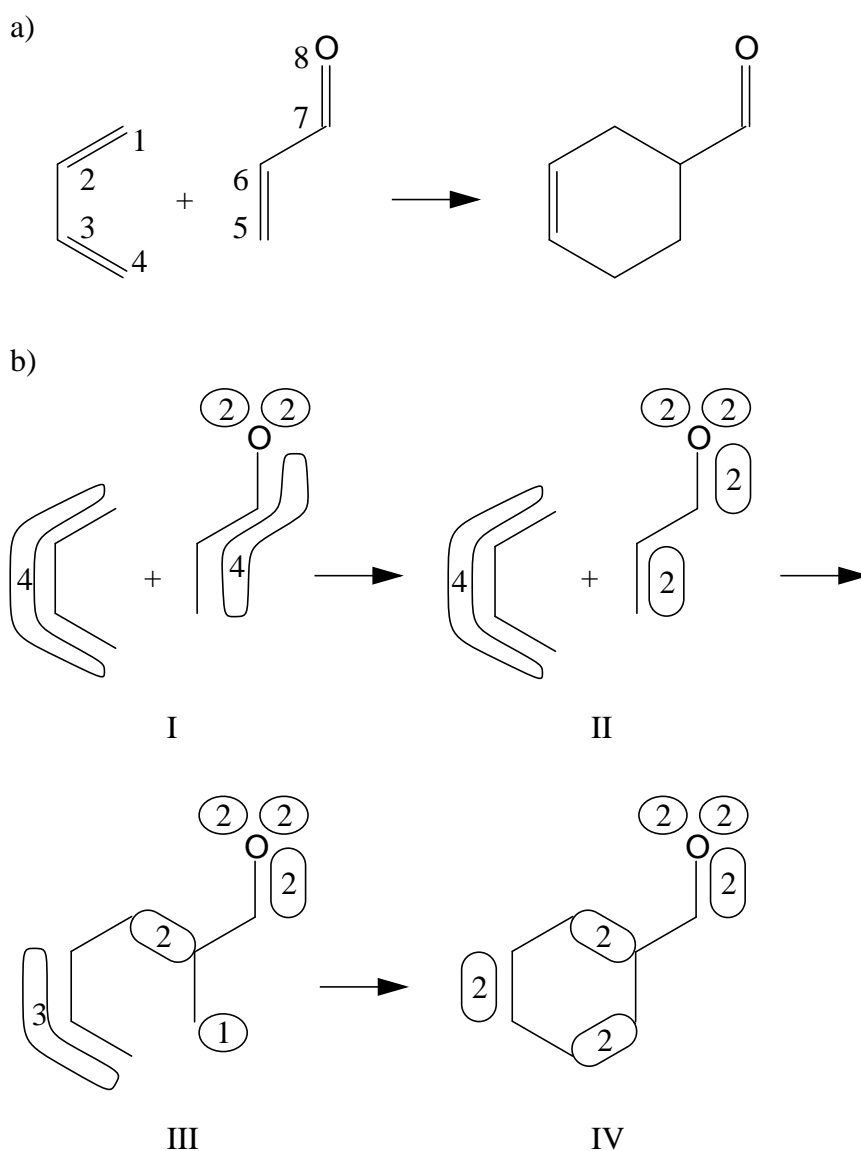


Abb. 6-12: Diels-Alder-Reaktion zwischen Buta-1,3-dien und Acrolein als Beispiel einer Reaktion, die sich aus mehreren Elementarschritten zusammensetzt. (a) VB-Kodierung, (b) RICOS-Kodierung.

6.2. Ändern von Wechselwirkungsgruppen und topologischen Gruppen während einer Reaktion

Im vorigen Abschnitt wurden Reaktionsschritte beschrieben, mit deren Hilfe das Bindungsgerüst von Molekülen geändert werden kann. Um die Zusammensetzung von Aggregaten ändern zu können, wurde ein Reaktionsschritt entwickelt, der es erlaubt, Atome oder Elektronensysteme aus einer Gruppe zu entfernen, oder zu ihr hinzuzufügen.

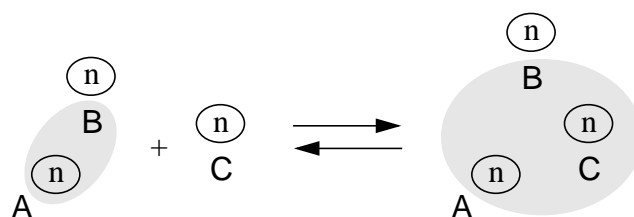


Abb. 6-13: Veränderung der Zusammensetzung einer Gruppe.

Die Reaktion zwischen Natriumchlorid und Silbernitrat kann durch Umordnen der ionischen Wechselwirkungen dargestellt werden (Abbildung 6-14). Das Eduktensemble besteht aus zwei Aggregaten, die durch ionische Wechselwirkungen zwischen dem Natriumkation und dem Chloridanion sowie zwischen dem Silberkation und dem Nitratanion zusammengehalten werden. Bei beiden Kationen wurden die nicht an den Gruppen beteiligten 10 π -Elektronensysteme nicht gezeichnet. Die Reaktion kann nun formuliert werden, daß aus beiden Wechselwirkungsgruppen jeweils das π -Elektronensystem des Kations entfernt wird und in die andere Gruppe eingefügt wird.

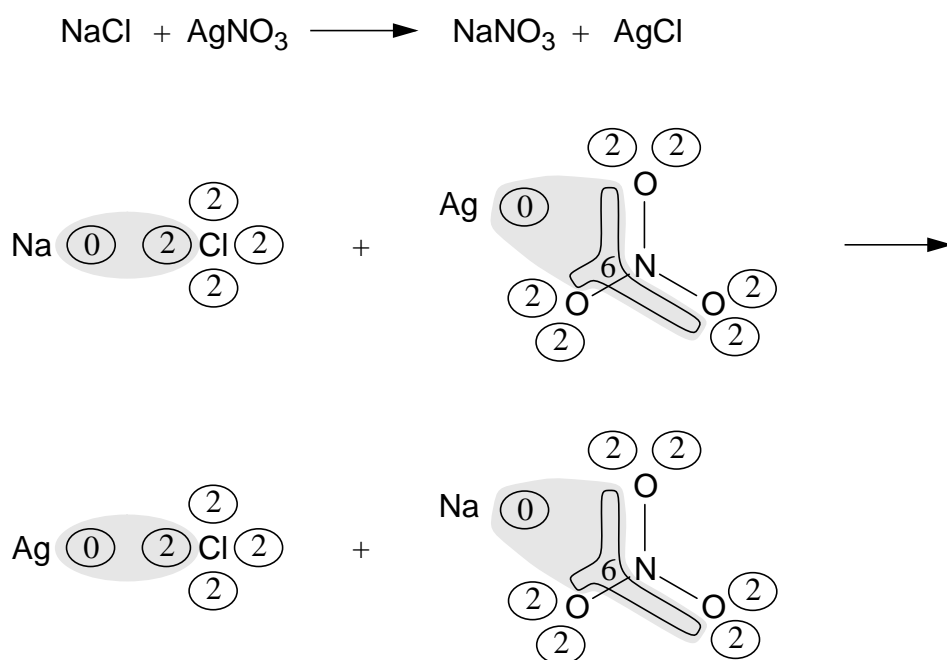


Abb. 6-14: Reaktion zwischen Natriumchlorid und Silbernitrat durch Umordnen der Wechselwirkungsgruppen.

6.3. Einfügen und Löschen von Molekülen und Aggregaten im Reaktionsensemble

Bei der Neuentwicklung des Programmsystems EROS wurden die Möglichkeiten zur formalen Beschreibung von Reaktionszentren stark erweitert. Hier wird erlaubt, daß Reagenzien, die benötigt werden, um eine Reaktion mit einem bestimmten Mechanismus zu formulieren, Teil der formalen Beschreibung in der Reaktionsregel sein können und erst während der Reaktion zu dem Eduktensemble addiert werden. Dies können Katalysatoren sein, oder aber auch kleine Moleküle, die als Reagenzien jedem Reaktionsgemisch zugeführt werden müssen, damit eine Reaktion ablaufen kann. Die Hydrolyse von Amidien kann sowohl über einen säure- als auch über einen basenkatalysierten Mechanismus modelliert werden [64]. Bei der Säurekatalyse müssen zum Eduktensemble während der Reaktion Protonen als Katalysator und Wasser als immer notwendiges Reagens addiert werden, bei der Basenkatalyse entsprechend Hydroxidionen und Wasser.

Um nun solche Verbindungen während einer Reaktion in ein Reaktionsensemble einzuschleusen, muß sich die Zusammensetzung eines Reaktionsensembles während einer Reaktion ändern dürfen. Dies wird ermöglicht, indem neue Aggregate oder Moleküle in einem Ensemble erzeugt werden. Um den Satz an potentiellen Reagenzien nicht einzuschränken, wurden zu diesem Zweck keine Moleküle vordefiniert. In die neuen Moleküle können Atome eingefügt werden, die π -Elektronensysteme mit allen Valenzelektronen und unbesetzten Orbitalen enthalten, die zur ihrer Beschreibung mit dem RICOS-Modell notwendig sind (siehe Abschnitt 4.1). Aus diesen Atomen werden dann mit Hilfe der in Abschnitt 6.1 beschriebenen Reaktionsschritte die benötigten Moleküle zusammengesetzt.

So werden das Wassermolekül und das Proton für die sauer katalysierte Amidhydrolyse am Beginn der Reaktion aus einem Sauerstoffatom und drei Wasserstoffatomen gebildet und zu dem gerade behandelten Amid addiert (Abbildung 6-15a). Für die basenkatalysierte Reaktion werden zwei Sauerstoffatome und drei Wasserstoffatome benötigt (Abbildung 6-15b), die zu Wasser und einem Hydroxidion kombiniert werden. Danach findet dann jeweils die eigentliche Hydrolysereaktion statt.

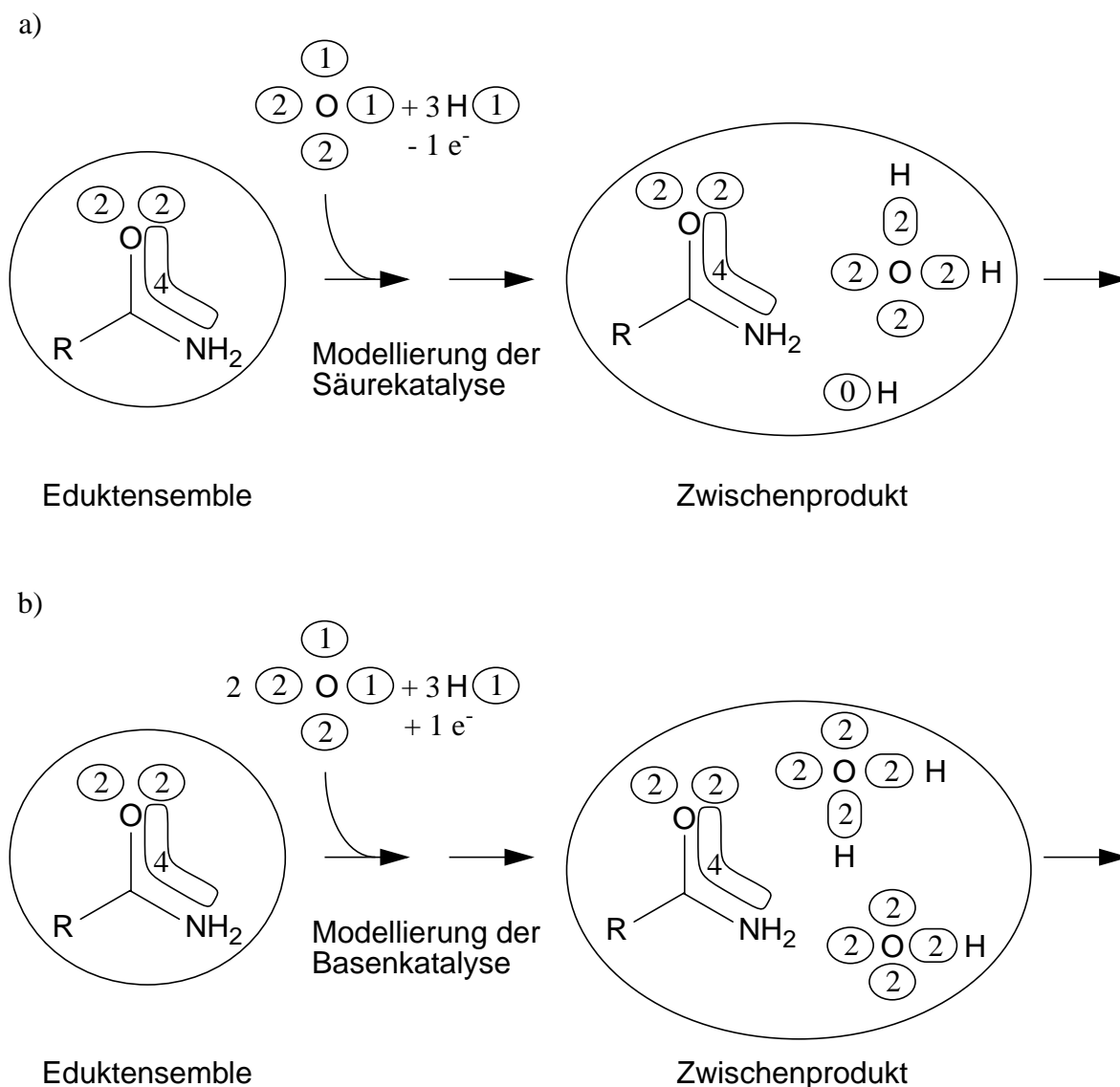


Abb. 6-15: Einschleusen von kleinen Molekülen in ein Eduktensemble, (a) sauer katalysierte Hydrolyse, (b) basenkatalysierte Hydrolyse.

Einzelne Aggregate oder Moleküle müssen auch aus dem Reaktionsensemble löschar sein. Denn es gibt Reaktionstypen, in denen in mehreren aufeinanderfolgenden Reaktionen jeweils nur bestimmte Produkte weiter reagieren können, andere dagegen aus dem Reaktionsensemble entfernt werden sollen. Im Massenspektrometer können nur geladene Moleküle gemessen werden, es können daher auch nur die Zerfälle geladener Moleküle modelliert werden. Deshalb werden entstandene Neutralfragmente als inert betrachtet und dürfen nicht mehr weiter reagieren. Sie müssen aus dem Reaktionsensemble am Ende einer Reaktion entfernt werden.

6.4. Erkennung von Reaktionsprodukten mit ungültigen Valenzen

Während der Modellierung einer Reaktion kann es vorkommen, daß Reaktionsprodukte mit ungültigen Valenzen gebildet werden. Die Anwendung der in den vorigen Kapiteln beschriebenen Elementarschritte kann zu folgenden Fehlern in den Reaktionsprodukten führen. Durch Oxidations- oder Reduktionsschritte kann die Zahl der Elektronen in einem Elektronensystem zu niedrig oder zu hoch werden. Zwischen zwei Atomen befinden sich Elektronensysteme, die nicht zueinander passen. Ein Atom trägt eine zu hohe oder zu niedrige Ladung.

Am Ende einer Sequenz von Elementarschritten können ungültige Produkte mit Hilfe eines sogenannten Valenztests gefunden und verworfen werden. Während des Valenztests werden die entstandenen Produkte anhand bestimmter Kriterien überprüft. Die Strenge dieser Kriterien hängt von der Art der verwendeten Anwendung ab. Für die Modellierung organischer Reaktionen werden die strengsten Kriterien angewandt, in anorganischen und massenspektroskopischen Prozessen sind sie weiter gespannt. Es gibt vier verschiedene Kategorien, die zwischen organischen, anorganischen und massenspektrometrischen Anwendungen unterscheiden. Da die physikochemischen Eigenschaften meist nur für organische Verbindungen parametrisiert sind, die außer den Elementen der ersten beiden Perioden nur Phosphoratome in der Oxidationsstufe III, Schwefelatome in der Oxidationsstufe II, und Chloratome in der Oxidationsstufe I enthalten, wird in der Kategorie für organische Verbindungen nur dieser Satz von Atomen erlaubt. Tabelle 6-1 zeigt die Zahl der Elektronen, die ein Elektronensystem enthalten darf. Sie sind für jeden Elektronensystemtyp und jede Kategorie aufgeführt. Zwischen der organischen und der anorganischen Kategorie besteht hier kein Unterschied, in massenspektroskopischen Prozessen sind jedoch auch radikalische σ -Elektronensysteme erlaubt.

Tab. 6-1: Zahl der Elektronen, die in einem Elektronensystem lokalisiert sein müssen mit n_A gleich Anzahl der Atome des Elektronensystems. Dabei stehen O für organische, A für anorganische Verbindungen, O-MS für organische und A-MS für anorganische Verbindungen in der Massespektrometrie.

| Art des Elektronensystems | Kategorien | | | |
|---|--------------------------------|--------------------------------|---------------------|---------------------|
| | O | A | O-MS | A-MS |
| reguläres σ -Elektronensystem | 2 | 2 | 1 - 2 | 1 - 2 |
| Mehrzentren- σ -Elektronensystem | gerade, 2 bis $2 \cdot n_A$ | gerade, 2 bis $2 \cdot n_A$ | 1 bis $2 \cdot n_A$ | 1 bis $2 \cdot n_A$ |
| π -Elektronensystem | 0 bis $2 \cdot n_A$ | 0 bis $2 \cdot n_A$ | 0 bis $2 \cdot n_A$ | 0 bis $2 \cdot n_A$ |
| koordinative Bindung | - | gerade, 2 bis $2 \cdot n_A$ | - | 1 bis $2 \cdot n_A$ |

Das zweite Kriterium, auf das die Elektronensysteme überprüft werden, ist die Kombination von Elektronensystemen, die je zwei Atome miteinander verbinden. Zwei Atome dürfen nur durch ein reguläres σ -Elektronensystem gebunden sein. Alle weiteren Elektronensysteme müssen π -Elektronensysteme sein. In Bindungen zwischen Nichtmetallatomen ist die maximale Bindungsmultiplizität drei, das heißt, zwei Atome dürfen höchstens an zwei gleichen π -Elektronensystemen beteiligt sein. Es können jedoch zwei Mehrzentrenelektronensysteme vorhanden sein, wie dies zum Beispiel zwischen beiden Boratomen in Diboran der Fall ist. Koordinative Bindungen zwischen zwei Atomen dürfen nicht mit anderen Elektronensystemen gemischt werden.

Das dritte Kriterium, das überprüft wird, betrifft die äquilibrierte Ladung an Atomen. In Abbildung 6-16 sind drei Verbindungen mit ähnlichen Elektronensystemen dargestellt. Die beiden Verbindungen aus Abbildung 6-16a tragen eine einfache positive Ladung, die Verbindung von Abbildung 6-16b eine zweifache positive Ladung. Mehrfach positiv geladene Spezies nehmen selten an organischen Reaktionen teil, sie sollten daher als Reaktionsprodukte verworfen werden. Eine Unterscheidung dieser Strukturen anhand der Elektronenverteilung in den Elektronensystemen gestaltet sich schwierig. Es ist aber möglich, sie aufgrund ihrer äquilibrierten Ladungen zu unterscheiden.

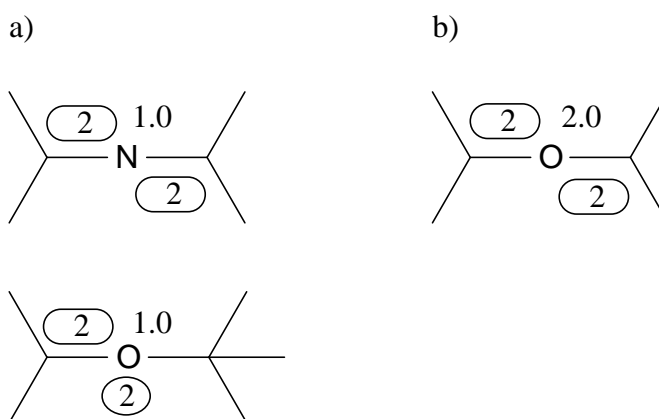


Abb. 6-16: Erlaubte (a) und nicht erlaubte (b) Kationen. Neben den Heteroatomen ist jeweils die äquilibrierte Ladung angegeben, alle übrigen Atome sind formal ungeladen.

Zur Bestimmung, ob ein Molekül erlaubt ist, werden die äquilibrierten Ladungen aller Atome des Moleküls betrachtet. Die verwendete äquilibrierte Ladung wird anhand der in Kapitel 5.2.1 vorgestellten Methode berechnet. Wenn alle Ladungen innerhalb eines vorgegebenen Bereichs liegen, ist diese Verbindung korrekt, wenn nicht, muß sie verworfen werden. Dabei ist der Bereich für organische Verbindungen enger gesteckt als für anorganische. Anhang B gibt die Unter- und Obergrenzen für alle Elemente des Periodensystems und alle Kategorien an. Allgemein gilt, daß sich die äquilibrierte Ladung der Elemente der zweiten Periode zwischen -1 und +1 bewegen muß. Die höchste erlaubte Ladung eines Metallatoms entspricht der höchsten möglichen Oxidationsstufe, die für dieses Element beobachtet wird.

7. Konvertierung zwischen RICOS und dem VB-Modell

Das RICOS-Modell erlaubt es, chemische Strukturen auf einfache Weise zu beschreiben, die im VB-Modell nur unter Schwierigkeiten dargestellt werden können. Neuentwickelte Programme, z. B. zur Reaktionsvorhersage, können von dieser verbesserten Darstellungsform chemischer Strukturen profitieren. Zusätzlich müssen aber auch bestehende Programmsysteme, die auf einer herkömmlichen Strukturdarstellung beruhen, zusammen mit diesen Neuentwicklungen benutzt werden können. Desweiteren gibt es große Sammlungen von Strukturen, Reaktionen oder Spektren, die zugänglich bleiben müssen. Aus diesen Gründen wurden Algorithmen entwickelt, um zwischen den beiden Strukturrepräsentationen konvertieren zu können. Diese Verfahren können einerseits dazu benutzt werden, Dateiformate, die auf gebräuchlichen Bindungslisten beruhen, zu lesen bzw. zu schreiben. Andererseits ist es aber auch möglich, Programmcode, der noch nicht auf die neue Darstellungsform umgestellt wurde, direkt in Neuentwicklungen einzubinden. Dies war z. B. im Fall der Berechnung physikochemischer Eigenschaften nötig (siehe Kapitel 5.5).

7.1. Umwandlung von Molekülen mit VB- Repräsentation in RICOS

Die Konvertierung von Strukturen in das RICOS-Format unterliegt einigen Einschränkungen, die ihre Ursache in Unzulänglichkeiten der VB-Darstellung haben. Die Beschreibung von Komplexverbindungen oder Verbindungen mit Mehrzentrenbindungen mit Hilfe herkömmlicher Bindungslisten ist schwierig und es gibt keine allgemeinen Regeln, wie diese Strukturen in Bindungslisten repräsentiert werden sollen. Aus diesem Grund können Verbindungen mit Komplex- oder Mehrzentrenbindungen nicht in das RICOS-Modell konvertiert werden. Desweiteren lassen sich Bindungslisten in Bezug auf Konjugation innerhalb von orthogonalen π -Systemen nicht immer eindeutig interpretieren. Daher müssen Annahmen über die Natur der Konjugation in delokalisierten Verbindungen gemacht werden, die zu einer möglichst sinnvollen Repräsentation in RICOS führen und zudem unabhängig von der vorgegebenen Resonanzstruktur und der Numerierung der Atome die gleiche RICOS-Repräsentation ergeben. Folgende Vorgaben müssen bei der Ableitung der Konjugation beachtet werden.

- Mehrfachbindungen, freie Elektronen und positive Ladungen benachbarter Atome sind miteinander konjugiert.
- Es findet keine Konjugation zwischen freien Elektronenpaaren benachbarter Atome statt. Beispielsweise sind in Verbindungen wie Wasserstoffperoxid oder Hydrazin die Bindungen zwischen beiden Heteroatomen Einfachbindungen, da aus der Kombination der p-Orbitale der zwei Atome sowohl das resultierende bindende als auch das antibindende Orbital mit je zwei Elektronen besetzt wäre.
- Jedes Atom darf zu jedem π -Elektronensystem nur einmal beitragen, d. h. überlappende π -Systeme werden nicht miteinander kombiniert.
- Um einer zu großen Bindungsordnung zwischen zwei benachbarten Atomen vorzubeugen,

wird die maximale Anzahl miteinander konjugierbarer π -Elektronensysteme zwischen zwei Nichtmetallatomen auf zwei und zwischen zwei Metallatomen auf drei beschränkt.

- Mehrdeutigkeiten in orthogonalen π -Elektronensystemen werden so aufgelöst, daß jeweils das größtmögliche π -Elektronensystem entsteht. Dies entspricht dem experimentellen Befund, daß konjugierte Derivate des Acetylens wie z. B. Tolan eine Konformation mit planarem Kohlenstoffgerüst annehmen [50]. Diese Geometrie läßt sich mit einem großen π -System, das sich über das gesamte Molekül erstreckt, und einem kleineren dazu orthogonalen π -System, das nur die beiden Kohlenstoffatome des Acetylens umfaßt, erklären.

Die Konvertierung eines Moleküls von der VB-Repräsentation in die RICOS-Repräsentation erfolgt in zwei Stufen. Im ersten Schritt werden Einfachbindungen in reguläre σ -Elektronensysteme, Mehrfachbindungen in kombinierte σ - und π -Elektronensysteme und freie Elektronen in π -Elektronensysteme umgewandelt. Anschließend werden, um den valenzchemischen Bedingungen aus Tabelle 4-1 Genüge zu tun, jedem Atom so viele leere π -Elektronensysteme zugefügt, daß die vorgeschriebene Anzahl an Elektronensystemen eingehalten wird. Im zweiten Schritt wird die Suche nach Konjugation durchgeführt, und benachbarte π -Elektronensysteme zu längeren π -Elektronensystemen zusammengefaßt.

Ausgehend von einem beliebigen Startatom, in der Regel dem ersten Atom des Moleküls, wird eine Tiefensuche auf dem σ -Gerüst, das den Molekülgraphen bildet, durchgeführt. Für Vater- und Sohnknoten werden jeweils eine Liste von gültigen Elektronensystemen erzeugt. Aus den Einträgen der Liste werden Kombinationen der Größe nach gebildet und nach den Vorgaben in Tabelle 7-1 miteinander kombiniert, sofern sie sich nicht überlappen oder bereits Teil desselben π -Elektronensystems sind. Es werden alle Kombinationen getestet, bis entweder die erlaubte Anzahl von π -Elektronensystemen zwischen beiden Atomen erreicht wurde oder keine neue Kombination mehr möglich ist. Die Elektronensysteme werden der Größe nach sortiert, um zu gewährleisten, daß Kombinationen, die in möglichst großen π -Elektronensystemen resultieren, bevorzugt werden. Wenn ein Atom sowohl ein freies Elektronenpaar als auch ein radikalisches Elektron besitzt, wird die Konjugation bevorzugt zu dem radikalischen Elektron ausgebildet. Als nächstes wird dasjenige Elektronensystem mit den meisten Elektronen herangezogen, um zu gewährleisten, daß in Elementen der höheren Perioden die freien Elektronenpaare zuerst berücksichtigt werden.

Tab. 7-1: Regeln zur Kombination benachbarter Elektronensysteme bei der Ableitung konjugierter π -Elektronensysteme. n_V bezeichnet die Anzahl der π -Elektronensysteme des Vaterknoten, n_S die des Sohns.

| n_V | n_S | Aktion |
|-------|-------|--|
| 0 | 0 - 3 | keine Aktion erforderlich |
| 0 - 3 | 0 | |
| 1 | 1 | kombiniere beide Systeme |
| 1 | 2,3 | kombiniere das einzelne System mit dem größten der zur Auswahl stehenden Systeme |
| 2, 3 | 1 | |
| 2 | 2 | kombiniere jeweils die beiden größten und die beiden kleinsten Systeme |
| 2 | 3 | kombiniere jeweils die beiden größten sowie die beiden zweitgrößten Systeme |
| 3 | 2 | |
| 3 | 3 | |

Für Thiophen sind beide Stufen der Konvertierung in Abbildung 7-1a dargestellt. Zunächst werden die beiden Doppelbindungen im Ring in zwei π -Elektronensysteme aus jeweils zwei Atomen mit zwei Elektronen umgewandelt. Das sp^2 -hybridisierte Schwefelatom erhält zwei gefüllte und fünf leere π -Elektronensysteme, um die vollständige Zahl von neun Elektronensystemen zu erzeugen. Im zweiten Schritt findet die Tiefensuche statt. Abbildung 7-1b zeigt zwei Schritte dieser Suche im Molekülgraphen von Thiophen. Im linken Teil der Abbildung ist Atom 2 der Vaterknoten, Atom 3 der Sohnknoten. Da beide Atome jeweils ein π -Elektronensystem besitzen, werden die beiden Systeme nach Tabelle 7-1 kombiniert. Der rechte Teil der Abbildung zeigt ein späteres Stadium der Suche. Hier sind die Atome 4 und 5 der Vater- bzw. Sohnknoten. Der Vaterknoten besitzt wiederum ein, der Sohnknoten diesmal aber sieben π -Elektronensysteme. Daher wird das Elektronensystem des Vaterknotens mit dem größten Elektronensystem des Sohnknotens vereinigt. Im Fall des Schwefels ist das größte Elektronensystem eines der beiden freien Elektronenpaare.

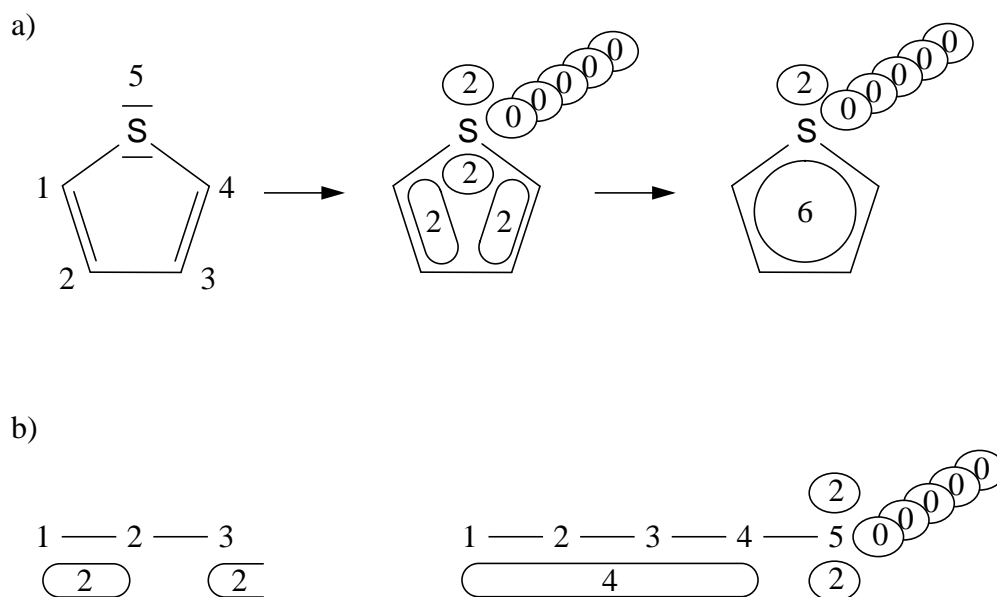


Abb. 7-1: Konvertierung der VB-Repräsentation von Thiophen nach RICOS. a) Beide Stufen der Konvertierung, b) 2 Schritte der Tiefensuche im Molekülgraphen von Thiophen.

Im letzten Schritt der Konvertierung werden Atom-, Molekül- oder Ensembleigenschaften, die unabhängig von der Repräsentation der Elektronen in Bindungen sind, von der VB-Kodierung auf die RICOS-Kodierung übertragen. Dies ist insbesondere für stereochemische Deskriptoren wichtig, die in einer späteren Erweiterung der von RICOS als Atomeigenschaft behandelt werden sollen und daher bei einer Konvertierung nicht verlorengehen dürfen. Aber auch andere Molekül- und Ensembleigenschaften wie etwa Namen oder Registernummern aus Datenbanken können so mitgeführt werden.

Abbildung 7-2 zeigt das Ergebnis der Konvertierung zweier Verbindungen mit orthogonalen π -Elektronensystemen. Für Tolan (Diphenylethin) wird, wie schon angesprochen, ein großes π -Elektronensystem, das sich über das gesamte Molekül erstreckt gebildet und eines, das nur die beiden Atome der Alkingruppe enthält. Das zweite Beispiel enthält zwei benachbarte freie Elektronenpaare, die nicht konjugiert werden.

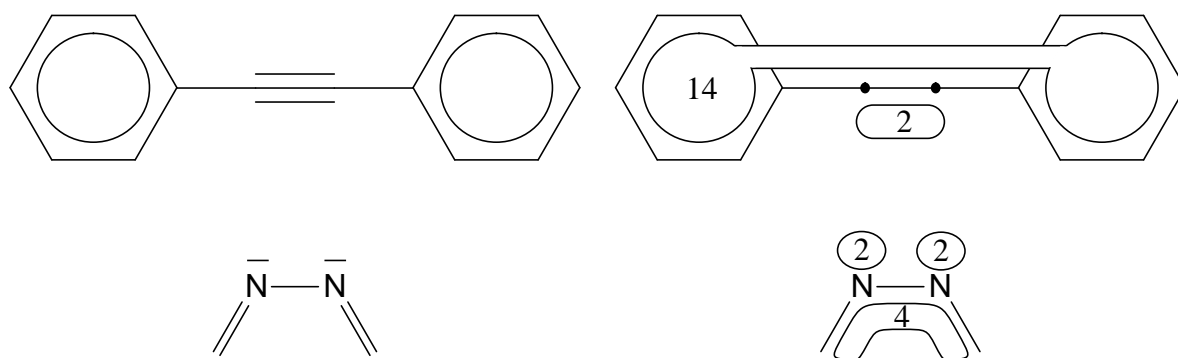


Abb. 7-2: Ergebnisse der automatischen Konvertierung zwischen dem VB-Modell und RICOS für Verbindungen mit orthogonalen π -Elektronensystemen.

7.2. Umwandlung von Molekülen in RICOS-Repräsentation in VB-Repräsentation

In der chemischen Literatur hat man sich mit der Bestimmung der Zahl der möglichen Kekulé-Strukturen in konjugierten Molekülen befaßt [65], [66], [67], [68]. Es wurden ebenfalls Algorithmen zur Bestimmung einer kanonischen Kekulé-Struktur für konjugierte Verbindungen entwickelt [66], [69], [70]. Diese Algorithmen sind aber für die Aufgabe, Elektronensysteme der RICOS-Repräsentation in Ladungen, freie Elektronen und Mehrfachbindungen zu übertragen, nicht geeignet. Denn sie beruhen auf der Annahme, daß die Zahl der Elektronen des zu lokalisierenden π -Systems der Zahl seiner Atome entspricht, daß also jedes Atom an mindestens einer Mehrfachbindung beteiligt ist. In RICOS sind jedoch auch freie Elektronenpaare und geladene Atome, die keine Valenzelektronen in das π -Elektronensystem beisteuern, Teil der delokalisierten π -Elektronensysteme. Es mußte daher ein neuer Algorithmus zur Verteilung der Elektronen in lokalisierte Bindungen entwickelt werden.

Auch hier mußten Einschränkungen in Bezug auf die Art der konvertierbaren Verbindungen gemacht werden. Es werden wiederum nur solche Verbindungen konvertiert, die reguläre σ - und π -Elektronensysteme enthalten, nicht jedoch Verbindungen mit Mehrzentrenbindungen oder koordinativen Bindungen. Da die interne Datenstruktur zur Kodierung des VB-Modells auch das Konzept der Aggregate und Wechselwirkungsgruppen enthält, ist eine Übertragung von Aggregaten, deren Wechselwirkungsgruppen aus Atomen bestehen, möglich.

7.2.1. Allgemeine Vorgehensweise

Alle Moleküle eines Ensembles oder Aggregats werden zur Konvertierung einem mehrstufigen Prozeß unterworfen. Zuerst wird eine Struktur erzeugt, in der jedes σ -Elektronensystem durch eine Einfachbindung ersetzt ist. Dabei wird das zu konvertierende Molekül auf Elektronensysteme geprüft, die von der Konvertierung ausgeschlossen sind. Dies sind koordinative Bindungen sowie σ -Elektronensysteme, die entweder nicht regulär sind, oder eine Elektronenzahl

ungleich zwei haben. Enthält ein Molekül ein solches Elektronensystem, wird das Verfahren an dieser Stelle abgebrochen.

Daraufhin werden sukzessive alle π -Elektronensysteme des Moleküls behandelt, und die in ihnen enthaltenen Elektronen lokalisiert. Dabei wird jeweils eine Resonanzstruktur gebildet, die die maximal mögliche Zahl an Doppelbindungen enthält, Ladungstrennung wird nach Möglichkeit vermieden. Das Ablaufschema für diesen Prozeß zeigt Abbildung 7-3. Zunächst wird im aktuellen Elektronensystem die Zahl der Atome mit der Zahl der Elektronen verglichen. Ist die Zahl der Elektronen höher als die Zahl der Atome, so enthält das π -Elektronensystem freie Elektronenpaare, ist sie niedriger, so enthält es ein Atom, dem kein Valenzelektron zugeordnet werden darf. Dies sind in der Regel positiv geladene Atome, ungeladen sind nur Atome der dritten Hauptgruppe. Die Differenz zwischen Atom- und Elektronenzahl entspricht der Zahl der Atome, auf denen freie Elektronen oder positive Ladungen lokalisiert werden müssen. Ist sie größer als Null, findet eine Suche nach den günstigsten Atomen statt, die entweder die freien Elektronen oder die positive Ladung tragen können mit der Nebenbedingung, daß am Ende der Konvertierung möglichst wenige Atome eine Formalladung tragen. Im nächsten Schritt werden die Doppelbindungen zwischen den übrigen Atomen verteilt. In einigen wenigen Fällen, in denen dies nicht möglich ist, wird eine Ladungstrennung durchgeführt, das heißt, die Zahl der Atome, die ein freies Elektronenpaar und eine positive Ladung tragen sollen, wird jeweils um eins erhöht. Danach wird der Prozeß beginnend mit der Suche nach geeigneten Atomen zur Lokalisierung der freien Elektronen und Ladungen wiederholt. Die gefundenen Lokalisationsorte werden während der Suche zwischengespeichert.

Wenn in diesem Prozeß alle Elektronen lokalisiert werden konnten, werden die Elektronen und Mehrfachbindungen auf das Bindungsgerüst aus Einfachbindungen übertragen. In einem letzten Schritt werden Atom-, Molekül-, Aggregat- und Ensembleeigenschaften kopiert.

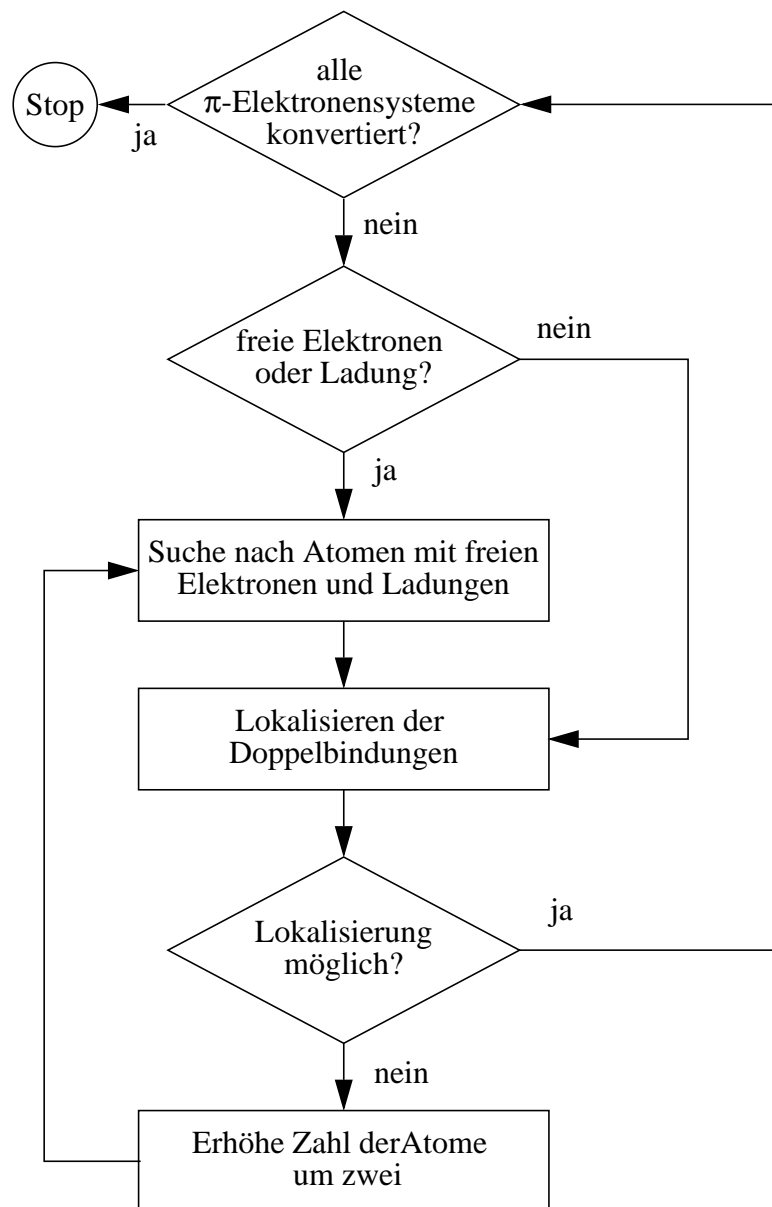


Abb. 7-3: Ablaufschema der Lokalisierung der Elektronen aller π -Elektronensysteme eines Moleküls.

7.2.2. Suche nach geeigneten Atomen zur Lokalisierung von Ladungen und freien Elektronen

Zu Beginn der Lokalisierung wird in jedem π -Elektronensystem geprüft, ob die Differenz zwischen der Zahl der Atome und der Zahl der Elektronen des Elektronensystems eine Verteilung freier Elektronen oder Ladungen erfordert. Ist die Ladung positiv, so entspricht sie der Zahl der zu verteilenden positiven Ladungen, ist sie negativ, der Zahl der Atome mit freien Elektronen. Ist die Zahl der Elektronen im System ungerade, so liegt ein Radikal vor. In Verbindungen, die in einem π -Elektronensystem sowohl Elemente der dritten Hauptgruppe als auch Elemente der

fünften bis siebten Hauptgruppe enthalten, ist es sinnvoll, nicht die maximal möglichen Doppelbindungen zu erzeugen. Dies würde in einer VB-Struktur mit zwei geladenen Atomen resultieren (Abbildung 7-4a). Um formal ungeladene Atome zu erhalten, werden für jedes solche Atompaar je ein Atom mit freien Elektronen und ein Atom ohne Valenzelektronen aus diesem π -Elektronensystem mehr bestimmt. Daraus resultieren VB-Strukturen wie in Abbildung 7-4b.

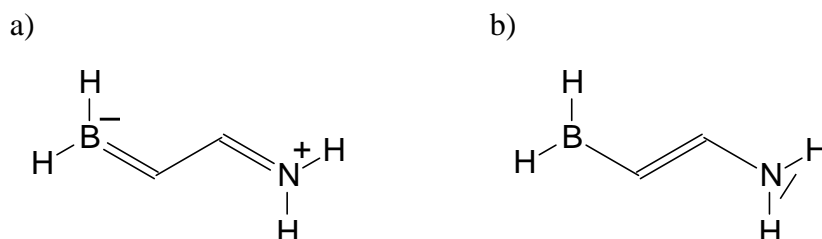


Abb. 7-4: Die Zuordnung zusätzlicher freier Elektronenpaare und Atome ohne Valenzelektronen in π -Elektronensystemen mit Atomen sowohl der dritten als auch der fünften bis siebten Hauptgruppe vermeidet Ladungstrennung.

Die Suche nach Atomen, die freie Elektronen oder positive Ladungen tragen sollen, erfolgt iterativ in zwei Stufen. Für jedes Elektronenpaar, Radikal oder positive Ladung werden zunächst alle möglichen Atome gesucht, die für die Platzierung in Frage kommen und danach das günstigste von ihnen ausgewählt.

Die Lage dieser Atome muß so ausgewählt werden, daß bei der nachfolgenden Verteilung der Doppelbindungen keine Biradikale entstehen. Anhand des Bindungsgerüsts, das dem Elektronensystem zu Grunde liegt, wird eine Tabelle angelegt, die für jedes Atom die Abstände zu allen nächst liegenden Endatomen bzw. Verzweigungen enthält, die von diesem Atom erreichbar sind, ohne einen Verzweigungspunkt zu überschreiten. Atome, die freie Elektronen oder positive Ladungen tragen können, werden in der folgenden Diskussion Zentrum genannt. Ein Atom ist dann ein mögliches Zentrum, wenn es einer der folgenden Regeln für die Vergabe dieser Punkte genügt:

1. Ein Atom ist ein Zentrum, wenn die Abstände zu allen Endatomen jeweils gerade und die Abstände zu allen Verzweigungen ungerade sind.
2. Verzweigungspunkte, die Regel 1 nicht genügen, sind auch dann Zentren, wenn sie mindestens zwei Endatome mit einem geraden Abstand besitzen.
3. Ausgehend von Verzweigungen, die nach Regel 1 oder 2 als Zentrum gefunden wurden, werden Atome auch dann berücksichtigt, wenn sie in geradem Abstand zu diesen Atomen liegen.
4. Atome, die mindestens noch zwei freie Elektronen tragen müssen, um formal neutral zu sein, und nicht in ungeradem Abstand zu einem Endatom liegen, werden ebenfalls als Zen-

tren markiert.

5. In isolierten Ringen werden alle Atome als Zentren markiert.

Die vier Verbindungen aus Abbildung 7-5 veranschaulichen diese Regeln. Atome, die nach Regel 1 gefunden werden, sind mit einem Punkt markiert, nach Regel 2 mit einem Quadrat, nach Regel 3 mit einem Dreieck und nach Regel 4 mit einer Raute. Die Zahlen neben jedem Atom geben den Abstand zur nächsten Verzweigung oder zum nächsten Endatom an. In **1** werden zwei Zentren gefunden, die beide Endatome sind und einen ungeraden Abstand zum Verzweigungspunkt besitzen. Das dritte Endatom ist zwei Atome vom Verzweigungspunkt entfernt und wird daher nicht als mögliches Zentrum erkannt. Der Verzweigungspunkt selbst hat einen ungeraden Abstand zu den beiden markierten Atomen und kommt daher ebenfalls nicht für ein Zentrum in Frage. Bei der Suche nach Zentren in **2** kommen Regel 1 und 3 zum Tragen. Ausgehend vom nach Regel 1 markierten Verzweigungspunkt werden auch die beiden Atome, die mit einem Quadrat markiert sind, gefunden. Die Zentren in Verbindung **3** werden durch Anwendung der Regel 2 und 3 gefunden. Zu beiden Verzweigungspunkten gibt es zwei Endpunkte mit geradem Abstand (Regel 3). Das zentrale Atom erfüllt zwar die Regel 1, wird aber aufgrund der Priorität von Regel 2 wieder aus der Liste der in Frage kommenden Atome entfernt, da in Ketten die Distanzen zwischen den beteiligten Atomen gerade sein müssen. Die Endpunkte werden wieder nach Regel 3 gefunden.

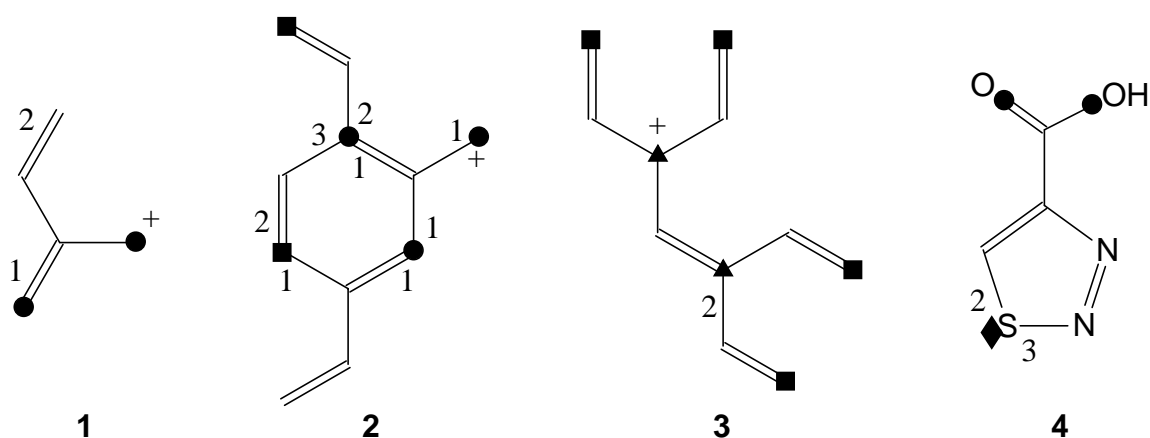


Abb. 7-5: Anwendung der Regeln zur Suche von Ladungszentren. Atome, die nach Regel 1 gefunden werden, sind mit einem Punkt markiert, nach Regel 2 mit einem Dreieck, nach Regel 3 mit einem Quadrat, nach Regel 4 mit einer Raute. Die Zahlen zeigen die Abstände der markierten Atome zum nächsten End- oder Verzweigungspunkt.

In Verbindung **4** führt die Anwendung der Regeln 1 und 4 zu drei möglichen Zentren. Die beiden Sauerstoffatome der Carboxylgruppe sind Endpunkte wie diejenigen von Verbindung **1**. Das Schwefelatom hat sowohl einen geraden als auch einen ungeraden Abstand zur Verzwei-

gung. Es erfüllt daher Regel 1 nicht. Da es aber mindestens ein freies Elektronenpaar benötigt, um formal ungeladen zu sein, wird es nach Regel 4 markiert.

Die Zahl der noch benötigten freien Elektronen n_{el} an einem Atom ergibt sich wie folgt:

$$n_{el} = Z - c - s - f \quad (7-1)$$

mit Z Ordnungszahl, c Zahl der Kernelektronen, s Zahl der σ -Elektronensysteme, f Zahl der freien Elektronen, die entweder schon gefunden wurden, oder durch einatomige π -Elektronensysteme vorgegeben sind. Für das betrachte Schwefelatom in Verbindung **4** ergibt sich:

$$n_{el} = 16 - 10 - 2 - 2 = 2 \quad (7-2)$$

Nachdem diese Punkte gefunden wurden, wird das günstigste Atom, das als Zentrum fungieren kann, gesucht. Dabei muß zwischen der Lokalisierung freie Elektronen und positiver Ladungen unterschieden werden, da freie Elektronen bevorzugt an Heteroatomen der höheren Hauptgruppen, positive Ladungen an Kohlenstoffatomen oder anderen Atomen der niedrigeren Hauptgruppen lokalisiert werden sollen. Es werden jeweils verschiedene Faktoren berücksichtigt, die gegeneinander gewichtet werden.

Bei der Platzierung von freien Elektronen sind folgende Faktoren ausschlaggebend: Wenn Atome mindestens zwei freie Elektronen benötigen, um formal ungeladen zu sein, haben sie höchste die Priorität, wobei Atome mit nur einem π -Elektronensystem favorisiert werden. Auch Atome höherer Hauptgruppen werden stärker gewichtet. Weniger starken Einfluß auf die Auswahl haben folgende Faktoren, die sich hauptsächlich in reinen Kohlenwasserstoffen auswirken. Zwei Quotienten, die Aussagen über die Umgebung des Atoms machen, sind das Verhältnis aus der Zahl ungeradzahlicher Verzweigungen zu allen Verzweigungen sowie aus der Zahl geradzahlicher Endatome zu allen Endatomen. Atome mit einem höheren Anteil an ungeradzahlichen Verzweigungspunkten sowie mit einem höheren Anteil an geradzahlichen Endatomen werden bevorzugt. Der letzte Faktor, der bei der Platzierung berücksichtigt wird, ist die sogenannte erweiterte Konnektivität (engl. *extended connectivity*), die auf dem Graphen der noch nicht zugeordneten Atomen des Elektronensystems berechnet wird. Mit ihr werden Atome eines Moleküls in Äquivalenzklassen eingeteilt, äquivalente Atome erhalten den gleichen Wert. Außerdem ist sie auch ein Maß dafür, wie weit im Zentrum eines Moleküls ein Atom liegt. Je zentraler die Lage des Atoms ist, desto höher ist der Wert der erweiterten Konnektivität. Sie wurde schon von Morgan [29] in seinem Algorithmus zur kanonischen Numerierung von Atomen in Molekülen verwendet. Atome am Rand des π -Elektronensystems, also mit einem niedrigen Wert für die erweiterte Konnektivität, werden bevorzugt. Zudem wird durch ihre Verwendung garantiert, daß nichtäquivalente Atome, die sich durch die bisher aufgeführten Faktoren nicht unterscheiden, unterschiedlich bewertet werden. Somit ist Numerierungsunabhängigkeit gewährleistet.

Bei der Plazierung von positiven Ladungen werden folgende Kriterien berücksichtigt: Positive Ladungen werden hauptsächlich an Kohlenstoffatomen in π -Elektronensystemen ohne Heteroatome lokalisiert, in denen es keine Atome gibt, die aufgrund ihrer Bindungsverhältnisse bevorzugt bei der Zentrenvergabe berücksichtigt werden müssen. Daher wurden die Kriterien so gewählt, daß die Atome in der Mitte des π -Elektronensystems höhere Prioritäten erhalten, da dies die spätere Verteilung der Doppelbindungen auf die restlichen Atome des π -Elektronensystems vereinfacht. Um das günstigste Atom zu finden, werden die beiden Quotienten, die das Verhältnis von End- und Verzweigungspunkten ausdrücken, und die erweiterte Konnektivität gegeneinander gewichtet. Diesmal werden jedoch Atome mit höherer Werten für die erweiterte Konnektivität bevorzugt.

Die Kohlenstoffatome der Kationen in Abbildung 7-6 werden zu Ladungszentren gewählt, da sie jeweils sowohl die höchste erweiterte Konnektivität besitzen als auch den höchsten Anteil an Verzweigungspunkten in ungeradem Abstand. Nach dieser Zuordnung besteht das π -Elektronensystem in Abbildung 7-6a) nur noch aus einem isolierten Ring, das π -Elektronensystem in Abbildung 7-6b) aus einer unverzweigten Kette. In beiden ist die anschließende Zuordnung von Doppel- und Einfachbindungen am einfachsten.

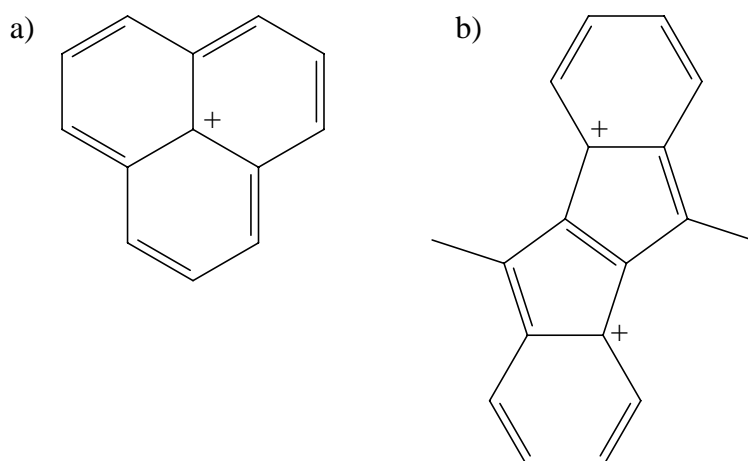


Abb. 7-6: Zuordnung der positiven Ladung zum Atom mit der jeweils höchsten erweiterten Konnektivität.

7.2.3. Suche nach Doppelbindungen

Durch die Bestimmung der Atome, die freien Elektronen bzw. positiven Ladungen aufnehmen, kann ein π -Elektronensystem in mehrere unzusammenhängende Abschnitte gespalten werden. Jeder dieser Abschnitte wird nun der Prozedur zur Verteilung von Doppelbindungen unterworfen. Sie ist in mehrere Schritte unterteilt.

Zunächst wird für jedes noch unbehandelte Atom eines π -Elektronensystems bestimmt, wieviele Nachbaratome es in diesem π -Elektronensystem hat. Atome mit nur einem Nachbaratom

werden im folgenden Endatome genannt, Atome mit drei Nachbaratomen Verzweigungspunkte.

Zunächst werden alle unverzweigten Ketten behandelt, da in ihnen die Verteilung der Doppelbindungen am einfachsten ist. Ausgehend von einem Endatom erhalten alle Atompaaire der Kette alternierende Doppel- und Einfachbindungen.

Im nächsten Schritt werden alle verzweigten Ketten und Ringsysteme mit exocyclischen Doppelbindungen behandelt. Die Vorgehensweise wird anhand der Abbildungen 7-7 und 7-8 veranschaulicht. An Verzweigungspunkten ist die Lage der Doppelbindung zwischen dem Verzweigungspunkt und einem seiner Nachbaratome nicht sofort ersichtlich. Daher wird die Bestimmung der Lage der Doppelbindungen von den Endatomen des π -Elektronensystems gestartet. Zu diesem Zweck wird zunächst jedem Atom ein Zähler mit dem Startwert 0 zugeordnet. Nun wird die Differenz zwischen der Zahl der Nachbaratome und dem Wert des Zählers gebildet. Wenn sie gleich eins ist, eignet sich das entsprechende Atom als Startatom zur Vergabe von Doppelbindungen. Im ersten Zyklus ist diese Bedingung für alle Endatome erfüllt. Von jedem Startatom ausgehend wird eine rekursive Tiefensuche auf dem Graphen des π -Elektronensystems nach alternierenden Doppel- und Einfachbindungen durchgeführt. Das Startatom hat nur einen Nachbarn, zu dem die erste Doppelbindung festgelegt wird. Beide Atome werden als behandelt markiert, indem der Zähler der beiden Atome, die an der Doppelbindung teilhaben, gleich der Zahl der Nachbaratome gesetzt wird. Daraufhin werden alle noch nicht behandelten Nachbarn des zweiten Atoms zu neuen Startatomen für die nächsten Doppelbindungen und dies solange wiederholt bis eines der folgenden Abbruchkriterien greift. Wenn ein Verzweigungspunkt als Startpunkt einer Doppelbindung erreicht wird, wird die Suche an dieser Stelle abgebrochen, und der Zähler für dieses Atom inkrementiert. Trifft die Suche auf ein Atom, für das die Differenz zwischen Zahl der Nachbaratome und dem Wert des Zählers gleich Null ist, so wurde das entsprechende Atom schon behandelt und der Prozeß wird ebenfalls abgebrochen.

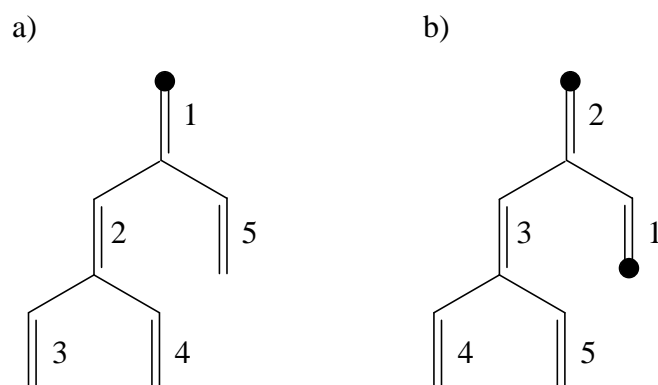


Abb. 7-7: Verteilung von Doppelbindungen in verzweigten π -Elektronensystemen in einem Zyklus. In a) und b) wurde eine unterschiedliche Reihenfolge der mit einem Punkt markierten Startatome verwendet. Neben den Doppelbindungen, ist die Reihenfolge, in der sie gefunden wurden, angegeben.

Im π -Elektronensystem aus Abbildung 7-7 reicht ein Zyklus, in dem alle Endatome als Startatome geprüft werden, aus, um die Lage aller Doppelbindungen zu bestimmen. Wenn das oberste Atom, das in a) mit einem Punkt markiert ist, als erstes Startatom gewählt wird, werden alle Doppelbindungen von diesem Startatom ausgehend in der neben den Doppelbindungen angegebenen Reihenfolge gefunden. Wenn ein anderes Atom als erstes Startatom gewählt wird, z. B. das in b) rechts markierte Atom, so kann von diesem Atom ausgehend nur die erste Doppelbindung festgelegt werden, denn dann trifft die Suche auf eine Verzweigung als Startpunkt und sie wird zunächst abgebrochen. Mit dem obersten Atom als nächstem Startatom können jedoch alle anderen Doppelbindungen festgelegt werden.

Für das verzweigte π -Elektronensystem aus 12 Atomen in Abbildung 7-8 reicht jedoch ein Zyklus nicht aus, um alle Doppelbindungen zu bestimmen. Im ersten Zyklus werden nacheinander die mit einem Punkt markierten Atome als Startatome verwendet und dabei die vier terminalen Doppelbindungen gefunden. Im nächsten Zyklus eignen sich diejenigen Verzweigungen als Startatome, deren Zähler im ersten Zyklus zweimal inkrementiert wurde. Dort ist nun die Differenz zwischen Nachbaratomen und Zähler gleich eins (Abbildung 7-8b). Zwei der Nachbaratome wurden schon behandelt, die Doppelbindung wird zum noch unbehandelten Atom ausgebildet, und dann wird die Tiefensuche wieder so weit durchgeführt, bis alle Abbruchkriterien erfüllt sind. Es werden so viele Zyklen durchlaufen, bis kein neues Startatom mehr zur Verfügung steht. Das ist entweder der Fall, wenn alle Atome des π -Elektronensystems abgearbeitet wurden (Abbildung 7-8c), oder wenn das π -Elektronensystem Ringsysteme ohne exocyclische Doppelbindungen enthält.

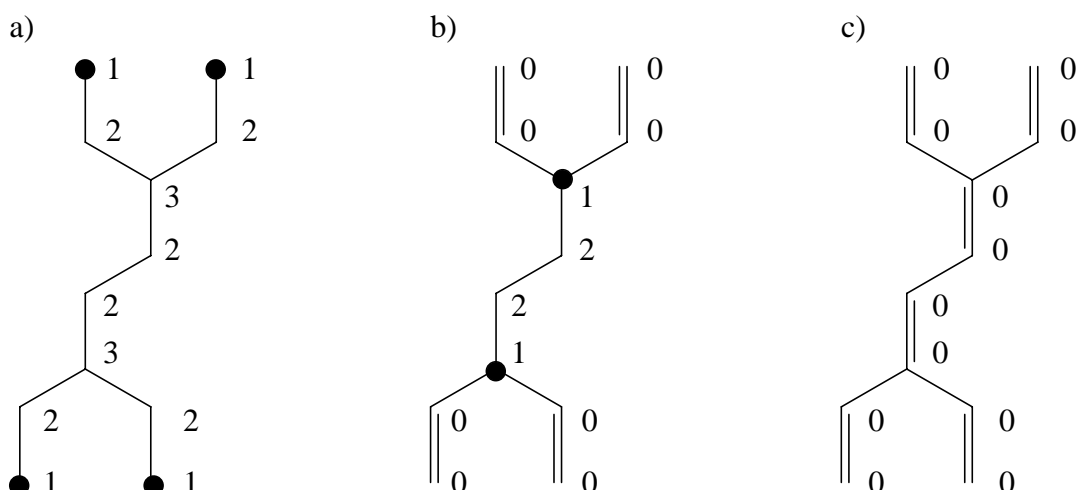


Abb. 7-8: Verteilung von Doppelbindungen in einem verzweigten π -Elektronensystem in mehreren Zyklen. Die Startatome der beiden Zyklen sind jeweils mit einem Punkt markiert. Neben jedem Atom ist jeweils die Differenz zwischen der Zahl der Nachbarn und dem Markierungszähler angegeben (Erklärung siehe Text).

Ringe mit exocyclischen Doppelbindungen werden ebenfalls mit dieser Methode behandelt. In Fulven wird das Methylenkohlenstoffatom als Startatom verwendet. Nachdem die erste Doppelbindung zum Verzweigungspunkt hin ausgebildet wird, kann die rekursive Suche für alternierende Einfach- und Doppelbindungen die beiden Doppelbindungen im Ring finden. Es ist in diesem Fall nur ein Zyklus notwendig, um alle Doppelbindungen zu platzieren.

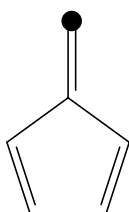


Abb. 7-9: Zuordnung von Doppelbindungen in π -Elektronensystemen mit exocyclischen Doppelbindungen am Beispiel von Fulven. Das Startatom ist mit einem Punkt markiert.

In Ringsystemen ohne exocyclische Doppelbindung gibt es keine Startatome mit nur einem Nachbarn, von denen die Lokalisierung ausgehen kann. Daher muß das Startatom zusammen mit dem ersten Nachbarn so ausgewählt werden, daß eine Lokalisierung möglich ist, ohne Biradikale zu erzeugen. Auch hier sind wieder zwei Fälle zu unterscheiden. In isolierten Ringen sind alle Atome als Startatom gleich gut geeignet. Daher werden als die beiden ersten Atome, zwischen denen eine Doppelbindung ausgebildet wird, das erste Atom des Rings sowie

dessen erstes Nachbaratom ausgewählt. Auf die übrigen Atome wird die rekursive Suche für alternierende Einfach- und Doppelbindungen angewendet, bis man wieder beim Startatom angelangt ist.

In kondensierten Ringen, oder Ringen, die durch Ketten miteinander verbunden sind, werden die ersten beiden Atome nach einer Heuristik ausgewählt. Das Startatom ist der Verzweigungspunkt mit der höchsten erweiterten Konnektivität, der zugleich Teil eines Rings ist. Das zweite Atom der Doppelbindung ist in Ringsystemen mit einer geraden Anzahl von Atomen das Nachbaratom des Startatoms im Ringsystem mit der niedrigsten erweiterten Konnektivität. In Ringsystemen mit einer ungeraden Anzahl von Atomen wird zunächst geprüft, ob ein exocyclisches Nachbaratom vorhanden ist, und falls nicht, ebenfalls das Nachbaratom im Ringsystem mit der niedrigsten Konnektivität gewählt. Dadurch wird das Ringsystem so geschnitten, daß einerseits eine möglichst lange Kette aus unbehandelten Atomen entsteht, auf der dann eine rekursive Suche nach Doppel- und Einfachbindungen durchgeführt wird. Andererseits wird durch die Auswahl einer Ringbindung in geradzahligen Ringsystemen bzw. einer exocyclischen Bindung in ungeradzahligen Ringsystemen gewährleistet, daß keine Biradikale entstehen. Die Wahl des Atoms mit der niedrigsten erweiterten Konnektivität verhindert ebenfalls die Bildung von Biradikalen.

Abbildung 7-10 zeigt die Anwendung der Heuristik zur Auswahl der Lage der ersten Doppelbindung in kondensierten Ringsystemen und Ringsystemen, die über Ketten miteinander verbunden sind. Die Zuordnungen a) und b) würden jeweils zu Biradikalen führen, die Zuordnungen c) und d) ergeben eine richtige Verteilung der Doppel- und Einfachbindungen. Wenn das π -Elektronensystem in a) ausgehend vom Startatom zum Nachbarn mit der höchsten erweiterten Konnektivität geschnitten würde, würden zwei Teile entstehen, die beide aus einer ungeraden Anzahl von Atomen bestehen. Dies wird vermieden, indem die erste Doppelbindung entlang des äußeren Rings festgelegt wird. In b) haben beide Verzweigungspunkte die gleiche erweiterte Konnektivität. Daher ist es numerierungsabhängig, welches Atom als Startatom gewählt wird. Beide führen jedoch zum selben Ergebnis.

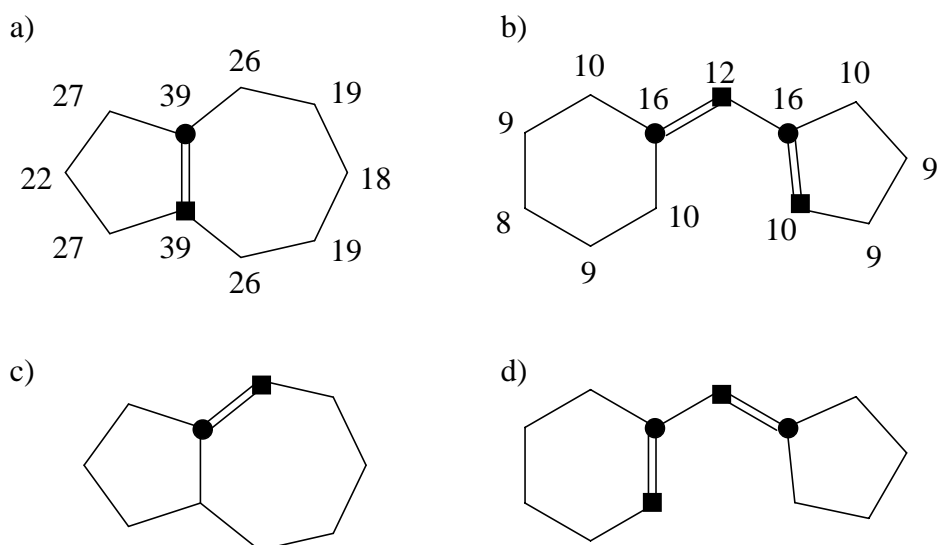


Abb. 7-10: Auswahl der Lage der ersten Doppelbindung in kondensierten Ringen und Ringen, die über Ketten verbunden sind. Das Startatom ist jeweils mit einem Punkt markiert, das Nachbaratom mit einem Quadrat. Die Zahlen neben den Atomen geben die erweiterte Konnektivität an.

Ausgehend von den Nachbarn beider Atome der Doppelbindung wird nun wieder die Suche nach alternierenden Doppel- und Einfachbindungen auf dem π -Elektronensystem durchgeführt, bis entweder alle Atome behandelt wurden oder Verzweigungspunkte als Startpunkte erreicht werden, die noch mehr als einen unbehandelten Nachbarn besitzen. In diesem Fall wird das nächste Atompaar anhand der Heuristik ausgewählt und dieser Zyklus so oft wiederholt, bis alle Atome behandelt wurden.

Die Anwendung dieser Methode auf ein komplexeres Beispiel soll anhand von Phthalocyanin verdeutlicht werden (Abbildung 7-11). Das π -Elektronensystem in der RICOS-Repräsentation erstreckt sich über alle vier Isoindoleinheiten. Es besteht aus 40 Atomen und 42 Elektronen und enthält also zwei freie Elektronenpaare. Diese werden an den grau unterlegten Stickstoffatomen aufgrund des in Kapitel 7.2.2 vorgestellten Verfahrens lokalisiert. Die restlichen Elektronen werden in Einfach- und Doppelbindungen umgewandelt. Das Elektronensystemgerüst ist symmetrisch. Um jedoch eine nicht radikalische Anordnung von Einfach- und Doppelbindungen zu erhalten, dürfen nicht alle äquivalenten Bindungen die gleiche Zuordnung erhalten, drei der Sechsringe müssen je drei Doppelbindungen erhalten, im vierten sind zwei exocyclische Doppelbindungen erforderlich. Die Zuordnung der Bindungsordnungen erfolgt in vier Zyklen, in denen Startatome anhand der erweiterten Konnektivität ausgewählt werden. Jedes Startatom ist mit einem Punkt markiert und die Richtung der ersten Doppelbindung mit einem Pfeil angegeben. Die Reihenfolge, in der die Doppelbindungen zugeordnet wurden, ist durch die Zahlen angegeben. Quadrate markieren die Endpunkte jedes Zyklus, an der die Suche nach alternie-

renden Bindungen abgebrochen wird. Im ersten Zyklus (Abbildung 7-11a) werden die drei Doppelbindungen im unteren Sechsring zugeordnet, im zweiten Zyklus (Abbildung 7-11b), diejenigen des oberen Sechsrings. Im dritten Zyklus (Abbildung 7-11c) werden zunächst die Doppelbindungen 7 bis 10 ermittelt. Vor dem mit einem Dreieck markierten Verzweigungspunkt wird die Suche zunächst abgebrochen. Dann wird das Nachbaratom des Startatoms behandelt und die Doppelbindungen 11 und 12 festgelegt. Als nächstes wird anhand der Differenz zwischen Zahl der Nachbarn und dem Markierungswert festgestellt, daß das mit dem Dreieck markierte Atom nur noch einen unbehandelten Nachbarn besitzt und ausgehend von diesem die Doppelbindungen 13 bis 16 zugeordnet. Im letzten Zyklus (Abbildung 7-11d) werden die Doppelbindungen im rechten Sechsring zugeordnet.

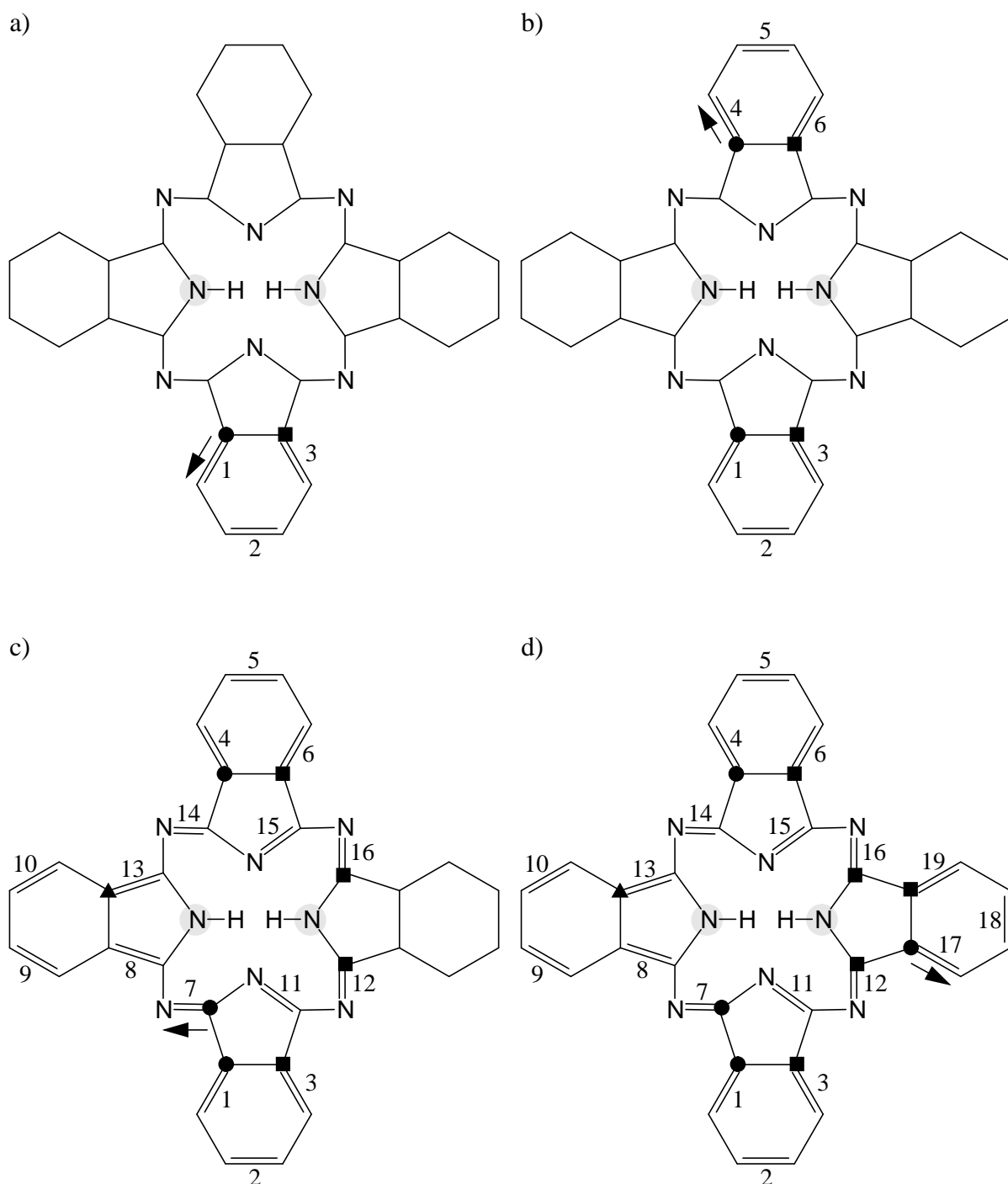


Abb. 7-11: Konvertierung des π -Elektronensystems in Phthalocyanin. Die Verteilung der Doppelbindungen erfolgt in vier Zyklen, die in a) bis d) dargestellt sind.

Um die Robustheit sowie den Zeitbedarf beider Konvertierungsverfahren zu testen, wurden 13 373 Verbindungen der Spektrendatenbank Specinfo [71] von einer VB-Struktur nach RICOS und wieder zurück konvertiert und beide VB-Repräsentationen miteinander verglichen.

Die Datenbank enthält organische Neutralverbindungen und Salze. Alle Ionen eines Salzes wurden für die Konvertierung in isolierte Moleküle eines Ensemble zusammengefaßt, ohne jedoch die Wechselwirkungen zu spezifizieren. Die Konvertierung wurde mit einer Sun Sparc Workstation 512 durchgeführt und benötigte für die Richtung nach RICOS 0.01s pro Ensemble und von RICOS in eine VB-Struktur 0.11s pro Ensemble. In allen Fällen konnte eine gültige VB-Struktur erzeugt werden, die jedoch nicht immer exakt mit der eingegebenen Struktur übereinstimmt, wenn sie äquivalente Atome in π -Elektronensystemen enthält oder mehrere mesomere Grenzformen der Verbindung möglich sind.

8. Realisierung der RICOS-Repräsentation mit einem objektorientierten Modell

Das vorgestellte Modell zur Repräsentation von chemischen Strukturen wurde mit objektorientierten Methoden implementiert. Diese Entscheidung wurde stark durch die Erfahrungen beeinflusst, die in der Arbeitsgruppe mit der Entwicklung der Programmsysteme EROS, FRANZ und MASSIMO gemacht wurden. Diese Programmsysteme wurden mit prozeduralen Sprachen wie Fortran und C programmiert. Ihre Weiterentwicklungen führten dazu, daß ihr Management immer schwieriger wurde. Jede Änderung der zu Grunde liegenden Datenstruktur, die erforderlich war, um neue Bedürfnisse zu befriedigen, zog eine große Anzahl von Änderungen im gesamten System nach sich. Die Integration der hier vorgestellten Repräsentation von chemischen Strukturen wäre nur unter sehr großem Aufwand möglich gewesen.

Daher wurde ein neues Design dieser Programmsysteme erstellt, das auf RCIOS als grundlegender Molekülrepräsentation basiert. Die folgenden Ziele waren für die Entscheidung, eine objektorientierte Klassenbibliothek zu entwickeln, verantwortlich. Die Bibliothek sollte einfach zu benutzen, flexibel und gut wartbar sein. Dies wurde durch den Einsatz von objektorientierten Techniken wie Abstraktion, Kapselung, Vererbung und Polymorphismus erreicht. Ein kurzer Überblick über diese Techniken wird in Kapitel 8.1 gegeben.

Die Klassenbibliothek wurde mit der Programmiersprache C++ implementiert. Sie wurde auf einer Sun-Workstation mit den Compilern von Sun, C++ V. 4.1, und von Gnu, g++ V. 2.7.0, entwickelt. C++ wurde gewählt, da es sich zügig zum *de facto* Standard für kommerzielle Programmierung entwickelt. Zudem kann C++ mit Fortran oder C gemischt werden. Es ist also in der Übergangphase möglich, nützliche Module der existierenden Programmsysteme zu integrieren, ohne sie sofort neu implementieren zu müssen. Außerdem erlaubt C++ auch die Optimierung zeitintensiver Teile der Bibliothek.

8.1. Grundbegriffe

Im folgenden werden die wichtigsten Grundbegriffe zu objektorientiertem Programmieren mit C++ erklärt. Lehrbücher, die sich mit dieser Thematik ausführlich beschäftigen, sind [72], [73], [74]. Die Entwicklung von objektorientierten Programmen verläuft in der Regel in einem Zyklus aus drei Stufen, Analyse, Design und Programmierung, der mehrfach durchlaufen und dabei verfeinert wird. Während der Analyse wird nach allgemeinen Konzepten, die das zu modellierende System beschreiben, gesucht. Für jedes gefundene Konzept werden dessen notwendige Eigenschaften und Verhaltensweisen sowie seine Beziehungen zu anderen Konzepten bestimmt. Hierbei wird nur die Sprache der zukünftigen Anwendung verwendet, jedoch noch nicht überlegt, wie diese Konzepte in einer Computersprache realisiert werden können. Im zweiten Schritt, dem Design, werden aus den Konzepten Klassen entwickelt und so verändert, daß sie einer Repräsentation mit dem Computer zugänglich werden. Zuletzt werden die Klassen mit einer objektorientierten Programmiersprache implementiert.

Während des Designs soll jedoch nicht nur das gestellte Problem modelliert werden, sondern es muß auch darauf geachtet werden, daß die entwickelten Klassen möglichst flexibel und erweiterungsfähig sind. Denn in der Regel ändern sich die Anforderungen an ein Softwaresystem im Laufe der Benutzungszeit. Dazu stehen die im folgenden beschriebenen Techniken zur Verfügung, die speziell dazu ausgelegt sind, diese Anforderungen zu gewährleisten. Sie werden an einem einfachen Modell für Moleküle verdeutlicht.

Eine *Klasse* besteht aus einem Datenbereich, in dem die Attribute des zu modellierenden Konzepts abgelegt werden, und einem Satz an Funktionen, die das Verhalten der Klasse bestimmen. Zudem werden die Daten und Funktionen in zwei logische Bereiche unterteilt, einen privaten Bereich, und einen öffentlichen Bereich. Der öffentliche Bereich ist allen Benutzern zugänglich und enthält das *Funktionsinterface*, welches das Verhalten der Klasse in Bezug auf andere Klassen festlegt. Der private Bereich enthält meist die Attribute der Klasse sowie die Implementation der Funktionen. Diese Trennung von Interface und Implementation nennt man *Kapselung*. Sie erlaubt es, die Implementation der Funktionalität einer Klasse während des Designprozesses zu verändern, ohne jedoch gleichzeitig das Interface mitzuverändern. So wird gewährleistet, daß sich ein Anwender darauf verlassen kann, daß sich ein Interface nur selten ändert. Ein zweiter Vorteil der Kapselung ist, daß sich der Anwender einer Klasse nicht mit den Implementationsdetails befassen muß. Er muß also weniger über eine Klasse lernen als in prozeduralen Programmiersprachen, wo eine Anwendung nur selten unabhängig von der Implementation möglich ist, und kann sich daher in kürzerer Zeit mit der Klassenbibliothek vertraut machen. Eine Klasse für Moleküle, die aus Atomen und Bindungen bestehen, wird in ihrem Interface Funktionen enthalten, die Auskunft über die Zahl der Atome oder Bindungen aus denen das Molekül besteht, geben. Dabei ist es nicht notwendig, zu wissen, wie diese Information bestimmt wird, ob sie aus einem Datenelement gelesen wird oder bei jeder Abfrage neu berechnet wird, solange die entsprechenden Funktionen immer gültige Werte liefern.

Es gibt drei Arten von Beziehungen zwischen Klassen, die zusammen entwickelt werden, *Komposition*, *Verwendung* und *Vererbung*.

In komplexen Systemen gibt es meist Grundbausteine, die immer wieder miteinander kombiniert werden, um verschiedene größere Einheiten zu bilden. Solche Beziehungen werden durch Komposition modelliert. Objekte einer Klasse, die aus mehreren Komponenten zusammengesetzt werden sollen, können Objekte von anderen Klassen enthalten, die diese Komponenten repräsentieren. Dabei können die Klassen, die die Grundbausteine repräsentieren, immer wieder verwendet werden, um neue Klassen zu bilden. So sind Atome die Grundbausteine von Bindungen und Molekülen. Sie könnten aber auch in anderem Zusammenhang als Bausteine eines Salzkristallgitters oder einer Legierung dienen.

Als Verwendung wird bezeichnet, wenn Objekte einer Klasse Objekte einer anderen Klasse zwar verwenden, um ein bestimmtes Verhalten zu erzielen, sie jedoch nicht durch Komposition enthalten.

Bei der Analyse des zu modellierenden Problems werden oft zunächst allgemeine Konzepte entwickelt, die als Grundlage für verfeinerte Modelle dienen. Vererbung ist die Methode, diese Beziehungen mit Klassen zu verwirklichen. So enthält eine allgemeine Klasse, eine sogenannte Basisklasse, die Daten und Funktionen, die dem allgemeinen Konzept entsprechen. Von ihr können durch Vererbung verschiedene Klassen abgeleitet werden, die zusätzliche Datenelemente und Funktionen zur Behandlung der speziellen Lösungen bieten. Sie erhalten gleichzeitig automatisch auch das Interface der Basisklasse. So können zwei Effekte erzielt werden. Einerseits wird schon entwickelter Code der Basisklasse wiederverwendet, andererseits ist es vorteilhaft, wenn miteinander verwandte Klassen das gleiche Interface besitzen, da sie dann einfacher zu verwenden sind. In Modellen, die Moleküle beschreiben, werden oft verschiedene Bindungstypen verwendet. Es ist dann sinnvoll, eine Basisklasse für Bindungen zu entwickeln, deren Interface die Behandlung der in der Bindung enthaltenen Atome und Elektronen beschreibt. Davon abgeleitete Klassen für spezielle Bindungstypen wie σ - oder π -Bindungen können nun zusätzliche Funktionen definieren.

Beim Design von Basisklassen werden häufig sogenannte *virtuelle* Funktionen entwickelt. Dies sind Funktionen, die von abgeleiteten Klassen überschrieben werden können. Das bedeutet, für eine abgeleitete Klasse kann eine virtuelle Funktion so implementiert werden, daß der spezielle Typ der abgeleiteten Klasse berücksichtigt wird, um die gewünschte Funktionalität zu erhalten. Die Implementation der Basisklasse für diese Funktion wird dann für alle Objekte der abgeleiteten Klasse nicht mehr berücksichtigt. Wenn in einer Anwendung Objekte der Basisklasse referenziert werden und deren virtuelle Funktionen aufgerufen werden, dann wird das gewünschte Ergebnis auch dann erzielt, wenn zum Zeitpunkt des Designs noch nicht alle abgeleiteten Klassen bekannt sind. Denn diese neu eingeführten Klassen können die virtuellen Funktionen der Basisklasse ebenfalls überschreiben. Die Anwendung selbst muß aber nicht geändert werden. Diese Technik wird *Polymorphismus* genannt. Sie ist einer der großen Pluspunkte von objektorientierten Programmiersprachen, da sie es erlaubt, daß neu eingeführte Subklassen in ein Anwendungsprogramm integriert werden können, ohne daß es stark geändert werden muß. Dies vereinfacht die Handhabung von großen Programmen sehr, da Änderungen nur noch an wenigen Stellen durchgeführt werden müssen. Eine Klassenfunktion einer Basisklasse für Bindungen, die verantwortlich ist, eine Bindung zu zeichnen, sollte virtuell sein und von jeder abgeleiteten Klasse überschrieben werden. σ - und π -Bindungen haben eine unterschiedliche Gestalt, die dann jeweils beim Aufruf dieser Funktion gezeichnet wird. Wenn nun ein Programm zur Darstellung von Molekülen entwickelt werden soll, ist es sinnvoll, beim Zeichnen von Bindungen nur Referenzen auf die Basisklasse zu verwenden. Falls später neue Bindungstypen entwickelt werden, müssen für diese die Zeichenroutinen ebenfalls überschrie-

ben werden. Aber das Programm zum Zeichnen der Moleküle braucht nicht verändert werden, da durch den Einsatz von Referenzen auf die Basisklasse und virtueller Funktionen zum Zeichnen von Bindungen auch die neu eingeführten Bindungen automatisch richtig gezeichnet werden.

Eine Besonderheit von C++ im Vergleich zu anderen objektorientierten Programmiersprache sind die sogenannten *Templates*. Sie stellen die Implementation eines häufig verwendeten Konzepts zur Verfügung, in dem jedoch der Datentyp, auf den das Konzept angewendet wird, nicht spezifiziert ist. Templates werden häufig zur Verwaltung von mehreren Objekten einer Klasse verwendet, z. B. in Arrays, Listen oder Graphen.

Die Abbildungen zur Erläuterung der Beziehungen zwischen Klassen in den folgenden Abschnitten verwenden die von Booch [73] entwickelten Symbole. Eine kurze Übersicht über diese Symbole zeigt Anhang C.

8.2. Allgemeine Klassen

In diesem Abschnitt werden allgemeine Klassen vorgestellt, die für die Entwicklung der Klassen zur Repräsentation chemischer Information als Grundlage dienen, selbst aber universell auch in anderen Bereichen einsetzbar sind. Sie gehören nicht zu den Standardklassen, die allgemein mit jedem Compiler mitgeliefert werden, und werden daher extra erläutert. Die meisten dieser Klassen sind sogenannte Containerklassen. Containerklassen dienen dazu, bestimmte Modelle zur Speicherung von Objekten zur Verfügung zu stellen. Sie enthalten Methoden zur Speicherung und zum Zugriff auf Objekte des gleichen Typs, sind jedoch allgemein auf beliebige Klassen anwendbar. Sie sind meist als Templates realisiert.

In Tabelle 8-1 ist eine Übersicht der verwendeten Containerklassen gegeben.

Außer diesen Containerklassen wurden noch zwei allgemeine Klassen für Datentypen verwendet. Da zum Zeitpunkt des Beginn der Entwicklung noch keine standardisierte Klasse zur Behandlung von Charakterarrays als Strings erhältlich war, wurde die String Klasse der C++-Klassenbibliothek von Gnu [75] verwendet. Eine weitere Klasse, *e_bit64*, implementiert die betriebsystemunabhängige Bitvektoren der Länge 64 und stellt die entsprechenden logischen Operatoren zur Verfügung.

Tab. 8-1: Übersicht über verwendete Containerklassen.

| Art des Containers | Klassenname | Beschreibung, Bedingung |
|--------------------------|---------------------------|---|
| einfach verkettete Liste | <i>e_Slist</i> <T> | Objekte vom Typ T |
| | <i>e_Splist</i> <T> | Zeiger auf Objekte vom Typ T |
| assoziativer Array | <i>e_Map</i> <K,V> | Objektpaare mit Objekten vom Typ K als Schlüssel und Objekten vom Typ V als Werte |
| Array | <i>vector</i> <T> | beliebige Objekte vom Typ T (nicht nur double, int) Beginn des Laufindex bei 0 |
| | <i>e_varlb_vector</i> <T> | Unter- und Obergrenze des Bereichs frei wählbar |
| zweidimensionaler Array | <i>e_matrix</i> <T> | beliebige Objekte vom Typ T Beginn der Laufindices bei 0 |
| dreidimensionaler Array | <i>e_3dim_array</i> <T> | beliebige Objekte vom Typ T Beginn der Laufindices bei 0 |

8.3. Klassenhierarchie der Hauptklassen

In Kapitel 4 wurden die Komponenten, aus denen sich ein Ensemble zusammensetzen kann, vorgestellt. Es besteht aus Aggregaten und Molekülen, die sich ihrerseits aus Atomen, Elektronensystemen und Wechselwirkungs- bzw. topologischen Gruppen zusammensetzen. Diese chemischen Begriffe wurden nun als Grundlage zur Entwicklung der Klassen genommen, mit denen RICOS implementiert werden soll. Tabelle 8-2 zeigt die Namen der Klassen, die für die jeweiligen Begriffe definiert wurden.

Tab. 8-2: Namen der Klassen, die zur Implementation von RICOS entwickelt wurden.

| Chemischer Begriff | Klassenname |
|----------------------------|---------------------|
| Ensemble | <i>ds_ensemble</i> |
| Aggregat | <i>ds_aggregate</i> |
| Molekül | <i>ds_molecule</i> |
| Atom | <i>ds_atom</i> |
| Elektronensystem | <i>ds_elecsys</i> |
| σ -Elektronensystem | <i>ds_sigmasys</i> |
| π -Elektronensystem | <i>ds_pisys</i> |
| koordinative Bindung | <i>ds_coordsys</i> |
| Gruppe | <i>ds_group</i> |
| Nachbaratompaaire | <i>ds_neighbors</i> |

Ensembles, Aggregate, Moleküle und Atome wurden jeweils direkt in Klassen umgesetzt. Für Elektronensysteme wurde eine gemeinsame Klasse, *ds_elecsys*, entwickelt, von der drei Klassen zur Beschreibung der speziellen Elektronensystemtypen abgeleitet wurden. Wechselwirkungsgruppen und topologische Gruppen wurden zu einer gemeinsamen Klasse zusammengefaßt. Zusätzlich zu den zuvor vorgestellten Konzepten wurde noch eine Klasse *ds_neighbors* definiert, die jeweils ein Paar benachbarter Atome zusammenfaßt. All diese Klassen haben jedoch Gemeinsamkeiten. Um diese nicht mehrfach implementieren zu müssen, wurden zwei Basisklassen gebildet, von denen alle Klassen aus Tabelle 8-2 abgeleitet wurden. Die Klasse *ds_chem_obj* beinhaltet die Funktionen, die die allgemeinen Aufgaben aller chemischen Klassen erfüllen, die Klasse *ds_chem_obj_w_groups* ist von ihr abgeleitet und enthält zusätzlich Funktionen, die sich mit der Behandlung von Gruppen befassen. Die gesamte Vererbungshierarchie ist in Abbildung 8-1 dargestellt. Die Basisklasse *ds_chem_obj* befindet sich an der Spitze, Subklassen sind jeweils durch einen Pfeil mit ihrer Basisklasse verbunden.

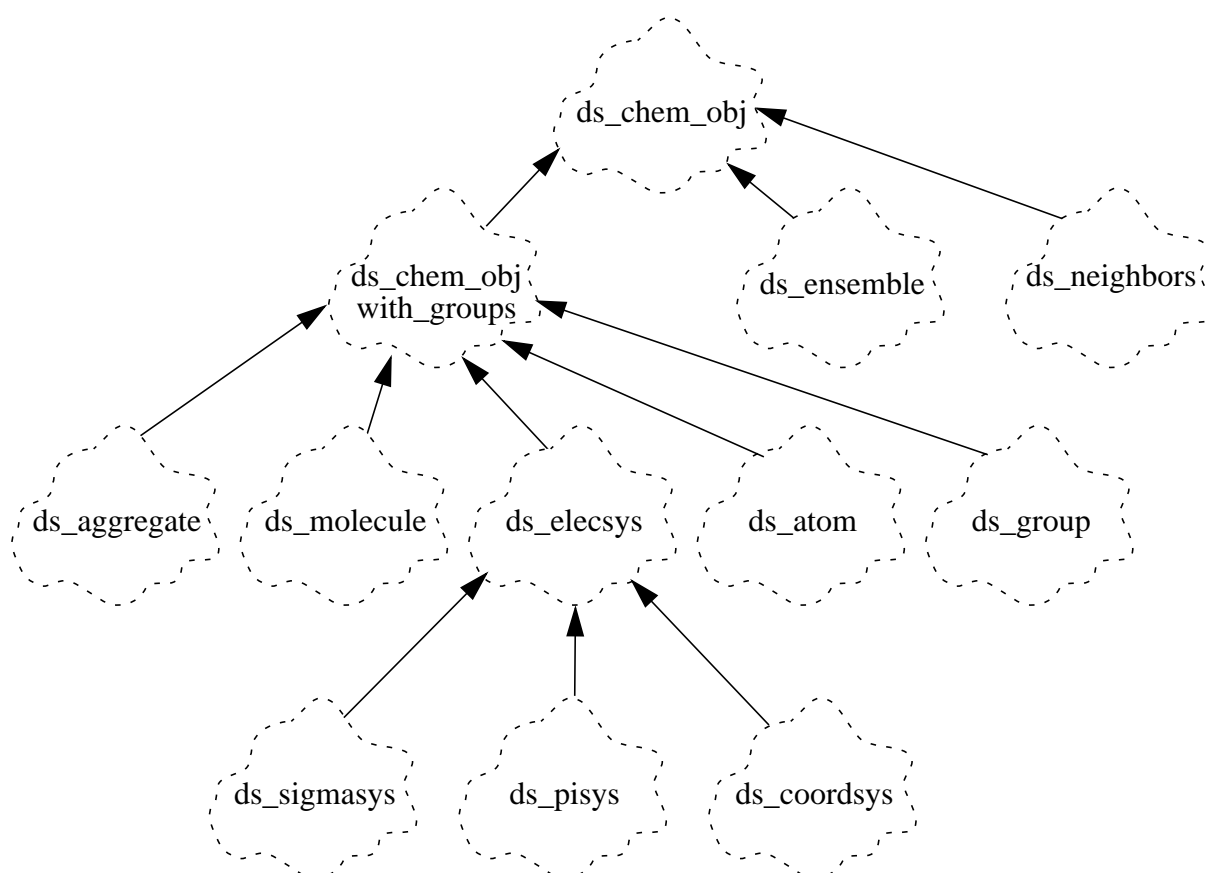


Abb. 8-1: Vererbungshierarchie der chemischen Klassen.

Die chemischen Klassen werden verwendet, um die Konnektivität und Assoziation von Molekülen zu beschreiben. Viele Programme nutzen aber nicht nur diese Information, sondern auch physikochemische Eigenschaften der Atome, Elektronensysteme oder Moleküle zur Lösung der gestellten Aufgaben. Es ist daher notwendig, eine Möglichkeit zur allgemeinen Behandlung von Eigenschaften eines chemischen Objektes zur Verfügung zu stellen. Da für alle chemischen Klassen Eigenschaften berechnet werden sollen, die Art und Menge der verwendeten Eigenschaften aber nicht begrenzt werden sollte, wurde die Klasse *ds_chem_obj* entwickelt, deren Aufgabe die dynamische Verwaltung dieser Eigenschaften ist und von der alle anderen chemischen Klassen abgeleitet werden. Dies bietet zwei Vorteile. Einerseits muß die Eigenschaftsverwaltung nur einmal implementiert werden und andererseits ist dadurch gewährleistet, daß der Zugriffsmechanismus für alle Klassen gleich ist, was eine einfache Handhabung erlaubt. Bei der Definition der chemischen Klassen wurden nur Attribute, die unbedingt zur Beschreibung eines Objekts dieser Klasse notwendig sind, als Klassenattribute abgelegt, alle anderen Eigenschaften werden über den allgemeinen Mechanismus verwaltet.

Auch Gruppen können chemischen Objekten unterschiedlichen Typs zugeordnet sein. Gruppen, die intramolekulare Wechselwirkungen darstellen, gehören zu dem entsprechenden Mole-

kül, Gruppen, die für intermolekulare Wechselwirkungen verantwortlich sind, werden dem entsprechenden Aggregat zugeordnet. Topologische Gruppen können die Umgebung von bestimmten Atomen oder Elektronensystemen beschreiben und werden diesen daher speziell zugeordnet. Um auch für die Behandlung von Gruppen einen allgemeinen Mechanismus zur Verfügung stellen zu können, wurde daher von der Klasse, *ds_chem_obj* die Klasse *ds_chem_obj_w_groups* abgeleitet, die für die Verwaltung von Gruppen zuständig ist.

8.4. Chemische Klassen

Bevor alle chemischen Klassen im Detail beschrieben werden, soll zunächst ein Überblick darüber gegeben werden, in welchen Beziehungen sie zueinander stehen. Abbildung 8-2 zeigt, wie sich die Objekte der Klassen gegenseitig referenzieren. Alle Klassen, die durch Linien miteinander verbunden sind, stehen direkt miteinander in Beziehung. Dabei markiert ein schwarzer Punkt die Klasse, die einen Verweis enthält, und ein leeres Quadrat die Klasse, auf die verwiesen wird. Die Zahlen neben den Punkten geben an, wieviele Objekte referenziert werden können. Ein Ensemble kann beliebig viele Aggregate oder Moleküle enthalten. Jedes Aggregat oder Molekül enthält einen Verweis auf das Ensemble, dem es angehört. Wenn sich ein Molekül in einem Aggregat befindet, enthält es zusätzlich eine Referenz darauf. Ein Molekül enthält Verweise auf seine Atome und Elektronensysteme, jedes Atom und Elektronensystem einen Verweis auf das Molekül zu dem es gehört. Da es häufig notwendig ist, alle Atome eines Elektronensystems bzw. alle Elektronensysteme, an denen ein Atom Teil hat, anzusprechen, sind sie gegenseitig referenziert, um einen schnellen Zugang zu dieser Information zu schaffen. Jedes Molekül kann zusätzlich einen Container mit allen Paaren aus Nachbaratomen besitzen. Die Beziehungen, die Gruppen zu den anderen Klassen haben, wurden in Abbildung 8-2 der Übersichtlichkeit wegen nicht dargestellt. Sie werden jedoch in den nächsten Abschnitten erläutert.

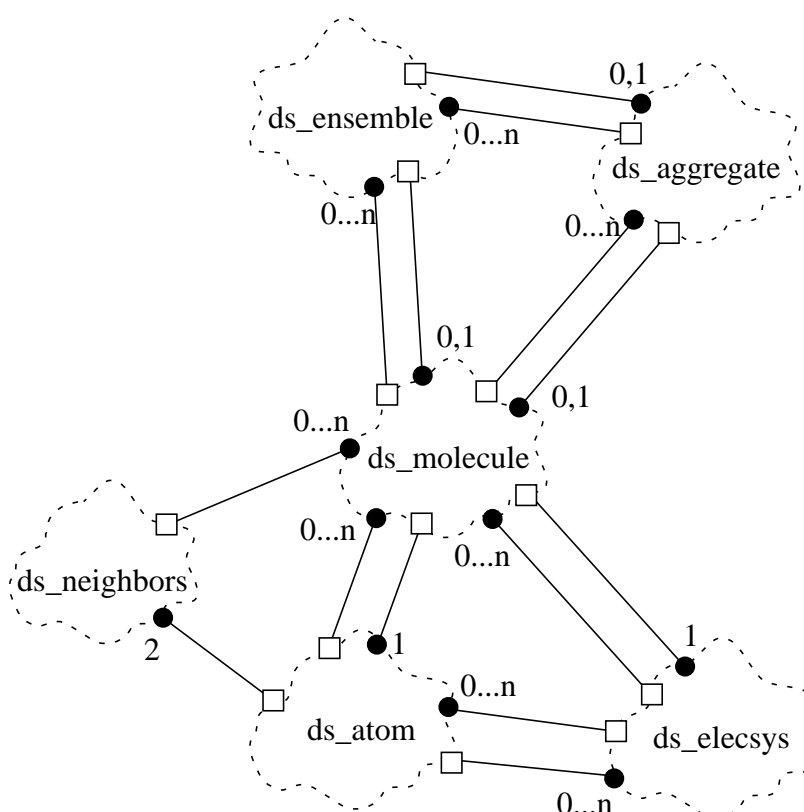


Abb. 8-2: Beziehungen zwischen den chemischen Klassen.

Für alle chemischen Klassen wurden Konstruktoren und Destruktoren implementiert sowie Funktionen zum Kopieren von Objekten chemischer Klassen bereitgestellt, sie werden im folgenden nicht mehr extra erwähnt.

8.4.1. Die Atomklasse

Es wird nicht nur die Funktionalität der Klasse *ds_atom* beschrieben, sondern stellvertretend für alle Chemieklassen auch deren interner Aufbau erläutert. Alle anderen Klassen sind aus entsprechenden Komponenten zusammengesetzt. Der Zugriff auf Objekte anderer Chemieklassen, die von der Atomklasse referenziert werden, erfolgt über Iteratoren. Deren Mechanismus wird am Beispiel des Iterators für Elektronensysteme erklärt. Die Iteratoren für die anderen Klassen sind analog definiert und implementiert und werden daher ebenfalls nicht weiter erläutert.

Die Klasse *ds_atom* besitzt Variablen zur Speicherung der Ordnungszahl, eines Laufindizes und eines Labels zur Kennzeichnung der Identität jedes Atoms. Sie enthält einen Container zur Speicherung der Elektronensysteme, an denen es beteiligt ist, sowie einen Container zur Speicherung der Gruppen von denen es ein Teil ist. Beide sind private, d. h. nicht ohne Zugriffsfunktionen zugängliche Bestandteile der Klasse *ds_atom*. Da ein Atom in der Regel nur an

wenigen Elektronensystemen beteiligt ist, diese sich in Reaktionen jedoch häufig ändern, wurde das Template für verkettete Listen zur Speicherung der Elektronensysteme gewählt. Zusätzlich zu den Funktionen, die den Zugriff auf die Klassenvariablen gewährleisten, hat *ds_atom* Funktionen, die es erlauben, Elektronensysteme an ein Atom zu addieren bzw. zu entfernen. Außerdem kann abgefragt werden, wieviele Elektronensysteme ein Atom enthält. Mit einer Instanz der Klasse *ds_electsys_iter* kann die Liste der Elektronensysteme eines Atoms traversiert werden. Sie stellt Funktionen zur Verfügung, die das erste und das aktuelle Element der Liste zurückgeben sowie je eine Funktion zum Vorwärtsgehen und zur Abfrage, ob der Iterator sich noch auf einem gültigen Element befindet. Im Gruppencontainer werden alle Gruppen festgehalten, von denen das Atom ein Bestandteil ist. Der Inhalt dieses Containers wird automatisch erzeugt und kann über einen Gruppeniterator abgefragt werden. Die Linien zwischen den Containerklassen und ihren Iteratoren, die in Abbildung 8-3 mit einem F (für Freund) markiert sind, kennzeichnen, daß die Iteratoren auch den privaten Bereich der Containerklassen benutzen dürfen. Zudem besitzt die Klasse *ds_atom* einen Container mit den Definitionen aller Atomeigenschaften, die während eines Programmlaufs zur Verfügung stehen. Dieser Container ist statisch deklariert, das bedeutet, er ist nur einmal für alle Objekte der Klasse *ds_atom* vorhanden. Alle anderen Chemieklassen besitzen ebenfalls einen Container mit den Definitionen der für sie vorhandenen Eigenschaften. Er wird daher im folgenden nicht mehr weiter erwähnt.

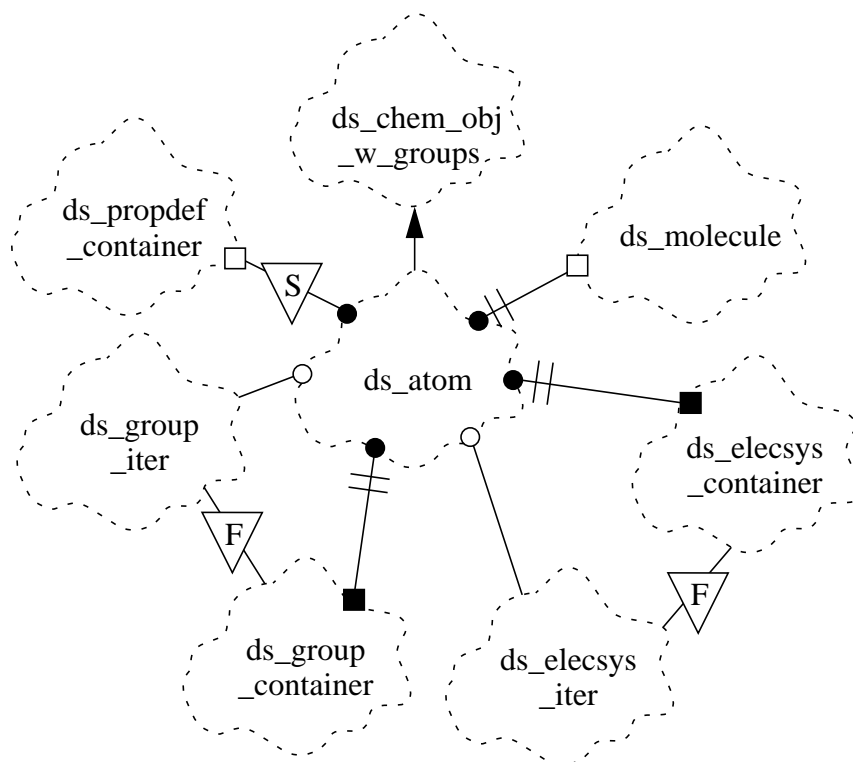


Abb. 8-3: Klassendiagramm der Atomklasse *ds_atom*.

8.4.2. Die Klassen für Elektronensysteme

Zur Behandlung von Elektronensystemen wurde eine Basisklasse *ds_elecsys* definiert, die die gemeinsamen Aufgaben aller Arten Elektronensysteme übernimmt. Sie enthält einen Laufindex und ein Label zur Identifizierung sowie je einen Container zur Speicherung der Atome, die in einer Instanz dieser Klasse enthalten sind, und der Gruppen, deren Bestandteil sie ist. Zudem enthält es eine Variable zur Speicherung des aktuellen Elektronensystemtyps. Von dieser Klasse sind die Klassen zur Behandlung der speziellen Elektronensysteme abgeleitet, *ds_sigmasys*, *ds_pisys* und *ds_coordsys*. *ds_sigmasys* ist um eine private Variable für den Geometriotyp, regulär, verbrückt oder geschlossen, und eine Referenz auf ein zentrales Atom, erweitert. Die Klassen *ds_pisys* und *ds_coordsys* haben keine zusätzlichen Attribute, sind aber notwendig, um virtuelle Funktionen der Basisklasse zu überschreiben und damit ein ihnen angepaßtes Verhalten zu ermöglichen.

8.4.3. Die Molekülklasse

Die Klasse zur Repräsentation von Molekülen, *ds_molecule*, enthält drei Container zur Speicherung von Atomen, Elektronensystemen und Nachbaratompaaaren. Sie stellt Funktionen und Iteratoren zur Verfügung, um Atome und Elektronensysteme eines Moleküls anzusprechen und ihre Zusammensetzung zu modifizieren. Aus der Konnektivität kann der vollständige Satz an Nachbaratompaaaren abgeleitet werden.

Sie enthält zusätzlich Container und Funktionen, die sich mit der Gültigkeit von Atom-, Elektronensystem-, Nachbaratompaar- und Moleküleigenschaften befassen. Die Verwaltung der Eigenschaften ist zwar grundsätzlich eine Aufgabe der Basisklasse *ds_chem_obj*. Da es aber einen zu großen Aufwand bedeutet, die Gültigkeit jeder Eigenschaft jeden Atoms etc. zu speichern, erhalten jeweils alle Eigenschaftswerte der gleichen Eigenschaft in allen Atomen etc. eines Moleküls einen gemeinsamen Status, der mit dem Molekül gespeichert wird.

8.4.4. Die Aggregatklasse

Die Klasse *ds_aggregate* enthält einen Container mit den in ihr enthaltenen Molekülen und stellt Funktionen und einen Iterator für deren Zugriff zur Verfügung. Es ist nicht zwingend erforderlich, daß ein Aggregat aus mindestens zwei Molekülen mit einer Wechselwirkung bestehen muß. Um eine konstante Hierarchie des Aufbaus eines Ensembles zu gewährleisten, wurde erlaubt, daß ein Aggregat auch nur aus einem Molekül besteht. Es darf dann allerdings keine Gruppen enthalten.

8.4.5. Die Ensembleklasse

Die Klasse *ds_ensemble* enthält je einen Container für Aggregate und Moleküle. Sie stellt einen Iterator zum Zugriff auf alle Aggregate sowie zwei Iteratoren für Moleküle zur Verfügung. Mit dem einen Iterator kann nur auf die direkt im Ensemble gespeicherten Moleküle

zugegriffen werden, mit dem anderen Iterator werden auch die Moleküle innerhalb von Aggregaten angesprochen.

8.4.6. Die Gruppenklasse

Wechselwirkungs- und topologische Gruppen werden in Objekten der Klasse *ds_group* gespeichert. *ds_group* enthält je einen Container für die an der Gruppe beteiligten Atome und Elektronensysteme sowie Funktionen zum Eintragen und Löschen von Atomen und Elektronensystemen. Es gibt zwei Möglichkeiten, wie die Atome bzw. Elektronensysteme, die eine Gruppe bilden, bestimmt werden. Eine Möglichkeit ist die Festlegung durch einen Benutzer, die zweite Möglichkeit ist, von *ds_group* eine Klasse abzuleiten, die eine Vorschrift zur Suche der in die Gruppe gehörigen Atome enthält. Dies ist für topologische Gruppen wie die Bestimmung des Satzes an Ringen in einem Molekül sinnvoll. Die Klasse *ds_group* ist von *ds_chem_obj_w_groups* abgeleitet, sie kann also ebenfalls Gruppen enthalten. Die Tiefe der Hierarchie ist momentan auf zwei begrenzt. Eine Gruppe kann einen Satz von Gruppen der gleichen Art enthalten, die sich aus Atomen und Elektronensystemen zusammensetzen müssen.

Jede Gruppe erhält einen Namen, durch den ihr Typ festgelegt wird. Wechselwirkungsgruppen unterscheiden sich von topologischen Gruppen durch die Wechselwirkung. Wenn deren Stärke spezifiziert werden soll, so kann der Wert in einer Gruppeneigenschaft abgelegt werden.

Die Klassenbibliothek enthält bisher zwei Subklassen, die spezielle topologische Gruppen implementieren. Mit der Klasse *ds_rings* wird der kleinste Satze an kleinen Ringen eines Moleküls abgeleitet. Dazu wurde ein Algorithmus von Sorkau [76] verwendet. Die Klasse *ds_atom_neighbors* macht alle Nachbaratome eines gegebenen Atoms zugänglich.

8.4.7. Die Nachbarschaftsklasse

Aus dem VB-Modell zur Beschreibung von Molekülen folgt, daß Atome, die an einer Bindung beteiligt sind, Nachbaratome sind. Die physikochemischen Eigenschaften, die Bindungen zugeordnet werden, sind häufig Eigenschaften, die sich auf zwei benachbarte Atome beziehen. In RICOS ist diese direkte Beziehung zwischen Elektronensystemen und Nachbarschaftsatomen nicht mehr gegeben, da die Zahl der Atome pro Elektronensystem nicht auf zwei festgelegt ist. Da aber auch in diesem Modell Eigenschaften benachbarter Atompaare zur Verfügung gestellt werden sollen, wurde für diese Zwecke die Klasse *ds_neighbors* entwickelt. Sie ist von *ds_chem_obj* direkt abgeleitet, Instanzen dieser Klasse können keine Gruppen zugewiesen werden. Die Klasse *ds_neighbors* enthält zwei Referenzen auf die beiden benachbarten Atome sowie Funktionen zu ihrem Zugriff.

8.5. Basisklassen

8.5.1. Chemische Objekte

Die Basisklasse *ds_chem_obj* stellt hauptsächlich Funktionen zum Setzen und Abfragen der Eigenschaften eines chemischen Objekts zur Verfügung. Jedes chemische Objekt enthält einen Eigenschaftscontainer (*ds_prop_container*) mit allen seiner Eigenschaften (Abbildung 8-4). Der Zugriff erfolgt über Objekte des Typs *ds_prop_entry*. Für jede Eigenschaft gibt es einen eindeutigen Eintrag, der aus dem Namen und dem Typ der Eigenschaft sowie dem Typ des aktuellen Chemieobjekts abgeleitet wird. Objekte der Klasse *ds_prop* enthalten den Wert der Eigenschaft. Um mit Eigenschaften sinnvoll arbeiten zu können, wurden zwei weitere Klassen entwickelt. Mit Objekten der Klasse *ds_prop_definit* wird jede Eigenschaft definiert. Die Definition enthält den Namen der Eigenschaft, die Art des Wertes, die Berechnungsmethode, Defaultwerte, für welche chemische Klasse sie definiert ist, und unter welchen Bedingungen sie ungültig wird. Tabelle 8-3 enthält die momentan gültigen Typen für Eigenschaftswerte. Einige der Berechnungsmethoden sind in Kapitel 5 erläutert.

Tab. 8-3: Gültige Typen für Eigenschaftswerte.

| Zugriffsschlüssel | Inhalt |
|-------------------|-----------------|
| PT_UNDEF | nicht definiert |
| PT_INT | int |
| PT_LONG | long int |
| PT_FLOAT | float |
| PT_DOUBLE | double |
| PT_CHAR | char |
| PT_STRING | String |
| PT_INTVEC | integer vector |
| PT_LONGVEC | long int vector |
| PT_FLOATVEC | float vector |
| PT_DOUBLEVEC | double vector |
| PT_CHARVEC | char vector |
| PT_STRINGVEC | String vector |
| PT_BIT64 | e_bit64 |
| PT_PROPEXTRY | ds_prop_entry |

In Objekten der Klasse *ds_pg_status* wird festgehalten, ob eine Eigenschaft momentan gültig ist oder nicht. Wenn der Wert einer ungültigen Eigenschaft abgefragt wird, wird anhand der Definition entweder eine Neuberechnung gestartet, oder falls keine Berechnungsmethode zur Verfügung steht, ein Defaultwert eingesetzt, bevor die Abfrage beantwortet wird. Alle Definitionen von Eigenschaften einer Art chemischer Objekte werden in einem Container (*ds_propdef_container*) gesammelt. Wie schon in Kapitel 8.4.3 erläutert, wird die Information, ob die Eigenschaften eines chemischen Objekts gültig sind, nicht immer mit dem chemischen Objekt gespeichert. Jedes chemische Objekt hat jedoch über virtuelle Funktionen die Möglichkeit, diese Information abzufragen, wenn der Wert einer Eigenschaft angefordert wird.

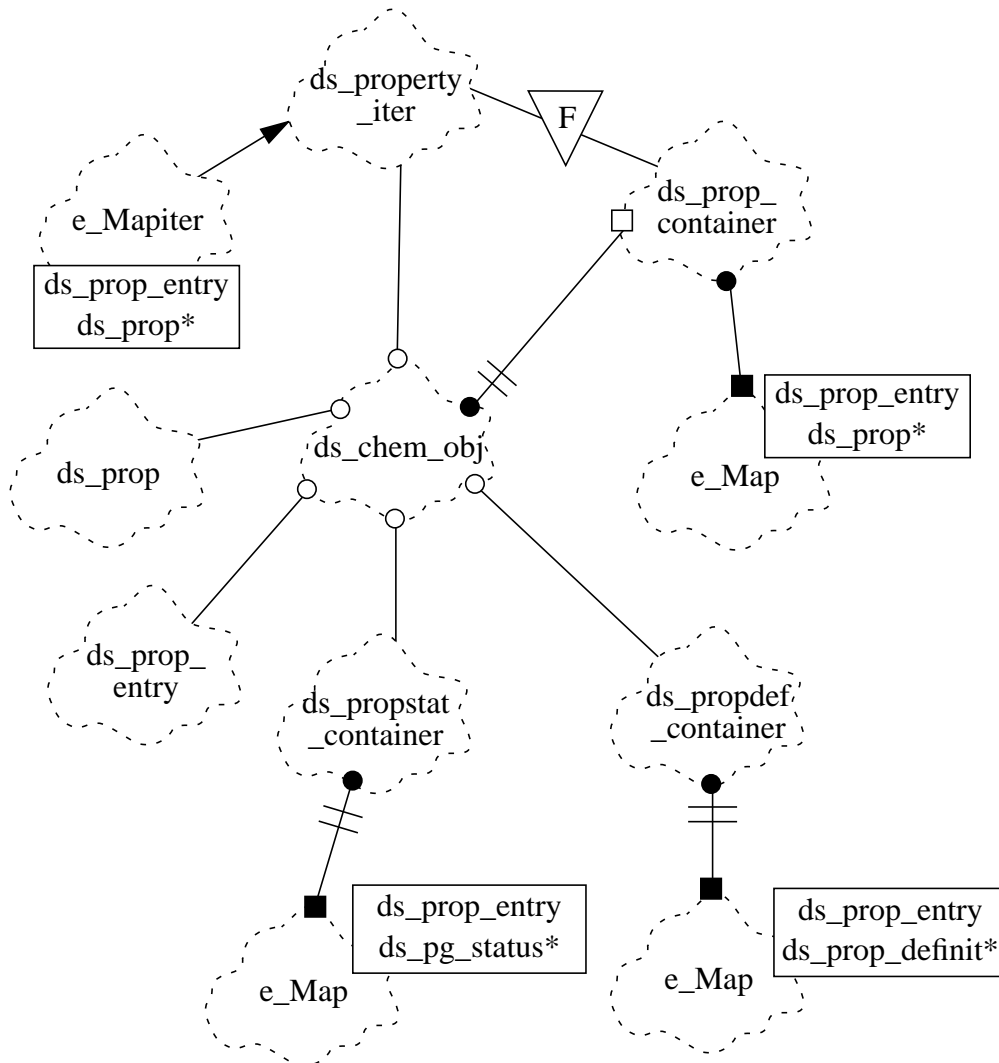


Abb. 8-4: Klassendiagramm für die Klassen zur Behandlung von Eigenschaften.

8.5.2. Chemische Objekte mit Gruppen

Die Klasse *ds_chem_obj_w_groups* ist von *ds_chem_obj* abgeleitet. Sie enthält im wesentlichen Funktionen zur Speicherung von Gruppen eines Chemieobjekts. Die Klassen zur Definition und Speicherung von Gruppen sind denen zur Behandlung von Eigenschaften in *ds_chem_obj* ähnlich. Die Klasse enthält einen Container zur Speicherung der Gruppen. Der Zugriff auf bestimmte Gruppen erfolgt über Klassenfunktionen mit einem Zugriffscode, einem Objekt der Klasse *ds_group_entry*. Es wird aus dem Namen der Gruppe und dem Typ des chemischen Objekts, mit dem die Gruppe verknüpft ist, generiert. Zudem verfügt die Klasse *ds_chem_obj_w_groups* über einen Iterator, mit dem alle Gruppen traversiert werden können. In der neuentwicklung des Programmsystems EROS wird es möglich sein, in den Reaktionsregeln neue Gruppentypen zu definieren und zu den schon vorhandenen Gruppensehreibungen zu addieren. Da die Reaktionsregeln hauptsächlich mit der der Scriptsprache Tcl [78] kodiert werden sollen, ist es in dieser Anwendung nicht möglich, neue Gruppenklassen durch Vererbung zu erzeugen. Daher wurde die Klasse *ds_group_definit* entwickelt. Sie speichert alle Informationen, die zur Beschreibung einer bestimmten Gruppe notwendig sind. Darunter fallen der Name, ein String mit einer Kurzbeschreibung und ein Attribut, mit dem festgelegt, wie eine Gruppe am ende einer Reaktionssequenz weiterbehandelt werden soll. Wenn die Atome einer Gruppe eines Moleküls nach einer Reaktion durch Spaltung auf mehrere Moleküle verteilt sind, gibt es dafür drei Möglichkeiten. Sie kann gelöscht werden, sie kann in zwei unabhängige Gruppen getrennt werden, die mit den beiden Reaktionsprodukten verbunden sind, oder sie bleibt bestehen, und aus den beiden Molekülen wird ein Aggregat gebildet. Alle Gruppensehreibungen werden in einem Container gesammelt. Sie können über den Namen oder den Zugriffscode abgefragt werden.

freien Elektronen eines Atoms gespeichert wird, sowie Funktionen zu deren Änderung. Die Klasse *ds_vbbond* dient zur Kodierung von Einfach- und Mehrfachbindungen zwischen Atompaaren. Ein Objekt dieser Klasse muß daher zwei Atome enthalten. Zudem stellt sie Funktionen zur Änderung der Bindungsmultiplizität bereit. In Abbildung 8-6 ist die Zusammensetzung eines Moleküls aus Objekten der Klassen *ds_vbatom* und *ds_vbbond* gezeigt.

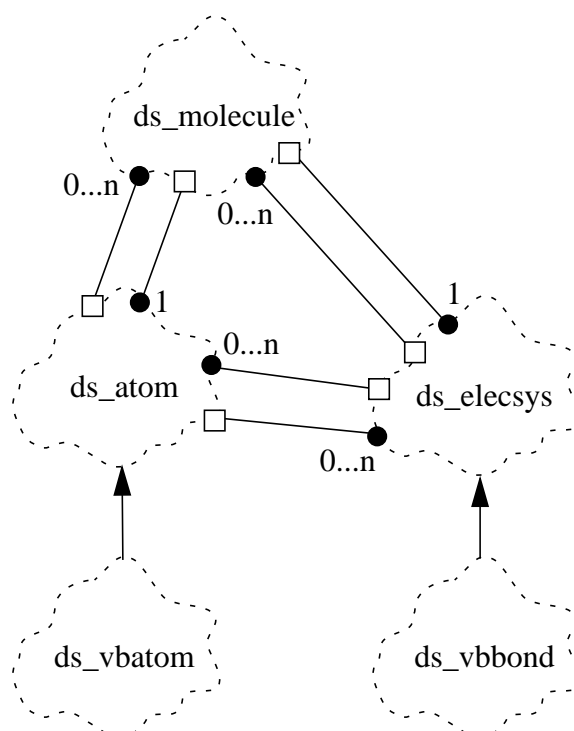


Abb. 8-6: Zusammensetzung eines Moleküls in VB-Repräsentation.

8.7. Reaktionen

Um die Modellierung von Reaktionen möglichst einfach zu gestalten, wurden für die in Kapitel 6 beschriebenen Elementarschritte zur Veränderung eines Ensembles und für die Valenzüberprüfung von Molekülen zwei Klassen entwickelt. Die Klasse *ds_modify* ist für die Reaktionsschritte verantwortlich, die Klasse *ds_valence* wird zum Test auf valenzchemische Gültigkeit eingesetzt.

Jede chemische Klasse enthält Funktionen, die es erlauben, andere chemische Objekte zu einem Objekt zu addieren oder zu entfernen. Diese Funktionen eignen sich jedoch nicht direkt für die Modifikation von chemischen Objekten in Reaktionen, da sie keine Konsistenzüberprüfung eines Ensembles umfassen. Als Schnittstelle zur Generierung von Reaktionen wurde daher die Klasse *ds_modify* entworfen. Jeder der in Kapitel 6 beschriebenen Elementarschritte wird durch eine Klassenfunktion implementiert. Neben der Durchführung der einzelnen Ele-

mentarschritte hat die Klasse *ds_modify* die Aufgabe, für Konsistenz innerhalb des Reaktionsensemble zu sorgen. Dies geschieht in mehreren Stufen. Während jeden Reaktionsschritts werden nur die an einem Reaktionsschritt beteiligten Atome, Elektronensysteme und Gruppen neu gebildet, modifiziert oder gelöscht. Die korrespondierenden Moleküle und Aggregate werden nicht umgeordnet, da in einer Reaktionssequenz häufig Zwischenprodukte gebildet werden, die später wieder gebrochen werden und die Zuordnung zu Molekülen sehr zeitaufwendig ist. Es wird jedoch mitverfolgt, welche Art von chemischen Objekten des Reaktionsensembles geändert wurden. Am Ende einer Reaktionssequenz werden ungültig gewordene Referenzen gelöscht und alle Atome, Elektronensysteme und Gruppen den richtigen Molekülen und Aggregaten zugeordnet. Die Behandlung von Wechselwirkungs- und topologischen Gruppen umfaßt drei Möglichkeiten, wie mit einer Gruppe verfahren werden kann. Gruppen, die durch die Änderung der Konnektivität der an ihnen beteiligten Atome oder Elektronensysteme ungültig werden, werden gelöscht. Ein Beispiel hierfür sind Ringsysteme, die während einer Reaktion aufgebrochen werden (Abbildung 8-7a). Gruppen, die Wechselwirkungen beschreiben, die von der durchgeführten Reaktion nicht betroffen sind, bleiben bestehen. Dabei kann es sein, daß sich intramolekulare Wechselwirkungen in intermolekulare Wechselwirkungen umgewandelt haben, was eine weitere Umordnung der Moleküle in Aggregaten nach sich zieht (Abbildung 8-7b). Die letzte Möglichkeit betrifft Gruppen, die verwendet werden, um bestimmte Atome zu markieren (Abbildung 8-7c). Wenn sich die Atome einer solchen Gruppe am Ende der Reaktion nicht mehr innerhalb eines Moleküls befinden, wird die Gruppe geteilt und auf die neu entstandenen Moleküle verteilt.

Zuletzt werden alle Eigenschaften aller chemischen Objekte darauf hin geprüft, ob sie ihre Gültigkeit durch einen der angewendeten Reaktionsschritte verloren haben und gegebenenfalls als ungültig markiert.

Die Klasse *ds_valence* führt einen Konsistenztest auf einem chemischen Objekt durch. Dies kann ein Ensemble, Aggregat, Molekül oder Atom sein. Sie überprüft das chemische Objekt anhand der in Kapitel 6.4 vorgestellten Kriterien für eine der vier Kategorien. Es wird geprüft, ob die Besetzung der Elektronensysteme mit Elektronen, die Kombination der Elektronensysteme sowie die äquilibrierten Ladungen in dem jeweils vorgegebenen Bereich liegen. Falls nicht, so wird für das entsprechende chemische Objekt ein Fehlerwert zurückgegeben, der von Art des aufgetretenen Fehlers abhängt.

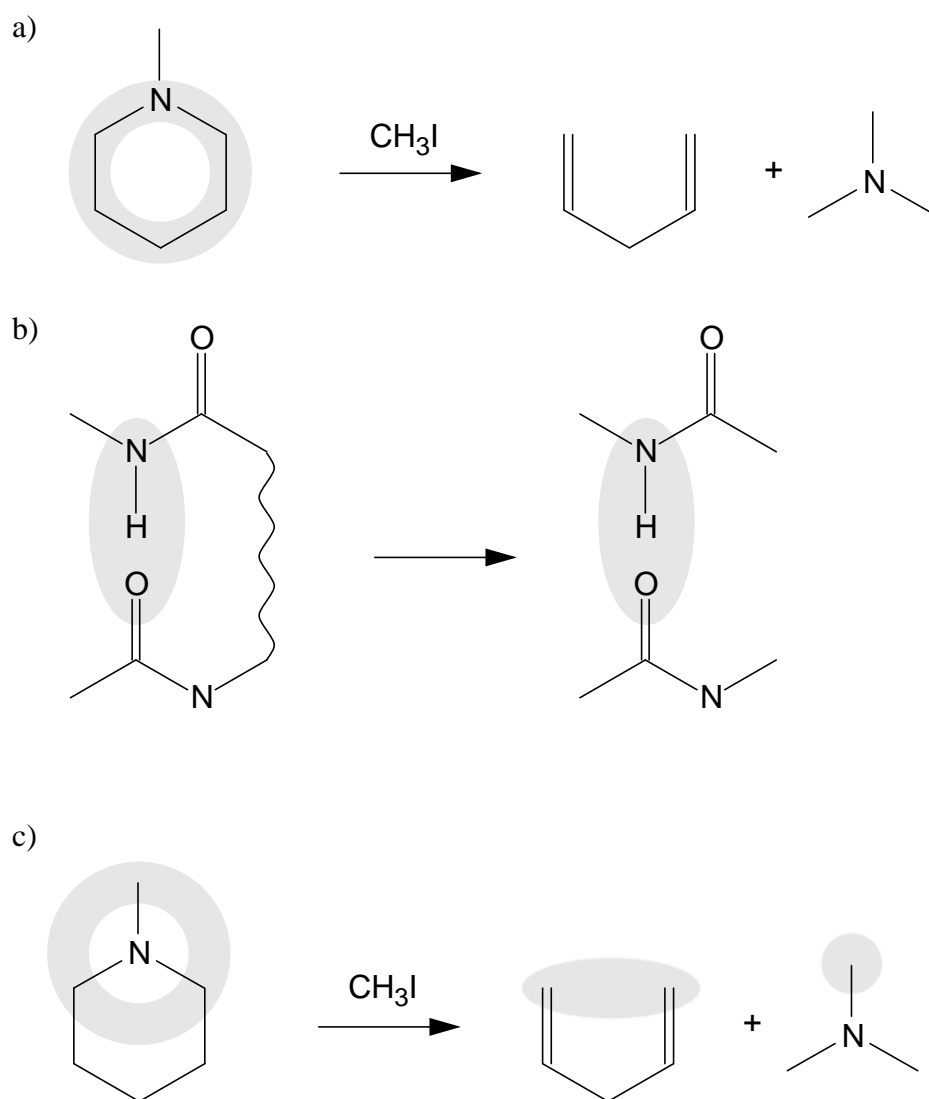


Abb. 8-7: Auswirkungen von Reaktionen auf die Zusammensetzung von Gruppen, (a) die Gruppe wird nach einer Reaktion ungültig und damit gelöscht, (b) intramolekulare Wechselwirkungen werden zu intermolekularen Wechselwirkungen, die die Bildung eines Aggregats zur Folge haben, (c) eine Gruppe wird in mehrere Gruppen gespalten.

9. EROS 7

9.1. Programmaufbau

Das Programmsystem EROS7, das von R. Höllering und L. Steinhauer in unserem Arbeitskreis entwickelt wurde, enthält außer der neuen Datenstruktur zur Repräsentation von Molekülen noch weitere Erweiterungen gegenüber früheren EROS-Versionen, die im folgenden vorgestellt werden sollen. Es bietet eine sehr flexible Steuerung der Generierung des Reaktionsnetzwerks und erlaubt zudem eine freie Gestaltung der formalen Reaktionstypen, die als Wissensbasis dienen. Abbildung 9-1 gibt einen Überblick über den schematischen Aufbau von EROS 7. Nach dem Einlesen der Edukte wird das Reaktionsnetzwerk aufgebaut. Anhand der Reaktionsregeln werden alle Edukte geprüft, ob sie den Bedingungen für die Reaktionszentren genügen. Ist dies der Fall, so wird die Reaktion durchgeführt, und abgespeichert. Um Produktkonzentrationen vorauszusagen, kann jede Regel eine Reaktivitätsfunktion enthalten, die die Geschwindigkeitskonstante der Reaktion abschätzt. Am Ende jeder Reaktionsebene werden kinetische Berechnungen durchgeführt, die anhand der Ausgangskonzentrationen und der modellierten Geschwindigkeitskonstanten die Konzentrationen der bis dahin erzeugten Verbindungen bestimmen.

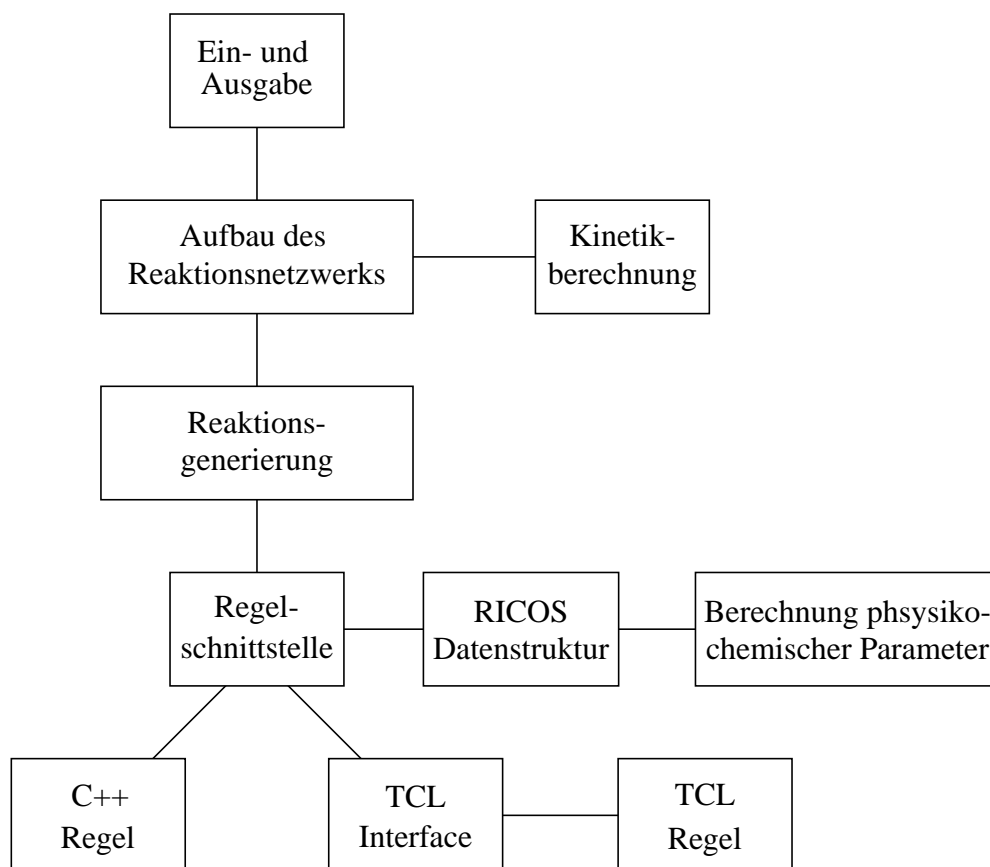


Abb. 9-1: Aufbau von EROS 7.

Wenn keine weitere Reaktion mehr durchgeführt werden kann, wird der Aufbau des Reaktionsnetzwerks beendet und die Ergebnisse ausgegeben. Zum Abbruch der Reaktionsgenerierung führen drei verschiedene Kriterien: Auf ein noch unbehandeltes Edukt kann keine weitere Reaktionsregel mehr angewendet werden, die Konzentration eines Produkts fällt unter eine vorgegebene Schwelle oder die erlaubte Zahl der Reaktionsebenen wird überschritten.

9.1.1. Aufbau des Reaktionsnetzwerks

Das System zur Steuerung des Aufbaus des Reaktionsnetzwerks in EROS 7 enthält verschiedene Modelle, mit denen festgelegt wird, welche der bereits erzeugten Moleküle in der nächsten Reaktionsebene zu neuen Eduktensembles zusammengefaßt werden. Dabei gibt es zwei Möglichkeiten: monomolekulare Zerfallsreaktionen und bimolekulare Reaktionen. In monomolekularen Zerfallsreaktionen darf jedes Eduktensemble nur ein Edukt enthalten. Wenn eine Reaktion zu zwei Produkten geführt hat, so werden diese getrennt voneinander als neue Edukte behandelt. In bimolekularen Reaktionen kann zwischen zwei Verfahren zur Kombination der Eduktmoleküle in ein Reaktionsensemble gewählt werden. Einerseits dürfen nur die Produkte einer Reaktion in der nächsten Reaktionsebene gemeinsam als Edukte ausgewählt werden. Andererseits werden die Eduktensembles der nächsten Reaktionsebene erzeugt, indem alle Kombinationen aus zwei der bisher aufgetretenen Moleküle gebildet werden. Dabei kann auch festgelegt werden, ob ein Edukt mit sich selbst oder nur mit einem anderen Reaktionspartner reagieren darf.

9.1.2. Reaktionsgenerierung und Reaktionsregeln

Die Reaktionsgenerierung ist in zwei Bereiche eingeteilt, der Datei, die die formalen Reaktionstypen enthält und der Datenstruktur. Die Regeldatei enthält eine Funktion bzw. Prozedur, die zum Programmsystem statisch oder dynamisch gelinkt werden kann, und über die Regelschnittstelle mit der Reaktionsgenerierung kommuniziert. Eine flexible Gestaltung der formalen Reaktionstypen ist nur dann möglich, wenn die Sprache, die zur Abfassung dieser Reaktionstypen verwendet wird, die geforderte Flexibilität bietet. Dies ist natürlich bei der Verwendung von C++, das auch zur Programmierung des gesamten EROS-Systems verwendet wurde, gegeben. Da es jedoch auch möglich sein soll, die Reaktionstypen unabhängig vom Programm zu formulieren und zu testen, mußte eine Scriptsprache gewählt werden, die leistungsfähig genug ist, um den an sie gestellten Anforderungen zu genügen. Als Scriptsprache wurde Tcl [78] gewählt, die sehr mächtig ist und sich daher für diese Aufgabe gut eignet.

Die Regeldatei besteht aus zwei Teilen, einem globalen, in dem Definitionen, die für alle Regeln gültig sind, gesetzt werden können, und den Reaktionsregeln selbst. Über die globalen Definitionen wird festgelegt, welches Modell zur Steuerung der Reaktionsgenerierung verwendet werden soll. Es kann zwischen verschiedenen Modellen zur Lösung der Differentialgleichungssysteme, die bei der Berechnung der Konzentrationen benötigt werden, gewählt

werden. Und es können Variablen, die das Kernsystem benötigt, mit Anfangswerten belegt werden, z. B. die Anfangskonzentration.

Jede Reaktionsregel ist in drei Teile gegliedert. Im ersten Teil wird die Reaktionssubstruktur, der Name und andere Attribute der Regel festgelegt. Im zweiten Teil werden die Bedingungen, die für die Reaktionssubstruktur erfüllt sein müssen, damit die Reaktion durchgeführt wird, festgelegt. Der dritte Teil besteht in der eigentlichen Generierung der Reaktion und optional in der Berechnung einer Reaktivitätsfunktion.

Die Reaktionssubstruktur kann aus beliebig vielen Teilsubstrukturen bestehen, die nicht benachbart sein müssen, und auch verzweigte Pfade enthalten dürfen. Auch die Anzahl der Atome einer Reaktionssubstruktur ist frei wählbar.

In den Reaktionsbedingungen können nicht nur die Atome der Reaktionssubstruktur geprüft werden, es ist auch möglich, andere Atome zu prüfen, die nicht notwendigerweise im Molekül vorhanden sein müssen, jedoch einen Einfluß auf die Reaktion haben können. So können Nachbaratome eines Teilzentrums der Reaktionssubstruktur überprüft werden, und falls sie gewissen Bedingungen nicht gehorchen, kann die Reaktion verworfen werden. Falls das entsprechende Teilzentrum jedoch keine Nachbaratome besitzt, kann die Reaktion dennoch durchgeführt werden.

Für die formale Beschreibung einer Reaktion stehen zwei Modelle zur Verfügung. Für Reaktionen, die sich auch gut mit einer Beschreibung als VB-Strukturen formulieren lassen, ist eine Emulation der VB-Elementarreaktionen vorhanden. Dies sind Reaktionen von neutralen Verbindungen aus Hauptgruppenelementen ohne Oktettaufweitung. Sie erlaubt es, die Bindungsordnung zwischen zwei Atomen zu erhöhen und zu erniedrigen und die Zahl der freien Elektronen zu verändern. Diese Funktionen werden intern auf das RICOS-Modell übertragen. Zur Formulierung der Reaktionstypen im RICOS-Modell stehen Funktionen zum Machen und Brechen von Elektronensystemen sowie zum Ändern der Zahl der Elektronen in einem Elektronensystem zur Verfügung.

9.2. Anwendungsbeispiele

Die Modellierung von Reaktionen mit EROS7 soll an zwei Anwendungsbeispielen gezeigt werden, der massenspektrometrischen Fragmentierung des in Kapitel 2.4.1 als Problemfall geschilderten Butan-2-ons und einer allgemeinen Regel zur Formulierung von thermischen Cycloadditionen.

9.2.1. Reaktionen in der Massenspektrometrie

In Kapitel 2 wurde gezeigt, daß sich VB-Strukturen zur Repräsentation von Reaktionen in der Massenspektrometrie nur bedingt eignen. Anhand des Reaktionsschemas von Butan-2-on wird erläutert, wie sich der Aufbau des Reaktionsnetzwerks durch den Einsatz von RICOS und der flexibleren Regelsprache ändert. Zur Generierung von Reaktionen in der Massenspektrometrie

eignet sich das Modell zum monomolekularen Zerfall von Verbindungen. Abbildung 9-2 illustriert die verwendeten Reaktionsregeln in RICOS-Schreibweise. Bei der Ionisation werden zwei Fälle unterschieden. Allgemein werden alle π -Elektronensysteme eines Neutramoleküls ionisiert (a). Nur in Carbonylen, in denen das Carbonyl- π -Elektronensystem aus zwei Atomen besteht, wird die n-Ionisation am Sauerstoffatom vorgezogen (b). σ -Ionisation findet nicht statt. Die α -Spaltungsreaktion wurde nun so kodiert, daß die beiden Atome A2 und A3, deren Bindung homolytisch gespalten wird, nicht Teil des radikalischen π -Elektronensystems sein dürfen. Für die induktive Spaltung muß das Kation am Rand eines π -Elektronensystems liegen. Die beiden Reaktionen zur Wasserstoffumlagerung konnten wieder in eine Regel vereinigt werden. Es wird jedoch weiterhin geprüft, ob der Abstand der beiden Schweratome bei der Beteiligung eines Heteroatoms zwischen 0 und 4 Atome beträgt und bei der Beteiligung zweier Kohlenstoffatome zwischen zwei und vier Atomen. Das radikalische Atom kann sowohl an einem Atom lokalisiert sein als auch Teil eines delokalisierten π -Elektronensystems sein. In diesem Fall wird das π -Elektronensystem vor der eigentlichen Umlagerungsreaktion so gespalten, daß das radikalische Elektron an dem zuvor ermittelten Reaktionszentrum liegt. Die formalen Kriterien der Reaktionstypen für die Carbonyl- und Wasserstoffeliminierung wurden nicht modifiziert, jedoch auf das RICOS-Modell angepaßt.

fache Massenspektrum von Butan-2-on kann dieser Vergleich jedoch auch manuell durchgeführt werden. Für alle Fragmente außer den beiden Verbindungen **7** und **13** findet sich ein entsprechendes Isotopenmuster im experimentellen Massenspektrum. Durch Verwendung der neuen Regeln hat sich die Zahl der Reaktionen von 24 auf 10 reduziert. Darin sind auch Wasserstoffumlagerungen enthalten, deren Produkte nicht weiterreagieren und die daher in FRANZ aus dem Reaktionsnetzwerk entfernt würden. Die beiden Reaktionen zu den Fragmenten **9** und **10**, die bisher Probleme bereitet hatten, treten nicht mehr auf. Sie sind in der Abbildung entsprechend markiert. Die beiden eingeklammerten Reaktionen zu **7** und **13** werden zwar durchgeführt, sie korrespondieren jedoch nicht mit Signalen im experimentellen Spektrum und wurden daher bei der Analyse mit FRANZ verworfen.

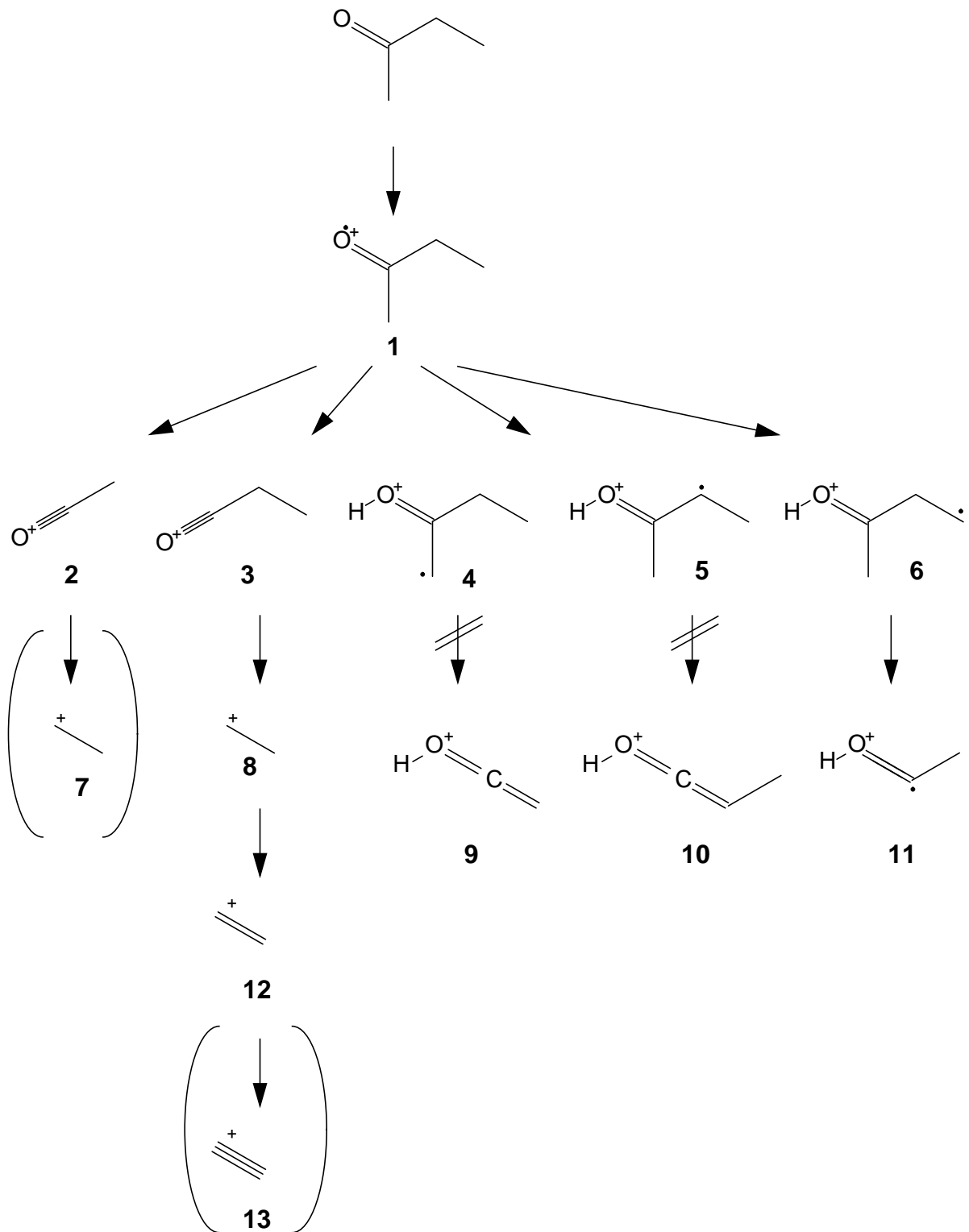


Abb. 9-3: Fragmentierungsschema von Butan-2-on mit EROS 7. Die Reaktionen mit durchgestrichenen Reaktionspfeilen werden nicht mehr durchgeführt, die eingeklammerten Reaktionen werden zwar erzeugt, führen jedoch zu Strukturen ohne korrespondierende Signale im experimentellen Spektrum.

9.2.2. Thermische Cycloadditionsreaktionen

Die Diels-Alder Reaktion ist die bekannteste [2+4] Cycloadditionsreaktion. Nach [80] wurde jedoch nicht nur über thermische [2+4] Cycloadditionen berichtet, sondern auch über Reaktionen vom Typ [4+6] und [2+8]. Es wurde nun eine allgemeine Reaktionsregel entwickelt, die auf Additionsreaktionen aller drei Typen angewendet werden kann.

Die Reaktionsbeschreibung einer allgemeinen Regel für thermische [n+m] Cycloadditionen zwischen neutralen Verbindungen lautet folgendermaßen: Es gibt vier unabhängige Atome in der Reaktionssubstruktur. Die Atome A1 und A2 müssen ein gemeinsames π -Elektronensystem besitzen, ebenso die Atome A3 und A4. Zwischen den Atomen A1 und A2 sowie A3 und A4 muß sich eine ungeradzahlige Kette von Atomen befinden, die verschieden lang sein darf. Die Summe der beiden Abstände muß eine durch vier teilbare Zahl sein, um symmetrieverbotene Reaktionen wie [2+2] Cycloadditionen zu verhindern. Sie darf nicht größer als 8 sein, um zu lange Ketten zu vermeiden. A1 und A3 müssen sich in zwei verschiedenen Molekülen befinden. Die Reaktion wird nun so durchgeführt, daß die vier Atome mit je einem Elektron aus den π -Elektronensystemen entfernt werden und zwischen den Atomen A1 und A3 sowie A2 und A4 ein reguläres σ -Elektronensystem gebildet wird, zu dem jedes Atom ein Elektron beisteuert. Reaktionen werden dann verworfen, wenn die Produkte Radikale oder Ladungen enthalten.

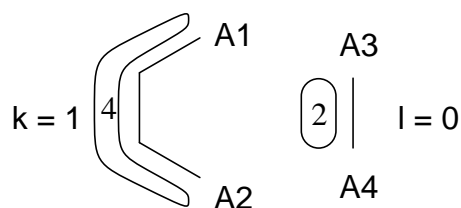
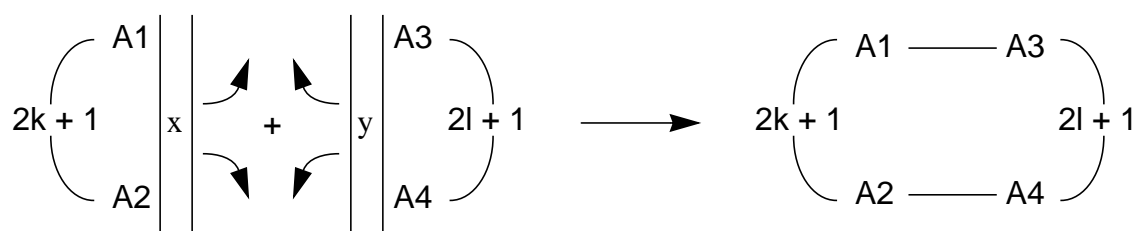


Abb. 9-4: Allgemeine Reaktionsregel für [n+m] Cycloadditionen und Abbildung der Reaktionssentren auf Buta-1,3-dien und Ethylen als Edukte.

Im unteren Teil der Abbildung ist das Mapping der Reaktionssentren auf Buta-1,3-dien und Ethylen als Edukte gezeigt.

Für ein Gemisch aus Octa-1,3,5,7-tetraen und Buta-1,3-dien wurden 19 Produkte mit der Reaktionsregel für die Cycloaddition erzeugt. Der verwendete Modus zur Reaktionsgenerierung erlaubte auch die Reaktion eines der beiden Reaktanden mit sich selbst. Die Produkte sind in Abbildung 9-5 dargestellt.

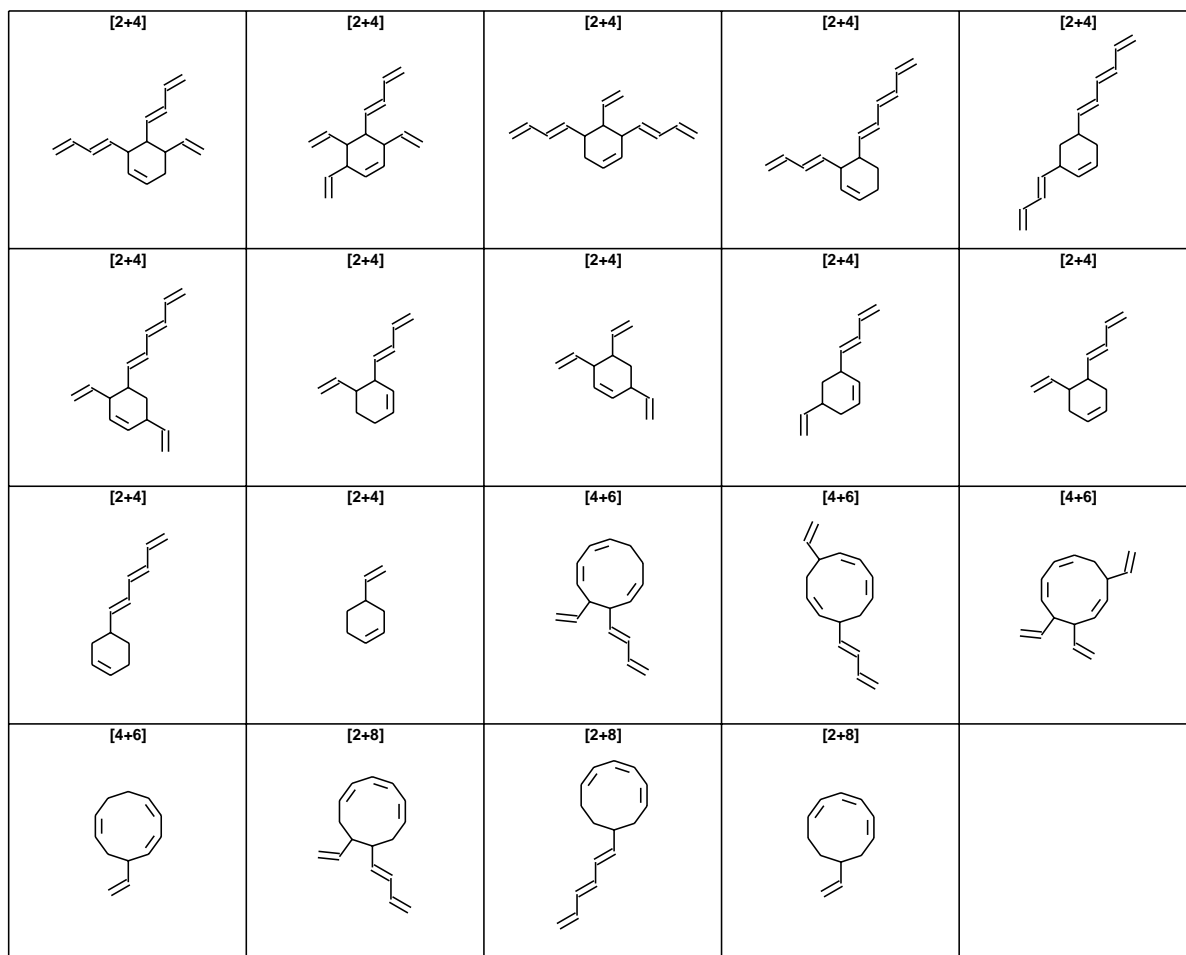


Abb. 9-5: Produkte der [2+4], [4+6] und [2+8] Cycloadditionsreaktionen zwischen Octa-1,3,5,7-tetraen und Buta-1,3-dien.

10. Zusammenfassung

Das mit RICOS vorgestellte Modell zur Repräsentation chemischer Strukturen erlaubt die Behandlung einer Vielzahl verschiedener chemischer Verbindungen. Neben organischen Verbindungen können auch Verbindungen mit Mehrzentrenbindungen sowie metallorganische Spezies dargestellt werden. Weiterhin bietet diese Form der Repräsentation die Möglichkeit, Mehrdeutigkeiten bei der Repräsentation organischer Verbindungen mit delokalisierten und orthogonalen π -Elektronensystemen aufzulösen. Sie ersetzt die klassische Datenstruktur der Bindungslisten, die es nur erlauben, chemische Strukturen mit zwischen zwei Atomen lokalisierten Bindungen zu beschreiben. Das Modell trennt σ - und π -Anteile einer Bindung in sogenannte σ - und π -Elektronensysteme und führt einen eigenen Bindungstyp für Bindungen in Metallkomplexen ein. Elektronen können in allen drei Bindungstypen zwischen mehr als zwei Atomen delokalisiert sein, Bindungen sind nicht mehr notwendigerweise zwischen zwei Atomen lokalisiert. Auch freie Elektronen werden nicht als Attribut eines Atoms betrachtet, sondern als π -Elektronensysteme eines Atoms mit maximal zwei Elektronen repräsentiert. Dies gestattet, verschieden angeregte Zustände der gleichen Verbindung zu beschreiben.

Neben der Repräsentation von Bindungen wurden zwei weitere Konzepte zur Beschreibung chemischer Spezies eingeführt, Wechselwirkungsgruppen und Aggregate. Wechselwirkungsgruppen beschreiben Wechselwirkungen zwischen Atomen und Elektronensystemen, die nicht kovalenter Natur sind und daher nicht in Bindungen zu fassen sind, wie z. B. ionische Wechselwirkungen. Aggregate kombinieren mehrere Moleküle, die über solche Wechselwirkungen miteinander verbunden sind.

Zur Generierung von Reaktionen in Programmen, die sich mit der Reaktionsvorhersage beschäftigen, wurden Elementarreaktionsschritte entwickelt, die es erlauben, die verschiedenen Bindungstypen ineinander umzuwandeln und damit eine Vielzahl verschiedener chemischer Reaktionstypen zu modellieren.

Die erste Anwendung der RICOS-Repräsentation liegt in der Neuentwicklung des Reaktionsvorhersageprogramms EROS. Es basiert auf RICOS als interner Strukturrepräsentation und erlaubt daher, Modelle zur Reaktionsvorhersage nicht nur organischer Reaktionen zu entwickeln, sondern in diese Modelle auch Reaktionen mit Metallkomplexen zu integrieren oder auch Zwischenstufen von Reaktionen explizit zu beschreiben. Am Beispiel von massenspektroskopischen Prozessen und einer allgemeinen Reaktionsregel zur Erzeugung aller formal möglichen Reaktionsprodukte thermischer Cycloadditionsreaktionen wurde diese Anwendung gezeigt.

Das vorgestellte Modell wurde mit objektorientierten Methoden implementiert. Die chemischen Konzepte wurden in Klassen übertragen, die die Konstitution und das Verhalten von Molekülen modellieren. Durch den Einsatz von objektorientierten Methoden konnte eine flexible Klassenbibliothek entwickelt werden, die es erlaubt, zukünftige Anforderungen an die

Strukturrepräsentation auf einfache Weise zu integrieren. Die Klassenbibliothek umfaßt außer den Klassen zur Beschreibung der Konstitution von Molekülen und deren Reaktionen auch eine flexible Methode, Verfahren zur Berechnung von physikochemischen und topologischen Eigenschaften der Atome und Bindungen in Molekülen in die Bibliothek zu integrieren und damit potentiellen Anwendern zur Verfügung zu stellen. Solche Eigenschaften können bei der Reaktionsvorhersage zum Abschätzen der Reaktivität oder in QSAR-Studien zur Modellierung von Aktivitäten eingesetzt werden.

11. Literaturverzeichnis

- [1] Corey, E. J.
„Computer-Assisted Design of Complex Organic Syntheses“
Science 166 (1969) 178-192
- [2] Corey, E. J.; Wipke, W. T.; Cramer III, R. D.; Howe, W. J.
„Computer-Assisted Synthetic Analysis for Complex Molecules. Methods and Procedures for Machine Generation of Synthetic Intermediates“
J. Am. Chem. Soc. 94 (1972) 440-459
- [3] Wipke, W. T.; Dyott, T. M.
„Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry“
J. Am. Chem. Soc. 96 (1974) 4825-4834
- [4] Gelernter, H.; Sridharan, N. S.; Hart, A. J.; Yen, S. C.; Fowler, F. W.; Shue, H. J.
„The Discovery of Organic Synthetic Routes by Computer“
Top. Curr. Chem. 41 (1973) 113-150
- [5] Salatin, T. D.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 1. Overview“
J. Org. Chem. 45 (1980) 2043-2051
- [6] Paderes, G.D.; Jorgensen, W. L.
„Computer-Assisted Evaluation of Oxidation Reactions“
J. Org. Chem. 54 (1989) 2058-2085
- [7] Gasteiger, J.; Ihlenfeldt, W. D.; Röse, P.
„A collection of computer methods for synthesis design and reaction prediction“
Recl. Trav. Chim. Pays-Bas 111 (1992) 270-290
- [8] Dugundji, J.; Ugi, I.
„An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs“
Top. Curr. Chem. 39 (1973) 19-64
- [9] Blair, J.; Gasteiger, J.; Gillespie, C.; Gillespie, P. D.; Ugi, I.
„Representation of the Constitutional and Stereochemical Features of Chemical Systems in the Computer Assisted Design of Syntheses“
Tetrahedron 30 (1974) 1845-1859
- [10] Bauer, J.
„IGOR2: a PC Program for Generating New Reactions and Molecular Structures“
Tetrahedron Comput. Methodol. 2 (1989) 269-280

- [11] Fontain, E.; Reitsam, K.
„The Generation of Reaction Networks with RAIN. 1. The Reaction Generator“
J. Chem. Inf. Comput. Sci. **31** (1991) 96-101
- [12] Gasteiger, J.; Jochum, C.; Marsili, M.; Thoma, J.
„The synthesis planning program EROS“
MATCH **6** (1979) 177-199
- [13] Kubinyi, H.
„Automatische Interpretation von Massenspektren“
In: Software-Entwicklung in der Chemie 2, Gasteiger, J. (Hrsg.)
Springer-Verlag Heidelberg (1988) 167-179
- [14] Brodmeier, T.; Gloor, A.; Cadish, M.; Burgin, R.; Pretsch, E.
„Hypermedia Tools for the Interpretation of Mass Spectra“
Anal. Chim. Acta **277** (1993) 297-304
- [15] Schulz, K. P.; Bauerschmidt, S.; Höllering, R.; Gasteiger, J.
„Automatic Elucidation of Reactions in the Mass Spectrometer“
In: AIP conference proceedings 330, Bernardi, F.; Rivail, J. L., Eds.
American Institute of Physics (1995) 726-733
- [16] Gasteiger, J.; Hanebeck, W.; Schulz, K. P.
„Prediction of Mass Spectra from Structural Information“
J. Chem. Inf. Comput. Sci. **32** (1992) 264-271
- [17] Ash, J. E.; Warr, W. A.; Willett, P., (Hrsgs.)
„Chemical structure systems computational techniques for representation, searching, and processing of structural information“
Ellis Horwood (1991)
- [18] Ugi, I; Gillespie, P.
„Beschreibung chemischer Systeme und ihrer Umwandlungen durch *be*-Matrizen und ihre Transformations-Eigenschaften“
Angew. Chem. **83**, (1973) 980-981
- [19] Hanebeck, W.
„Simulation und Rekonstruktion von Reaktionen im Massenspektrometer“
Dissertation, Technische Universität München (1991)
- [20] McLafferty, F. W.
„Interpretation of Mass Spectra“
University Science Books, Mill Valley CA (1980)

- [21] Höllering, R.
„Reaktionen halogenerter Alkane im Massenspektrometer: Vergleich einer automatischen Spektrenauswertung mit MS-MS-Spektren“
Diplomarbeit, TU München (1992)
- [22] Simon, V.
„Bewertungsmechanismen für Reaktionen und Anwendungen neuronaler Netze zur Reaktionsvorhersage“
Dissertation, Technische Universität München (1993)
- [23] Schulz, K. P.
„Computergestützte Untersuchungen über Zusammenhänge zwischen Struktur und Massenspektrum“
Dissertation, Technische Universität München (1992)
- [24] Bouchoux, G.
„Keto-enol Tautomers and Distonic Ions: The Chemistry of $[C_nH_{2n}O]$ Radical Cations. Part I“
Mass Spectrom. Reviews. 7 (1988) 1-39
- [25] McAdoo, D. J.
„Rearrangements of Ionized Carbonyl Compounds“
Org. Mass. Spectrom. 23 (1988)
- [26] Traeger, J. C.; McLoughlin, R. G.; Nicholson, A. J. C.
„Heat of Formation for Acetyl Cation in the Gas Phase“
J. Am. Chem. Soc. 104 (1982) 5318-5322
- [27] McAdoo, D. J.; McLafferty, F. W.; Parks, T. E.
„Isomerization and Decomposition Reactions of $C_4H_8O^+$ Ions.“
J. Am. Chem. Soc. 94 (1972) 1601-1609
- [28] Dittmar, P. G.; Stobaugh, R. E.
„The Chemical Abstracts Service Chemical Registry System. I. General Design“
J. Chem. Inf. Comput. Sci. 16 (1976) 111-121
- [29] Morgan, H. L.
„The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service“
J. Chem. Doc. 5 (1965) 107-112
- [30] Mockus, J.; Stobaugh, R. E.
„The Chemical Abstracts Service Chemical Registry System. VII. Tautomerism and Alternating Bonds“
J. Chem. Inf. Comput. Sci. 20 (1980) 18-22

- [31] Jochum, C.
„Building a Structure-Oriented Numerical Factual Database“
In: Chemical Structures - The International Language of Chemistry, Warr, W. A. (Hrsg.)
Springer-Verlag, Heidelberg (1988) 187-193
- [32] Welford, S. M.
„Die Datenstruktur des Beilstein für organische Verbindungen“
In: Software-Entwicklung in der Chemie 1, Gasteiger, J. (Hrsg.)
Springer-Verlag, Heidelberg (1987) Vol. 1, 5-11
- [33] Gasteiger, J.
„A Representation of π -Systems for Efficient Computer Manipulation“
J. Chem. Inf. Comput. Sci. 19 (1979) 111-115
- [34] Gasteiger, J.
„Automatische Erzeugung pericyclischer Reaktionen“
Z. Naturforsch. 34b (1979) 67-75
- [35] Barnard, J. M.
„Problems of Substructure Search and their Solution“
In: Chemical Structures - The international Language of Chemistry, Warr, W. A. (Hrsg.)
Springer-Verlag, Heidelberg (1989) 113-126
- [36] Roth, R.; Böhmer, H.-U.; Deplanque, R.
„Registration of substances in the Gmelin Factual Database“
Anal. Chim. Acta 265 (1992) 301-304
- [37] Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J.
„Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited“
J. Chem. Inf. Comput. Sci. 32 (1992) 244-255
- [38] Bebak, H.; Buse, C.; Donner, W.T.; Hoever, P.; Jacob, H.; Klaus, H.; Pesch, J.; Roemelt, J.; Schilling, P.; Woost, B.; Zirz, C.
„The Standard Molecular Data Format (SMD Format) as an integration tool in computer chemistry“
J. Chem. Inf. Comput. Sci. 29(1989) 1-5
- [39] Barnard, J. M.
„Draft Specification for Revised Version of the Standard Molecular Data (SMD) Format“
J. Chem. Inf. Comput. Sci. 30 (1990) 81-96

- [40] Rayner, J. D.
„A Concise Connection Table Based on Systematic Nomenclatural Terms“
J. Chem. Inf. Comput. Sci. 25 (1985) 108-111
- [41] Ugi, I.; Stein, N.; Knauer, M.; Gruber, B.; Bley, K.; Weidinger, R.
„New Elements in the Representation of the Logical Structure of Chemistry by Qualitative Mathematical Models and Corresponding Data Structures“
Top. Curr. Chem. 166 (1993) 199-233
- [42] Stein, N.
„New Perspectives in Computer-Assisted Formal Synthesis Design - Treatment of Delocalized Electrons“
J. Chem. Inf. Comput. Sci. 35 (1995) 305-309
- [43] Dietz, A.
„Yet Another Representation of Molecular Structure“
J. Chem. Inf. Comput. Sci. 35 (1995) 787-802
- [44] Clark, T.
„A Handbook of Computational Chemistry“
John Wiley & Sons (1985)
- [45] Weininger, D.
„SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules“
J. Chem. Inf. Comput. Sci. 28 (1988) 31-36
- [46] Die Definition von SMILES hat sich seit 1988 stark geändert. Die neueste Definition ist im World Wide Web unter folgender URL zu finden:
<http://www.daylight.com/dayhtml/smiles/index.html>
- [47] Greenwood, N. N.; Earnshaw, A.
„Chemie der Elemente“
VCH, Weinheim (1988)
- [48] Hay, P. J.
„Generalized Valence Bond Studies of the Electronic Structure of SF₂, SF₄, and SF₆“
J. Am. Chem. Soc. 99 (1977) 1003-1012
- [49] March, J.
„Advanced Organic Chemistry“
John Wiley and Sons, New York (1985) S. 272ff.

- [50] Mavridis, A.; Moustakali-Mavridis, I.
„A Reinvestigation of Tolane“
Acta Cryst. **B33** (1977) 3612-3615
- [51] Weiss, E.; Lucken, A. A. C.
„Die Kristall- und Elektronenstruktur des Methyllithiums“
J. Organometal. Chem. **2** (1964) 197-205
- [52] Switzer, M. E.; Wang, R.; Rettig, M. F.; Maki, A. H.
„On the Electronic Ground States of Manganocene and 1,1'-Dimethylmanganocene“
J. Am. Chem. Soc. **96** (1974) 7669-7674
- [53] Cotton, F. A.; Mague, J. T.
„The Existence of the Re_3Cl_9 Cluster in Anhydrous Rhenium(III) Chloride and Its Persistence in Solutions of Rhenium(III) chloride“
Inorg. Chem. **3** (1964) 1402-1407
- [54] Ihlenfeldt, W. D.
„Computer gestützte Syntheseplanung durch Erkennung synthetisch nutzbarer Ähnlichkeit von Molekülen“
Dissertation, Technische Universität München (1991)
- [55] Ihlenfeldt, W. D.; Gasteiger, J.
„Hash Codes for the Identification and Classification of Molecular Structure Elements“
J. Comp. Chem. **15** (1994) 793-813
- [56] Gasteiger, J.; Hutchings, M. G.
„Quantification of Effective Polarizability. Applications to Studies of X-Ray Photoelectron Spectroscopy and Alkylamine Protonation“
J. Chem. Soc., Perkin Trans. **2** (1984) 559-564
- [57] Hutchings, M. G.; Gasteiger, J.
„Residual Electronegativity - An Empirical Quantification of Polar Influences and its Application to the Proton Affinity of Amines“
Tetrahedron Lett. **24** (1983) 2541-2544
- [58] Gasteiger, J.; Saller, H.
„Berechnung der Ladungsverteilung in konjugierten Systemen durch eine Quantifizierung des Mesomeriekonzeptes“
Angew. Chem. **97** (1985) 699-701
- [59] Gasteiger, J.; Marsili, M.
„Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges“
Tetrahedron **36** (1980) 3219-3288

- [60] Rafeiner, K.; Gasteiger, J.
„Calculation of Heats of Formation for Organic Radicals and Cations“
In: Software Development in Chemistry 5, Gasteiger, J. (Hrsg.)
Springer-Verlag Heidelberg (1991) 69-75
- [61] Fröhlich, A.
„Modelle zur Beschreibung elektronischer Effekte in π -Systemen und ihr Einsatz in der Modellierung der elektrophilen aromatischen Substitution“
Dissertation, Technische Universität München (1993)
- [62] Elschenbroich, C.; Salzer, A.
„Organometallchemie“
Teubner, Stuttgart (1988) S. 271
- [63] Young, J. F.; Osborne, J. A.; Jardine, F. H.; Wilkinson, G.
„Hydride Intermediates in Homogenous Hydrogenation Reactions of Olefins and Acetylenes using Rhodium Catalysts“
Chem. Comm. 131 (1965) 131-132
- [64] Gasteiger, J.; Hondelmann, U.; Röse, P.; Witzenbichler, W.
„Computer-assisted Prediction of the Degradation of Chemicals: Hydrolysis of Amides and Phenylureas“
J. Chem. Soc. Perkin Trans. 2 (1995) 193-204
- [65] Randic, M.
„Enumeration of the Kekulé Structures in conjugated Hydrocarbons“
J. Chem. Soc. Faraday Trans. II (1976) 232-243
- [66] Dzonova-Jerman-Blazic, B.; Trinajstic, N.
„Computer-aided Enumeration and Generation of the Kekulé Structures in Conjugated Hydrocarbons“
Comput. Chem. 6 (1982) 121-132
- [67] Klein, O. J.; Randic, M.
„Innate Freedom of a Graph“
J. Comput. Chem. 8 (1987) 516-521
- [68] Randic M.; Trinajstic, N.
„Conjugation and Aromaticity of Corannulenes“
J. Am. Chem. Soc. 106 (1984) 4428-4434
- [69] Kearsley, S. K.
„A Quick Robust Method for Assigning a Kekulé Structure“
Comput. Chem. 17 (1993) 1-10

- [70] Hansen, P. Z.
„Assigning a Kekulé Structure to a Conjugated Molecule“
Comput. Chem. 19 (1995) 21-26
- [71] Specinfo, erhältlich von Chemical Concepts Weinheim.
- [72] Stroustrup, B.
„Die C++-Programmiersprache“
Addison-Wesley, Bonn (1992)
- [73] Booch, G.
„Object Oriented Design with Applications“
Benjamin/Cummings, Redwood City, CA (1991)
- [74] Martin, R. C.
„Designing Object-Oriented C++ Applications Using the Booch Method“
Prentice-Hall, Englewood Cliffs, N.J. (1995)
- [75] Das Gnu Projekt der Free Software Foundation
<http://www.gnu.ai.mit.edu>
- [76] Sorkau, E.
„Ringerkennung in chemischen Strukturen mit dem Computer“
Wiss. Z. Tech. Hochsch. Leuna-Merseburg 27 (1985) 765-770
- [77] Gasteiger, J.; Hutchings, M. G.; Saller, H.; Löw, P.
„Prediction of Chemical Reactivity and Design of Organic Synthesis“
In: *Chemical Structures - The International Language of Chemistry*, Warr, W. A. (Hrsg.)
Springer-Verlag, Heidelberg (1988) 344-359
- [78] Ousterhout, J. K.
„Tcl and the Tk Toolkit“
Addison-Wesley (1984)
- [79] Höllering, R.
„Das wissensbasierte Reaktionsvorhersagesystem EROS7.“
Manual (1997).
- [80] Fleming, Ian
„Grenzorbitale und Reaktionen organischer Verbindungen“
VCH Weinheim (1988)

12. Anhang

A Elektronen- und Orbitalkonfiguration für die Elemente im Grundzustand

| Element | n | Besetzung | | | | | | | | | |
|---------|----|-----------|---|---|---|---|---|---|---|---|---|
| H | 1 | 1 | | | | | | | | | |
| He | 1 | 2 | | | | | | | | | |
| Li | 4 | 1 | 0 | 0 | 0 | | | | | | |
| Be | 4 | 1 | 1 | 0 | 0 | | | | | | |
| B | 4 | 1 | 1 | 1 | 0 | | | | | | |
| C | 4 | 1 | 1 | 1 | 1 | | | | | | |
| N | 4 | 2 | 1 | 1 | 1 | | | | | | |
| O | 4 | 2 | 2 | 1 | 1 | | | | | | |
| F | 4 | 2 | 2 | 2 | 1 | | | | | | |
| Ne | 4 | 2 | 2 | 2 | 2 | | | | | | |
| Na | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Mg | 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Al | 9 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Si | 9 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| P | 9 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| S | 9 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| Cl | 9 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | |
| Ar | 9 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | |
| K | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Ca | 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| Sc | 11 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ti | 11 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | 11 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cr | 11 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Mn | 11 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Fe | 11 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |

| Element | n | Besetzung | | | | | | | | | | |
|---------|----|-----------|---|---|---|---|---|---|---|---|---|---|
| Co | 11 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Ni | 11 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Cu | 11 | 1 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| Zn | 11 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| Ga | 9 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| Ge | 9 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| As | 9 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| Se | 9 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| Br | 9 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| Kr | 9 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | | |
| Rb | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| Sr | 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| Y | 11 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Zr | 11 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Nb | 11 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mo | 11 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Tc | 11 | 1 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Ru | 11 | 1 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Rh | 11 | 1 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Pd | 11 | 0 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| Ag | 11 | 1 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| Cd | 11 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| In | 9 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| Sn | 9 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| Sb | 9 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| Te | 9 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| I | 9 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| Xe | 9 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | | |
| Cs | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |

| Element | n | Besetzung | | | | | | | | | | | |
|---------|----|-----------|---|---|---|---|---|---|---|---|---|---|---|
| Ba | 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| La | 11 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ce | 14 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pr | 14 | 2 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Nd | 14 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pm | 14 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Sm | 14 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Eu | 14 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Gd | 14 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Tb | 14 | 2 | 0 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Dy | 14 | 2 | 0 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Ho | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| Er | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 |
| Tm | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 |
| Yb | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| Lu | 14 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| Hf | 11 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ta | 11 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| W | 11 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Re | 11 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Os | 11 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ir | 11 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pt | 11 | 1 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Au | 11 | 1 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hg | 11 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tl | 9 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pb | 9 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bi | 9 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Po | 9 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Element | n | Besetzung | | | | | | | | | | | |
|---------|----|-----------|---|---|---|---|---|---|---|---|---|---|---|
| At | 9 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Rn | 9 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Fr | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ra | 9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ac | 11 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Th | 14 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pa | 14 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| U | 14 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Np | 14 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pu | 14 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Am | 14 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Cm | 14 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Bk | 14 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Cf | 14 | 2 | 0 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Es | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| Fm | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 |
| Md | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 |
| No | 14 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| Ln | 14 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |

B Valenztafel zur Erkennung ungültiger Strukturen

Die Tabelle enthält die Untergrenzen und Obergrenzen, die für die äquilibrierte Ladung eines Atoms während der Valenzüberprüfung in den verschiedenen Kategorien eingehalten werden

müssen. Dabei stehen O für organische, A für anorganische Verbindungen sowie O-MS für organische und AO-MS für anorganische Verbindungen in massenspektrometrischen Prozessen.

| Element | Untergrenze | | | | Obergrenze | | | |
|---------|-------------|----|------|------|------------|---|------|------|
| | O | A | O-MS | A-MS | O | A | O-MS | A-MS |
| H | 0 | -1 | 0 | -1 | 1 | 1 | 1 | 1 |
| He | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Li | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Be | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| B | 0 | -1 | 0 | -1 | 0 | 1 | 0 | 1 |
| C | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 |
| N | -1 | -3 | -1 | -3 | 1 | 1 | 1 | 1 |
| O | -1 | -2 | -1 | -2 | 1 | 1 | 1 | 1 |
| F | -1 | -1 | -1 | -1 | 0 | 1 | 0 | 1 |
| Ne | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Na | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Mg | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Al | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Si | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 |
| P | -1 | -3 | -1 | -3 | 1 | 1 | 1 | 1 |
| S | -1 | -2 | -1 | -2 | 1 | 1 | 1 | 1 |
| Cl | -1 | -1 | -1 | -1 | 1 | 4 | 1 | 4 |
| Ar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Ca | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Sc | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Ti | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| V | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 5 |
| Cr | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Mn | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 7 |

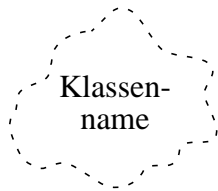
| Element | Untergrenze | | | | Obergrenze | | | |
|---------|-------------|----|------|------|------------|---|------|------|
| | O | A | O-MS | A-MS | O | A | O-MS | A-MS |
| Fe | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Co | 0 | -1 | 0 | -1 | 0 | 4 | 0 | 4 |
| Ni | 0 | -1 | 0 | -1 | 0 | 4 | 0 | 4 |
| Cu | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Zn | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Ga | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Ge | -1 | -1 | -1 | -1 | 0 | 4 | 0 | 4 |
| As | -1 | -3 | -1 | -3 | 0 | 5 | 0 | 5 |
| Se | -1 | -2 | -1 | -2 | 0 | 6 | 0 | 6 |
| Br | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 |
| Kr | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Rb | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Sr | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Y | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Zr | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Nb | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 5 |
| Mo | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Tc | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 7 |
| Ru | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 7 |
| Rh | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Pd | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Ag | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 5 |
| Cd | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| In | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Sn | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Sb | -1 | -1 | -1 | -1 | 0 | 5 | 0 | 5 |
| Te | -1 | -1 | -1 | -1 | 0 | 6 | 0 | 6 |

| Element | Untergrenze | | | | Obergrenze | | | |
|---------|-------------|----|------|------|------------|---|------|------|
| | O | A | O-MS | A-MS | O | A | O-MS | A-MS |
| I | -1 | -1 | -1 | -1 | 0 | 1 | 0 | 1 |
| Xe | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 8 |
| Cs | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Ba | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| La | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Ce | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Pr | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Nd | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Pm | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Sm | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Eu | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Gd | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Tb | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Dy | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Ho | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Er | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Tm | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Yb | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Lu | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Hf | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Ta | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 5 |
| W | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Re | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 7 |
| Os | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 8 |
| Ir | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Pt | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Au | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 5 |

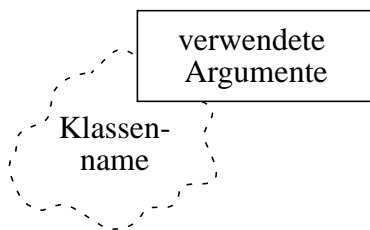
| Element | Untergrenze | | | | Obergrenze | | | |
|---------|-------------|----|------|------|------------|---|------|------|
| | O | A | O-MS | A-MS | O | A | O-MS | A-MS |
| Hg | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Tl | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Pb | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Bi | -1 | -1 | -1 | -1 | 0 | 5 | 0 | 5 |
| Po | -1 | -1 | -1 | -1 | 0 | 6 | 0 | 6 |
| At | -1 | -1 | -1 | -1 | 0 | 1 | 0 | 1 |
| Rn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Fr | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Ra | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| Ac | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Th | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Pa | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 5 |
| U | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Np | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 7 |
| Pu | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 7 |
| Am | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 6 |
| Cm | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Bk | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 |
| Cf | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Es | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Fm | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Md | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| No | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Ln | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| Ku | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Lp | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |

C Symbole, die in den Klassendiagrammen verwendet werden.

Klassendiagramme



einfache Klassen



Templateklassen, im Rechteck sind jeweils die Namen der verwendeten Klassen angegeben

Klassenbeziehungen

| | |
|------------|------------------------------|
| A ————— B | Assoziation zwischen A und B |
| A —————> B | Vererbung (B erbt von A) |
| A ●———— B | A enthält B |
| A ○———— B | A verwendet B |
| A ●————■ B | A enthält B als Wert |
| A ●————□ B | A enthält B als Referenz |

Eigenschaft einer Klassenbeziehung



Freund-Beziehung



statische Variable

Exportkontrolle

| | |
|---------------------|------------|
| keine Kennzeichnung | öffentlich |
| | privat |

13. Publikationsliste

([@] Teil der Dissertation)

- [1] Bauerschmidt, S.; Hanebeck, W.; Schulz, K.-P.; Gasteiger, J.
„Elucidation of Reactions in the Mass Spectrometer“
Anal. Chim. Acta, 265 (1992) 169-182.
- [2] Gasteiger, J.; Hanebeck, W.; Schulz, K.-P.; Bauerschmidt, S.; Höllering, R.
Automatic Analysis and Simulation of Mass Spectra.
In *Computer-Enhanced Analytical Spectroscopy*, Vol. 4,
C. L. Wilkins, (Hrsg.)
Plenum Press, New York, 1993, 97-133.
- [3] Schulz, K.-P.; Bauerschmidt, S.; Höllering, R.; Gasteiger, J.
Automatic Elucidation of Reactions in the Mass Spectrometer.
In *AIP conference proceedings 330*: E. C. C. C. 1
F. Bernardi, J. L. Rivail (Hrsgs.)
American Institute of Physics, Woodbury, NY, 1995; 726-733.
- [4][@] Bauerschmidt, S.; Gasteiger, J.
Overcoming the Limitations of a Connection Table Description:
A Universal Representation of Chemical Species.
J. Chem. Inf. Comput. Sci. akzeptiert

14. Lebenslauf

| | |
|------------------------|---|
| Vor- und Zuname: | Susanne Bauerschmidt |
| Geburtsdatum und -ort: | 16.06.65 in München |
| Eltern: | Bruno Bauerschmidt Elfriede Bauerschmidt, geb. Prädel |
| Staatsangehörigkeit: | deutsch |
| Familienstand: | ledig |
| Schulbildung: | 1971-75 Grundschule an der Wendelsteinstraße Vaterstetten 1975-84 Gymnasium Vaterstetten |
| Auslandsaufenthalt: | 1984-85 Au-Pair Aufenthalt in Edinburgh Stevenson College of Further Education Cambridge Proficiency Certificate |
| Studium: | November 1985 bis Dezember 1991 Studium der Chemie an der Technischen Universität München 1991 Diplomarbeit bei Prof. Dr. J. Gasteiger |
| Promotion: | Begonnen im Dezember 1991 bei Prof. Dr. J. Gasteiger an der TU München, im November 1993 Wechsel mit dem Doktorva- ter an die Universität Erlangen-Nürnberg |