Bielefeld University
Faculty of Technology

Graph Rewriting and NMR Spectroscopy in Chemical Reaction Network Modelling

[Graph Transformationen und NMR-Spektroskopie in Chemischen Reaktions Netzwerken]

Masterarbeit

4. November 2026

Kilian Seng
4330700

Erstprüfer: Prof. Dr. Daniel Merkle
Zweitprüfer: Dr. Marcel Friedrichs

Inhaltsverzeichnis

Abbildungsverzeichnis **IV**

1 Einleitung **2**
1.1 ILP und Graphtransformation . . . 2
1.2 NMR-Spektroskopie . . . 3
1.3 Zielsetzung . . . 3

2 Theoretische Grundlagen **4**
2.1 Graphen Theorie . . . 4
2.1.1 Das chemische Molekül als nicht einzigartig gelabelter Graph . . . 5
2.1.2 Morphismen . . . 5
2.1.2.1 Morphismus . . . 6
2.1.2.2 Monomorphismus . . . 6
2.1.2.3 Subgraph Isomorphismus . . . 6
2.1.2.4 Isomorphismus . . . 6
2.1.3 Double-Push-Out (DPO) . . . 7
2.1.3.1 Pushout . . . 8
2.1.3.2 Pullback . . . 8
2.1.3.3 Pushout-Komplement . . . 10
2.1.3.4 Double-Push-Out . . . 12
2.1.4 Gerichteter Multi-Hypergraph . . . 14
2.1.5 Zyklen . . . 15
2.1.6 Integer-Linear-Programming (ILP) . . . 15
2.1.7 Petrinetze . . . 17
2.2 NMR-Spektroskopie . . . 21
2.2.1 Spektren . . . 21
2.2.2 Kernspin . . . 21
2.2.3 NMR aktive Kerne . . . 22
2.2.4 NMR-Spektrometer . . . 23
2.2.5 Chemische Verschiebung . . . 23
2.2.6 Signalintensität . . . 25
2.2.7 Multiplettaufspaltung . . . 26

2.2.8 Spektrumanalyse . . . 27
2.2.9 Schreibweise . . . 28
2.2.10 Kohlenstoff NMR . . . 29
2.3 Dichte Funktional Theorie (DFT) . . . 29

3 Methoden und Tools **31**
3.1 Graptransformation Derivation Graph . . . 31
3.2 networkx . . . 33
3.3 ORCA . . . 34
3.4 NMR-Vergleich . . . 34
3.5 Gurobipy . . . 36
3.6 Petrinetzzertifikat . . . 36

4 Ergebnisse **37**
4.1 Ein-Molekül-Reaktionen . . . 37
4.1.1 Koffein . . . 37
4.1.2 Vanillin und 3,4-Dihydroxyaldehyd . . . 38
4.2 Graphtransformation: Ein-Molekül-Reaktionen . . . 39
4.2.1 Koffein . . . 39
4.2.2 Vanillin und 3,4-Dihydroxyaldehyd . . . 42
4.3 NMR: Ein-Molekül-Reaktionen . . . 45
4.3.1 Koffein . . . 45
4.3.2 Vanillin und 3,4-Dihydroxyaldehyd . . . 49
4.4 NMR *SI*: Ein-Molekül-Reaktionen . . . 53
4.4.1 Koffein . . . 53
4.4.2 Vanillin und 3,4-Dihydroxyaldehyd . . . 54
4.5 ILP: Ein-Molekül-Reaktionen . . . 55
4.5.0.1 Koffein . . . 56
4.5.1 Vanillin und 3,4-Dihydroxyaldehyd . . . 57
4.6 Fazit: Ein-Molekül-Reaktionen . . . 58
4.7 Multi-Molekül-Reaktionen . . . 59
4.8 Olefinmetathese und Diels-Alder . . . 59
4.8.1 Olefinmetathese . . . 59
4.8.2 Diels-Alder . . . 60
4.8.3 Butadien . . . 61
4.9 Graphtransformation: Multi-Molekül-Reaktionen . . . 61
4.9.1 Olefinmetathese und Diels-Alder . . . 61
4.10 NMR: Multi-Molekül-Reaktionen . . . 66

4.10.1 Olefinmetathese und Diels-Alder 66
4.11 NMR *SI*: Multi-Molekül-Reaktionen 66
4.11.1 Olefinmetathese und Diels-Alder 66
4.12 ILP: Multi-Molekül-Reaktionen 68
4.12.1 Olefinmetathese und Diels-Alder 68
4.12.1.1 Kantenwahrscheinlichkeit 68
4.12.1.2 Knoten Auswahl 70
4.13 Petrinetzzertifikat: Multi-Molekül-Reaktionen 72
4.13.1 Petrinetzzertifikat: Olefinmetathes und Diels-Alder 72
4.13.1.1 Produkt 72
4.13.1.2 Durchschnitt 73
4.13.1.3 Knoten Auswahl 73
4.14 Fazit: Multi-Molekül-Reaktionen 73

5 Zusammenfassung und Ausblick **75**

Quellenverzeichnis **77**

Abbildungsverzeichnis

2.1 Beispiele der Morphismen dargestellt als rote Pfeile. (a) Ein Morphismus, der kein Monomorphismus ist, (b) ein Monomorphismus, der kein Subgraph Isomorphismus ist, (c) ein Subgraph Isomorphismus, der kein Isomorphismus ist und (d) ein Isomorphismus [7]. 7

2.2 Beispiel für ein kommutatives Diagramm mit einem Pushout D. Die Knoten $A, B, C, D, D'' \in Obj(\mathbf{C})$ stellen selber Graphen dar. Die Kanten $g, g', g'', f, f', f'', d'' \in Mor(\mathbf{C})$ stellen Morphismen dieser Graphen dar. [7]. 9

2.3 Beispiel für ein kommutatives Diagramm mit einem Pullback A. Die Knoten $A, B, C, D, A'' \in Obj(\mathbf{C})$ stellen selber Graphen dar. Die Kanten $g, g', g'', f, f', f'', a'' \in Mor(\mathbf{C})$ stellen Morphismen dieser Graphen dar. [7]. 10

2.4 Beispiel für ein tatsächliches Pushout-Pullback-Konstrukt (oben-links), ein zu kleines Pushout (oben-rechts), ein zu großes Pushout (unten-links) und ein zu kleines Pullback (unten-rechts). 11

2.5 Ein Beispiel für den Double-Push-Out. Die Knoten K, L, R, D, G und H stellen Graphen dar. Die Kanten l, r, l', r', m, m' und d stellen Morphismen auf diesen Graphen dar. Bei einer Transformation von L zu R stellt L die Vorbedingung, R die Folgebedingung, G den Vorzustand und H den Folgezustand dar. Die Knoten K und D werden als Kontext bezeichnet [7]. 12

2.6 Wenn in einer Regel r ein Kantenlabel verändert wird, kann die Kante im Kontext nicht auftreten (a) oder als Kombination beider Kanten gelabelt werden (b). In Beispiel (a) wird die Kante mit Label a in der Transformation entfernt und eine neue Kante mit Label b hinzugefügt. In Beispiel (b) wird das Label einer Kante a zum Label b geändert. In Beispiel (a) wäre es möglich K über den Labelraum Ω zu definieren. In Beispiel (b) wird der Labelraum $\Omega \times \Omega$ nötig [7]. 14

2.7 Drei Beispiele für Hyperkanten in Hypergraphen. Die linke Hyperkante beschreibt den Übergang von den Graphen g_1 und g_2 zu g_3. Die mittlere Kante beschreibt den Übergang von dem Graph g_4 zu g_5. Bei einem einfachen Übergang wird eine einfache Kante verwendet. Die rechte Hyperkante beschreibt den Übergang von dem Graphen g_6 zu g_7. Es werden zwei Exemplare von g_6 benötigt, was durch den doppelten Pfeil dargestellt wird [7]. 14

2.8 Ein einfaches Petrinetz bestehend aus den Zuständen p_1, p_2, p_3, p_4, p_5, p_6, p_7, dargestellt als Kreise und den Übergängen t_1, t_2, t_3, t_4, t_5, t_6, dargestellt als Querbalken [14]. 18

2.9 Ein einfaches Petrinetz bestehend aus den Zuständen p_1, p_2, p_3, p_4, p_5, p_6, p_7, dargestellt als Kreise und den Übergängen t_1, t_2, t_3, t_4, t_5, t_6, dargestellt als Querbalken. Die Zustände p_1, p_5 und p_7 haben ein Token, weshalb nur Übergang t_2 ausgelöst werden kann [14]. 19

2.10 Ein einfaches Petrinetz bestehend aus den Zuständen p_1, p_2, p_3, p_4, p_5, p_6, p_7, dargestellt als Kreise und den Übergängen t_1, t_2, t_3, t_4, t_5, t_6, dargestellt als Querbalken. Die Zustände p_2, p_5, p_7 und p_3 haben ein Token, weshalb die Übergänge t_1, t_3 und t_5 ausgelöst werden können [14]. 20

2.11 Schematischer Aufbau eines NMR-Spektrometers [5]. 23

2.12 Symmetrieoperationen am Beispiel Chlorethan (A) und Cyclohexan (B). Bei Chlorethan kann eine Spiegelachse vertikal von Chlor (Cl) zum unteren Wasserstoff (grün) gezogen werden. Ebenso ist eine Rotationsachse zwischen den beiden Kohlenstoffen (Kreismitte) möglich. In Cyclohexan stehen sechs Wasserstoff orthogonal (axial) und sechs waagerecht (äquatorial) zum 6-Ring. Durch ein Ändern der Sesselkonfiguration, können beide ineinander überführt werden [5]. 25

2.13 ^{1}H-NMR von 1,1,2-Trichlorpropan. Der Methylwasserstoff (H_c rot) hat ein Dublett bei 1.68 ppm mit Intensität 3, der zentrale Wasserstoff (H_b grün) hat bei 4.35 ppm ein Dublett in Quartett mit Signalstärke 1 und der Endwasserstoff (H_a blau) erzeugt ein Dublett bei 5.96 ppm mit Signalstärke 1 [5]. 28

4.1 Synthese von Koffein in Pflanzen aus den Nukleotiden Adenin und Guanin. Erst werden die Purinringstrukturen oxidiert und nach der Abspaltung der Ribose die Stickstoffe methyliert. Die Methylierung erfolgt in der 7, 3 und zuletzt der 1 Position [24]. 38

4.2 Synthese von Vanilline in der Vanille Orchidee. Die Synthese erfolgt aus der Aminosäure Phenylalanin. Dabei wird zuerst die Aminogruppe eliminiert, anschließend der Phenylrest hydroxyliert und die aus der Eliminierung entstandene Doppelbindung oxidiert (rot). Der letzte Schritt ist die Methylierung des 4-Hydroxyrests [25]. 39
4.3 Darstellung von Xanthin in dem Graphtransformationstool. 40
4.4 Darstellung der Regel zum Methylieren eines protonierten Stickstoffs in dem Graphtransformationstool. 40
4.5 Darstellung aller Synthesepfade von Koffein aus Xanthin in dem Graphtransformationstool. Die Graphen $p_{0,0}$ bis $p_{0,5}$ stellen dabei die Zwischenprodukte 1-Methylxanthin, 3-Methylxanthin, 7-Methylxanthin, 1,3-Dimethylxanthin oder Theophylline, 1,7-Dimethylxanthin oder Paraxanthin und 3,7-Dimethylxanthin oder Theobromin dar. 41
4.6 Darstellung von Zimtsäure in dem Graphtransformationstool. 42
4.7 Darstellung der Regel zum Hydroxylieren des Phenylrests in dem Graphtransformationstool. 42
4.8 Darstellung der Regel zum oxidieren der Zimtsäure Doppelbindung zu einem Aldehyd in dem Graphtransformationstool. 43
4.9 Darstellung aller Synthesepfade von 3,4-Dihydroxybenzaldehyd aus Zimtsäure in dem Graphtransformationstool. Nur ein Teil aller Zwischenprodukte wird dabei tatsächlich zu dem gewünschten Produkt. 44
4.10 Katalysezyklus der Olefinmetathese [42]. 60
4.11 Die Diels-Alder-Reaktion [43]. 60
4.12 Butadien in dem Graphtransformationstool. 62
4.13 Olefinmetathese in dem Graphtransformationstool. 62
4.14 Diels-Alder-Reaktion in dem Graphtransformationstool. 62
4.15 Benzolsynthese in dem Graphtransformationstool. Die Syntheseschritte lauten erst Olefinmetathese von zwei Butadien zu Hexatrien und Ethylen, dann Olefinmetathese von Hexatrien und Butdien zu Ethylen und Octatetraen und zuletzt eine Ringschluss Olefinmetathese von Octatetraen zu Benzol und Ethylen. 65

Tabellenverzeichnis

2.1 Tabelle der chemischen Isotope, der in der organischen Chemie relevanten Elemente, mit natürlicher relativer Häufigkeit und ob sie NMR aktiv sind [5]. 22
2.2 Relative Intensitäten der einzelnen Peaks in den verschiedenen Multipletts, sowie die dafür nötige Anzahl chemisch-äquivalenter benachbarter Wasserstoffe [5]. 27

4.1 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Xanthin mit einem Literatur Spektrum [26]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.14 ppm. . 46
4.2 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 1-Methylxanthin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.39 ppm. 46
4.3 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 3-Methylxanthin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.35 ppm. 46
4.4 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 7-Methylxanthin mit einem Literatur Spektrum [28]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.73 ppm. 47
4.5 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Theophyllin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.61 ppm. 47
4.6 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Paraxanthin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.19 ppm. 47

4.7 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Theobromin mit einem Literatur Spektrum [29]. Die Methylsignale (etwa 30 ppm) wurden aus einem anderen Spektrum genommen, da diese Angabe fehlte [30]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.63 ppm. 48
4.8 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Koffein mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.77 ppm. . 48
4.9 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Zimtsäure mit einem Literatur Spektrum [31]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 4.02 ppm. . 50
4.10 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für m-Coumarsäure mit einem Literatur Spektrum [32]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.53 ppm. 50
4.11 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für p-Coumarsäure mit einem Literatur Spektrum [33]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 4.71 ppm. 50
4.12 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Benzaldehyd mit einem Literatur Spektrum [34]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 4.80 ppm. 51
4.13 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Kaffeesäure mit einem Literatur Spektrum [35]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.18 ppm. 51
4.14 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 3-Hydroxybenzaldehyd mit einem Literatur Spektrum [36]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.83 ppm. 51
4.15 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 4-Hydroxybenzaldehyd mit einem Literatur Spektrum [37]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.35 ppm. 52

4.16 Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 3,4-Dihydroxybenzaldehyd mit einem Literatur Spektrum [38]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt -3.81 ppm. 52
4.17 Resultierende normalisierte Wahrscheinlichkeiten beim Vergleich aller Zwischenprodukte der Koffeinsynthese, sowie Xanthin und Koffein mit den natürlichen Zwischenprodukten 7-Methyxanthin und Theobromin. 54
4.18 Resultierende normalisierte Wahrscheinlichkeiten beim Vergleich aller Zwischenprodukte der Vanillesynthese mit den natürlichen Zwischenprodukten p-Coumarsäure und 4-Hydroxybenzaldehyd. 55
4.19 Der für die Koffeinsynthese definierte Hypergraph in Gurobi mit den für jede Kante definierten Wahrscheinlichkeiten. 56
4.20 Lösung von Gurobi für die Koffeinsynthese. 57
4.21 Der für die Vanillinsynthese definierte Hypergraph in Gurobi mit den für jede Kante definierten Wahrscheinlichkeiten. 57
4.22 Lösung von Gurobi für die Vanillinsynthese. 58
4.23 Resultierende Wahrscheinlichkeiten für den Vergleich aller aus Butadien Entwickelten Moleküle mit Hexatrien und Ethylen, Octatetraen und Ethylen und Benzol und Etylen. 67
4.24 Lösung 1 für die Knotenauswahl beim ILP vom Butadien Beispiel mit $z_3 = 3$. 71
4.25 Lösung 2 für die Knotenauswahl beim ILP vom Butadien Beispiel mit $z_3 = 4$. 72
4.26 Lösung 10 für die Kantenwahrscheinlichkeit als Produkt beim ILP vom Butadien Beispiel mit $z = 197.08$. 72
4.27 Lösung 5 für die Kantenwahrscheinlichkeit als Durchschnitt beim ILP vom Butadien Beispiel mit $z = 798.54$. 73

Abkürzungsverzeichnis

ILP: Integer-Linear-Programming
DPO: Double-Push-Out
NMR: Nucleus-Magnetic-Resonance
MS: Massenspektroskopie
IR: Infrarot
UV/Vis: Ultraviolet und Visuelles Licht
DFT: Dichte-Funktional-Theorie
TMS: Tetramethylsilan
ppm: Parts per million
Mød: MedØlDatschgerl
GML: Game maker language
SMILES: Simplified Molecular Input Line Entry System
Exp: Experimentell

Abstract

Cheminformatic has been established as a valuable tool in the development and analysis of large chemical networks. Opening the possibility to tackle system sizes that would not be feasible for a human researcher. The possibilities both for the development of reactions as well as the the selection and verification of large reaction networks are presented in this thesis. This thesis aims to find sensible synthesis in chemical reaction networks using NMR-spectra and their characteristics.

1 Einleitung

Die Chemieinformatik ist ein aufstrebendes Forschungsfeld, das zum Ziel hat die Beschränkungen der chemischen Industrie mit der Rechenstärke der Informatik anzugehen. So sollen neue Reaktion und Moleküle entwickelt, bestehende Prozesse optimiert und Eigenschaften analysiert werden [1]. Die Chemieinformatik eignet sich dabei besonders in der Untersuchung großer Systeme , wie der Entwicklung neuer Medikamente [2]. Diese Arbeit thematisiert eine Möglichkeit chemische Reaktionen näher zu analysieren und auf Basis dessen einen Prozess zu optimieren. Dabei werden zueinander ähnliche Moleküle über ihre zugehörigen NMR-Spektren verglichen. Basierend auf diesen Gemeinsamkeiten und Unterschieden wird eine Synthese bestimmt.

1.1 ILP und Graphtransformation

Das Ziel bei der Synthese chemischer Moleküle ist die Maximierung der Ausbeute und die Minimierung des nötigen Aufwands. Bekannte Reaktionen können dabei auf eine Vielzahl von Startmolekülen angewandt werden. Aber auch für ein gewähltes Startmolekül ist sowohl in der Auswahl der verwendeten Reaktionen, als auch der gewählten Reihenfolge, große Varianz möglich. Um aus diesem Netzwerk an Reaktionen eine optimale Syntheseplanung zu entwickeln eignet sich das Integer-Linear-Programming (ILP). Dabei werden alle über die einzelnen Reaktionsschritte bekannten Informationen in einem Reaktionsnetzwerk zusammengefasst. Zu den geeigneten Informationen zählen bevorzugte Konfigurationen, Energiebarrieren oder Kosten für Katalysatoren und Reaktanten. All diese Informationen können mittels ILP gewichtet zu einem geeigneten Syntheseplan optimiert werden [3]. Es ist dabei möglich das Reaktionsnetzwerk ausschließlich aus bereits einmal durchgeführten Reaktionen, mit bekannten Edukten und Produkten zusammenzusetzen. In der Chemie werden Reaktionen aber meistens für Kategorien von Molekülen mit bestimmten Eigenschaften angegeben. Dabei kann das Vorliegen einer funktionellen Gruppe oder spezieller Bindungs- und Atomkonfigurationen entscheidend sein, ob eine Reaktion auf die Edukte anwendbar ist und welche Produkte zu erwarten sind. Diese Entscheidung kann bei kleinen Reaktionsnetzwerken händisch durchgeführt werden, steigt in ihrem Aufwand aber schnell an. Eine zeitsparende Alternative ist die Entwicklung des Reaktionsnetzwerks durch Graphtransformationen. Dabei werden Moleküle als Graphen und chemische Reaktionen als Graphentransformation abstrahiert [4]. So sind Reaktionsnetzwerke

mit mehr Reaktionstypen, sowie mehr und größere Moleküle möglich.

1.2 NMR-Spektroskopie

Die NMR-Spektroskopie ist eine der wichtigsten chemischen Werkzeuge zur Strukturaufklärung unbekannter Moleküle. Die Spektren sind dabei charakteristisch für eine Struktur und ermöglichen so eine Bestätigung des Erfolgs einer geplanten Synthese. Über die vorliegenden Zusammenhänge ist zudem auch eine weitgehende Analyse der Nachbarschaftsbeziehungen in unbekannten Molekülen möglich. Diese Analyse wird meistens mit weiteren Methoden kombiniert, wobei die verschiedenen NMR-Typen einen wesentlichen Anteil ausmachen [5]. Insbesondere in Artikeln über die Synthese neuer Moleküle wird so meistens das NMR-Spektrum angegeben, um eine externe Verifikation der Synthese durch unabhängige Chemiker zu ermöglichen. Auch bei der eigenen Synthese ist der Vergleich des NMR-Spektrums mit bekannten Spektren der Literatur eine zentrale Möglichkeit der Verifikation. Dieser Vergleich erfolgt meistens empirisch und ist zeitaufwändig, was die Automatisierung dieser Analyse nahelegt.

1.3 Zielsetzung

Das Ziel dieser Masterarbeit ist die Kombination des mathematischen Integer-Linear-Programming, der informatischen Struktur des Hypergraphen und der chemischen Methode des NMR-Spektrums. Dabei soll für bekannte Reaktionen und Moleküle ein chemisches Reaktionsnetzwerk in der Form eines Hypergraphen entwickelt werden. Für die so gefundenen Moleküle sollen NMR-Spektren bestimmt werden. Diese NMR-Spektren können dabei sowohl durch quantenchemische Programme berechnet werden, als auch aus den Literaturangaben zu experimentell bestimmten Spektren stammen. Diese NMR-Spektren sollen miteinander verglichen werden und ihre Ähnlichkeit, sowie ihre charakteristischen Unterschiede quantifiziert werden. Dieser Grad der Ähnlichkeit soll anschließend genutzt werden, um einen Reaktionspfad zu bestimmen, der diese Ähnlichkeit maximiert. Das angestrebte Ziel ist dabei die automatisierte der Verifizierung von Molekülen.

2 Theoretische Grundlagen

Für diese Arbeit werden sowohl Werkzeuge der Graphentheorie (Kapitel 2.1), als auch ein Verständnis der NMR-Spektroskopie (Kapitel 2.2) und der quantenchemischen Berechnung selbiger benötigt (Kapitel 2.2.10), welche im Folgenden näher erläutert werden.

2.1 Graphen Theorie

Die Graphen Theorie befasst sich mit dem Abbilden von Netzwerken und Entitätenbeziehungen. Dabei werden die Netzwerkteilnehmer und Entitäten als Menge von Knoten (V) und ihre Verbindungen oder Beziehungen als Menge von Kanten (E) abstrahiert. Ein Knoten v ist dabei ein Element der Menge V und eine Kante ein Element der Teilmenge $E \subseteq P_2(V)$. Das Tupel (u, v) aus zwei Knoten u und v, mit $u, v \in V$ ist ein Element der Menge $(u, v) \in E$. Ein einfacher Graph (G) besteht aus einer Menge von Knoten und Kanten $G = (V, E)$ [6]. Bei einem gerichteten Graph gilt dabei, dass die Kante (u, v) von Knoten u zu Knoten v ein geordnetes Paar darstellt. Es gilt $(u, v) \neq (v, u)$, weshalb beide in der Menge von Kanten vorkommen können. In einem ungerichteten Graphen gilt, dass die beiden Tupel dieselbe Kante beschreiben. Es gilt also $(u, v) = (v, u)$. Ein Multigraph ist darüber definiert, dass anders als in einem einfachen Graphen, eine identische Kante mehrmals enthalten ist. Die Kanten sind also keine Menge. Die beiden Kanten haben dann die selbe Repräsentation $(u, v) \stackrel{r}{=} (u, v)$, aber sind nicht die gleiche Kante $(u, v) \neq (u, v)$. Es besteht auch die Möglichkeit, dass eine Kante von einem Knoten zu sich selber führt, also (v, v), was als Schleife bezeichnet wird. Sowohl gerichtete, als auch ungerichtete Graphen können gelabelt sein. Sie werden dann als gelabelter Graph bezeichnet. Dabei lässt sich zwischen Kanten und Knotenlabeln unterscheiden. Für Kantenlabel (2.1) und Knotenlabel (2.2) in einem klassischen gelabelten Graphen gilt, dass jeder Knoten oder jede Kante ein Label aus der Menge der Label Ω zugeordnet haben muss.

$$\forall e \in E : \exists label(e) \in \Omega_E \tag{2.1}$$

$$\forall v \in V : \exists label(v) \in \Omega_V \tag{2.2}$$

2.1.1 Das chemische Molekül als nicht einzigartig gelabelter Graph

Ein chemisches Molekül bestehend aus kovalenten Bindungen kann als ein ungerichteter Graph $G = (V_G, E_G)$ betrachtet werden. Dabei können die Kanten nach den Bindungstypen: einfach, doppelt, dreifach, vierfach und aromatisch gelabelt werden. Die Knoten können nach dem Elementtyp und mit der Elementarladung gelabelt werden [7]. Die Label für Knoten und Kanten in einem Molekül müssen nicht einzigartig sein. Es ist möglich das Kantenlabel beispielsweise einer Doppelbindung durch zwei Einfachbindungen zwischen den selben Knoten zu ersetzen. Diese Darstellung kann aber keine aromatischen Bindungen repräsentieren. Etwaige sterische Informationen sind dabei mit einem Graphen nicht darstellbar. Eine chemische Reaktion kann dabei als eine Graphenoperation beschrieben werden, welche in Kapitel 2.1.3 näher beschrieben wird. Um zu überprüfen, ob die nötigen funktionellen Gruppen für eine Reaktion vorhanden sind, muss ein Monomorphismus gefunden werden. Um zu bestimmen, ob ein erzeugtes Molekül bereits gefunden wurde, muss ein Isomorphismus gefunden werden. Beide werden in Kapitel 2.1.2 definiert. Die Standard Darstellung von chemischen Molekülen als Graphen folgt dabei der Lewis-Repräsentation [8, 5].

- Oktettregel: In organischen Molekülen hat Wasserstoff 2 und alle anderen Hauptgruppenelemente 8 Elektronen.
- In kovalenten Bindungen werden 2 mal die Bindungsstärke Elektronen verwendet. Diese zählen für beide an der Bindung beteiligten Atomen.
- Einfachbindungen werden als Kanten zwischen den benachbarten Atomen gezeichnet. Eine Einfachbindung repräsentiert 2 Elektronen.
- Ein Atom kann maximal einen Bindungsgrad von der Anzahl seiner Valenzelektronen haben. Zum Beispiel hat Wasserstoff maximale eine Einfachbindung und Chlor bis zu 7.
- Aromatische Bindungen zählen als 1.5 Elektronen, treten aber meistens in gerader Anzahl auf.

2.1.2 Morphismen

Die Definition der Morphismen wurde aus der Doktorarbeit von Jakob Andersen entnommen [7]. Es seien $G = (V_G, E_G)$ und $H = (V_H, E_H)$, zwei einfache, ungerichtete Graphen.

2.1.2.1 Morphismus

Ein Graphmorphismus $m : G \to H$ ist dann eine Abbildung von G nach H. In dieser Abbildung müssen alle Knoten und Kanten in G auch eine Abbildung in H haben. Also gilt: Wenn $e = (u, v) \in E_G \implies \exists m(e) = (m(u), m(v)) \in E_H$. Die Abbildung muss nicht injektiv sein. Ein Beispiel für einen Morphismus, der kein Monomorphismus ist, ist wenn der Graph G aus zwei verbundenen Knoten und der Graph H aus einem Knoten mit einer Schleife besteht (Abbildung 2.1). Die Kante kann zwischen beiden Graphen injektiv abgebildet werden, aber beide Knoten aus Graph G werden auf den selben Knoten in Graph H abgebildet.

2.1.2.2 Monomorphismus

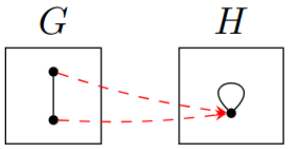
Der Monomorphismus ist ein Morphismus, bei dem alle Knoten in H nur von einem Knoten in G abgebildet werden. Die Abbildung muss Injektiv sein. Es gilt: $\forall u, v \in V_G, u \neq v \implies m(u) \neq m(v)$. Ein Beispiel für einen Monomorphismus, der aber kein Subgraph Isomorphismus ist, ist wenn der Graph G aus drei in Reihe verbundenen Knoten besteht und der Graph H aus drei Knoten besteht, die alle untereinander verbunden sind. Dadurch würde zwischen Knoten, die Teil der Abbildung sind eine Kante bestehen, die wiederum nicht abgebildet wird.

2.1.2.3 Subgraph Isomorphismus

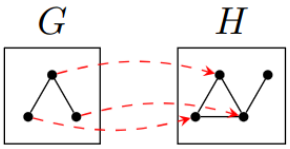
Der Subgraph Isomorphismus ist ein Monomorphismus, bei dem zwischen allen abgebildeten Knoten keine nicht abgebildeten Kanten bestehen. Es gilt: $(u, v) \in E_G \iff (m(u), m(v)) \in E_H$. Ein Beispiel für einen Subgraph Isomorphismus, der kein Isomorphismus ist, ist wenn ein Graph G aus drei linear verbundenen Knoten besteht und Graph H aus einem Graphen mit vier Knoten besteht,wobei drei linear verbunden sind und der letzte beliebig mit den drei anderen verbunden ist.

2.1.2.4 Isomorphismus

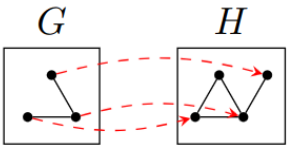
Der Isomorphismus ist ein Subgraph Isomorphismus, bei dem alle Knoten und Kanten jeweils genau einen Knoten und eine Kante haben auf den sie abgebildet werden können. Die Abbildung ist also bijektiv.
Um Morphismen in einem chemisch-gelabelten Graphen zu finden, müssen sie zu der richtigen Kantenzahl und Konformation, auch das richtige Label aufweisen. Wir definieren einen chemisch-gelabelten Graphen $G = (V_G, E_G, l_G^V, l_G^E)$ mit $l_G^V : V_G \to \Omega_V$ als eine Abbildung l_G^V der Knoten V_G auf die Element Label Ω_V und $l_G^E : E_G \to \Omega_E$ als eine Abbildung l_G^E der Kanten E_G auf die Bindungslabel $\Omega_V = \{-, =, :, \#\}$ als Einfach, Doppel,



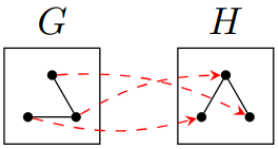
(a) A morphism.



(b) A monomorphism.



(c) A subgraph isomorphism.



(d) An isomorphism.

Abbildung 2.1: Beispiele der Morphismen dargestellt als rote Pfeile. (a) Ein Morphismus, der kein Monomorphismus ist, (b) ein Monomorphismus, der kein Subgraph Isomorphismus ist, (c) ein Subgraph Isomorphismus, der kein Isomorphismus ist und (d) ein Isomorphismus [7].

Aromatische und Dreifachbindung respektive. Die Abbildungen sind dabei surjektiv, da jeder Knoten und jede Kante gelabelt sind. Bei einem Morphismus $m : G \to H$ kommt die Label-Assoziation hinzu, welche einen Knoten und eine Kante jeweils einem Gegenstück in der Abbildung zuordnet (Gleichung 2.3, 2.4) [7].

$$A_V(m) = \langle l_G^V(v), l_H^V(m(v)) \rangle | v \in V_G \tag{2.3}$$

$$A_E(m) = \langle l_G^E(e), l_H^E(m(e)) \rangle | e \in E_G \tag{2.4}$$

Da die Label als einfache Strings definiert sind gilt: $s_1 = s_2, \forall (s_1, s_2) \in A_V(m) \cup A_E(m)$ [7]. Für jedes assoziierte Paar in der Knoten und Kantenmenge müssen beide Label gleich sein.

2.1.3 Double-Push-Out (DPO)

Bisher wurde der chemische Graph statisch betrachtet. Um chemische Reaktionen zu definieren, müssen Regeln für das Umschreiben des Graphen definieren werden. Die verwendete Methode ist der Double-Push-Out-Ansatz. Zuerst wird für die Graphen eine Kategorie definiert. Die Kategorie $\mathbf{C}$ besteht aus:

- einer Klasse von Objekten $Obj(\mathbf{C})$
- einer Klasse von Morphismen $Mor(\mathbf{C})$, der zwei Elemente $A, B \in Obj(\mathbf{C})$ aufeinander Abbildet $m : A \to B$

- einem Morphismus-Kompositions-Operator $\circ : Mor(\mathbf{C}) \times Mor(\mathbf{C}) \to Mor(\mathbf{C})$ für den gilt:
  - die Komposition ist abgeschlossen also: $\forall m, n \in Mor(\mathbf{C})$ mit $m : A \to B, n : B \to C, \exists n \circ m \in Mor(\mathbf{C})$ mit $n \circ m : A \to C$
  - es existiert ein Identitätsmorphismus $\mathrm{id}_A \in Mor(\mathbf{C}$ mit $\mathrm{id}_A : A \to A$.
  - die Komposition ist assoziativ also: $\forall m, n, g \in Mor(\mathbf{C})$ mit $m : A \to B, n : B \to C$, $g : C \to D$ gilt $(n \circ m) \circ g = n \circ (m \circ g)$
- zwei Morphismen können den selben Start und Endgraphen haben

Auf dieser Kategorie von Graphen lassen sich drei unterliegende Graphentypen definieren.

2.1.3.1 Pushout

Bei zwei gegebenen Morphismen $g : A \to C$ und $f : A \to B$ erzeugen die Morphismen $g' : B \to D$ und $f' : C \to D$ genau dann ein Pushout D von f und g (Abbildung 2.2), wenn:

- die Morphismen kommutativ sind: $g' \circ f = f' \circ g$
- für alle Morphismen Paare $f'' : C \to D''$ und $g'' : B \to D''$ mit $f'' \circ g = g'' \circ f$ genau ein einzigartiger Morphismus $d'' : D \to D''$ existiert, sodass $f'' = d'' \circ f'$ und $g'' = d'' \circ g'$.

2.1.3.2 Pullback

Bei zwei gegebenen Morphismen $g' : B \to D$ und $f' : C \to D$ erzeugen die Morphismen $g : A \to C$ und $f : A \to B$ genau dann ein Pullback A von f' und g' (Abbildung 2.3), wenn:

- die Morphismen kommutativ sind: $g' \circ f = f' \circ g$
- für alle Morphismen Paare $f'' : A'' \to B$ und $g'' : A'' \to C$ mit $f' \circ g'' = g'' \circ f''$ genau ein einzigartiger Morphismus $a'' : A'' \to A$ existiert, sodass $f'' = f \circ a''$ und $g'' = g \circ a''$.

Im Folgenden werden zum Pushout und Pullback vier Beispiele näher untersucht (Abbildung 2.4). Das oben-linke Beispiel stellt ein korrektes Pushout-Pullback-Konstrukt dar. Bei jedem Morphismus wird genau eine Kante und ein zu dieser Kante gehöriger Knoten eingefügt. Im oben-rechten Beispiel wird für der Pushout nur insgesamt eine Kante

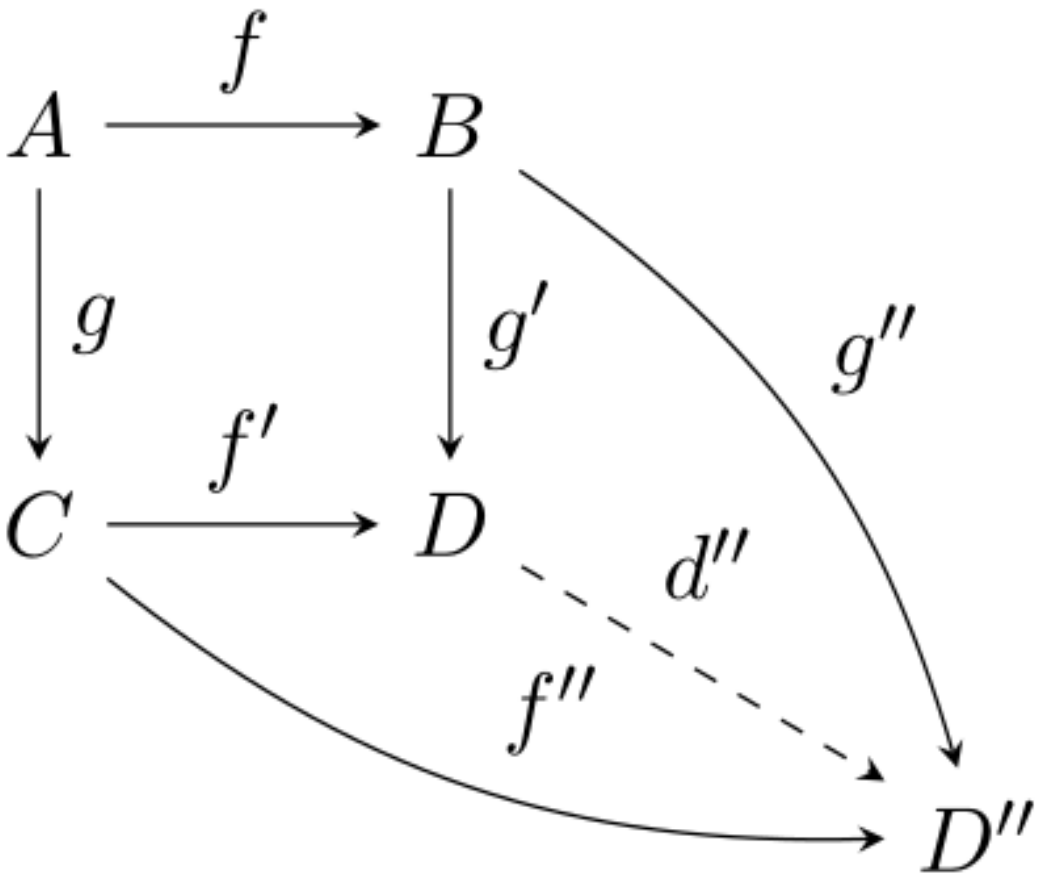


Abbildung 2.2: Beispiel für ein kommutatives Diagramm mit einem Pushout D. Die Knoten $A, B, C, D, D'' \in Obj(\mathbf{C})$ stellen selber Graphen dar. Die Kanten $g, g', g'', f, f', f'', d'' \in Mor(\mathbf{C})$ stellen Morphismen dieser Graphen dar. [7].

hinzugefügt. Die beiden zuvor erzeugten Kanten werden also kombiniert, weshalb zwei unterschiedliche Morphismusmen $d'' : D \to D''$ mit D" als tatsächliches Pushout existieren würden. Bei dem unten-linken Beispiel wird für den Pushout-Kandidat eine Kante zu viel eingefügt, wodurch der Pushout-Kandidat ein Dreistern-Graphen ist. Dadurch existieren zwei verschiedene Morphismen (grün und blau) zu einem weiteren Vierstern-Graphen als Pushout-Kandidat. Ebenso existiert kein Morphismus von dem Dreistern-Graphen zu dem tatsächlichen Pushout. Der Pushout ist also zu groß. In dem Beispiel unten.rechts wird für den Pullback-Kandidat angenommen, dass wenn die Kante entfernt wird, auch die zugehörigen Knoten entfernt werden. Dadurch ist kein Morphismus von dem tatsächlichen Pullback zu diesem Pullback-Kandidat möglich. Ebenso würden, wenn jeweils zwei Knoten mit einer Kante mit g und f hinzugefügt würden, diese im Pushout-Kandidat kombiniert werden. Beides widerspricht den etablierten Regeln.

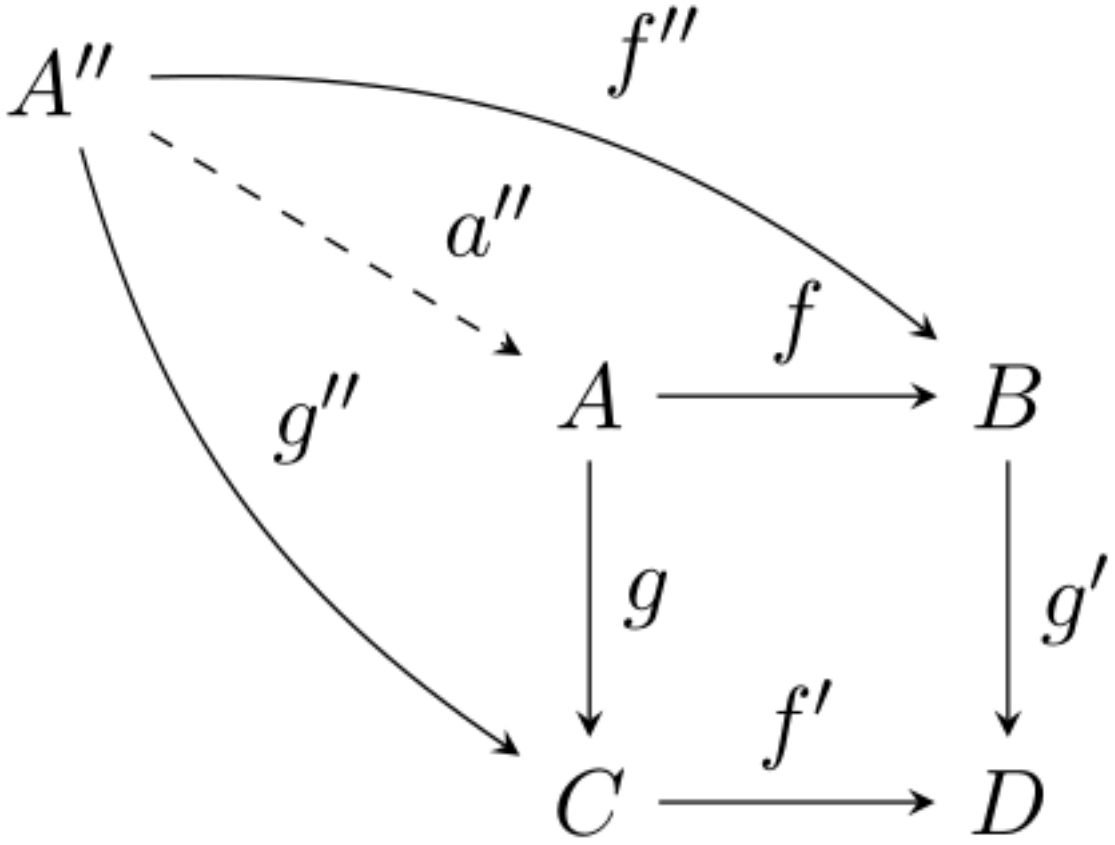


Abbildung 2.3: Beispiel für ein kommutatives Diagramm mit einem Pullback A. Die Knoten $A, B, C, D, A'' \in Obj(\mathbf{C})$ stellen selber Graphen dar. Die Kanten $g, g', g'', f, f', f'', a'' \in Mor(\mathbf{C})$ stellen Morphismen dieser Graphen dar. [7].

2.1.3.3 Pushout-Komplement

Bei zwei gegebenen Morphismen $g' : B \to D$ und $f : A \to B$ erzeugen die Morphismen $g : A \to C$ und $f' : C \to D$ genau dann ein Pushout-Komplement C von f und g', wenn f' und g' der Pushout von f und g ist oder f und g der Pullback von f' und g' ist.

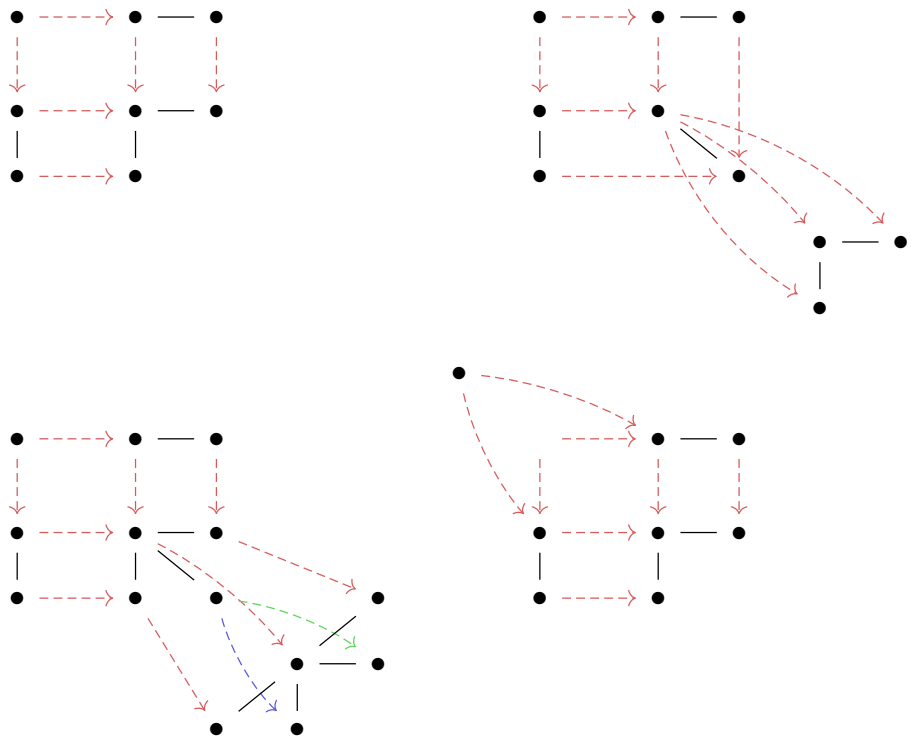


Abbildung 2.4: Beispiel für ein tatsächliches Pushout-Pullback-Konstrukt (oben-links), ein zu kleines Pushout (oben-rechts), ein zu großes Pushout (unten-links) und ein zu kleines Pullback (unten-rechts).

2.1.3.4 Double-Push-Out

Eine Graphtransformation lässt sich mit dem Double-Push-Out (DPO) Ansatz realisieren. Der DPO ist dabei eine Kombination aus den Morphismen l, r, l', r', m, m' (Abbildung 2.5) und einer Transformationsregel $p = (L \xleftarrow{l} K \xrightarrow{r} R)$. Die Regel besteht aus dem Kontext K, den zwei Morphismen l und r und den Bedingungen L und R. Bei einer angenommen Transformation von links nach rechts stellt L die Vorbedingung, R die Folgebedingung, G den Vorzustand und H den Folgezustand dar. Die Graphransformation eines Graphen G zu einem Graphen H mit der Regel p und dem Morphismus m wird über den Ausdruck $G \overset{p,m}{\Longrightarrow} H$ dargestellt. Dafür sind die folgenden Schritte nötig.

- Finde einen Morphismus $m : L \to G$ von der Vorbedingung L zum Startgraphen G.
- Bestimme aus den Morphismen $m : L \to G$ und $l : K \to L$ das Pushout-Komplement D mit den Morphismen $l' : D \to G$ und $d : K \to D$.
- Bestimme aus den Morphismen $r : K \to R$ und $d : K \to D$ den Pushout H mit den Morphismen $r' : D \to H$ und $m' : R \to H$.

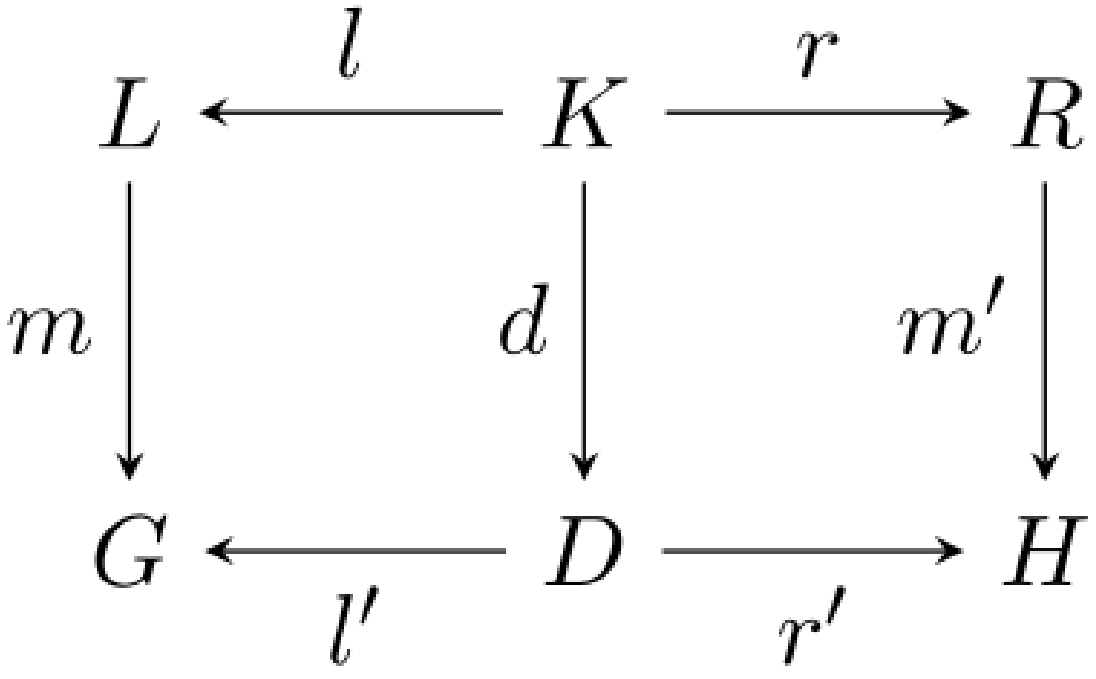


Abbildung 2.5: Ein Beispiel für den Double-Push-Out. Die Knoten K, L, R, D, G und H stellen Graphen dar. Die Kanten l, r, l', r', m, m' und d stellen Morphismen auf diesen Graphen dar. Bei einer Transformation von L zu R stellt L die Vorbedingung, R die Folgebedingung, G den Vorzustand und H den Folgezustand dar. Die Knoten K und D werden als Kontext bezeichnet [7].

Die rechte Seite R und die linke Seite L sind dabei austauschbar. Wenn die Graphentransformation $G \overset{p,m}{\Longrightarrow} H$ in eine Richtung abzubilden ist, existiert auch eine inverse Graphtransformation $H \overset{p^{-1},m'}{\Longrightarrow} G$, welche die inverse Regel $p^{-1} = (R \overset{r}{\leftarrow} K \overset{l}{\rightarrow} L)$ verwendet. Die Existenz des Morphismus m' ist schon für die ursprüngliche Graphtransformation gegeben. Im chemischen Kontext lassen sich die meisten Reaktionen als Gleichgewichtsreaktionen definieren, bei welchen sowohl Hin- als auch Rückreaktion simultan stattfinden [9]. Für einen gegebenen Morphismus $m : L \to G$ kann also D bestimmt werden, wenn:

- es keine Kanten in $E_G \setminus m(E_L)$ die an einen Knoten in $m(V_L \setminus l(V_K))$ gebunden sind, also es keine Kanten in $E_H \setminus m(E_R)$ gibt, die an einen Knoten in $m(V_R \setminus r(V_K))$ gebunden sind. Also gilt für alle Knoten die nicht in einer der beiden Regelseiten, aber nicht im Kontext sind, also nach der Transformation nicht mehr vorhanden sind, nicht mit Kanten verbunden sein können, die Teil der Vorbedingung waren. Wenn p also einen Knoten entfernt, müssen auch von ihm ausgehenden Kanten entfernt werden.

- es keine distinkten Knoten $u, v \in V_L$ gibt, mit $m(u) = m(v)$ und $u \in l(V_k), v \notin l(V_k)$ und es keine distinkten Kanten $e, f \in E_L$ gibt, mit $m(e) = m(f)$ und $e \in l(E_k), f \notin l(E_k)$. Wenn also eine Kante oder ein Knoten in der Vorbedingung ist, aber durch die Transformation entfernt wird, so kann dieser nicht durch den Morphismus m dem gleichen Knoten zugeordnet werden, wie ein Knoten oder eine Kante die nicht entfernt wurde.

Im Folgenden werden die Morphismen l und r als Monomorphismen festgelegt, da so nur eine Möglichkeit für D besteht. Dafür würde es reichen, dass einer der beiden Morphismen ein Monomorphismus ist. Die Beschränkung beider Morphismen ist nötig, damit die Invertierbarkeit gewahrt bleibt. Ebenso werden die Morphismen m und m' als Monomorphismen festgelegt, damit keine Knoten fusioniert werden können [7]. Für den DPO-Ansatz bei einer Transformation an einem chemisch-gelabelten Graphen wird eine Menge von Labeln Ω angenommen. Die Graphen L und R sind dabei mit Ω gelabelt, während der Kontext K beide Label vereint und somit mit $\Omega \times \Omega$ gelabelt ist. Es wäre möglich das Umbenennen von Labeln zu verhindern wie in Abbildung 2.6 Beispiel (a) zusehen, aber diese Definition erlaubt das Umbenennen nach Beispiel (b) für eine effizientere Darstellung. Wird dies auf den chemisch-gelabelten Graphen angewendet, so stellen die Knotenlabel weiterhin die chemischen Elemente mit etwaigen Ladungen dar, jedoch ist für die Kantenlabel neben den chemischen Bindungen noch ein **nil** Label nötig, um das Entfernen einer Bindung zu beschreiben. Bei einer Graphtransformation müssen weiterhin die Regeln für einen chemisch-gelabelten Graphen gewahrt werden.

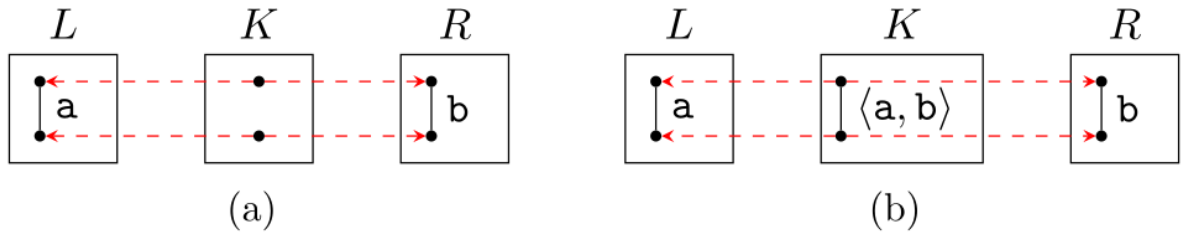


Abbildung 2.6: Wenn in einer Regel r ein Kantenlabel verändert wird, kann die Kante im Kontext nicht auftreten (a) oder als Kombination beider Kanten gelabelt werden (b). In Beispiel (a) wird die Kante mit Label a in der Transformation entfernt und eine neue Kante mit Label b hinzugefügt. In Beispiel (b) wird das Label einer Kante a zum Label b geändert. In Beispiel (a) wäre es möglich K über den Labelraum Ω zu definieren. In Beispiel (b) wird der Labelraum $\Omega \times \Omega$ nötig [7].

2.1.4 Gerichteter Multi-Hypergraph

Ein gerichteter Multi-Hypergraph $\mathcal{H} = (V, E)$ ist ein Tupel. Dabei ist V eine Menge von Knoten und E eine Menge von gerichteten Hyperkanten. Eine Hyperkante $e \in E$ ist selber ein Tupel $e = (U, W)$ aus einer Multimenge von Startknoten (tail) U und einer Multimenge von Zielknoten (head) W. Beide Multimengen enthalten ausschließlich Knoten und sind daher Teilmengen der Menge von Knoten $U, W \subseteq V$ [4]. Da es sich bei den Startkonten und Zielknoten um Multimengen handelt kann eine Kante den gleichen Knoten mehrmals in head und tail haben (Abbildung 2.7). Ein s,t-Superpfad in einem

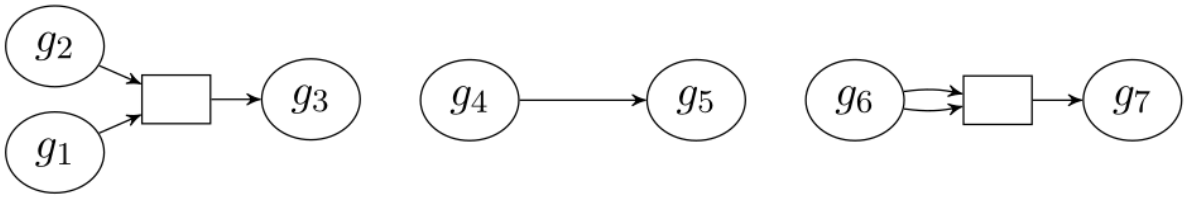


Abbildung 2.7: Drei Beispiele für Hyperkanten in Hypergraphen. Die linke Hyperkante beschreibt den Übergang von den Graphen g_1 und g_2 zu g_3. Die mittlere Kante beschreibt den Übergang von dem Graph g_4 zu g_5. Bei einem einfachen Übergang wird eine einfache Kante verwendet. Die rechte Hyperkante beschreibt den Übergang von dem Graphen g_6 zu g_7. Es werden zwei Exemplare von g_6 benötigt, was durch den doppelten Pfeil dargestellt wird [7].

Hypergraphen $\mathcal{H} = (V, E)$ für die unterscheidbaren Knoten $s, t \in V$ ist definiert als eine Kantenuntermenge $F \subseteq E$, sodass für die Kanten geordnet als e_1 bis e_k gilt [10]:

- $tail(e_1) = s$
- $\forall 1 < i \leq k : tail(e_i) \subseteq s \bigcup_{1 \leq j < i} head(e_j)$

- $head(e_k) = t$

Ein Hyperpfad P ist ein minimaler Superpfad. Also gilt: $\forall e \in P : P \setminus e$ ist kein Superpfad.

2.1.5 Zyklen

Für einen endlichen ungerichteten Graphen $G = (V, E)$ ist ein einfacher Pfad definiert als endliche Sequenz $V_P = \langle v_1, ..., v_k \rangle$ mit $v_1, ..., v_k \in V$ und $E_P = \{(v_1, v_2), ..., (v_{k-1}, v_k)\}$ mit $(v_i, v_{i+1}) \in E$ mit $i \in 1, ..., k-1$. Ein solcher Pfad ist ein Zyklus, wenn gilt $(v_k, v_1) \in E$. Eine Sehne in einem Pfad ist eine Kante $e_c \in E$ zwischen zwei Knoten des Pfades $e_i, e_j \in V_P$, welche wiederum nicht Teil des Pfades ist $(e_i, e_j) \notin E_P$. Ein sehnenloser Zyklus ist ein sehnenloser Pfad, bei dem der erste und letzte Knoten über eine Kante verbunden sind [11].

2.1.6 Integer-Linear-Programming (ILP)

Ein Lineares-Programmier (LP) Problem ist ein mathematisches Problem, bei dem für eine Menge an Variablen $\{x_1, ..., x_n\}$ Werte gefunden werden, so dass eine Zielfunktion maximiert oder minimiert wird. Ein LP kann mathematisch als eine Zielfunktion und eine Menge von Bedingungen beschrieben werden.

- $max\ z = \sum_j c_j \times x_j$
- Bedingung 1: $\sum_j a_{i,j} \times x_j \geq b_i$
- Bedingung 2: $0 \leq x_j$

Dabei sind alle Werte außer einer Variable x_j Konstanten. Bei einem Integer-LP Problem ist mindestens eine der Komponenten, meistens die Variable auf Integer also ganzzahlige Werte beschränkt. Es ist möglich, ein Maximierungsproblem in ein Minimierungsproblem zu überführen, indem die Zielfunktion mit einem negativen Wert multipliziert wird. Ebenso können mehrere Zielfunktionen kombiniert werden, indem sie addiert oder subtrahiert werden. Die Kombination aus Multiplikation mit einem großen Faktor und Addition erlaubt es eine primäre Zielfunktion mit einer sekundären oder tertiären Zielfunktion zu kombinieren. Um das ILP zu lösen, werden alle Variablen und Konstanten auf die gleiche Seite der Gleichung gebracht und Ungleichungen durch positive Konstanten s ausgedrückt.

- $max\ z - \sum_j c_j \times x_j = 0$
- Bedingung 1: $\sum_j a_{i,j} \times x_j + s_i = b_i$
- Bedingung 2: $0 \leq x_j$

- Bedingung 3: $0 \leq s_i$

Diese Menge an Gleichungen kann in eine Matrix $\mathbf{M}$ überführt werden, bei der die Zeilen die Gleichungen und die Spalten die Variablen x_j, si darstellen (Gleichung 2.5).

$$\begin{bmatrix} z & -c_1 & ... & -c_j & 0 & ... & 0 & |0 \\ 0 & -a_{1,1} & ... & -a_{1,j} & s_1 & ... & 0 & |b_1 \\ 0 & ... & ... & ... & ... & ... & 0 & |... \\ 0 & -a_{i,1} & ... & -a_{i,j} & 0 & ... & s_i & |b_i \end{bmatrix} \times \begin{bmatrix} x_1 \\ ... \\ x_j \\ s_1 \\ ... \\ s_i \end{bmatrix} \tag{2.5}$$

Das lineare Gleichungssystem kann durch das Invertieren der Matrix gelöst werden (Gleichung 2.6):

$$\mathbf{A} \times x = b \iff \mathbf{A}^{-1} \times b = x \tag{2.6}$$

Die daraus resultierende Lösung ist nicht unbedingt ganzzahlig. Um eine ganzzahlige, optimale Lösung zu finden, wird ein Baum mit der über die Matrixinversion erhaltenen Lösung als Wurzel aufgespannt. Durch eine Abfolge von Schritten kann eine ganzzahlige Lösung ermittelt werden.

- Bestimme $\mathbf{A}^{-1}$ aus $\mathbf{A}$ und b.
- Bestimme x aus $\mathbf{A}^{-1}$ und b.
- Sind alle Variablen in x nur Integer?
- Wähle eine Variable x_i oder s_i aus der Menge von x und s aus.
- Teile in zwei neue lineare Gleichungssysteme auf, für welche erneut eine Lösung gesucht wird.
  - Größer: Füge $x_i \geq ceiling(x_i)$ als neue Bedingung hinzu.
  - Kleiner: Füge $x_i \leq floor(x_i)$ als neue Bedingung hinzu.
- Ein Pfad muss nicht weiter verfolgt werden, wenn:
  - Unmöglich: Es gibt keine optimale Lösung mit der Bedingung.
  - Optimal: Der neue Knoten ist ganzzahlig.
  - Begrenzung: Die Parameterbegrenzung liegt außerhalb der Anfangsbedingungen.

Die Entscheidung, welche Variable als nächstes eingeschränkt wird, geht nach der folgenden Reihenfolge.

- Am nächsten an .5
- Stärkster Faktor in der Zielfunktion
- niedrigster Index

Um zu entscheiden, welcher Knoten als nächstes untersucht wird, können verschiedene Baumsuchalgorithmen verwendet werden [12].

2.1.7 Petrinetze

Ein Petrinetz ist ein Werkzeug zur Untersuchung der Erfüllbarkeit parallel ablaufender Prozesse [13]. Es besteht aus Knoten und Kanten. Die Knoten lassen sich dabei in Zustände p und Übergänge t unterteilen. Die Zustände werden dabei als Kreise und die Übergänge als Querbalken dargestellt (Abbildung 2.8).
In einem Petrinetz kann auf jedem Zustand ein Token platziert werden, um einen Gesamtzustand des Systems zu repräsentieren. Ein Übergang kann ausgelöst werden, wenn alle Zustände, die auf ihn zeigen, mindestens einen Token beinhalten. In dem Beispielgraphen (Abbildung 2.9) kann nur der Übergang t_2 auslösen, da ein Token in p_1 ist. Bei den Übergängen t_3 und t_5 liegt nur ein Token in p_5 und p_7 vor. Damit die Übergänge auslösen können, muss ein Token in p_3 vorliegen.
Nachdem Übergang t_2 ausgelöst hat, liegt ein neuer Token in p_2 und p_3 (Abbildung 2.10). Der Token in p_3 ist nötig, damit die Übergänge t_3 und t_5 ausgelöst werden können. Ebenso kann ein erneutes sequenzielles auslösen von t_1 und t_2 eine unbegrenzte Anzahl weiterer Token in p_3 erzeugen. Dieser Teil des Petrinetzes ist also ein Erzeuger. Die erzeugten Token in p_3 können entweder in t_3 und t_4 oder t_5 und t_6 verbraucht werden. Die Übergangsequenzen sind also im Konflikt zueinander. Die Feststellung, welche sequenziellen Übergänge und Gesamtzustände aus einem initialen Marking oder der initialen Tokenverteilung μ möglich sind, wird *Reachability* genannt. Ein Zustand μ' ist direkt erreichbar, wenn er durch das Auslösen eines möglichen Übergangs erzeugt wird. Ein Zustand μ ist allgemein erreichbar, wenn er durch einen Zustand direkt erreichbar ist, welcher selber allgemein oder direkt erreichbar ist. Eine *Reachability* Menge $R(M)$ kann für ein Petrinetz $M = (P, T, I, O, \mu)$ mit der Zustandsmenge P, der Übergangsmenge T, der Menge von Input und Output Funktionen für Übergänge I und O und dem initialen Marking μ definiert werden, als die Menge aller Markings, die durch das markierte Petrinetz erreicht werden können [14]. Mit Blick auf den zuvor definierten Hypergraphen (Kapitel 2.1.4) können die Zustände in einem Petrinetz als Knoten in einem Hypergraphen und die Übergänge als Hyperkanten verstanden werden. Dann erlaubt es der Einsatz von Petrinetzten die Erfüllbarkeit einer Abfolge von Hyperkantenübergängen zu untersuchen. Insbesondere bei

Reaktionsnetzwerken kann so durch ein initiales Marking der Startmoleküle oder eine erzeugende Kante ohne Input gewährleisten, dass nur Reaktionen stattfinden, für die alle nötigen Edukte vorliegen.

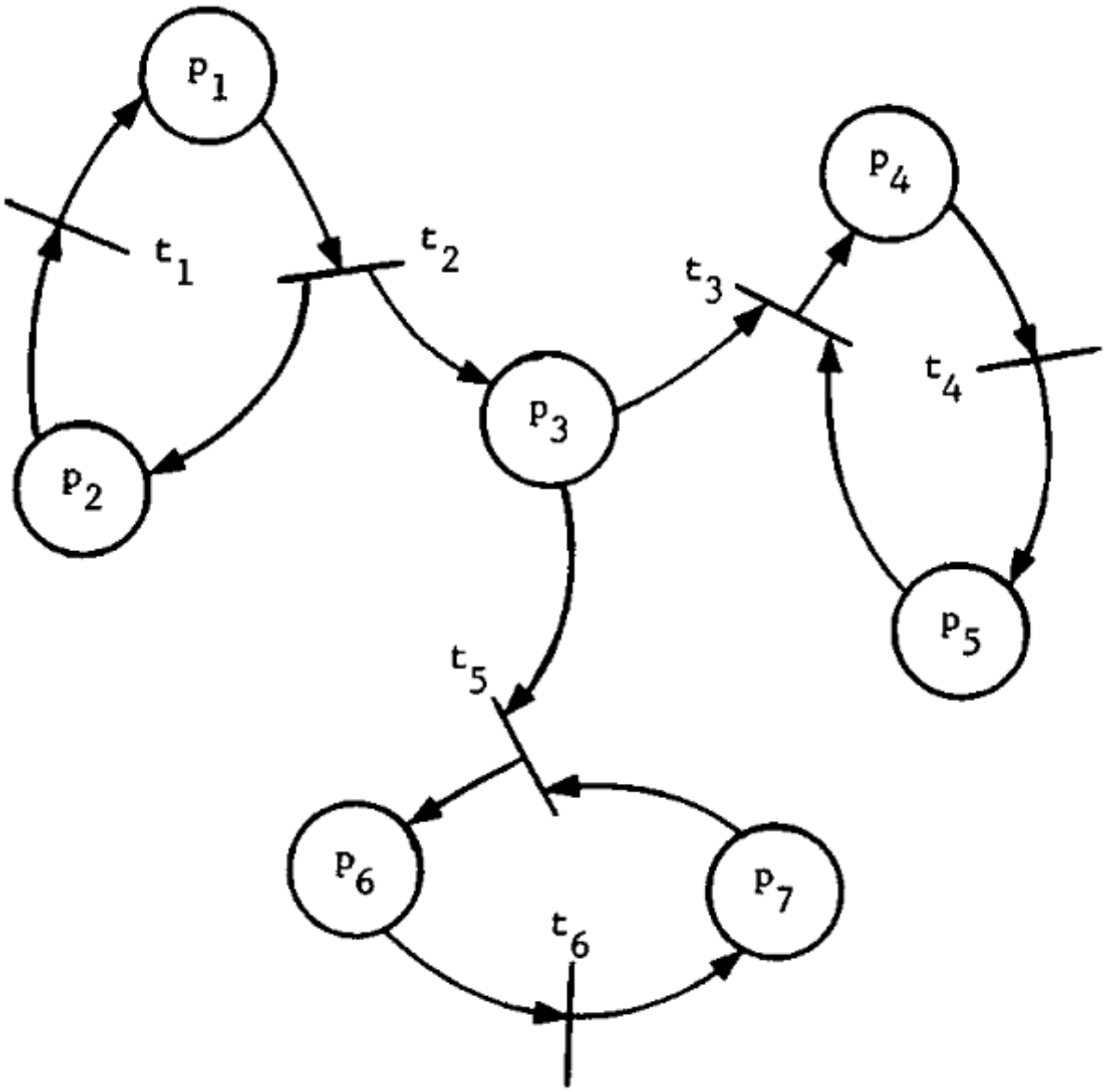


Abbildung 2.8: Ein einfaches Petrinetz bestehend aus den Zuständen p_1, p_2, p_3, p_4, p_5, p_6, p_7, dargestellt als Kreise und den Übergängen t_1, t_2, t_3, t_4, t_5, t_6, dargestellt als Querbalken [14].

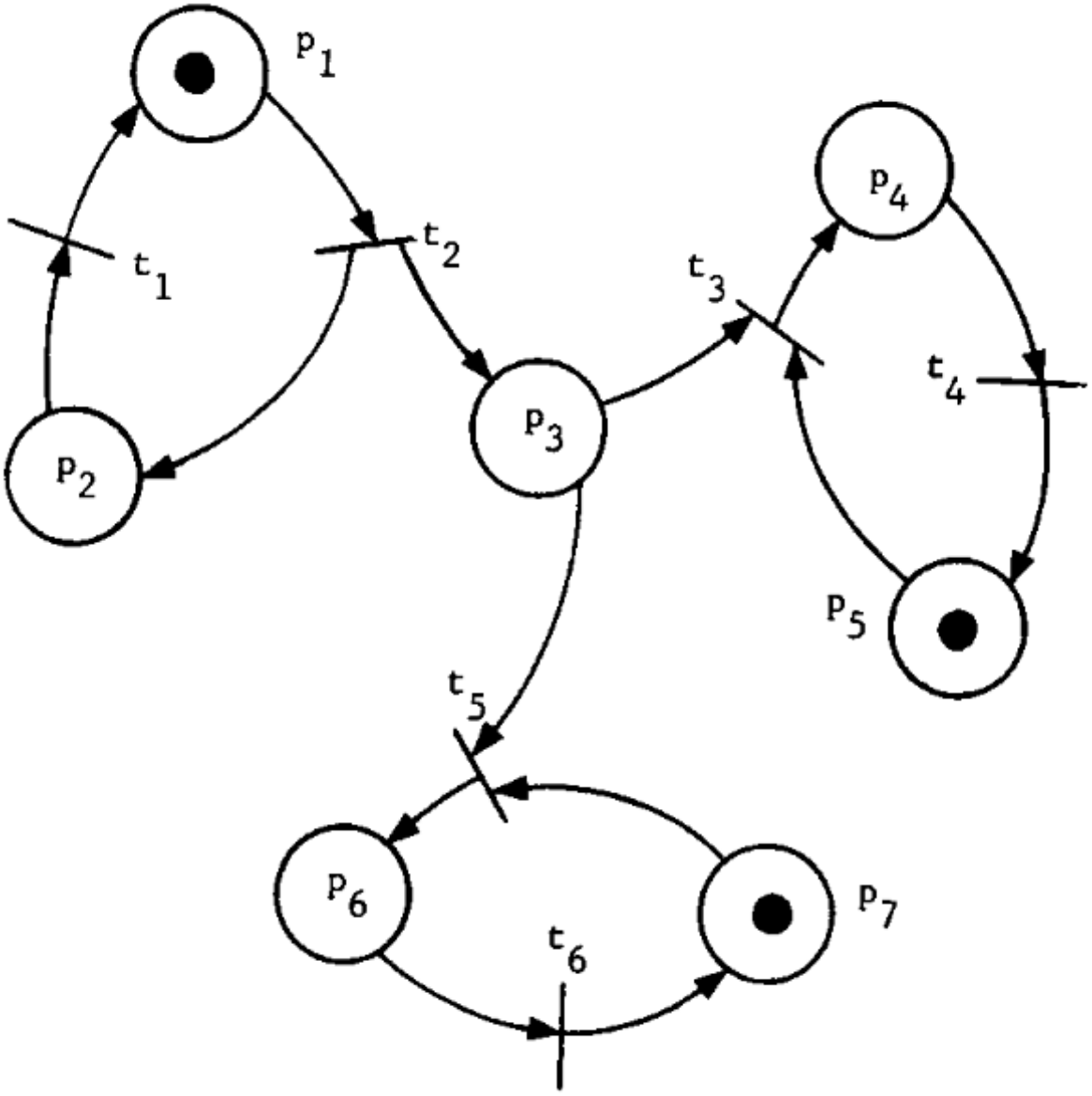


Abbildung 2.9: Ein einfaches Petrinetz bestehend aus den Zuständen p_1, p_2, p_3, p_4, p_5, p_6, p_7, dargestellt als Kreise und den Übergängen t_1, t_2, t_3, t_4, t_5, t_6, dargestellt als Querbalken. Die Zustände p_1, p_5 und p_7 haben ein Token, weshalb nur Übergang t_2 ausgelöst werden kann [14].

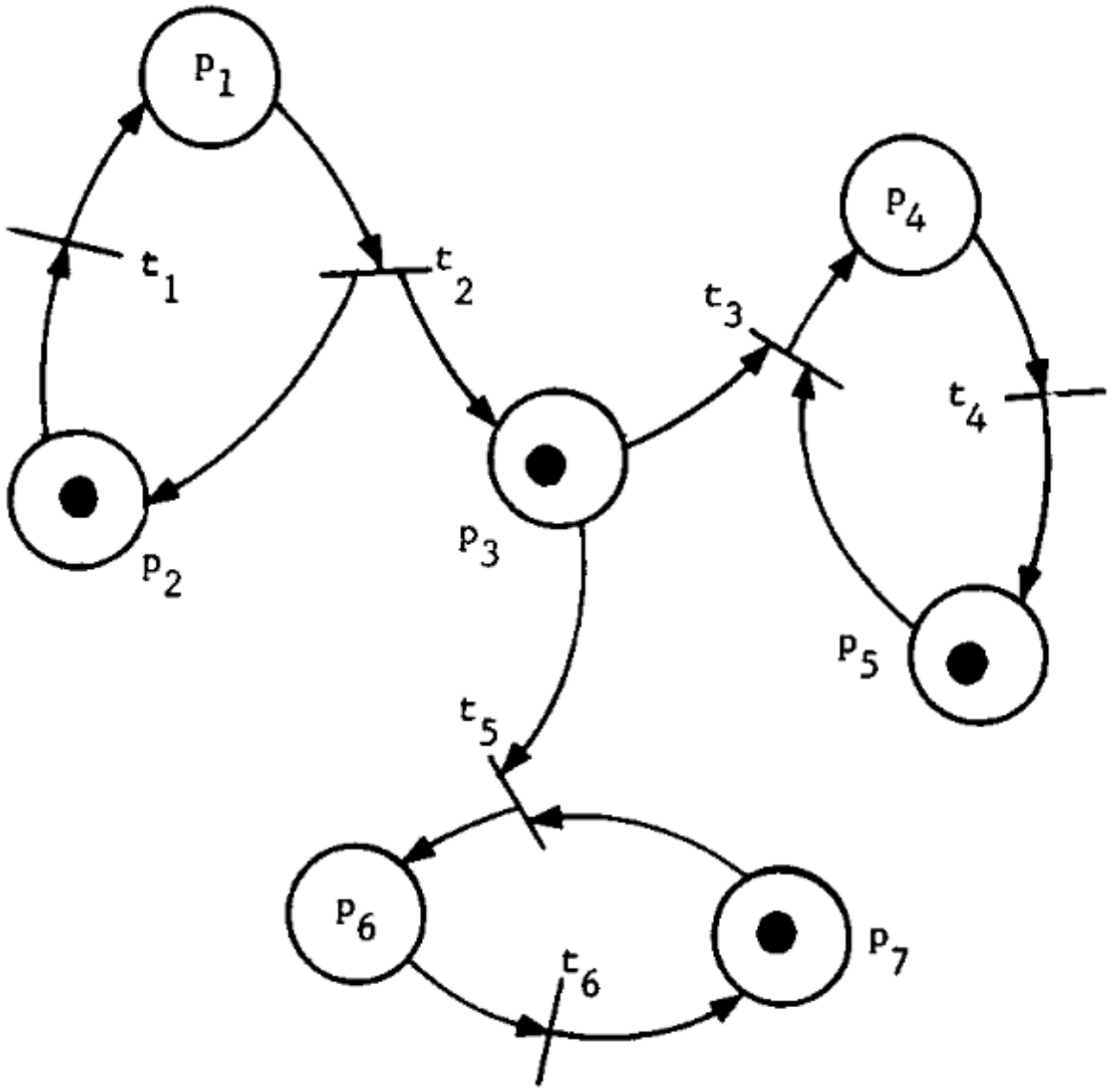


Abbildung 2.10: Ein einfaches Petrinetz bestehend aus den Zuständen p_1, p_2, p_3, p_4, p_5, p_6, p_7, dargestellt als Kreise und den Übergängen t_1, t_2, t_3, t_4, t_5, t_6, dargestellt als Querbalken. Die Zustände p_2, p_5, p_7 und p_3 haben ein Token, weshalb die Übergänge t_1, t_3 und t_5 ausgelöst werden können [14].

2.2 NMR-Spektroskopie

Um die im Hypergraphen entwickelten organischen Moleküle zu untersuchen eignet sich NMR-Spektroskopie. Für die Erklärung der Grundlagen der NMR-Spektroskopie wurde das Lehrbuch Organische Chemie von Peter Vollhardt [5] verwendet.

2.2.1 Spektren

Die Spektroskopie (lat. Erscheinung, Sichtbarwerden) befasst sich mit der Darstellung realer Eigenschaften durch eine Auftragung einer kontinuierlichen oder kleinschrittig gestaffelten Größe gegen die Intensität. In der organischen Chemie werden vor allem vier Spektroskopie Typen verwendet. Die Massenspektroskopie (MS) hat in der Auftragung der x-Achse kleinschrittig gestaffelte Werte. Bei der MS wird ein Molekül in kleinere, geladene Fragmente geteilt. Diese Fragmente können über ein meist ganzzahliges Masse-Ladungs-Verhältnis m/z und ihre Häufigkeit charakterisiert werden. Zerstörungsfrei hingegen sind die kontinuierlichen, auf elektromagnetischen Wellen basierenden Absorptions- oder Resonanzspektren. Die verbreitetsten Absorptionsspektren sind dabei UV/Vis und IR. Die verwendete elektromagnetische Strahlung liegt dabei im UV/Vis- (200 - 800 nm) und IR- (800 nm - 1 mm) Bereich. Die meist verbreitetste Resonanzspektroskopie hingegen ist die Kern-Magnet-Resonanz-Spektroskopie (Nucleus-Magnetic-Resonance, kurz NMR-Spektroskopie). UV-Resonanzspektroskopie ist möglich, wobei die Menge von Molekülen, die dafür verwendbar sind, sehr gering ist. Die elektromagnetischen Wellen für NMR sind im Radiowellenbereich (über 100 mm) [5].

2.2.2 Kernspin

Die im Atomkern befindlichen Neutronen und Protonen können sich in einem von zwei für NMR relevanten Zuständen befinden. Diese Zustände des Kernspins sind die α- und β-Konfiguration. Dieser Kernspin ist keine visuelle Rotation, hat aber die Eigenschaft, wie Ladung in Bewegung, ein Magnetfeld zu erzeugen. Die verschiedenen Zustände unterscheiden sich dabei in der theoretischen Drehrichtung und damit auch in der Richtung des induzierten Magnetfeldes. Unter normalen Umständen treten beide Konfigurationen im Verhältnis 1:1 auf. Ein entgegen gerichtetes, externes Magnetfeld führt dazu, dass diese Konfiguration energetisch gehindert wird. Dadurch befinden sich deutlich weniger Kerne in diesem Zustand. Die kinetische Energie der Teilchen erlaubt jedoch einen Konfigurationswechsel, wenn der Energieunterschied klein genug ist. Es ist ebenfalls möglich diesen Konfigurationswechsel durch Radiowellen einer zu dem Kern gehörigen, charakteristischen Wellenlänge zu erzwingen, wobei die Rückänderung der Konfiguration, auch Relaxation genannt, nach kurzer Zeit eintritt. Diese Relaxation emittiert dabei eine für den Kern

charakteristische Wellenlänge. Sowohl die absorbierte, als auch die emittierte Wellenlänge hängen von der Energiedifferenz der beiden Konfigurationen ab, welche wiederum von der Umgebung des Kerns, aber auch von dem angelegten Magnetfeld abhängen. Die Frequenz der Strahlung ist dabei linear-proportional zur Stärke des Magnetfeldes.

2.2.3 NMR aktive Kerne

Da für NMR-Spektren der Kernspin nötig ist, kann diese Methode nur für Spektren verwendet werden, die Kernresonanz aufweisen. Das sind alle Kerne, die keine gerade Anzahl von Protonen und gerade Anzahl von Neutronen haben. Für die organische Chemie besonders wichtige Kerne sind Wasserstoff, Kohlenstoff, Sauerstoff und Stickstoff (Tabelle 2.1). Für Wasserstoff ist dabei ^{1}H besonders gut geeignet. Es tritt mit hoher Häufigkeit von 99.985 % auf und hat mit einer ungefähren Resonanzfrequenz von 300 MHz eine deutlich verschiedene Frequenz zu dem mit 0.015 % sehr seltenen ^{2}H (oder D Deuterium), welches bei etwa 45.7 MHz resoniert. Bei Stickstoff bietet sich das Isotop ^{14}N ebenso gut wie Deuterium an. Für Kohlenstoff ist ^{12}C mit relativer Häufigkeit von 98.89 % nicht geeignet, da dieses Isotop 6 Neutronen und 6 Protonen hat und somit nicht NMR aktiv ist. Darum wird das deutlich seltenere ^{13}C für NMR-Spektren verwendet. Für Sauerstoff ist ^{17}O die einzige Option, wobei die relative Häufigkeit mit 0.037 % noch geringer ist. Die

Tabelle 2.1: Tabelle der chemischen Isotope, der in der organischen Chemie relevanten Elemente, mit natürlicher relativer Häufigkeit und ob sie NMR aktiv sind [5].

Isotop N	Protonen/Neutronen	NMR-Aktivität	Natürliche Häufigkeit in %
^{1}H	1/0	Ja	99.985
^{2}H (D)	1/1	Ja	0.0015
^{3}H (T)	1/2	Ja	0
^{12}C	6/6	Nein	98.89
^{13}C	6/7	Ja	1.11
^{14}N	7/7	Ja	99.63
^{15}N	7/8	Ja	0.37
^{16}O	8/8	Nein	99.759
^{17}O	8/9	Ja	0.037
^{18}O	8/10	Nein	0.037

verschiedenen Elemente und Isotope haben meistens unterschiedliche Frequenzbereiche, mit denen sie angeregt werden können.

2.2.4 NMR-Spektrometer

Um ein NMR-Spektrum aufzunehmen, werden wenige Milligramm einer zu untersuchenden Probe in 0.3 bis 0.5 mL Lösungsmittel aufgelöst. Das Lösungsmittel selbst sollte dabei möglichst wenig der zu untersuchenden Kerne enthalten. Bei Wasserstoff wird zum Beispiel ein deuteriertes Lösungsmittel verwendet. Die Lösung wird in ein elektrisch induziertes Magnetfeld eingebracht (Abbildung 2.11). Ein Radiofrequenzpulsgenerator sendet einen kurzen Impuls an Radiowellen in die Probe, der das gesamte zu untersuchende Spektrum umfasst. Diese Radiowellen treffen auf die Röhre, wobei ein überwiegender Teil nicht absorbiert wird. Die absorbierten Radiowellen regen die Kernspins an, welche nacheinander wieder relaxieren. Die durch die Relaxation emittierten Radiowellen werden in alle Richtungen abgestrahlt und von einem Radiofrequenzdetektor gemessen. Dieser ist orthogonal zum Radiofrequenzpulsgenerator positioniert, damit möglichst keine der durchgehenden Strahlung direkt auf den Detektor trifft. Der Radiofrequenzdetektor misst eine gedämpfte Schwingung mit verschiedenen Frequenzen. Diese Frequenzen können mit der Fouriertransformation von einem zeitabhängigen Oszillationsspektrum zu einem frequenzabhängigen Peakspektrum umgerechnet werden.

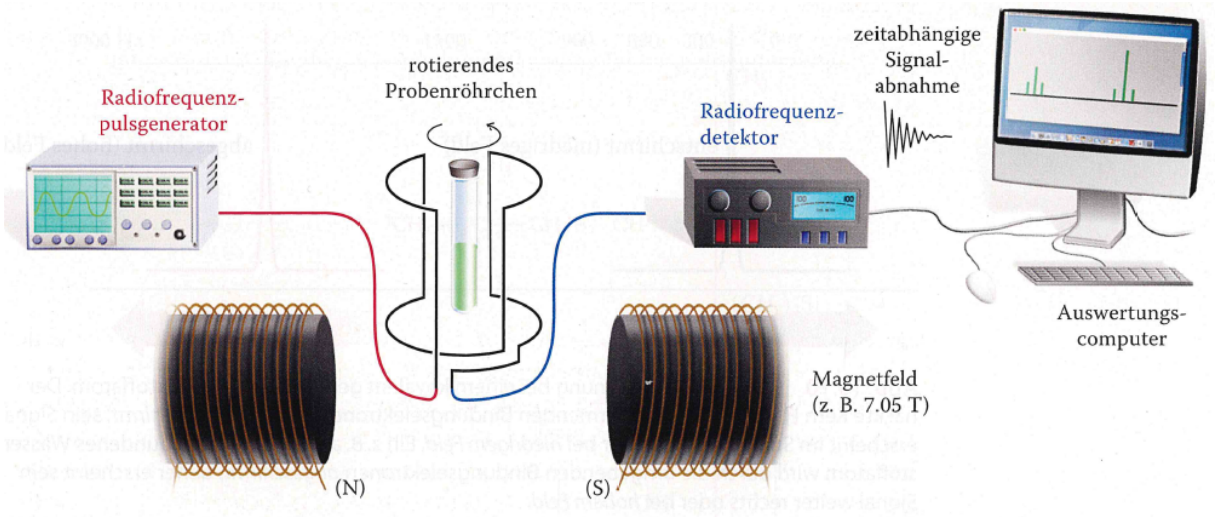


Abbildung 2.11: Schematischer Aufbau eines NMR-Spektrometers [5].

2.2.5 Chemische Verschiebung

Da die Rotation des geladenen Atomkerns ein Magnetfeld induziert liegt nahe, dass die Bewegung von Elektronen ebenfalls zu einer solchen Induktion führt. Das von den Elektronen induzierte Magnetfeld ist nach der Lenz'schen Regel dem induzierenden Magnetfeld entgegen gerichtet. Es lässt sich sagen, dass bei Kernen mit hoher umliegender Elektronendichte ein schwächeres effektives Feld anliegt. Um solche Kerne anzuregen, wird eine

geringere Frequenz benötigt. Da diese Frequenz zwar für den Kern charakteristisch ist, aber mit dem angelegten Magnetfeld variiert, wird ein interner Standard verwendet. Für Wasserstoff- und Kohlenstoff-Spektren ist dieser Tetramethylsilan (TMS $(CH_3)_4Si$). Die Frequenz ν wird daher als Differenz zu der Frequenz dieses internen Standards ν_{TMS} betrachtet. Da aber die Differenz eines Kerns zum internen Standard proportional zur Feldstärke zunimmt, wird die Differenz durch die Frequenz des Spektrometers ν_0 geteilt, um die chemische Verschiebung δ in ppm zu erhalten (Gleichung 2.7).

$$\delta[ppm] = \frac{\nu[Hz] - \nu_{TMS}[Hz]}{\nu_0[MHz]} \tag{2.7}$$

Kerne, bei denen die gleiche Elektronendichte vorliegt, haben ein gleichstarkes effektive Magnetfeld, absorbieren die gleiche Wellenlänge und ihre Peaks liegen somit im entstehenden Spektrum bei der gleichen ppm Zahl. Sie werden darum als chemisch-äquivalent bezeichnet. Um zu bestimmen, ob zwei Kerne in einem Molekül chemisch-äquivalent zueinander sind, müssen beide Kerne durch Symmetrieoperationen, wie Rotationen um Einfachbindungen oder Spiegelungen aufeinander abzubilden sein. Als Beispiel wird Chlorethan (H_3C-CH_2Cl) betrachtet (Abbildung 2.12). Die Wasserstoffe in der Methylgruppe (H_b) sind, projiziert auf eine Ebene, alle 120° auseinander. Wenn die Gruppe also sequenziell 120° um die C-C-Einfachbindung rotiert wird, können alle Wasserstoffe aufeinander abgebildet werden. Wenn die Methylgruppe so rotiert wird, dass ein Wasserstoff parallel zum Chlor steht, kann durch die H-C-Bindung und die C-C-Bindung eine Spiegelebene aufgespannt werden. Diese Spiegelebene kann die beiden chlornahen Wasserstoffe H_a aufeinander abbilden. Ein komplizierteres Beispiel ist Cyclohexan (Abbildung 2.12). Die energetisch-günstigste Konfiguration für Cyclohexan ist die Sesselkonfiguration, bei welcher aus einem Sechseck zwei der Ecken in entgegengesetzte Richtungen geklappt wurden. In dieser Form sind alle Kohlenstoffe äquivalent zueinander, da die Winkel und benachbarten Atome gleich sind. Jeder Kohlenstoff hat dabei einen Wasserstoff, welcher planar oder äquatorial zum Sechsring liegt (rot) und einen Wasserstoff, welcher orthogonal oder axial zum Sechsring steht. Ohne Rotation lassen sich diese beiden Wasserstoffe voneinander unterscheiden. Durch Rotation um die zu den Spitzen führenden C-C-Einfachbindungen lässt sich die Richtung der beiden Spitzen umkehren, wodurch alle äquatorialen zu axialen Wasserstoffen werden und umgekehrt. Diese Rotation findet bei Raumtemperatur schnell genug statt, um im NMR-Spektrum nur ein Signal zu erzeugen. Eine Restgruppe an einem der Kohlenstoffe oder eine Temperatur von -90 °C würde es aber erlauben, zwischen äquatorialen und axialen Positionen zu unterscheiden.

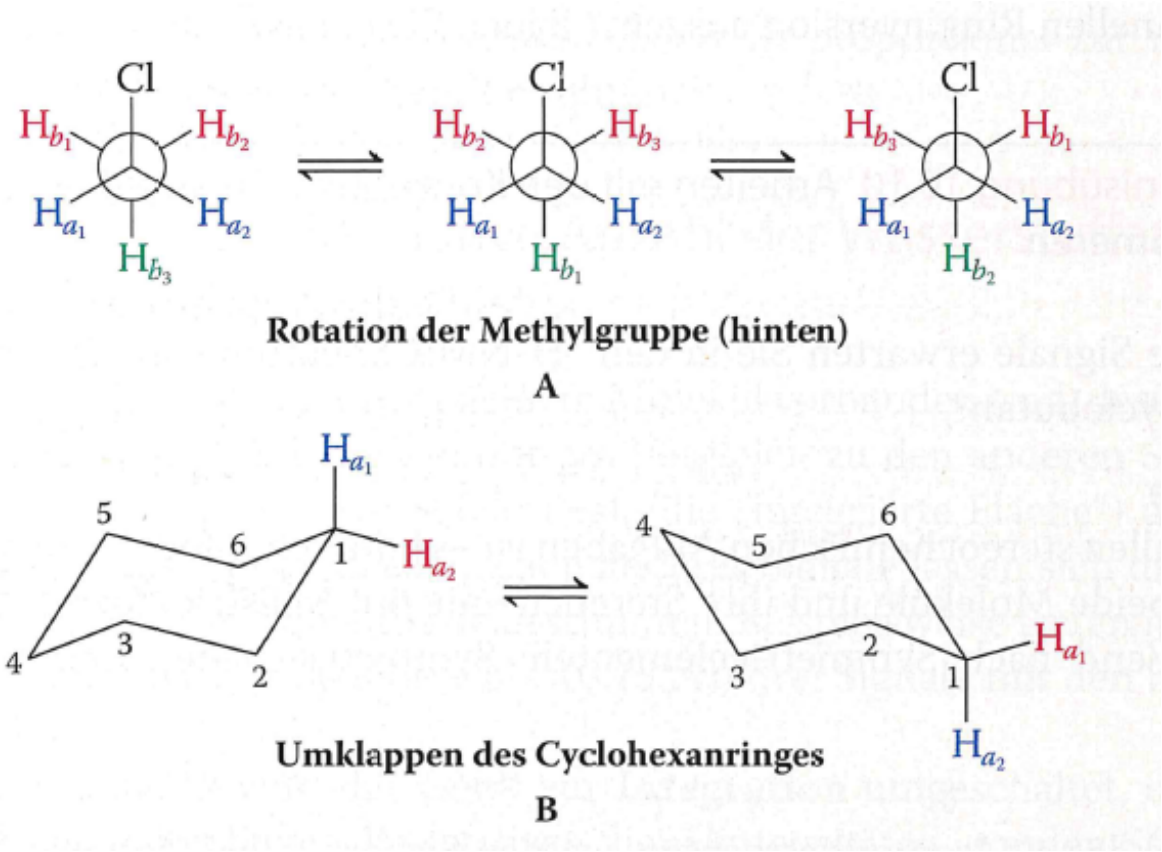


Abbildung 2.12: Symmetrieoperationen am Beispiel Chlorethan (A) und Cyclohexan (B). Bei Chlorethan kann eine Spiegelachse vertikal von Chlor (Cl) zum unteren Wasserstoff (grün) gezogen werden. Ebenso ist eine Rotationsachse zwischen den beiden Kohlenstoffen (Kreismitte) möglich. In Cyclohexan stehen sechs Wasserstoff orthogonal (axial) und sechs waagerecht (äquatorial) zum 6-Ring. Durch ein Ändern der Sesselkonfiguration, können beide ineinander überführt werden [5].

2.2.6 Signalintensität

Nachdem die chemische Verschiebung der Peaks diskutiert wurde, wird nun thematisiert, was an den ppm-Verschiebungswerten zu sehen ist. Das NMR-Spektrum besteht aus einer berauschten Grundlinie, aus der an manchen Stellen hohe Intensitätswerte hervorstechen. Diese Werte werden Peaks genannt. Die ppm-Werte können dabei den einzelnen Kernen zugeordnet werden. Ein Peak kann aber auch in einem Peakcluster angeordnet sein, welche in Kapitel 2.2.7 thematisiert werden. Die Peaks selber haben in den meisten Fällen eine breite Basis, die mit steigender Intensität immer spitzer zuläuft. Besonders bei sauren Wasserstoffen ist es jedoch möglich, dass der Peak auseinander gezogen wird und nur eine deutlich breitere und niedrigere Gaußglocke zu erkennen ist. In manchen Messungen ist dieser Peak nur als ein Anheben der Grundlinie zu erkennen. Die Peakhöhe und -breite kann dabei durch äußere Umstände variieren. Um für das Signal trotzdem vergleichbare Werte zu bekommen, wird die Fläche unter den Peaks integriert. Da die Flächen nur

relativ zueinander konstante Werte ergeben, wird die kleinste Fläche als 1 definiert. Es wird davon ausgegangen, dass dieser Peak genau einem Wasserstoff im zu untersuchenden Molekül entspricht. Dabei gilt auch, dass die Peaks von chemisch-äquivalenten Kernen an der gleichen Verschiebung auftreten und dadurch überlagern. Die Peakfläche entspricht somit der Anzahl an Kernen, die diese Verschiebung aufweisen. Sollte bei dieser Analyse herauskommen, dass größere Peaks eindeutig Brüche darstellen, wird die kleinste Fläche in ihrem Wert erhöht, bis alle Flächen ganzzahlig sind. In dem ^{1}H-Spektrum vorherigen Beispiel Chlorethan, gibt es zwei Peaks. Der kleinere Peak entspricht den chlornahen Wasserstoffen und der größere der Methylgruppe. Wenn für diesen Peak nun eine relative Fläche von 1 angenommen werden würde, so würde der Peak der Methylgruppe eine Fläche von 1.5 aufweisen. Die Fläche des kleinsten Peaks wird also auf die nächstgrößere, natürliche Zahl erhöht. Dadurch sind beide Peakflächen mit 2 und 3 ganzzahlig, was auch der tatsächlichen Häufigkeit in Chlorethan entspricht.

2.2.7 Multiplettaufspaltung

Da das Anlegen eines Magnetfeldes nur eine Kernspinkonfiguration bevorzugt, jedoch nicht alle Kerne in diese überführt ist immer noch ein Großteil der Kernspins zufällig ausgerichtet. Liegt ein angeregter Kern nun neben nicht angeregten Kernen, so beeinflusst deren Magnetfeld den angeregten Kern. Bei der Ausrichtung der nicht angeregten Kerne mit dem Magnetfeld, wird das Signal zu höherer Frequenz und bei entgegengesetzter Ausrichtung zu niedrigere Frequenz verschoben. Beide Signale treten nun im Spektrum auf, wobei sie zu dem selben Kern gehören. Sie werden deshalb als ein Peak zusammengefasst. Chemisch-äquivalente Kerne spalten das Multiplett mit gleicher Stärke auf. Dabei kann binomialverteilt jede Kombination an mit und entgegen gerichteten Kernen auftreten (Tabelle 2.2). Liegt in der näheren Umgebung ein weiterer, nicht zu den anderen benachbarten Kernen chemisch-äquivalenter Kern, so kann dieser das Spektrum mit einer anderen Frequenz aufspalten. Da die Frequenz der Aufspaltung verschieden ist, kommt es zu einem kombinierten Multiplett. Zum Beispiel gib es bei 1,1-Dichlorpropan zu einer Aufspaltung der zentralen 2-Wasserstoffe zu einer Dublett in Quartett Aufspaltung. Die Methylgruppe spaltet dabei in ein Quartett auf und der chlornahe Wasserstoff spaltet dieses Quartett wiederum in ein Dublett auf. Je nach Auspaltungsfrequenz können die beiden Multipletts verschieden ineinander verschränkt sein. Es können zwei nebeneinanderliegende Quartetts (Quartett in Dublett) oder jedes Signal des Quartetts aufgespalten (Dublett in Quartett) sein. Bei Wasserstoffkernen folgt die Multiplett Zahl der $N+1$-Regel. Bei anderen Kernen kann es aber auch zu einer $2N+1$-Verteilung kommen. Da die Aufspaltung durch Interaktion der Kerne miteinander entsteht und der Interaktionsabstand der Kerne zueinander gleich ist, ist die Aufspaltungsfrequenz symmetrisch. Es ist also möglich zwei benachbar-

te Kerne darüber zu identifizieren, dass der Abstand innerhalb des Multipletts gleich ist. Diese Interaktion ist dabei über bis zu 3 Bindungen gut erkennbar. Bei geminalen, also an den gleichen Kohlenstoff gebundenen Wasserstoffen, können Aufspaltungen von 0 - 18 Hz auftreten und bei vicinalen, also an benachbarten Kohlenstoff gebundenen Wasserstoffen, kann diese 6 - 8 Hz betragen.

Tabelle 2.2: Relative Intensitäten der einzelnen Peaks in den verschiedenen Multipletts, sowie die dafür nötige Anzahl chemisch-äquivalenter benachbarter Wasserstoffe [5].

Anzahl chemisch-äquivalenter Wasserstoffe N	Multiplett Name	relative Peakhöhe
0	Singulett (s)	1
1	Dublett (d)	1 : 1
2	Triplett (t)	1 : 2 : 1
3	Quartett (q)	1 : 3 : 3 : 1
4	Quintett (quin)	1 : 4 : 6 : 4 : 1
5	Sextett	1 : 5 : 10 : 10 : 5 : 1
6	Septett	1 : 6 : 15 : 20 : 15 : 6 : 1

2.2.8 Spektrumanalyse

Im Folgenden wird die Analyse eines NMR-Spektrums am Beispiel von 1,1,2-Trichlorpropan erläutert (Abbildung 2.13). Nach der Rotationssymmetrie lassen sich alle Wasserstoffe der Methylgruppe aufeinander abbilden. Durch die an der 1-C-Position gebundenen Chloratome gibt es aber keine Symmetrie von den Wasserstoffen der Endpositionen aufeinander. Es gibt also 3 verschiedene Wasserstofftypen. Ebenso lässt sich sagen, dass das Signal für die Methylgruppe mit 3 Wasserstoffen eine Gesamtintensität von 3 aufweist. Da die beiden anderen Wasserstofftypen nur jeweils ein Proton umfassen, haben sie jeweils eine Intensität von 1. Die H_c-Methylwassestoffe haben zu dem an den 2-Kohlenstoff gebundenen H_b-Wasserstoff einen Abstand von 3 Bindungen. Die beiden Signale spalten einander also jeweils mit der Kopplungskonstante $J_{bc} = 6.8Hz$ auf. Das H_c-Signal wird dabei als ein Dublett aufgespalten, da nur ein chemisch-äquivalenter Kern benachbart ist. Der Wasserstoff der 1-C-Position H_a und der Wasserstoff der 2-C-Position sind zueinander ebenfalls 3 Bindungen entfernt und spalten einander mit der Kopplungskonstante $J_{ab} = 3.6Hz$ auf. Da zu dem H_a nur ein chemisch-äquivalentes H_b benachbart ist, wird das H_a-Signal zu einem Dublett aufgespalten. Der H_b hat sowohl H_a, als auch drei H_c in seiner Nachbarschaft. Darum wird es von den drei H_c zu einem Quartett und von dem H_a zu einem Dublett aufgespalten. Es entsteht ein Dublett in Quartett. Da Chlor elektronenziehend ist, wird mit zunehmender Chlorsubstitution das Proton immer weiter entschirmt und das

Signal erscheint bei höherer Verschiebung. Der H_a hat mit zwei benachbarten Chloratomen die größte, H_b mit einem benachbarten Chloratom die mittlere und H_c mit keinen benachbarten Chloratomen die niedrigste chemische Verschiebung.

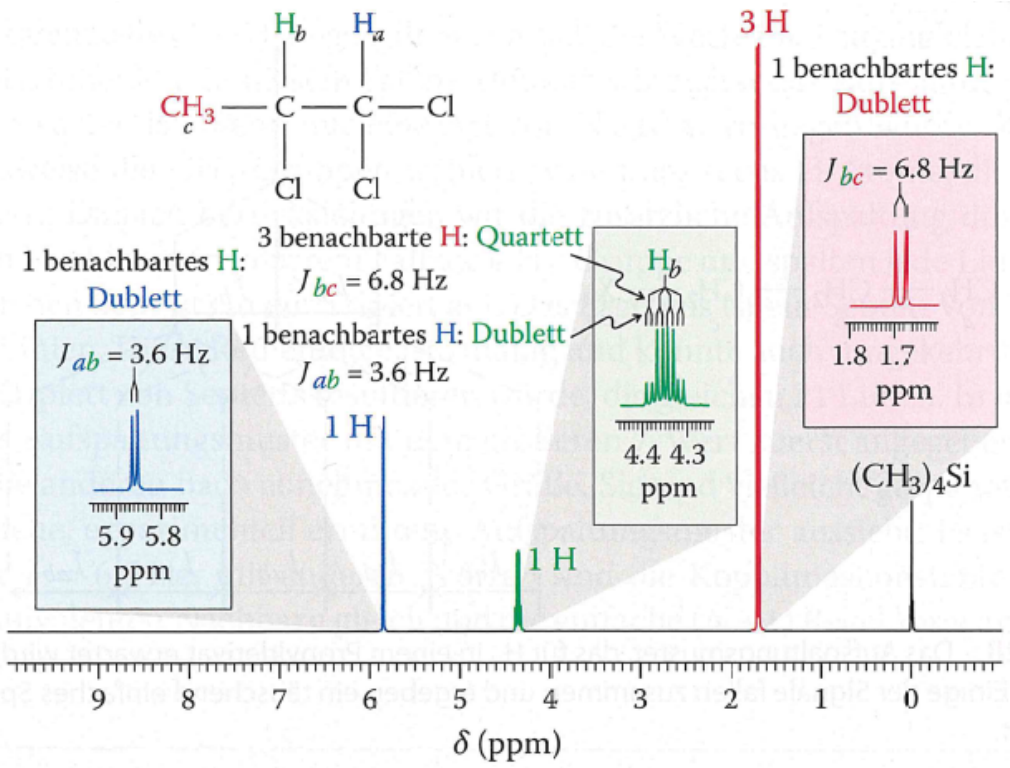


Abbildung 2.13: ^{1}H-NMR von 1,1,2-Trichlorpropan. Der Methylwasserstoff (H_c rot) hat ein Dublett bei 1.68 ppm mit Intensität 3, der zentrale Wasserstoff (H_b grün) hat bei 4.35 ppm ein Dublett in Quartett mit Signalstärke 1 und der Endwasserstoff (H_a blau) erzeugt ein Dublett bei 5.96 ppm mit Signalstärke 1 [5].

2.2.9 Schreibweise

In einem wissenschaftlichen Kontext wird nicht immer das gesamte Spektrum als Bild angegeben. Um trotzdem eine Angabe zu machen, die das Spektrum ausreichend beschreibt, muss ein aussagekräftiger String verwendet werden. Für Chlorethan wäre dieser String zum Beispiel ^{1}H-NMR (300 MHz): δ 1.488 (-CH_3, 3, t, $^3J = 7.232Hz$), 1.488 (-CH_2-, 2, q, $^3J = 7.232Hz$). Für jeden Kern wird dabei die Verschiebung in ppm, eine kurze Zuordnung zum Kern, die Intensität, die Multiplettaufspaltung und die Stärke der Aufspaltung genannt. Für die Stärke der Aufspaltung wird zudem angegeben, über wie viele Bindungen diese Interaktion stattfindet. Für das ganze Spektrum wird auch angegeben, auf welche Frequenz oder Feldstärke das Spektrometer eingestellt wurde.

2.2.10 Kohlenstoff NMR

Der neben Wasserstoff am häufigsten untersuchte Kern ist Kohlenstoff. Für sein NMR wird ^{13}C verwendet und ein Lösungsmittel mit ^{12}C eingesetzt. Das Spektrum verhält sich dabei größtenteils genauso wie das von Wasserstoff, mit der Ausnahme dass ^{13}C fast nie nebeneinander vorliegt, also keine ^{13}C-^{13}C-Multiplett-Interaktion stattfindet. ^{13}C kann aber durchaus mit ^{1}H interagieren. Diese Kopplung wird häufig gezielt neutralisiert, da der ^{13}C Kern weniger leicht angeregt wird und so ein schlechteres Signal/Rausch-Verhältnis vorliegt. Kleinere Peaks können so leichter im Rauschen untergehen. Darum wird oft ein ^{1}H Breitband entkoppeltes ^{13}C Spektrum aufgenommen. Die Notation dafür ist $^{13}C\{^{1}H\}$. Um dies zu erreichen, wird ein breites Wasserstofffrequenzsignal während der gesamten Messung eingestrahlt. Das resultierende Spektrum weist für jeden chemisch-äquivalenten Kohlenstoff nur einen Peak auf. Die Peak Fläche wird dabei je nach ausgeblendeten Kern unterschiedlich beeinflusst, weshalb die Intensität nur noch als Richtwert und nicht mehr als absolute Größe betrachtet werden kann.

2.3 Dichte Funktional Theorie (DFT)

Die Dichte Funktional Theorie (DFT) vereinfach die quantenchemischen Atomorbitale als eine über den Raum verteilte Elektronendichtefunktion $\rho[r]$. Für ein gegebens Molekül lässt sich daraus die kinetische Energie T (Gleichung 2.8), die Energie der Elektronen-Protonen-Wechselwirkung V_{eN} (Gleichung 2.9), sowie die Energie der Elektronen-Elektronen-Wechselwirkung V_{ee} (Gleichung 2.10) als Integral dieser Dichtefunktion über den gesamten Raum r bestimmen.

$$T[\rho] = C \int \rho(r)^{\frac{5}{5}} dr \tag{2.8}$$

$$V_{eN}[\rho] = -j_0 Z_0 \int \frac{\rho(r)}{r} dr \tag{2.9}$$

$$V_{ee}[\rho] = \frac{1}{2} j_0 \int \frac{\rho(r_1)\rho(r_1)}{|r_1 - r_2|} dr_1 dr_2 \tag{2.10}$$

Die Gesamtenergie des Moleküls lässt sich als Summe aller Teilenergien ausdrücken (Gleichung 2.11).

$$E_{TF}[\rho] = T[\rho] + V_{eN}[\rho] + V_{ee}[\rho] \tag{2.11}$$

Nach dem Variationsprinzip ist die tatsächliche Elektronenkonfiguration genau dann erreicht, wenn die Gesamtenergie für die feste Elektronenzahl und Kernzusammensetzung den niedrigsten Wert erreicht. Während die Elektronendichte variiert wird, ist auch die Position der Kerne festgelegt. Die beste gefundene Konfiguration ist dabei immer energetisch ungünstiger, als die tatsächliche Konfiguration, wobei der gefundene Minimalwert

das System am besten beschreibt [15]. Um die Bestimmung der optimalen Dichteverteilung zu vereinfachen, wird diese Funktion $\rho(r)$ nach dem Fourierprinzip als Linearkombination von Basisfunktionen η_K, mit einem Faktor $d_K^{\mu\nu}$ ausgedrückt (Gleichung 2.12).

$$\rho_{\mu\nu}(r) = \sum_K d_K^{\mu\nu} \eta_K \tag{2.12}$$

Die hierfür verwendeten Basisfunktionen können variieren, sind aber meistens Slater oder Gauss-Funktionen [16].

3 Methoden und Tools

Für die Entwicklung der Reaktionsnetzwerke (Kapitel 3.1), die Beschränkung der entstehenden Graphen (Kapitel 3.2), die Berechnung der NMR-Spektren (Kapitel 3.3) und die Lösung (Kapitel 3.5) und Verifikation (Kapitel 3.6) des ILP wurden verschiedene Programme verwendet, welche im Folgenden näher erläutert werden. Ebenso wird die Methode definiert, mit der die NMR-Spektren untereinander verglichen werden (Kapitel 3.4).

3.1 Graptransformation Derivation Graph

Mød kurz für MedØlDatschgerl ist ein pythonbasiertes Graphtransformtions Tool, welches mit dem DPO-Ansatz aus Graphtransformationsregeln und aus Startgraphen neue Graphen bestimmt. Aus den durch Graphtransformation gefundenen neuen Graphen wird ein Hypergraph definiert. Die Graphen in Mød können dabei erzeugt werden, indem sie über einen SMILES-String geladen oder über einen GML-String definiert werden. Im Folgenden wird das selbe Molekül Cyclohexan über seinen SMILES- und einen GML-String definiert.

```
1       cyclohexan = smiles ("C1CCCCC1")
2       cyclohexan = Graph.fromGMLString(
3       """
4       graph [
5       node [ id 0 label "C" ]
6       node [ id 1 label "H" ]
7       node [ id 2 label "C" ]
8       node [ id 3 label "H" ]
9       node [ id 4 label "C" ]
10      node [ id 5 label "H" ]
11      node [ id 6 label "C" ]
12      node [ id 7 label "H" ]
13      node [ id 8 label "C" ]
14      node [ id 9 label "H" ]
15      node [ id 10 label "C" ]
16      node [ id 11 label "H" ]
17      edge [ source 0 target 1 label "-" ]
18      edge [ source 0 target 2 label "-" ]
19      edge [ source 2 target 3 label "-" ]
```

```
20      edge [ source 2 target 4 label "-" ]
21      edge [ source 4 target 5 label "-" ]
22      edge [ source 4 target 6 label "-" ]
23      edge [ source 6 target 7 label "-" ]
24      edge [ source 6 target 8 label "-" ]
25      edge [ source 8 target 9 label "-" ]
26      edge [ source 8 target 10 label "-" ]
27      edge [ source 10 target 11 label "-" ]
28      edge [ source 10 target 0 label "-" ]
29      ]
30      """
31      , name = "Cyclohexan)
32
```

Wenngleich die Definition des Moleküls als String deutlich einfacher ist, erlaubt der GML-String die Definition nicht chemischer Graphen und auch von Molekülzuständen, die keinen SMILES-String zugeordnet haben. Eine Regel kann nur über einen GML-String definiert werden. Für eine Regel müssen dabei linke und rechte Seite sowie der Kontext angegeben werden. Um zum Beispiel den Ring des Cyclohexans aufzuspalten um n-Hexan zu erhalten, würde eine Regel benötigt, welche eine C-C-Einfachbindung aufspaltet und die Kohlenstoffe mit Wasserstoff sättigt.

```
1       cracking = Rule.fromGMLString(
2       """
3       rule [
4       left [
5       edge [ source 1 target 2 label "-" ]
6       ]
7       context [
8       node [ id 1 label "C" ]
9       node [ id 2 label "C"]
10      ]
11      right [
12      node [ id 3 label "H" ]
13      node [ id 4 label "H" ]
14      edge [ source 1 target 3 label "-" ]
15      edge [ source 2 target 4 label "-" ]
16      ]
17      ]
18      )
19
```

Alle definierten Graphen werden in der Variable inputGraphs und alle definierten Regeln in der Variable inputRules gespeichert. Mit der Variable inputGraphs kann ein Derivationgraph DG erzeugt werden. Auf die in dem Derivationgraph enthaltenen Moleküle

können die Regeln mit einer benutzerdefinierten Abfolge angewendet werden. Die in der Variable inputRules zusammengefassten Regeln können dabei allgemein oder eine spezifische Regel mit ihrem Variablennamen angewendet werden. Es ist ebenfalls möglich nach jeder Regelanwendung eine restriction-Funktion aufzurufen, in welcher unerwünschte Eigenschaften der durch die Regeln erzeugten Graphen abgelehnt werden können.

```
dg = DG(graphDatabase=inputGraphs)
dg.build().execute(
addSubset(inputGraphs)
>> rightPredicate[
restriction
](
repeat(inputRules))
)
)

```

Der durch die Regelanwendung erzeugte Hypergraph kann über verschiedene print-Befehle in einer latex-generierten PDF-Datei dargestellt werden. Dafür muss ein FlowPrinter, DGPrinter oder GraphPrinter Objekt erzeugt werden [17].

```
flowPrinter = FlowPrinter()
flowPrinter.printUnfiltered = False
postSection("Loaded Graphs")
for a in inputGraphs:
a.print()
postSection("Loaded Rules")
for a in inputRules:
a.print()
postSection("Product Graphs")
for a in dg.vertices:
a.graph.print()

```

Mød wurde verwendet, um die Einwirkung von natürlich auftretenden Reaktionen auf natürliche oder synthetische Moleküle zu untersuchen.

3.2 networkx

Networkx ist ein pythonbasiertes Netzwerk- und Graphenanalysetool [18]. Eine Funktion ist dabei das Identifizieren aller sehnenlosen Zyklen. Während der Laufzeit des Graphtransformationstools wurden verschiedene Analysen dieser sehnenlosen Zyklen als Beschränkung verwendet.

3.3 ORCA

ORCA ist ein quantenchemisches Rechnungsprogramm auf DFT Basis. Es erlaubt unter anderem Geometrieoptimierung und NMR-Spektrenberechnung von chemischen Molekülen [19]. Für diese Arbeit wurde ORCA 6.1.1 verwendet. Für die durch das Graphtransformationstool generierten Moleküle wurden die Graphen und SMILES-Strings bestimmt. Die zugehörigen XYZ-Koordinatendateien wurden aus der RDKit-Bibliothek über ihren SMILES-String bezogen [20] oder mit dem chemischen Zeichenprogramm Avogadro2 [21] gezeichnet. In beiden Fällen wurde die Geometrie von ORCA mit den der Dichtefunktion PBE, D4 Dispersionskorrektur und dem Basissatz DEF2-SVP optimiert. Anschließend wurden für die optimierte Struktur mit der Dichtefunktion TPSS, dem Basissatz pcSseg-3 und mit autoaux und tightscf die chemischen Verschiebungen berechnet.

```
1       !PBE D4 DEF2-SVP OPT
2
3       !TPSS pcSseg-3 autoaux tightscf NMR
4
```

Die selben Berechnungen für TMS durchgeführt und in der finalen Bestimmung der NMR-Spektren angegeben. Die Kohlenstoff Verschiebung von TMS lag dabei bei 179.73 ppm. Die Bestimmung chemisch-äquivalenter Kerne wurde von Hand durchgeführt. Für jedes mit dem Graptransformationstool erzeugte Molekül wurde ein ^{13}C-NMR berechnet.

3.4 NMR-Vergleich

Um die von dem quantenchemischen Rechnungsprogramm berechneten NMR-Spektren zu Vergleichen und ihre Ähnlichkeit zu quantifizieren wurde die Binning-Methode mit anschließender Crosscorrelation angewandt [22]. Jeder Peak j des Spektrums setzt sich aus seiner Intensität I_j und seiner Verschiebung ν_j zusammen. Zuerst wird für ein gegebenes NMR-Spektrum die beiden Peaks mit der größten chemischen Verschiebung ν_{max} und der kleinsten chemischen Verschiebung ν_{min} bestimmt und eine Bingröße b festgelegt. Der größte Wert ν_{max} wird dabei aufgerundet und um eine Bingröße b erhöht (Gleichung 3.1). Der kleinste Wert ν_{min} wird abgerundet und um eine Bingröße b verringert (Gleichung 3.2).

$$\nu_{max} = ceiling(max) + b \tag{3.1}$$

$$\nu_{min} = floor(min) - b \tag{3.2}$$

Aus der Obergrenze ν_{max} und Untergrenze ν_{min} wird mit der Bingröße die nötige Anzahl an Bins n bestimmt (Gleichung 3.3).

$$n = ceiling(\frac{\nu_{max} - \nu_{min}}{b}) \tag{3.3}$$

In welchem Bin i ein Peak j liegt wird bestimmt, indem der kleinste Bin bestimmt wird, dessen unter Grenze kleiner als die chemische Verschiebung des Peaks ν_j ist (Gleichung 3.4)

$$i = floor((\nu_j - \nu min)/b) \tag{3.4}$$

Die Intensität des Bins I_i ist dabei die Summe aller in ihm liegenden Peaks (Gleichung 3.5).

$$I_i = \sum_{b \times i \leq \nu_j - \nu_{min} \leq b \times (i+1)} I_j \tag{3.5}$$

Die Intensität wird über alle Bins normalisiert.
Um zwei Spektren A und B zu vergleichen wird anstelle der Randwerte des einen Spektrums gemeinsame Randwerte für beide Spektren nach dem gleichen System bestimmt. Diese gemeinsamen Randwerte werden verwendet um die Peaks auf die Bins aufzuteilen. Anschließend werden die Bins für jedes Spektrum einzeln normiert. Der Überlapp eines Bins des ersten Spektrums i mit einem Bin des zweiten Spektrums j ist dabei das Produkt der Werte beider Bins. Der Teilüberüberlapp für einen Bin O_i ist die Summe aus dem Überlapp mit dem Bin mit gleichem Index und dem mit 0.5 gewichteten Überlapp der direkt benachbarten Bins (Gleichung 3.6).

$$O_i = 0.5 \times (I_i \times I_{j-1}) + (I_i \times I_j) + 0.5 \times (I_i \times I_{j+1}) : i = j \tag{3.6}$$

Der Gesamtüberlapp O ist die Summe aller Teilüberlappe O_i (Gleichung 3.7)

$$O_{AB} = \sum_i O_i \tag{3.7}$$

Aus dieser Crosskorrelation beider Spektren O_{AB} und der Selbstkorrelation beider Spektren O_{AA} und O_{BB}, wird der Ähnlichkeitsindex SI berechnet (Gleichung 3.8)

$$SI_{ij} = \frac{O_{AB}}{\sqrt{O_{AA} \times O_{BB}}} \tag{3.8}$$

Für alle Reaktionen eines Derivationgraphes wurden Referenzspektren bestimmt. Jedes mit dem quantenchemisches Rechnungsprogramm berechnete ^{13}C-NMR, wurde mit den Referenzspektren verglichen. Die für ein Referenzspektrum normierten Ähnlichkeitsindizes wurden als Wahrscheinlichkeit angenommen, dass beide NMR-Spektren die Gleiche

Substanz abbilden.

3.5 Gurobipy

Gurobipy [23] ist ein pythonbasierter ILP-Solver. Zur Bestimmung der optimalen Parameter wird ein ILP-Baum verwendet. Die bestimmten Wahrscheinlichkeiten wurden mit dem Derivationgraph kombiniert. Darauf basierend wurde Gruobipy verwendet, um einen diese Wahrscheinlichkeiten maximierenden Reaktionpfad in dem Derivationgraph zu identifizieren.

3.6 Petrinetzzertifikat

Eine weitere Funktion von Mød ist das Zertifizieren eines bestimmten Reaktionpfades des Hypergraphen mittels Petrinetzen. Dafür müssen die mittels Gurobi bestimmten Parameter für den Derivationgraph in Mød festgelegt werden. Ein mittels ILP bestimmter Reaktionspfad kann so nähergehend untersucht werden, indem Reaktionen auf Erfüllbarkeit basierend auf den gegebenen Startmolekülen überprüft werden.

4 Ergebnisse

Mit den erläuterten Werkzeugen und Methoden werden drei Reaktionsnetzwerke analysiert. Dabei werden zwei natürlich auftretende Ein-Molekül-Reaktionen (Kapitel 4.1) und eine Multi-Molekül-Reaktion (Kapitel 4.7) aus der Organischensynthesechemie betrachtet. Die Beispiele sind dabei nicht erschöpfend und dienen lediglich als Demonstration der Methode und als ein mögliches Anwendungsschema.

4.1 Ein-Molekül-Reaktionen

Für eine einfache Demonstration der Kombination von ILP wurden simple Reaktionen mit hoher natürlicher Selektivität gesucht. Dabei wurden die natürlichen Synthesen der Stoffe Koffein und Vanillin näher untersucht. Diese Synthesen erfolgen im Wesentlichen durch Reaktionen an einem zentralen Molekül, wobei die Reaktionspartner nicht weiter relevant bleiben. Sie werden im Folgenden als Ein-Molekül-Reaktionen bezeichnet.

4.1.1 Koffein

Koffein ist ein dreifachmethyliertes Xanthin, welches vor allem in Schwarztee und Kaffeebohnen vorkommt. Es wurde erstmals in den 1820 Jahren isoliert. Heute sind koffeinhaltige Lebensmittel aus dem Alltag nicht mehr wegzudenken. Der Wirkstoff erfüllt aber auch in zahlreichen weiteren Anwendungsfeldern eine signifikante Rolle. In der Natur wird Koffein aus den Nukleotidbasen Adenin und Guanin über die Oxidierung zu Xanthosin und anschließende Methylierung der Stickstoffe des Rings hergestellt (Abbildung 4.1). Die Methylierung erfolgt erst in der 7, dann in der 3 und zuletzt in der 1 Position. Diese geordnete Methylierung erfolgt mit hoher Selektivität und eignet sich so für eine nähere Untersuchung [24].

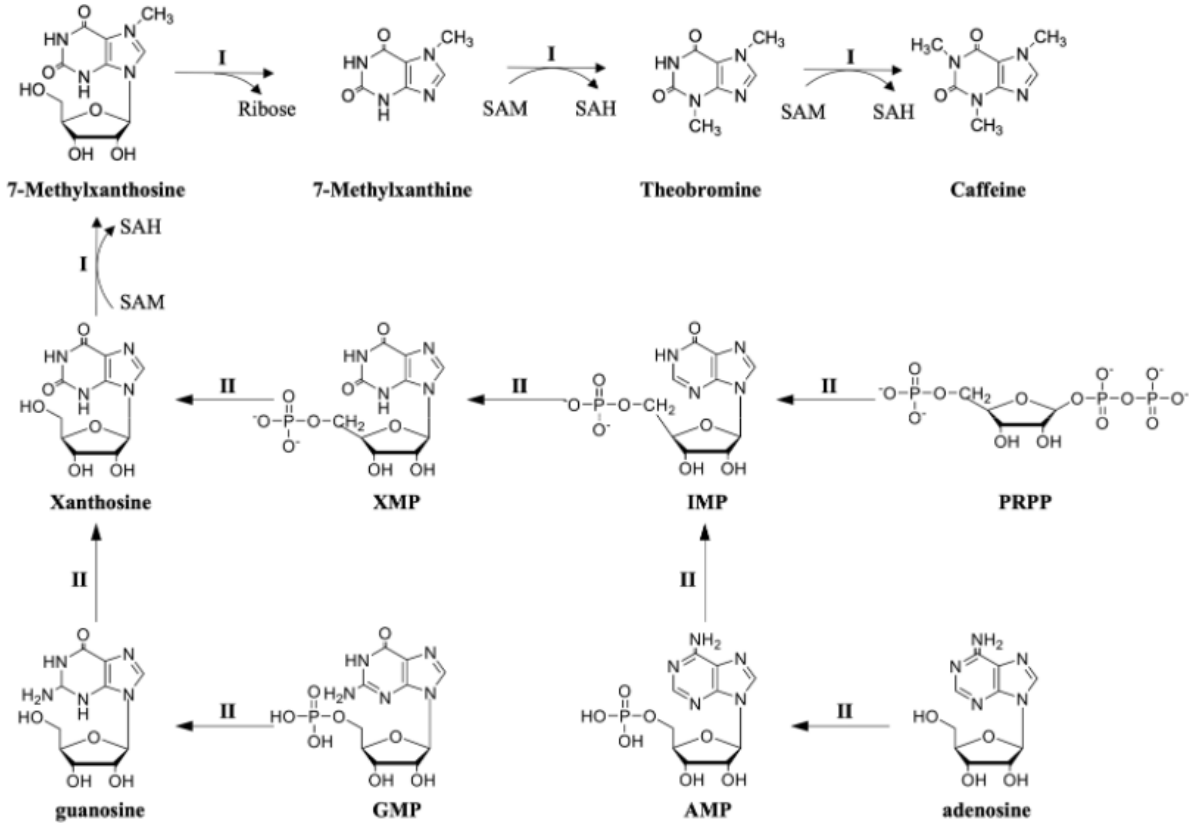


Abbildung 4.1: Synthese von Koffein in Pflanzen aus den Nukleotiden Adenin und Guanin. Erst werden die Purinringstrukturen oxidiert und nach der Abspaltung der Ribose die Stickstoffe methyliert. Die Methylierung erfolgt in der 7, 3 und zuletzt der 1 Position [24].

4.1.2 Vanillin und 3,4-Dihydroxyaldehyd

Vanillin ist eines der beliebtesten aromatischen Moleküle mit hoher Verwendung in Lebensmitteln Der überwiegende Teil wird synthetisch hergestellt. In der Natur wird Vanillin in Vanille Orchideen aus Phenylalanin synthetisiert. Zuerst wird dabei der Aminrest in einer Eliminierungsreaktion entfernt, wodurch Zimtsäure entsteht. Die Zimtsäure wird dann in der 3- und 4-Position des Phenylrests hydroxiliert und die Doppelbingung zu einem Aldehyd oxidiert (Abbildung 4.2). Die Hydroxylierung und die Oxidation zu einem Aldehyd sind in ihrer Reihenfolge selektiv und eignet sich für eine nähere Untersuchung [25].

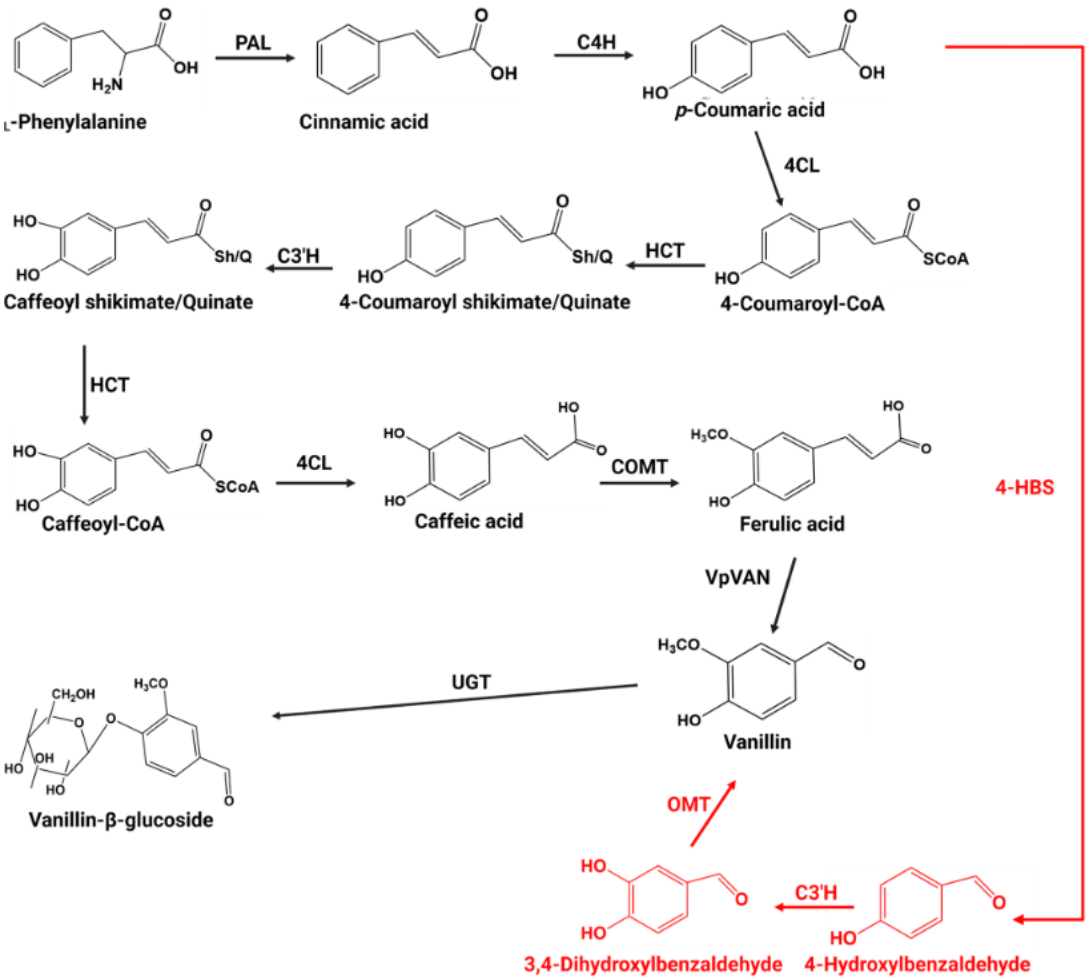


Abbildung 4.2: Synthese von Vanilline in der Vanille Orchidee. Die Synthese erfolgt aus der Aminosäure Phenylalanin. Dabei wird zuerst die Aminogruppe eliminiert, anschließend der Phenylrest hydroxyliert und die aus der Eliminierung entstandene Doppelbindung oxidiert (rot). Der letzte Schritt ist die Methylierung des 4-Hydroxyrests [25].

4.2 Graphtransformation: Ein-Molekül-Reaktionen

Für die Ein-Molekül-Reaktionen wurden zuerst in dem Graphtransformationstool Transformationsregeln definiert. Diese Regeln wurden dann auf ein Startmolekül angewandt, welches Teil des natürlichen Syntheseprozesses ist. Der betrachtete Syntheseausschnitt wurde dabei mit möglichst wenigen und möglichst einfachen Regeln dargestellt.

4.2.1 Koffein

Für die Synthese von Koffein wurde mit dem Purinring Xanthin gestartet (Abbildung 4.3). Ebenso wurde für die Transformationsregel zum Austausch eines an einen Stickstoff gebundenen Wasserstoff durch eine Methylgruppe definiert. (Abbildung 4.4). Durch das wiederholte anwenden der Regel auf Xanthin konnte ein Reaktionsnetzwerk entwickelt

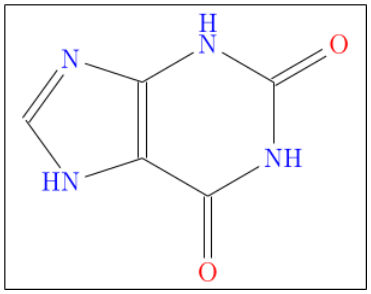


Abbildung 4.3: Darstellung von Xanthin in dem Graphtransformationstool.

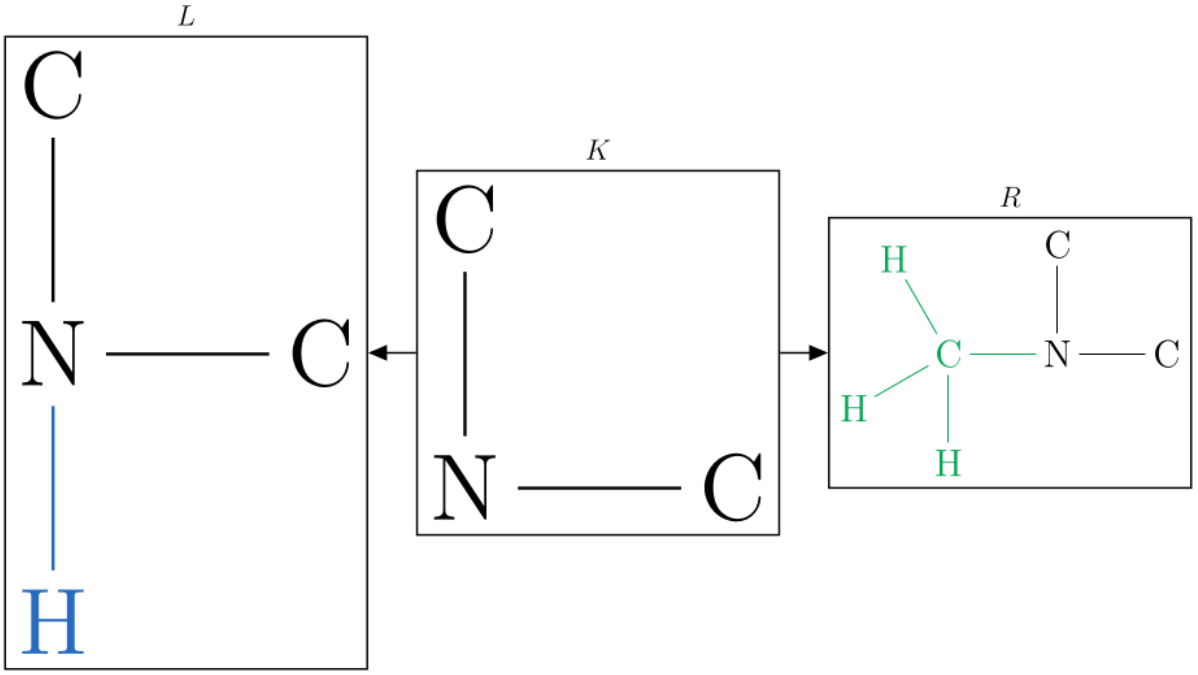


Abbildung 4.4: Darstellung der Regel zum Methylieren eines protonierten Stickstoffs in dem Graphtransformationstool.

werden, an dessen Ende die Synthese von Koffein steht (Abbildung 4.5). Dabei wurden die folgenden Zwischenprodukte bestimmt.

- $p_{0,0}$: 1-Methylxanthin
- $p_{0,1}$: 3-Methylxanthin
- $p_{0,2}$: 7-Methylxanthin
- $p_{0,3}$: 1,3-Dimethylxanthin oder Theophyllin
- $p_{0,4}$: 1,7-Dimethylxanthin oder Paraxanthin
- $p_{0,5}$: 3,7-Dimethylxanthin oder Theobromin

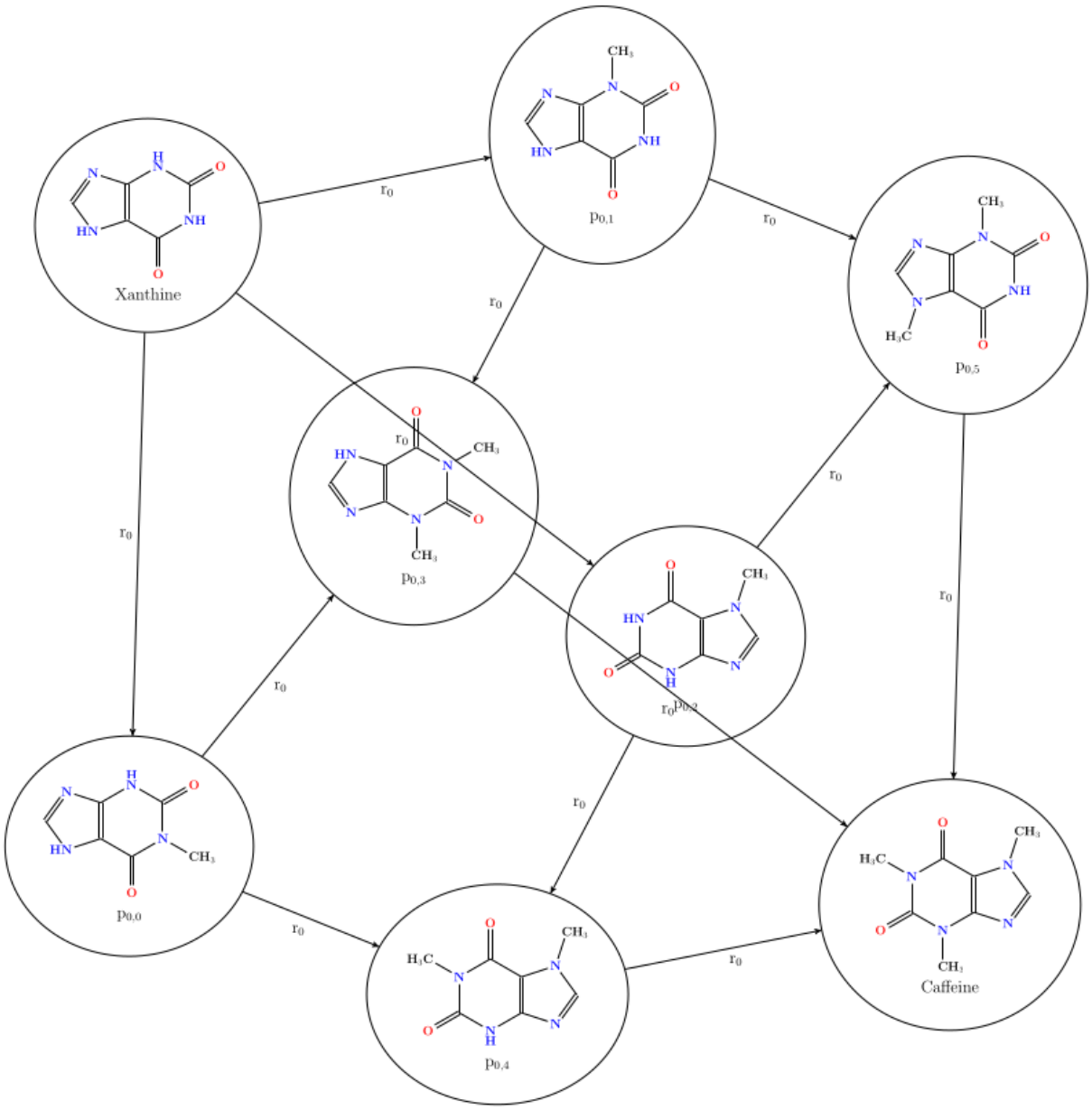


Abbildung 4.5: Darstellung aller Synthesepfade von Koffein aus Xanthin in dem Graphtransformationstool. Die Graphen $p_{0,0}$ bis $p_{0,5}$ stellen dabei die Zwischenprodukte 1-Methylxanthin, 3-Methylxanthin, 7-Methylxanthin, 1,3-Dimethylxanthin oder Theophylline, 1,7-Dimethylxanthin oder Paraxanthin und 3,7-Dimethylxanthin oder Theobromin dar.

4.2.2 Vanillin und 3,4-Dihydroxyaldehyd

Für die Synthese von Vanillin wurde mit Zimtsäure gestartet (Abbildung 4.6). Das natürliche Produkt ist dabei E-Zimtsäure.

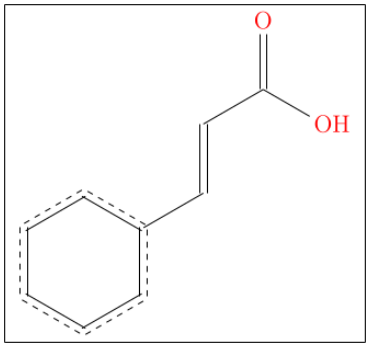


Abbildung 4.6: Darstellung von Zimtsäure in dem Graphtransformationstool.

Zur Umsetzung der Zimtsäure zu dem Vanillin Zwischenprodukts, 3,4-Dihydroxybenzaldehyd, wurde eine Regel zur Hydroxylierung der aromatischen Kohlenstoffe (Abbildung 4.7) und zur Oxidierung einer Doppelbindung mit Carboxylgruppe zu einem Aldehyd (Abbildung 4.8) definiert. Durch das wiederholte Anwenden beider Regeln auf Zimtsäure

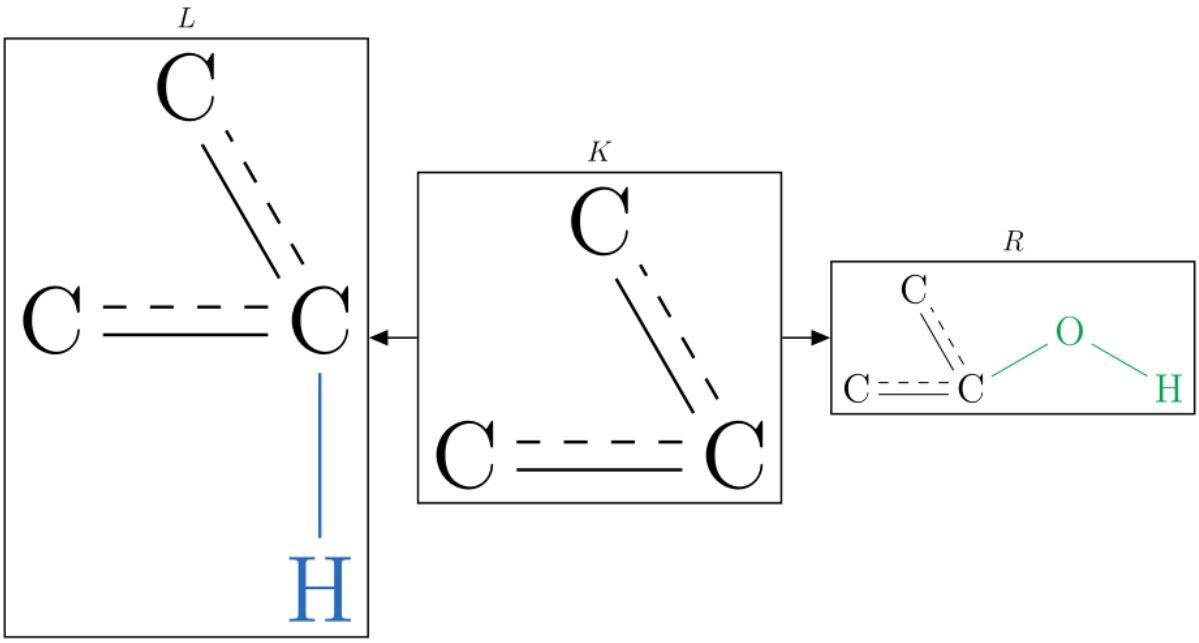


Abbildung 4.7: Darstellung der Regel zum Hydroxylieren des Phenylrests in dem Graphtransformationstool.

konnte ein Reaktionsnetzwerk entwickelt werden, an dessen Ende die Synthese von 3,4-Dihydroxybenzaldehyd steht (Abbildung 4.9). Die breite Regelanwendung bedeutet die

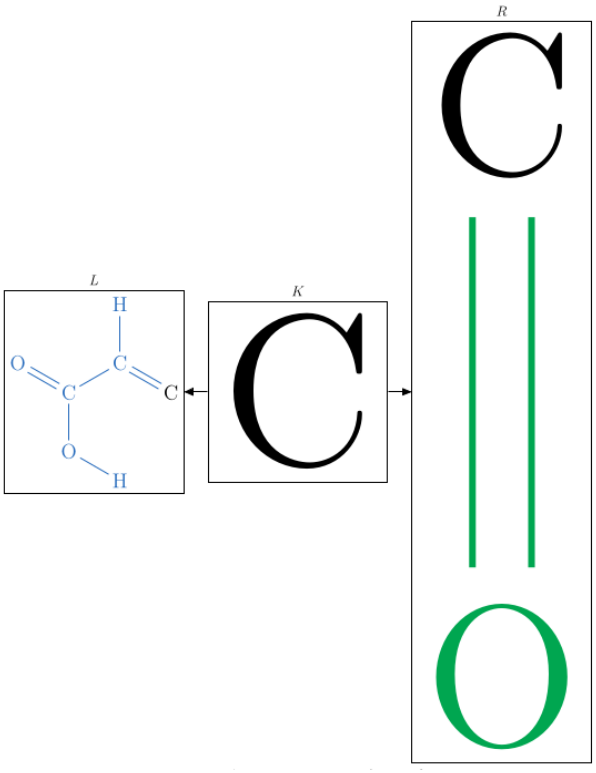


Abbildung 4.8: Darstellung der Regel zum oxidieren der Zimtsäure Doppelbindung zu einem Aldehyd in dem Graphtransformationstool.

Ermittlung vieler Moleküle, welche nicht mehr zu dem Zielmolekül umgesetzt werden können. Darum wurden die Zwischenprodukte auf die Moleküle beschränkt, die tatsächlich Zwischenprodukte darstellen. Dabei wurden die folgenden Zwischenprodukte bestimmt.

- $p_{0,4}$: 3-Hydroxzimtsäure oder m-Coumarsäure
- $p_{0,5}$: 4-Hydroxzimtsäure oder p-Coumarsäure
- $p_{0,6}$: Benzaldehyd
- $p_{0,13}$: 3,4-Dihydroxzimtsäure oder Kaffeesäure
- $p_{0,16}$: 3-Hydroxybenzaldehyd
- $p_{0,17}$: 4-Hydroxybenzaldehyd

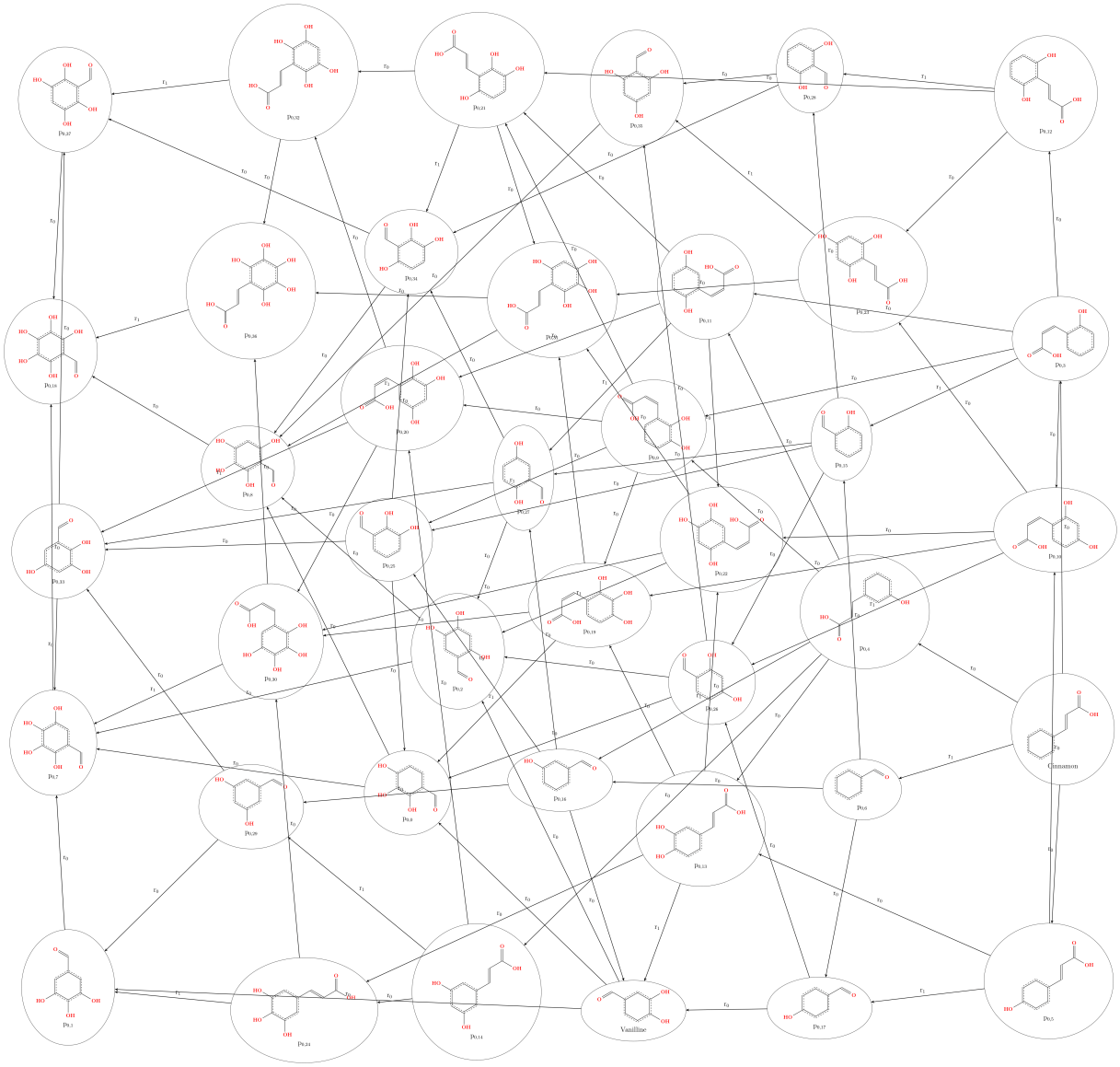


Abbildung 4.9: Darstellung aller Synthesepfade von 3,4-Dihydroxybenzaldehyd aus Zimtsäure in dem Graphtransformationstool. Nur ein Teil aller Zwischenprodukte wird dabei tatsächlich zu dem gewünschten Produkt.

4.3 NMR: Ein-Molekül-Reaktionen

Für alle Zwischenprodukte wurde mit dem quantenchemischen Rechnungsprogramm ein ^{13}C-NMR berechnet und ein dazugehöriges Literaturspektrum recherchiert.

4.3.1 Koffein

Für alle Xanthinringe lässt sich sagen, dass ausschließlich die Protonen der Methygruppen chemisch-äquivalent zueinander sind. In einem ^{13}C-NMR weisen also alle Kohlenstoffe ein eigenes Signal mit Intensität 1 auf. Die mit dem quantenchemischen Rechnungsprogramm berechneten NMR-Spektren wurden den recherchierten Literaturspektren gegenübergestellt (Tabellen 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8) und ihre Differenz zueinander bestimmt.
Die Methylwasserstoffe lassen sich dabei den Signalen um etwa 30 ppm zuordnen. Das 1-Methylsignal weist mit 28.6 bis 30.18 von dem quantenchemischen Rechnungsprogramm berechnet und 26.7 bis 27.7 in den experimentellen Spektren den niedrigsten Wert auf. Die mittlere Verschiebung weist das 3-Methylsignal mit 31.0 bis 32.1 von dem quantenchemischen Rechnungsprogramm berechnet und 28.8 bis 29.7 in den experimentellen Spektren auf. Die größte Verschiebung ist dem 7-Methylsignal mit 36.7 bis 36.9 von dem quantenchemischen Rechnungsprogramm berechnet und 32.9 bis 33.9 in den experimentellen Spektren zuzuordnen. Die Signale des Purinrings sind für den größten Teil konstant über alle Moleküle, wobei Xanthin mit 151.03 von dem quantenchemischen Rechnungsprogramm berechnet, zu 140.5 im Experiment eine besonders große Differenz aufweist. Die größten Unterschiede zwischen den berechneten und experimentellen Spektren sind über alle Xanthine bei dem Kohlenstoff neben dem 7-Stickstoff. Über alle Moleküle gemittelt lässt sich bestimmen, dass die von dem quantenchemischen Rechnungsprogramm berechneten Spektren im Durchschnitt um 2.5 ppm nach oben verschoben sind.

Tabelle 4.1: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Xanthin mit einem Literatur Spektrum [26]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.14 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
112.57	108.8	3.77
138.61	137.2	1.413
151.03	140.5	10.53
152.87	152.4	0.47
155.64	156.1	-0.46
		3.14

Tabelle 4.2: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 1-Methylxanthin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.39 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
30.18	27	3.18
112.28	106.4	5.88
137.88	140.6	-2.72
150.37	147	3.37
153.13	151.2	1.93
157.91	155.2	2.71
		2.39

Tabelle 4.3: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 3-Methylxanthin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.35 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
31.34	28.8	2.54
112.87	106.9	5.97
138.12	140.5	-2.38
153.46	149.4	4.06
154.78	151.2	3.58
155	154.7	0.30
		2.35

Tabelle 4.4: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 7-Methylxanthin mit einem Literatur Spektrum [28]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.73 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
36.69	33.03	3.66
113.78	106.9	6.88
143.18	143.01	0.17
151.13	149.3	1.83
153.94	151.35	2.59
157.1	155.85	1.25
		2.73

Tabelle 4.5: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Theophyllin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.61 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
29.23	27.7	1.53
32.05	29.7	2.35
112.36	106.4	5.96
137.43	140.5	-3.07
152.62	147.8	4.82
155.4	151.2	4.20
156.89	154.4	2.49
		2.61

Tabelle 4.6: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Paraxanthin mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.19 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
28.59	26.7	1.89
36.8	32.9	3.9
113.44	106.5	6.94
142.72	143	-0.28
151.81	147.4	4.41
153.04	151.1	1.94
158.8	155.3	3.5
		3.19

Tabelle 4.7: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Theobromin mit einem Literatur Spektrum [29]. Die Methylsignale (etwa 30 ppm) wurden aus einem anderen Spektrum genommen, da diese Angabe fehlte [30]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.63 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
30.96	29.3	1.66
36.7	33.9	2.8
114.14	108.1	6.04
142.73	142.6	0.13
153.39	150	3.39
155.92	153.1	2.82
156.49	154.9	1.59
		2.63

Tabelle 4.8: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Koffein mit einem Literatur Spektrum [27]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.77 ppm.

Berechnet	Exp	Differenz: Berechnet - Exp
28.8	27.5	1.3
31.72	29.3	2.42
36.86	33.1	3.76
113.66	106.6	7.06
142.13	142.8	-0.67
153.86	148.1	5.76
155.29	151	4.29
158.29	154.5	3.79
		3.77

4.3.2 Vanillin und 3,4-Dihydroxyaldehyd

Für 3-hydroxylierte Zwischenprodukte gilt, dass keine Kerne chemisch-äquivalent zueinander sind. Wenn nicht oder nur in der 4-Position hydroxyliert ist, sind jeweils die 2 mit 6 und 3 mit 5 Kohlenstoffe im Phenylrest, sowie die jeweils an diese gebundenen Wasserstoffe zueinander chemisch-äquivalent. In einem ^{13}C-NMR weisen also alle Kohlenstoffe in 3-hydroxilierten Moleküle ein eigenes Signal mit Intensität 1 auf. In nicht oder 4-hydroxylierten Molekülen verschmelzen die Signale der Kohlenstoffe in der 2 und 6 Position sowie die in der 3 und 5 Position. Es resultieren zwei Signalen mit Intensität 2. Die mit dem quantenchemischen Rechnungsprogramm berechneten NMR-Spektren wurden den recherchierten Literaturspektren gegenübergestellt (Tabellen 4.9, 4.10, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16) und ihre Differenz zueinander bestimmt.
Bei den Zwischenprodukten der 3,4-Dihydroxybenzaldehyd-Synthese lässt sich im Vergleich zu den Zwischenprodukten der Koffein-Synthese kein so konstantes Muster erkennen. Das ist unter anderem darauf zurückzuführen, dass der deutliche Elektronegativitäten-Unterschied zwischen Sauerstoff und Wasserstoff einen großen Einfluss auf das Ringsystem erwirkt. Dadurch sind die Signale des Kerngerüsts deutlich weniger konstant. Das Endprodukt 3,4-Dihydroxybenzaldehyd sticht dabei trotzdem heraus, da es im Gegensatz zu allen anderen Spektren durchschnittlich eine größere berechnete chemische Verschiebung als experimentell bestimmte chemische Verschiebung aufweist. Über alle Spektren gemittelt haben die mit dem quantenchemischen Rechnungsprogramm berechneten Spektren trotzdem eine um 2.8 ppm größere chemische Verschiebung.

Tabelle 4.9: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Zimtsäure mit einem Literatur Spektrum [31]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 4.02 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
1	117.33	120.33	3
2	128.39	133.48	5.09
2	128.98	134.2	5.22
1	130.77	135.91	5.14
1	134.04	140.47	6.43
1	147.14	151.41	4.27
1	172.6	171.56	-1.04
			4.02

Tabelle 4.10: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für m-Coumarsäure mit einem Literatur Spektrum [32]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.53 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
1	115.3	118.6	3.3
1	118.6	121.09	2.49
1	119.2	121.69	2.49
1	120.8	122.36	1.56
1	131	134.46	3.46
1	137.1	141.29	4.19
1	146.5	153.69	7.19
1	159.1	164.15	5.05
1	170.5	172.55	2.05
			3.53

Tabelle 4.11: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für p-Coumarsäure mit einem Literatur Spektrum [33]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 4.71 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
1	115.41	117.18	1.77
2	115.83	118.65	2.82
1	125.36	132.76	7.40
2	130.17	135.75	5.58
1	144.27	149.67	5.40
1	159.67	166.09	6.42
1	168.05	171.61	3.56
			4.71

Tabelle 4.12: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Benzaldehyd mit einem Literatur Spektrum [34]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 4.80 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
2	128.9	133.45	4.55
2	129.5	135.86	6.36
1	134.2	139.15	4.95
1	136.4	141.47	5.07
1	192	195.09	3.09
			4.80

Tabelle 4.13: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für Kaffeesäure mit einem Literatur Spektrum [35]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 2.18 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
1	147.9	155.79	7.89
1	146.5	149.88	3.38
1	129.3	132.99	3.69
1	123.1	131.94	8.84
1	121.7	118.15	-3.55
1	116.4	116.16	-0.24
1	141.5	149.28	7.78
1	114.6	110.95	-3.65
1	176.2	171.66	-4.55
			2.18

Tabelle 4.14: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 3-Hydroxybenzaldehyd mit einem Literatur Spektrum [36]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.83 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
1	115	122.95	7.95
1	122	123.33	1.33
1	123	124.61	1.61
1	131	133.96	2.96
1	138	143.41	5.41
1	158	164.39	6.39
1	194	195.16	1.16
			3.83

Tabelle 4.15: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 4-Hydroxybenzaldehyd mit einem Literatur Spektrum [37]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt 3.35 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
2	116.5	118.46	1.96
1	130.3	135.08	4.78
2	132.7	137.69	4.99
1	163.8	168.28	4.48
1	191	191.53	0.53
			3.35

Tabelle 4.16: Vergleich des mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum für 3,4-Dihydroxybenzaldehyd mit einem Literatur Spektrum [38]. Die Werte sind in ppm angegeben. Die durchschnittliche Differenz beträgt -3.81 ppm.

Height	Exp	Berechnet	Differenz: Berechnet - Exp
1	113.1	114.22	1.12
1	117.79	115.21	-2.58
1	133.4	124.59	-8.81
1	134.66	130.13	-4.53
1	149.32	145.44	-3.88
1	158.08	151.27	-6.81
1	191.41	190.26	-1.15
			-3.81

4.4 NMR SI: Ein-Molekül-Reaktionen

In bisherigen Analysen zu der Genauigkeit von berechneten NMR-Spektren im Vergleich zu experimentellen Spektren wurde festgestellt, dass einfache DFT Methoden die Effekte von schweren Atomen wie dem in TMS enthaltenden Silizium nicht akkurat bestimmen können. Insbesondere kann davon ausgegangen werden, dass die Frequenz des TMS-Kohlenstoff zu gering ausfällt [39, 40]. Das führt wiederum dazu, dass die chemische Verschiebung der Signale im mit dem quantenchemischen Rechnungsprogramm berechneten Spektrum zu hoch ausfällt. Dieser Wert ist systematisch, weshalb er als ungefähres Mittel der Unterschiede zwischen den mit dem quantenchemischen Rechnungsprogramm berechneten und experimentellen Spektren quantifiziert wurde. Auf alle Peaks der experimentellen Spektren wurden 2.5 ppm hinzuaddiert. Die Bingröße b variiert in Literaturangaben zwischen 1 und 0.1 ppm [22, 41] mit einer Genauigkeit von 0.1 ppm. Deshalb wurde die Bingröße b (Gleichung 4.1) variiert und die resultierenden Ähnlichkeitsindizes SI_b über alle sinnvollen Werte gemittelt.

$$b \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\} \tag{4.1}$$

4.4.1 Koffein

Den Erwartungen entsprechend weist das mit dem quantenchemischen Rechnungsprogramm berechnete Spektrum von 7-Methylxanthin eine hohe Ähnlichkeit zu dem experimentellen Spektrum von 7-Methylxanthin auf. Das Spektrum von Paraxanthin ist jedoch noch etwas ähnlicher. Die Spektren von Theobromin haben im Vergleich dazu die größten Übereinstimmungen, wobei Koffein am zweit ähnlichsten ist. Allgemein scheint eine Identifikation des 7-Methylrest gut zu funktionieren, jedoch keine einfache Unterscheidung der restlichen Methylgruppen möglich zu sein. Da in allen Molekülen das Xanthin-Grundgerüst vorliegt und die Addition einer Methylgruppe dieses nicht zu sehr verändert, haben alle Moleküle eine signifikante Ähnlichkeit zueinander (Tabelle 4.17).

Tabelle 4.17: Resultierende normalisierte Wahrscheinlichkeiten beim Vergleich aller Zwischenprodukte der Koffeinsynthese, sowie Xanthin und Koffein mit den natürlichen Zwischenprodukten 7-Methyxanthin und Theobromin.

Molekül	7-Methylxanthin	Theobromin
Xanthin	0.08	0.14
1-Methylxanthin	0.11	0.09
3-Methylxanthin	0.08	0.11
7-Methylxanthin	0.16	0.12
Theophylline	0.11	0.09
Paraxanthin	0.19	0.10
Theobromin	0.07	0.18
Koffein	0.19	0.17

4.4.2 Vanillin und 3,4-Dihydroxyaldehyd

Bei der Vanillin Synthese liegen die chemischen Verschiebungen des Grundgerüsts weiter auseinander, weshalb manche Moleküle, wie 3-Hydroxybenzaldehyd nur eine sehr geringe Ähnlichkeit zu den natürlichen Zwischenprodukten aufweisen. Die p-Coumarsäure hat die größte Ähnlichkeit zu sowohl seinem eigenen experimentellen Spektrum, sowie zu dem von 4-Hydoxybenzaldehyd. Die zweitgrößte Ähnlichkeit zu dem experimentellen Spektrum von p-Coumarsäure haben Kaffeesäure und 4-Hydroxybenzaldehyd. Bei dem experimentellen Spektrum von 4-Hydroxybenzaldehyd hat das richtige Molekül die zweit größte Ähnlichkeit (Tabelle 4.18).
Im Vergleich zu den Xanthinen sind die Abweichungen der berechneten zu den experimentellen Spektren bei den Vanillevorprodukten deutlich weniger konstant. So hat Kaffeesäure mit 2.35 ppm die geringste Abweichung zwischen beiden zugehörigen Spektren. Trotzdem ist die absolute Verschiebung der einzelnen Peaks zueinander relativ groß. Darum ist die Varianz der Werte ist deutlich größer als bei den Spektren der Xanthine, weshalb die nicht normalisierte Crosskorrelation allgemein geringer ausfällt. Das berechnete Spektrum von Kaffeesäure hat dabei zu vielen Peaks in dem Literaturspektrum von p-Coumarsäure eine Distanz von unter 1.0 ppm, weshalb der Überlapp größer als erwartet ausfällt. Diese überlappenden Peaks weichen besonders stark von den Peaks im eigenen Referenzspektrum ab. Die berechneten Spektren von p-Coumarsäure und 4-Hydroxybenzaldehyd haben nur wenige überlappende Peaks mit beiden Literaturspektren. Diese überlappenden Peaks sind haben jedoch eine Intensität von 2, welche durch die Multiplikation der Intensitäten stärker gewichtet werden. Allgemein haben alle para-hydroxylierten Phenylgruppen größeren Überlapp mit der p-Coumarsäure. Bei 4-Hydroxybenzaldehyd scheint der Überlapp größer bei Molekülen ohne 3-Hydroxygruppe zu sein.

Tabelle 4.18: Resultierende normalisierte Wahrscheinlichkeiten beim Vergleich aller Zwischenprodukte der Vanillesynthese mit den natürlichen Zwischenprodukten p-Coumarsäure und 4-Hydroxybenzaldehyd.

Molekül	p-Coumarsäure	4-Hydroxybenzaldehyd
Zimtsäure	0.06	0.11
p-Coumarsäure	0.25	0.29
m-Coumarsäure	0.08	0.08
Benzaldehyd	0.06	0.12
Kaffeesäure	0.22	0.10
3-Hydroxybenzaldehyd	0.01	0.02
4-Hydroxybenzaldehyd	0.22	0.17
3,4-Dihydroxybenzaldehyd	0.11	0.11

4.5 ILP: Ein-Molekül-Reaktionen

Der Hypergraph der komplett explorierten Synthese wurde an Gurobi übergeben und um eine Kante, die in das Startmolekül hineinfließt und aus dem Zielmolekül herausfließt, erweitert. Die Wahrscheinlichkeiten wurden als konstante Werte $p_{i,e}$ den Kanten e des Hypergraphen zugeordnet. Die Wahrscheinlichkeit des Moleküls wurde den Kanten zugeordnet, welche in den Knoten fließen. Die Kante 4 von Zimtsäure zu Benzaldehyd hat also die Wahrscheinlichkeit $p_{1,4} = 0.06$, dass diese Reaktion dem ersten Referenzspektrum mit p-Coumarsäure entspricht. Für jede Kante werden vier ganzzahlige Variablen x_e, b_e, $n_{1,e}$ und $n_{2,e}$ definiert. Die Variable b_e ist dabei eine Wahrheitsvariable die definiert, ob eine Kante verwendet wird. Die Variable x_e beschreibt wie oft die selbe Kante verwendet wird. Die Variablen $n_{1,e}$ und $n_{2,e}$ sind ebenfalls Wahrheitsvariablen und legen fest, welche zwei Reaktionen den Referenz NMR-Spektren entsprechen und deswegen zu der Zielfunktion dazuzählen. Dabei ist $n_{1,e}$ die Zuordnungsvariable zu dem ersten Referenzspektrum mit 7-Methylxanthin oder p-Coumarsäure und $n_{2,e}$ die Zuordnungsvariable zu dem zweiten Referenzspektrum mit 3,7-Methylxanthin oder 4-Hydroxybenzaldehyd. Die Variablen unterliegen den Beschränkungen, dass:

- $x_e \geq b_e$
- $b_e \geq n_{1,e}$
- $b_e \geq n_{2,e}$
- s: ([], [Startmolekül]), $x_s = 1$
- f: ([Zielmolekül], []), $x_f = 1$
- $\forall v \in V$ inflow = outflow

- $n_{1,e} + n_{2,e} \leq 1$
- $\sum_e n_{1,e} = 1$
- $\sum_e n_{2,e} = 1$

Unter diesen Beschränkungen soll ein Pfad durch den Hypergraphen bestimmt werden, welcher ein geordnetes Paar von Molekülen auswählt. Dieses Paar von Molekülen soll alle Zwischenprodukte der Synthese beinhalten und addiert die größte Wahrscheinlichkeit aufweisen. Die ausgehenden Kanten sind nicht in die Zielfunktion eingeflossen (Gleichung 4.2).

$$max\ z = \sum_e 1000 \times (p_{1,e} \times n_{1,e} + p_{2,e} \times n_{2,e}) - \frac{x_e}{max(p_{1,e}, p_{2,e})} \tag{4.2}$$

4.5.0.1 Koffein

Der Hypergraph für die Koffeinsynthese wird in Tabelle 4.19 dargestellt. Die beste gefun-

Tabelle 4.19: Der für die Koffeinsynthese definierte Hypergraph in Gurobi mit den für jede Kante definierten Wahrscheinlichkeiten.

Index	Edukt	Produkt	$n_{\text{7-Methylxanthin},e}$	$n_{\text{3,7-Dimethylxanthin},e}$
1		Xanthin	0.08	0.14
2	Xanthin	1-Methylxanthin	0.11	0.09
3	Xanthin	3-Methylxanthin	0.08	0.11
4	Xanthin	7-Methylxanthin	0.16	0.12
5	1-Methylxanthin	Theophyllin	0.11	0.09
6	1-Methylxanthin	Paraxanthin	0.19	0.10
7	3-Methylxanthin	Theophyllin	0.11	0.09
8	3-Methylxanthin	Theobromin	0.07	0.18
9	7-Methylxanthin	Paraxanthin	0.19	0.10
10	7-Methylxanthin	Theobromin	0.07	0.18
11	Theophyllin	Koffein	0.19	0.17
12	Paraxanthin	Koffein	0.19	0.17
13	Theobromin	Koffein	0.19	0.17
14	Koffein		0.0	0.0

dene Lösung ist für die Kanten 1, 4, 10, 13 und 14 jeweils ein flow $x_e = 1$ und alle anderen Kanten ein flow $x_e = 0$ (Tabelle 4.20). Dies entspricht der natürlich auftretenden Lösung.

Tabelle 4.20: Lösung von Gurobi für die Koffeinsynthese.

Index	Edukt	Produkt	flow x_e
1		Xanthin	1
2	Xanthin	1-Methylxanthin	0
3	Xanthin	3-Methylxanthin	0
4	Xanthin	7-Methylxanthin	1
5	1-Methylxanthin	Theophyllin	0
6	1-Methylxanthin	Paraxanthin	0
7	3-Methylxanthin	Theophyllin	0
8	3-Methylxanthin	Theobromin	0
9	7-Methylxanthin	Paraxanthin	0
10	7-Methylxanthin	Theobromin	1
11	Theophyllin	Koffein	0
12	Paraxanthin	Koffein	0
13	Theobromin	Koffein	1
14	Koffein		1

4.5.1 Vanillin und 3,4-Dihydroxyaldehyd

Der Hypergraph für die Vanillinsynthese wird in Tabelle 4.21 dargestellt. Die beste ge-

Tabelle 4.21: Der für die Vanillinsynthese definierte Hypergraph in Gurobi mit den für jede Kante definierten Wahrscheinlichkeiten.

Index	Edukt	Produkt	$n_{\text{p-Coumarsäure},e}$	$n_{\text{4-Hydroxybenzaldehyd},e}$
1		Zimtsäure	0.06	0.11
2	Zimtsäure	p-Coumarsäure	0.25	0.29
3	Zimtsäure	m-Coumarsäure	0.08	0.08
4	Zimtsäure	Benzaldehyd	0.06	0.12
5	p-Coumarsäure	Kaffeesäure	0.22	0.10
6	p-Coumarsäure	4-Hydroxybenzaldehyd	0.22	0.17
7	m-Coumarsäure	Kaffeesäure	0.22	0.10
8	m-Coumarsäure	3-Hydroxybenzaldehyd	0.01	0.02
9	Benzaldehyd	3-Hydroxybenzaldehyd	0.01	0.02
10	Benzaldehyd	4-Hydroxybenzaldehyd	0.22	0.17
11	Kaffeesäure	3,4-Dihydroxybenzaldehyd	0.11	0.11
12	3-Hydroxybenzaldehyd	3,4-Dihydroxybenzaldehyd	0.11	0.11
13	4-Hydroxybenzaldehyd	3,4-Dihydroxybenzaldehyd	0.11	0.11
14	3,4-Dihydroxybenzaldehyd		0.0	0.0

fundene Lösung ist für die Kanten 1, 2, 6, 13 und 14 jeweils ein flow $x_e = 1$ und alle anderen Kanten ein flow $x_e = 0$ (Tabelle 4.22). Dies entspricht der natürlich auftretenden Lösung. Dabei ist jedoch anzumerken, dass die Lösung des ILP für das erste Moleküle 4-Hydroxybenzaldehyd und für das zweite Molekül p-Coumarsäure gewählt hat. Die Zuordnung erfolgte also genau anders herum. Zudem ist davon auszugehen, dass die zweitbeste Lösung mit erst Kaffeesäure und dann p-Coumarsäure genau den selben Zielwert aufweist.

Tabelle 4.22: Lösung von Gurobi für die Vanillinsynthese.

Index	Edukt	Produkt	flow x_e
1		Zimtsäure	1
2	Zimtsäure	p-Coumarsäure	1
3	Zimtsäure	m-Coumarsäure	0
4	Zimtsäure	Benzaldehyd	0
5	p-Coumarsäure	Kaffeesäure	0
6	p-Coumarsäure	4-Hydroxybenzaldehyd	1
7	m-Coumarsäure	Kaffeesäure	0
8	m-Coumarsäure	3-Hydroxybenzaldehyd	0
9	Benzaldehyd	3-Hydroxybenzaldehyd	0
10	Benzaldehyd	4-Hydroxybenzaldehyd	0
11	Kaffeesäure	3,4-Dihydroxybenzaldehyd	0
12	3-Hydroxybenzaldehyd	3,4-Dihydroxybenzaldehyd	0
13	4-Hydroxybenzaldehyd	3,4-Dihydroxybenzaldehyd	1
14	3,4-Dihydroxybenzaldehyd		1

4.6 Fazit: Ein-Molekül-Reaktionen

Bei den bisher betrachteten Ein-Molekül-Reaktionen wurde immer nur ein festgelegtes Startmolekül in das System gegeben. Dieses Startmolekül wurde durch Transformationsregeln verändert, wobei nach dem natürlichen Vorbild genau ein Zielmolekül synthetisiert werden sollte. In dem Koffein Beispiel gab es keine Enantiomere oder E/Z-Isomere. In dem Vanille Beispiel gab es keine Enantiomere. Es gab mit Zimtsäure, Coumarsäure und Kaffeesäure zwar E/Z-Isomere, welche aber durch die natürliche Vorgabe von E-Zimtsäure bereits vorgegeben waren. Außerdem entstand keine neue Doppelbindung mit einem E/Z-Isomer durch Graphtransformation. Der Schluss von dem Graphen auf eine Geometrie war deswegen immer eindeutig.
Die Berechnung der NMR-Spektren war trotzdem fehleranfällig. Dabei konnten bestimmte Molekülveränderungen besser in dem NMR-Berechnungstool abgebildet werden. Eine Methylierung erzeugte nur eine geringe Veränderung im Xanthingrundgerüst, weshalb die Zuordnung vor allem in dem Signal der Methylgruppe lag. Die Hydroxylierung und Oxidierung zum Aldehyd resultierten in deutlichen Veränderungen der Phenylgruppe, weshalb das Grundgerüst selber verändert wurde.
Für die Synthesen waren immer genau drei Reaktionen nötig, welche mittels des ILP nur bezüglich ihrer Reihenfolge variiert werden konnten. Das Problem ließe sich daher ebenfalls durch die einfache Addition zweier Moleküle entlang eines ausgewählten Pfades lösen. Der Raum möglicher Lösungen ist mit 6 pro Hypergraph stark begrenzt. Trotzdem eigneten sich beide Beispiele mit der natürlich auftretenden Selektivität der Reihenfolge, der größtenteils guten Differenzierbarkeit der NMR-Spektren und Kompaktheit der Synthese als Graphtransformationsregeln als Demonstration des Konzeptes. Es ist anzumerken, dass die erwartete Synthese über die Auswahl der Referenzspektren festgelegt wurde. Diese Synthese wurde über NMR-Spektren abstrahiert. Es lässt sich so sagen,

dass die NMR-Spektren in dieser Analyse die Struktur hinreichend codieren, aber nicht eine unbekannte Synthese vollständig hergeleitet wurde.
Für ausgewählte Literaturspektren von nicht in der natürlichen Synthese erwarteten Molekülen wurden ebenfalls der SI mit den berechneten Spektren ermittelt. Dabei konnte ebenfalls eine erhöhte Korrelation mit dem zugehörigen berechneten Spektrum festgestellt werden. Das ILP konnte dabei den neuen Syntheseplan ebenfalls zuverlässig zuordnen. Die Auswahl eines Syntheseplans basierend auf der Ähnlichkeit von NMR-Spektren ist daher gerechtfertigt und stellt eine Verbesserung gegenüber der Auswahl ohne selbige dar.

4.7 Multi-Molekül-Reaktionen

Die bisherigen Ein-Molekül-Reaktionen lassen sich durch ein ILP lösen, erfordern die Verwendung jedoch nicht. Um die Grenzen des ILP weiter auszureizen, wird ein komplexerer Hypergraph entwickelt. Dieser Hypergraph wird mit wenigen, einfachen Transformationsregeln und mit einem Ausgangsmolekül erzeugt. Das Ziel ist ein hochvernetztes Reaktionsnetzwerk.

4.8 Olefinmetathese und Diels-Alder

Die zwei Reaktionen Olefinmetathese in Kombination mit der Diels-Alder-Reaktion erlaubt mit dem richtigen Ausgangsmolekülen die Entwicklung eines hochvernetzten Reaktionsnetzwerks und unterliegenden Hypergraphen.

4.8.1 Olefinmetathese

Die Olefinmetathese ist eine vergleichbar neuere Reaktion [42]. Mit einem organischen Ruthenium-Katalysator werden dabei die Reste zweier Doppelbindungen rekombiniert. Der Ruthenium-Katalysator hat dabei eine offene Bindungsstelle an die eine Hälfte der C-C-Doppelbindung als Carben bindet (A und E). Ein Olefin (B und F) bindet an diesen Komplex (C und G). Aus diesem Komplex wird ein Olefin abgespalten, welches entweder das eingehende Olefin oder eine beliebige Kombination beider Olefine (D) ist (Abbildung 4.10).
In der Forschung wurde dieser katalytische Prozess zu immer gezielteren Ausbeuten durch spezialisierte Katalysatoren weiterentwickelt. Dabei wurde neben der generellen Selektivität der Produkte auch eine gezielte Selektivität hin zu E-Isomeren erzielt. Für die Anwendung auf ein Reaktionsnetzwerk ist eine Produktspezialisierung jedoch weniger interessant. Die E/Z-Selektivität ist dabei insbesondere im Hinblick auf das chemische Molekül als Graphen nicht umsetzbar, wenngleich sie bei der Geometrierealisierung berücksichtigt

werden kann. Die Möglichkeiten der Olefinmetathese sind jedoch begrenzt, wenn es um die Queervernetzung geht. Bei einer Olefinmetathese bleibt die Gesamtzahl an Einfach und Doppelbindungen gleich.

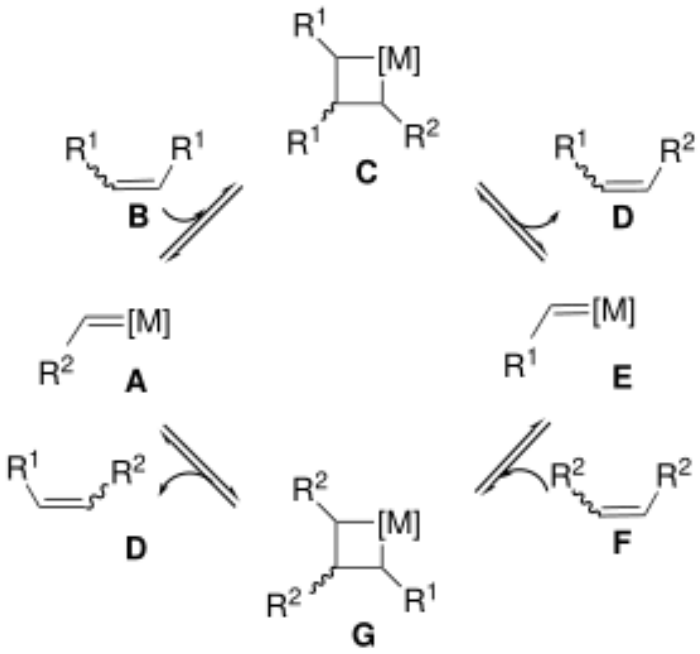


Abbildung 4.10: Katalysezyklus der Olefinmetathese [42].

4.8.2 Diels-Alder

Die Diels-Alder-Reaktion ist eine ohne Katalysator natürlich auftretende Reaktion, bei der zwei Doppelbindungen enthaltenden Moleküle verbunden werden. Dabei wird eine Vinylgruppe und zwei mit einer Einfachbindung verbundenen Vinylgruppen zu einem Cyclohexen Ring verbunden (Abbildung 4.11). Wenn diese Bedingungen in einer Lösung vorhanden sind, kommt es zu dieser Kombination, mit nur geringer Rückreaktion [43]. Die

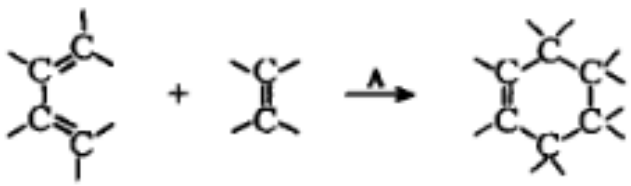


Abbildung 4.11: Die Diels-Alder-Reaktion [43].

Diels-Alder-Reaktion ist einfach und sorgt dafür, dass Doppelbindungen zu Einfachbin-

dungen werden. In der Kombination mit der Olefinmetathese werden so viele verschiedene Produkte möglich.

4.8.3 Butadien

Das gewählte Startmolekül ist Butadien. Das Vorliegen von zwei Doppelbindungen erlaubt es der Olefinmetathese Kohlenstoffketten mit alternierenden Einfach und Doppelbindungen aufzubauen. Diese können wiederum mit der Diels-Alder-Reaktion untereinander kombiniert werden und so zu einer großen Menge an unterschiedlichen Molekülen führen, welche alle untereinander kombinierbar sind.

4.9 Graphtransformation: Multi-Molekül-Reaktionen

Bei der Definition der Transformationsregeln für Multi-Molekül-Reaktionen müssen auf der linken und rechten Seite, sowie im Kontext die gleichen Atome vorliegen. Bei den meisten mehrstufigen, chemischen Synthesen wird ein zentrales Molekül betrachtet. Es kommt zwar vor, dass für eine Reaktion zwei Moleküle zusammen geführt werden, aber Teile des Moleküls, die in Reaktionen abgespalten werden, werden in der restlichen Synthese nicht weiter erwähnt. Bei einer Multi-Molekül-Reaktion sollen diese abgespaltenen Moleküle potenziell wieder miteinander reagieren. Es ist also wichtig, dass eine Transformationsregel in Hinsicht auf Multi-Molekül-Reaktionen alle Atome vollständig weiterführt.

4.9.1 Olefinmetathese und Diels-Alder

In dem Graphtransformationstool wurde mit dem Molekül Butadien (Abbildung 4.12) gestartet. Auf dieses wurden wiederholt die Olefinmetathese (Abbildung 4.13) und die Diels-Alder Reaktion (Abbildung 4.14) angewendet. Bei dieser Reaktion ist es anders als bei den bisherigen Reaktionen möglich, dass der Ergebnisraum unbegrenzt wachsen kann. Darum sind Beschränkungen des Ergebnisraums nötig. Die einfachste Beschränkung der Ergebnisse ist eine Obergrenze für die Anzahl an Atomen. Damit Butadien als Startmolekül diese Begrenzung erfüllt, müssen mindestens 4 Kohlenstoffe zugelassen werden. Damit die Diels-Alder-Reaktion stattfinden kann müssen 6 Kohlenstoffe erlaubt sein. Die Regeln zeigen einen diversen Raum an gefundenen Molekülen bei 10 Kohlenstoffen. Da dies die Computerleistung stark beansprucht, muss der Raum weiter eingeschränkt werden. Die Olefinmetathese ist bekannt dafür, Ringe mit großer Ringspannung aufzubrechen [42]. Besonders strapaziert sind kleine Ringe mit der Größe 3 und 4, sowie große Ringe. Ein Kohlenstoff Ring der Größe 6 ist am stabilsten [44]. Da die Olefinmetathese mit hoher Selektivität Ringe mit hoher Spannung aufbricht, ist davon auszugehen, dass zu große und zu kleine Ringe nicht Teil des Ergebnisraums sind. Insbesondere ist nur der 6 Ring

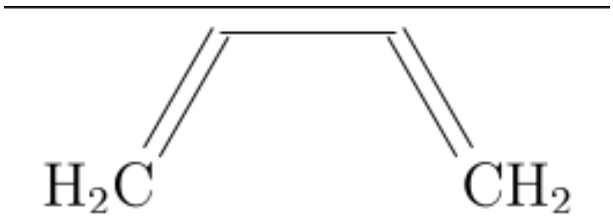


Abbildung 4.12: Butadien in dem Graphtransformationstool.

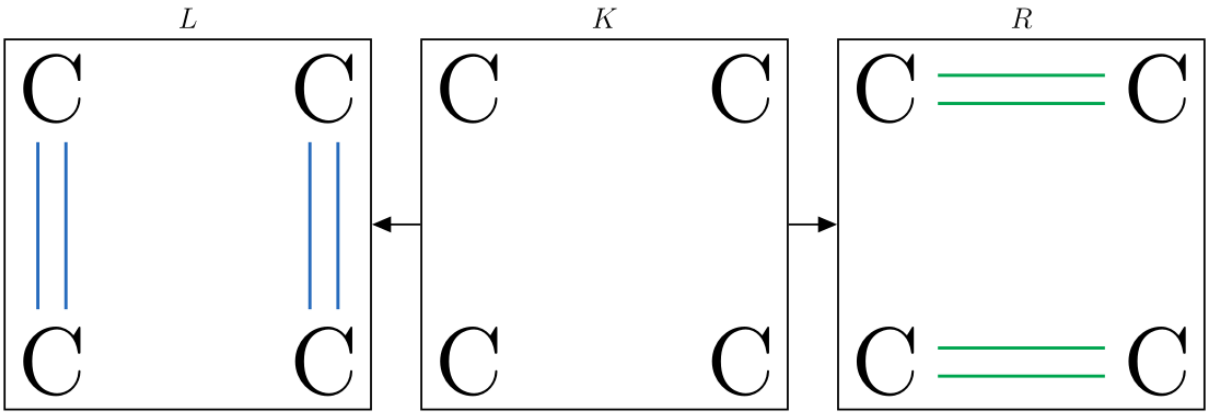


Abbildung 4.13: Olefinmetathese in dem Graphtransformationstool.

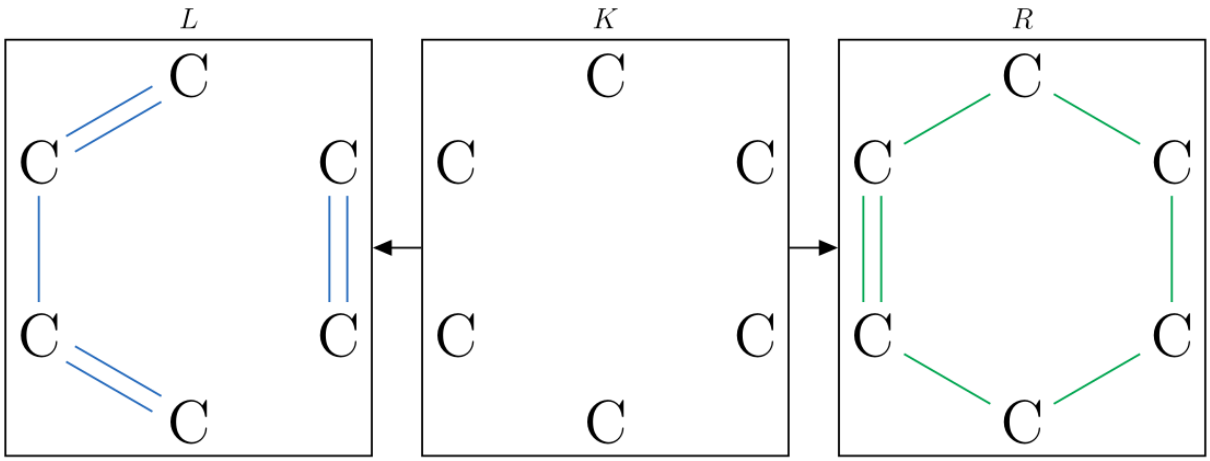


Abbildung 4.14: Diels-Alder-Reaktion in dem Graphtransformationstool.

als die stabilste Ringgröße zu erwarten. Deshalb wurden für jedes gefundene Molekül alle sehnenlosen Ringe gefunden und alle Ringe, die nicht die Größe 6 hatten aus dem Ergebnis Raum entfernt. Eine weitere Quelle der Ringspannung sind überlappende Ringe. Norbornan wird zum Beispiel aus drei überlagernden Ringen gebildet. Da auch in diesen Molekülen hohe Ringspannung vorliegt, wurden alle Moleküle mit Ringen, die mehr als zwei gemeinsame Atome haben aus dem Ergebnisraum entfernt. Die letzte Begrenzung für Ringspannung ist, dass kein Atom in vier Ringen sein kann.
Mit der Beschränkung wurden durch die Regelanwendung 19 Moleküle mit etwa 200 Transformationen aus Butadien entwickelt. Die Moleküle und ihre SMILES-Strings werden im Folgenden aufgeführt.

- Butadien: C=CC=C
- Ethylen: C=C
- Hexatrien: C=CC=CC=C oder C=C/C=C/C=C
- 4-Vinyl-Cyclohex-1-en: C1CCC(C=C)CC=1
- 4-Vinyl-Octa-1,7-Dien: C(CCC(C=C)CC=C)=C
- Octatetraen: C=CC=CC=CC=C oder C=C/C=CC=C/C=C
- 4-Butadienyl-Cyclohex-1-en: C1CCC(C=CC=C)CC=1
- 3-Vinyl-Cyclohex1en: C=CC1C=CCCC1
- Cyclohexen: C1CCCCC=1
- 4,5-Divinyl-Cyclohex-1-en: C=CC1CC=CCC1C=C
- 3,4-Divinyl-Cyclohex-1-en: C1CCC(C=C)C(C=C)C=1
- 3,5-Divinyl-Cyclohex-1-en: C(C1CC(C=C)C=CC1)=C
- Benzol: C1C=CC=CC=1
- 5-Butan-1-enyl-Cyclo-1,3-dien: C(CCC1C=CC=CC1)=C
- 3-Butadienyl-Cyclohex-1-en: C1C(C=CC=C)CCCC=1
- 3-Vinyl-Octa-1,7-Dien: C=CC(C=C)CCCC=C
- Okta-1,7-dien: C=CCCCCC=C
- 3,6-Divinyl-Cyclohex1en: C=CC1C=CC(C=C)CC1

- 1,2,3,4,4a,5,6,7,8,8a-Octahydronaphthalin: C1CC2CCCCC2CC=1
- 1,4,4a,8a-Tetrahydronaphthalin: C1CC2C=CC=CC2CC=1

Wenn Butadien mit einem generischen Olefinmetathese-Katalysator zum reagieren gebracht wird ist zu erwarten, dass alle Moleküle in verschiedenen Enantiomeren und E/Z-Isomeren synthetisiert werden. Daher ist es ohne das tatsächliche Durchführen dieses Experiments schwierig einen Reaktionspfad herauszusuchen, der von dem ILP bestätigt werden soll. Die Synthese von Benzol (Abbildung 4.15) stellt eine gute Option dar, da die hohe Symmetrie im Vergleich zu den anderen Spektren zu einem sehr markanten Spektrum führt. Die direkteste Synthese stellen dabei die Olefinmetathese von zwei Butadien zu Hexatrien und Ethylen, die Olefinmetathese von Hexatrien mit Butadien zu Octatetraen und eine Ringschluss Olefinmetathese von Octatretraen zu Benzol und Ethylen dar. Es werden drei Butadien eingesetzt, um mit drei Olefinmetathesen ein Benzol und drei Ethylen zu erhalten.

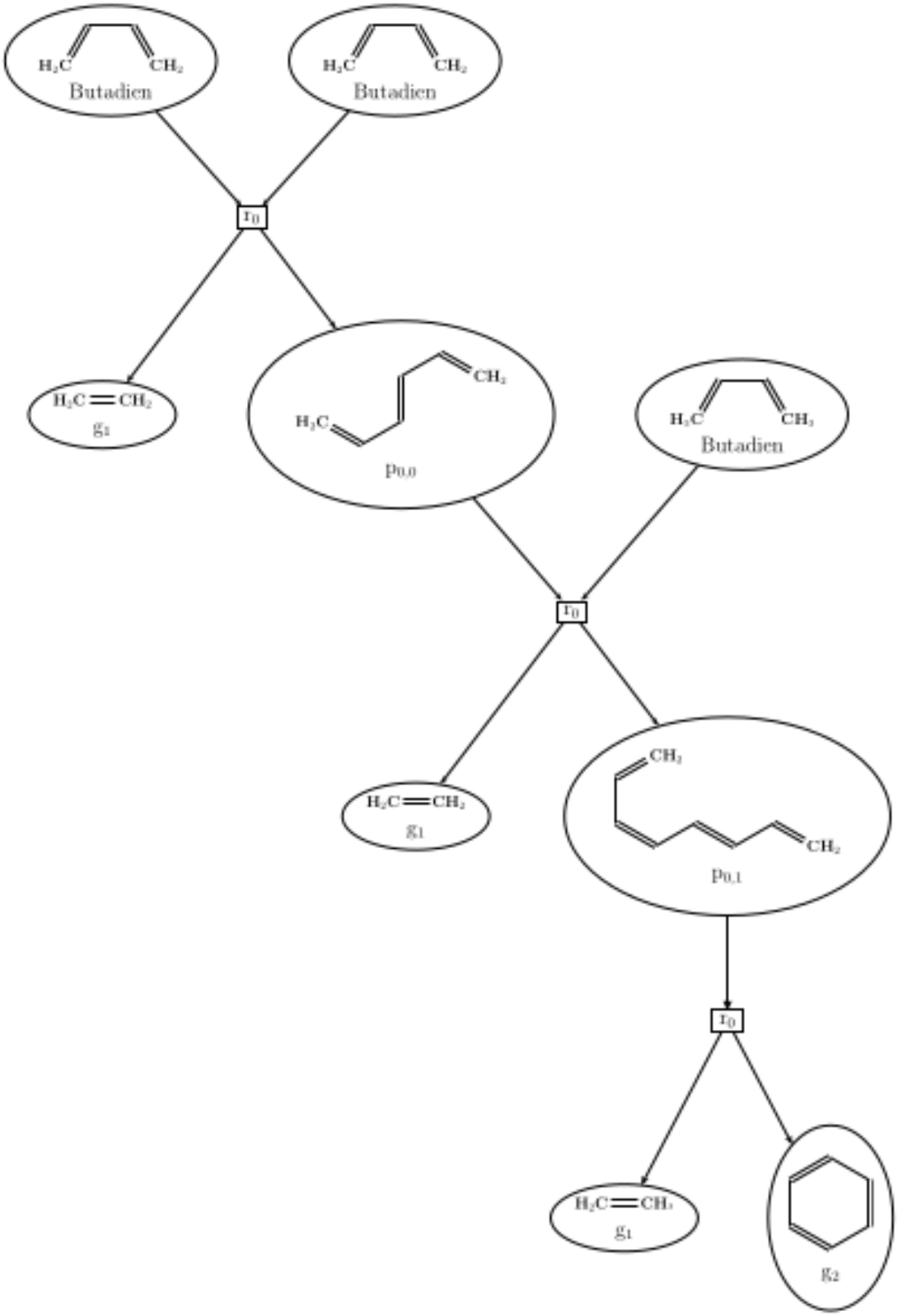


Abbildung 4.15: Benzolsynthese in dem Graphtransformationstool. Die Syntheseschritte lauten erst Olefinmetathese von zwei Butadien zu Hexatrien und Ethylen, dann Olefinmetathese von Hexatrien und Butdien zu Ethylen und Octatetraen und zuletzt eine Ringschluss Olefinmetathese von Octatetraen zu Benzol und Ethylen.

4.10 NMR: Multi-Molekül-Reaktionen

4.10.1 Olefinmetathese und Diels-Alder

Die generierten Graphen weisen anders als bei den vorherigen Beispielen verschiedene Enantiomere und E/Z-Isomere auf. Es ist daher notwendig diese zu unterscheiden. Für die Moleküle Hexatrien und Octatetraen wurden gezielt die Konfigurationen gewählt, bei der alle Doppelbindungen in der E-Konfiguration sind. Bei den Verbindungen mit Cyclohexanring und angehängten Vinylgruppen, wurde die Position so gewählt, dass die Reste äquatorial zum Ring stehen. Bei den beiden Hydronaphtalinen wurde die Konfiguration gewählt, bei der beide Ringe in der äquatorialen Konfiguration liegen. Verallgemeinert wurde für jedes Molekül die Konfiguration mit der größten Symmetrie und dadurch mit den meisten chemisch-äquivalenten Kohlenstoffen gewählt.

4.11 NMR SI: Multi-Molekül-Reaktionen

In der Literatur werden NMR-Spektren bevorzugt für einen vollständig aufgereinigten Stoff angegeben. Daher wurden die NMR-Spektren für jedes Molekül mit dem kombinierten NMR-Spektrum aller Produkte der ausgewählten Reaktionen verglichen.

4.11.1 Olefinmetathese und Diels-Alder

Die Referenzspektren der Benzolsynthese sind im Folgenden aufgeführt.

- Hexatrien und Ethylen
- Octatetraen und Ethylen
- Benzol und Ethylen

Für den Vergleich der NMR-Spektren wurden die gleichen Bingrößen b wie für die bisherigen Ein-Molekül-Reaktionen (Gleichung 4.1) verwendet. Der Ähnlichkeitsindex SI wurde als Mittelwert der SI_b für alle Bingrößen berechnet. Die finale Wahrscheinlichkeit entspricht dem SI normalisiert für jedes Referenzspektrum (Tabelle 4.23). Das Referenzspekrtrum mit Hexatrien und Ethylen weist eine große Ähnlichkeit mit den beiden tatsächlichen Molekülen auf. Die Ähnlichkeit zum Referenzspektrum liegt für Ethylen bei 0.18 und für Hexatrien bei 0.31. Die Moleküle Butadien mit 0.11 und Octatetraen mit 0.21 weisen ebenfalls eine hohe Ähnlichkeit zum Referenzspektrum auf. Da es sich bei Octatetraen, Hexatrien und Butadien um Kohlenstoffketten mit alternierenden Einfach- und Doppelbindungen handelt, ist diese Ähnlichkeit zu erwarten. Alle drei haben ein Signal bei etwa 122 ppm und ein Signal bei etwa 145 ppm. Die längeren Ketten haben zusätzlich

Tabelle 4.23: Resultierende Wahrscheinlichkeiten für den Vergleich aller aus Butadien Entwickelten Moleküle mit Hexatrien und Ethylen, Octatetraen und Ethylen und Benzol und Etylen.

Molekül	Hexatrien und Ethylen	Octatetraen und Ethylen	Benzol und Etylen
Butadien	0.11	0.11	0.0
Ethylen	0.18	0.15	0.17
Hexatrien	0.31	0.21	0.0
4-Vinyl-Cyclohex-1-en	0.0	0.0	0.0
4-Vinyl-Octa-1,7-Dien	0.04	0.04	0.0
Octatetraen	0.21	0.31	0.0
4-Butadienyl-Cyclohex-1-en	0.02	0.02	0.03
3-Vinyl-Cyclohex1en	0.0	0.0	0.0
Cyclohexen	0.0	0.0	0.0
4,5-Divinyl-Cyclohex-1-en	0.0	0.0	0.01
3,4-Divinyl-Cyclohex-1-en	0.0	0.0	0.0
3,5-Divinyl-Cyclohex-1-en	0.0	0.0	0.16
Benzol	0.0	0.0	0.51
5-Butan-1-enyl-Cyclo-1,3-dien	0.0	0.0	0.09
3-Butadienyl-Cyclohex-1-en	0.03	0.03	0.0
3-Vinyl-Octa-1,7-Dien	0.0	0.0	0.0
Okta-1,7-dien	0.0	0.0	0.0
3,6-Divinyl-Cyclohex1en	0.06	0.07	0.0
1,2,3,4,4a,5,6,7,8,8a-Octahydronaphthalin	0.0	0.0	0.0
1,4,4a,8a-Tetrahydronaphthalin	0.03	0.04	0.03

ein Signal bei etwa 140 ppm, weshalb der Überlapp bei Hexatrien und Octatetraen etwas größer ist. Das Referenzspektrum mit Octatetraen und Ethylen weist eine große Ähnlichkeit mit beiden tatsächlichen Molekülen auf. Die Ähnlichkeit zum Referenzspektrum liegt für Ethylen bei 0.15 und für Octatetraen bei 0.31. Die Moleküle Butadien mit 0.11 und Hexatrien mit 0.21 weisen ebenfalls eine hohe Ähnlichkeit zum Referenzspektrum auf. Die Gründe dafür sind die gleichen wie beim ersten Referenzsspektrum, da der Ähnlichkeitsindex SI zweier Moleküle symmetrisch ist. Das Referenzspektrum mit Benzol und Etyhlen hat die größte Ähnlichkeit mit den beiden tatsächlich im Spektrum enthaltenden Molekülen. Ethylen hat einen SI von 0.17 und Benzol von 0.51. Es gibt aber auch einige weitere Stoffe wie 5-Butan-1-enyl-Cyclo-1,3-dien und 3,5-Divinyl-Cyclohex-1-en, welche mit 0.09 und 0.16 ebenfalls eine deutliche Ähnlichkeit aufweisen. Diese Ähnlichkeit liegt daran, dass der eine Benzolpeak bei etwa 132.96 ppm mit manchen Peaks aus den anderen beiden Spektren überlagert. Beide Moleküle haben einen Peak bei 132.93 und 133.29 ppm respektive. Da der Peak von 5-Butan-1-enyl-Cyclo-1,3-dien näher liegt ist die Ähnlichkeit größer. Diese Vergleiche demonstrieren die Limitationen eines simpleren Ansatz wie dem Binning mit Crosskorrelation. Ähnliche Moleküle mit vergleichbaren chemischen Verschiebungen können einen hohen Ähnlichkeitsindex SI aufweisen, auch wenn die Anzahl der Peaks nicht zum Molekül passt. Die tiefer gehenden chemischen Informationen können so verloren gehen.

4.12 ILP: Multi-Molekül-Reaktionen

Für die Definition des ILP Problems wurde der Hypergraph über alle mit dem Graphtransformationstool entwickelten Reaktionen definiert. Dabei wurde eine Kante, die den Ausgangsstoff in das System einführen kann und für jedes andere Molekül eine Kante, die es aus dem System entfernt, hinzugefügt. Die Kanten, bei der die Edukte und Produkte identisch zueinander sind, wurden aufgrund ihrer Redundanz entfernt.

4.12.1 Olefinmetathese und Diels-Alder

Für das ILP-Model wurde ein Ansatz entwickelt, welcher wie in den vorangegangenen Beispielen die Molekülwahrscheinlichkeiten den Kanten zuordnet. Zudem wurde ein Model entwickelt, welches zuerst die Wahrscheinlichsten Moleküle ermittelt und deren Produktion maximiert.

4.12.1.1 Kantenwahrscheinlichkeit

Bei diesem Model werden die Wahrscheinlichkeiten als konstante Werte $p_{i,e}$ den Kanten e des Hypergraphen zugeordnet. In der Olefinmetathese fließt jede Kante entweder in keine, ein oder zwei Edukte. Ausgehende Kanten, die ins Nichts fließen, sollten in der Zielfunktion nicht berücksichtigt werden. Die chemischen Reaktionen sind also nur die Kanten mit ein oder zwei Produkten. Da den Produkten Wahrscheinlichkeitswerte zugewiesen sind liegt nahe, dass die Gesamtwahrscheinlichkeit dem Produkt beider Einzelwahrscheinlichkeiten entspricht. Da die Wahrscheinlichkeiten kleiner als 1 sind, wird dies bei Kanten mit zwei Edukten zu kleineren Wahrscheinlichkeiten führen als bei Kanten mit nur einem Edukt. Ein weiterer Ansatz ist es die Gesamtwahrscheinlichkeit als deb Mittelwert beider Produkte zu wählen. Dieser Ansatz sorgt dafür, dass eine Kante mit einem sehr unwahrscheinlichen und einem sehr wahrscheinlichen Produkt trotzdem eine mittlere Gesamtwahrscheinlichkeit hat. Zur Demonstration der beiden Ansätze zur Berechnung der Kantenwahrscheinlichkeit werden die Kantenwahrescheinlichkeiten von zwei Kanten im Hypergraphen für das dritte Referenzspektrum bestimmt.

- 42: Octatetraen → Benzol + Ethylen
- 44: 4-Butadienyl-Cyclohex-1-en → 5-Butan-1-enyl-Cyclo-1,3-dien

Die Wahrscheinlichkeiten für Kanten mit nur einem Produkt entsprechen der Wahrscheinlichkeit für dieses Produkt. Demnach hat die Kante 44 für das dritte Referenzspektrum die Wahrscheinlichkeit $p_{3,44}$ für Durchschnitt und Produkt. Kante 42 hingegen hat die Wahrscheinlichkeit 0.0876 für das Produkt (Gleichung 4.3) und 0.34 für den Durchschnitt

(Gleichung 4.4).

$$p_{3,42} = 0.51 \times 0.17 = 0.0876 \tag{4.3}$$

$$p_{3,42} = \frac{0.51 + 0.17}{2} = 0.34 \tag{4.4}$$

Für jede Kante werden fünf ganzzahlige Variablen x_e, b_e, $n_{1,e}$, $n_{2,e}$ und $n_{3,e}$ definiert. Die Variable b_e ist dabei eine Wahrheitsvariable, die definiert ob eine Kante verwendet wird. Die Variable x_e beschreibt wie oft die selbe Kante verwendet wird. Die Variablen $n_{1,e}$, $n_{2,e}$ und $n_{3,e}$ sind ebenfalls Wahrheitsvariablen und legen fest, welche drei Reaktionen den Referenzspektren entsprechen und deswegen zu der Zielfunktion dazuzählen. Dabei ist $n_{1,e}$ die Zuordnungsvariable zu dem ersten Referenzspektrum mit Hexatrien und Ethylen, $n_{2,e}$ die Zuordnungsvariable zu dem zweiten Referenzspektrum mit Octatetraen und Ethylen und $n_{3,e}$ die Zuordnungsvariable zum dritten Referenzspektrum mit Benzol und Ethylen. Da das ILP die Syntheseschritte durch den Vergleich der NMR-Spektren findet, werden alle synthetisierten Moleküle bestimmt. Um dahingehend die Lösung nicht einzuschränken ist kein Zielmolekül festgelegt. Die Variablen unterliegen den Beschränkungen, dass:

- $x_e \geq b_e$
- $b_e \geq n_{1,e}$
- $b_e \geq n_{2,e}$
- $b_e \geq n_{3,e}$
- s: ([], [Startmolekül]), $x_s = 3$
- $\forall v \in V$ inflow = outflow
- $\forall i, j \in E$ i: head $\rightarrow$ tail und j: tail $\rightarrow \sum b_i + b_j \leq 1$
- $n_{1,e} + n_{2,e} + n_{3,e} \leq 1$
- $\sum_e n_{1,e} = 1$
- $\sum_e n_{2,e} = 1$
- $\sum_e n_{3,e} = 1$
- wenn die größte Wahrscheinlichkeit einer nicht ausgehenden Kante $p_{max,e} = 0$ ist, wird diese stattdessen auf $p_{max,e} = 0.0001$ gesetzt um einen Wertfehler zu verhindern

Unter diesen Beschränkungen soll ein Pfad durch den Hypergraphen bestimmt werden, welcher die drei wahrscheinlichsten Kanten für die drei Referenzspektren auswählt und dabei möglichst wenige unwahrscheinliche Kanten beinhaltet. Es ist nicht üblich, Kanten

mit einer Wahrscheinlichkeit $p_{max,e} = 0.00$ auf $p_{max,e} = 0.0001$ zu setzen, da die Wahl des Wertes das Ergebnis beeinflussen kann. In dem ILP-Model ist es sehr restriktiv Kanten mit einer Wahrscheinlichkeit von $p_{max,e} = 0.00$ komplett zu verbieten, da die Differenz zu Kanten mit Wahrscheinlichkeit $p_{max,e} = 0.01$ nicht groß ist, jedoch $p_{max,e} = 0.00$ ausgeschlossen wird. Ein Ansatz der diese Entscheidung umgeht ist die Auswahl von Zielknoten (Kapitel 4.12.1.2). Die ausgehenden Kanten sind nicht in die Zielfunktion eingeflossen (Gleichung 4.5).

$$max\ z = \sum_{e} 1000 \times (p_{1,e} \times n_{1,e} + p_{2,e} \times n_{2,e} + p_{3,e} \times n_{3,e}) - \frac{x_e}{max(p_{1,e}, p_{2,e}, p_{3,e})} \quad (4.5)$$

Für beide Kantenwahrscheinlichkeitsdefinitionen wurden jeweils die 10 besten Lösungen gesucht.

4.12.1.2 Knoten Auswahl

Ein Ansatz, der unabhängig von der Auswahl der Kantenwahrscheinlichkeitsberechnung ist, ist durch die NMR-Spektren Zielmoleküle zu wählen und einen Pfad zu finden, der diese synthetisiert. Dafür wurde jedem Molekül v die Wahrscheinlichkeit $p_{i,v}$ direkt zugeordnet. Für jede Kante des Hypergraphen werden zwei Variablen x_e und b_e definiert. Die Variable b_e ist dabei eine Wahrheitsvariable, die definiert ob eine Kante verwendet wird. Die Variable x_e beschreibt wie oft die selbe Kante verwendet wird. Für jeden Knoten des Hypergraphen werden drei Variablen $n_{1,e}$, $n_{2,e}$ und $n_{3,e}$ definiert. Diese drei Variablen sind ebenfalls Wahrheitsvariablen und legen fest, welche drei Moleküle den drei Referenzspektren entsprechen. Dabei ist $n_{1,e}$ die Zuordnungsvariable zu dem ersten Referenz NMR-Spektrum mit Hexatrien und Ethylen, $n_{2,e}$ die Zuordnungsvariable zu dem zweiten Referenz NMR-Spektrum mit Octatetraen und Ethylen und $n_{3,e}$ die Zuordnungsvariable zum dritten Referenz NMR-Spektrum mit Benzol und Ethylen. Die Variablen unterliegen den Beschränkungen, dass:

- $x_e \geq b_e$
- Hilfsvariable: $m_v = \sum_e b_e : v \in head(b_e)$
- $m_v \geq n_{1,v}$
- $m_v \geq n_{2,v}$
- $m_v \geq n_{3,v}$
- s: ([], [Startmolekül]), $x_s = 3$
- $\forall v \in V$ inflow = outflow

- $\forall i, j \in E$ i: head $\rightarrow$ tail und j: tail $\rightarrow \sum b_i + b_j \leq 1$
- $n_{1,v} + n_{2,v} + n_{3,v} \leq 1$
- $\sum_v n_{1,v} = 1$
- $\sum_v n_{2,v} = 1$
- $\sum_v n_{3,v} = 1$

Dabei soll die Wahrscheinlichkeit der ausgewählten Moleküle maximiert werden (Gleichung 4.6). Nachdem drei Zielmoleküle ausgewählt wurden wird der Fluss in diese Moleküle als Bedingung gesetzt.

$$max\ z_1 = \sum_v \sum_i n_{i,v} \times p_{i,v} \tag{4.6}$$

Alle Lösungen, die positiven Fluss $b_e = 1$ in die Zielmoleküle haben, werden danach sortiert, welche Lösung die geringste Anzahl an Reaktionen hat (Gleichung 4.7). Wenn mehrere Lösungen die gleiche Anzahl an Reaktionen haben, wird die ausgewählt, die den geringsten flow hat (Gleichung 4.8).

$$max\ z_2 = \sum_e b_e \tag{4.7}$$

$$max\ z_3 = \sum_e x_e \tag{4.8}$$

Es wurden die beiden besten Lösungen für das ILP bestimmt.

Tabelle 4.24: Lösung 1 für die Knotenauswahl beim ILP vom Butadien Beispiel mit $z_3 = 3$.

Flow	Kante	Edukt	Produkt
3	1	Ethylen	
1	4	2 Butadien	Hexatrien + Ethylen
1	23	Butadien + Hexatrien	Octatetraen + Ethylen
1	41	Benzol	
1	42	Octatetraen	Benzol + Ethylen
3	213		Butadien

Beide Lösungen die aus der Auswahl der besten Knoten resultieren ergeben chemisch Sinn. Die erste Lösung (Tabelle 4.24) ist die erwartete Lösung. Die zweite Lösung (Tabelle 4.25) verwendet viele der gleichen Kanten, aber produziert mit Kante 18, Octatetraen aus zwei Hexatrien. Diese Variation ist ebenfalls erwartbar.

Tabelle 4.25: Lösung 2 für die Knotenauswahl beim ILP vom Butadien Beispiel mit $z_3 = 4$.

Flow	Kante	Edukt	Produkt
3	1	Ethylen	
2	4	2 Butadien	Hexatrien + Ethylen
1	18	2 Hexatrien	Octatetraen + Butadien
1	41	Benzol	
1	42	Octatetraen	Benzol + Ethylen
3	213		Butadien

4.13 Petrinetzzertifikat: Multi-Molekül-Reaktionen

Durch den einfachen Optimierungsprozess des ILP wird die Zielfunktion maximiert und eine Menge an Kanten zurückgegeben. Diese Kanten haben keine Reihenfolge oder Zusammenhang abseits der Definition des ILP. Darum kommt es oft vor, dass die Reaktionen in Sequenz ein Molekül für eine Reaktion benötigen, welches erst durch eine andere Reaktion erzeugt wird. Beide Reaktioen ließen sich so nur chemisch durchführen, wenn die jeweils andere zuvor stattgefunden hat. Um solche Lösungen des ILP-Models auszuschließen wurde für jede Lösung ein Petrinetzzertifikat erstellt.

4.13.1 Petrinetzzertifikat: Olefinmetathes und Diels-Alder

4.13.1.1 Produkt

Die 10 Lösungen, welche durch die Bestimmung der Kantenwahrscheinlichkeit als Produkt beider Produkte erhalten wurden, wurden mit einem Petrinetz, basierend auf dem Hypergraphen, auf Erfüllbarkeit untersucht. Dabei wurden zu Beginn drei Token in Butadien angenommen. Nach der Untersuchung mit einem Petrinetz konnte nur eine Lösung zertifiziert werden (Tabelle 4.26). Dabei wird wie erwartet Hexatrien erzeugt. Aus Hexatrien erfolgt die Synthese von 3,5-Divinyl-Cyclohex-1-en. Es findet keine weitere Reaktion statt.

Tabelle 4.26: Lösung 10 für die Kantenwahrscheinlichkeit als Produkt beim ILP vom Butadien Beispiel mit $z = 197.08$.

Flow	Kante	Edukt	Produkt
1	1	Ethylen	
1	4	2 Butadien	Hexatrien + Ethylen
1	36	3,5-Divinyl-Cyclohex-1-en	
1	37	Butadien + Ethylen	3,5-Divinyl-Cyclohex-1-en
3	213		Butadien

Diese Lösung entspricht nicht der erwarteten Synthese von Benzol aus Butadien (Abbildung 4.15).

4.13.1.2 Durchschnitt

Die 10 Lösungen, welche durch die Bestimmung der Kantenwahrscheinlichkeit als Durchschnitt beider Produkte erhalten wurden, wurden mit einem Petrinetz, basierend auf dem Hypergraphen, auf Erfüllbarkeit untersucht. Dabei wurden zu Beginn drei Token in Butadien angenommen. Nach der Untersuchung mit einem Petrinetz konnte nur eine Lösung zertifiziert werden (Tabelle 4.27).

Tabelle 4.27: Lösung 5 für die Kantenwahrscheinlichkeit als Durchschnitt beim ILP vom Butadien Beispiel mit $z = 798.54$.

Flow	Kante	Edukt	Produkt
3	1	Ethylen	
1	4	2 Butadien	Hexatrien + Ethylen
1	23	Butadien + Hexatrien	Octatetraen + Ethylen
1	41	Benzol	
1	42	Octatetraen	Benzol + Ethylen
3	213		Butadien

Diese Lösung entspricht der erwarteten Synthese von Benzol aus Butadien (Abbildung 4.15).

4.13.1.3 Knoten Auswahl

Beide Lösungen der Knotenauswahlmethode erfüllen ein Petrinetzzertifikat.

4.14 Fazit: Multi-Molekül-Reaktionen

Im Vergleich zu einem kleinen Netzwerk mit festgelegten Start- und Zielmolekülen, ist die Formulierung der Zielfunktionen und Einschränkungen für ein Multi-Molekül-Reaktions-Netzwerk deutlich komplexer. Dazu kommt die Entscheidung nur die Ausgangsmoleküle zu definieren. Dabei kommt nur das ILP mit der Kantenwahrscheinlichkeit als Produkt gar nicht auf die erwartete Lösung. In beiden anderen Modellen ist die direkte Synthese von Benzol die erste Lösung mit Petrinetzzertifikat. Das ILP mit der Kantenwahrscheinlichkeit als Durchschnitt hatte trotzdem einige Lösungen, welche deutlich von der erwarteten Lösung abwichen und teilweise keine Kanten mit dieser teilten. Das ILP bei dem drei Knoten ausgewählt wurden die in der Synthese vorkommen müssen war dabei noch effizienter. Die zweitbeste Lösung bei diesem ILP ist ebenfalls eine realistische Möglichkeit unter den Reaktionsbedingungen, auch wenn sie in der Kohlenstoffeffizienz geringer ist. Die Lösung würde eigentlich vier Butadien zum Start benötigen. Da bei den Reaktionen aber ein neues Butadien entsteht, sind die ILP Bedingungen dennoch erfüllt. Da Butadien als Startmolekül theoretisch unbegrenzt vorliegt, ist die Synthese trotzdem chemisch möglich. Das überschüssige Butadien kann zu Beginn der nächsten Synthese erneut verbraucht

werden. Zur Verifizierung des Butadien Beispiels wurden NMR-unabhängige Werte verwendet. Die betrachtete Synthese von Benzol aus Butadien stellt demnach auch ohne die Wahrscheinlichkeiten die optimale Lösung dar, wenn das Zielmolekül festgelegt wird. Ohne Wahrscheinlichkeiten oder mit zufälligen Wahrscheinlichkeiten und einem nicht definierten Zielmolekül, gibt es jedoch entweder eine Lösung ohne Reaktionen oder eine Lösung mit zufälligen Zielmolekül. Auch hier resultiert die Verwendung der NMR-basierten Wahrscheinlichkeiten in einer zielgerichteteren Lösungsauswahl.
Auch die Synthese von Benzol aus Butadien zeigt, dass die in den NMR-Spektren enthaltenen Informationen eine Rekonstruktion der vorgegeben Reaktionen ermöglicht. Mehr als in den natürlich auftretenden Synthesen muss angemerkt werden, dass die durch NMR-Spektren repräsentierten Reaktionen nicht auf in der Literatur vorgefundenen Synthesen oder im Labor gemessenen Spektren basiert.

5 Zusammenfassung und Ausblick

In dieser Masterarbeit wurde die Möglichkeit untersucht die Syntheseplanung und -analyse mittels NMR-Spektren zu automatisieren. Dabei wurden die charakteristischen Verschiebungen in ^{13}C-NMR-Spektren verwendet, um die Wahrscheinlichkeit einzelner Moleküle und Reaktionen zu bestimmen. Diese Wahrscheinlichkeiten wurden über eine ILP-Berechnung optimiert. Noch öfter als ^{13}C-NMR-Spektren in der organischen Chemie sind die ^{1}H-NMR-Spektren. Diese haben zu den charakteristischen Verschiebungen noch zusätzliche Informationen durch Multiplettaufspaltungen. Da Wasserstoff im Gegensatz zu Kohlenstoffen als Protonen abgespalten werden können, kommt es vor allem bei Wasserstoff-Stickstoff- und Wasserstoff-Sauerstoff-Bindungen zu großen Schwankungen in der chemischen Verschiebung. Hinzu kommt, dass das quantenchemischen Rechnungsprogramm in der Berechnung der ^{1}H-NMR-Spektren deutlich inkonstanter von den experimentellen Spektren abweicht. Zudem kommt es bei der Berechnung der Peakintensitäten und Multiplettaufspaltung zu größeren Fehlern, weshalb es für die verwendete Analyse nicht geeignet ist. Die Crosskorrelation bietet eine sehr einfache Methode zum Vergleich von NMR-Spektren. Da jedoch bei den ^{1}H-NMR-Spektren die Multiplettaufspaltung oft mehr Informationen enthält als nur Intensität und chemische Verschiebung, geht bei der Analyse eine zentrale Komponente verloren. Beim Vergleich zwischen Spektren der gleichen Substanz, ist die Crosskorrelation auch bei ^{1}H-NMR-Spektren geeignet, da nur leichte Verschiebungen der Peaks zu erwarten sind. Wenn jedoch zwei verschiedene Substanzen verglichen werden, kann die Verschiebung schon größer sein, weshalb die Ähnlichkeit gering ist, auch wenn die Spektren als Ganzes betrachtet ähnlich sind. Für eine solche Analyse könnte ein Machine-Learning-Model trainiert werden [45], welches diese Details berücksichtigt. Zum Training dieses Models müssten jedoch deutlich mehr Beispiele verwendet werden und der NMR-Vergleich müsste eine deutlich zentralere Rolle in dieser Arbeit einnehmen. Mehr Beispiele können zudem verwendet werden, um den Schweratomeffekt in der NMR-Berechnung zu quantifizieren, sowie allgemeine Fehler in der Verschiebung zu ermitteln und so den idealen Verschiebungswert zu ermitteln. Die Kombination beider Ein-Molekül-Reaktionen legt jedoch nahe, dass dieser Fehler nicht durch eine Verschiebung mit einer Konstante zu beschreiben ist. In der NMR-Spektroskopie ist es auch möglich zwei Substanzen in einem NMR-Spektrum abzubilden. Dabei kann der Vergleich der Intensität der zugeordneten Peaks bestimmt werden, welches Stoffmengenverhältnis beide Substanzen zueinander haben [46]. Es ist also möglich anstelle von mehreren NMR-Spektren für

eine Synthese, nur ein gemeinsames NMR-Spektrum zu bestimmen und auf Basis dieses Spektrums neben den vorliegenden Substanzen, auch deren relative Stoffmenge zu bestimmen. Sowohl die natürlichen Synthesen im Organismus, als auch die Olefinmetathese und Diels-Alder-Reaktion für Butadien, eignen sich als Reaktion in einem Topf. Dadurch könnten die Startmoleküle mit den Enzymen und Katalysatoren in eine Lösung gebracht werden und nach einer festen Zeit NMR-Spektren aufgenommen werden. Basierend auf diesen Spektren könnten die relativen Stoffmengen und basierend auf diesen die Gleichgewichtskonstanten für diese Reaktionen bestimmt werden. Zudem erlaubt eine solche Analyse die Bestimmung von Nebenprodukten, welche in dem bisherigen Ansatz nicht definiert wurden.
Die Lösungen des ILP erfüllen die Erwartungen. Es ist aber denkbar, ein Model zu verwenden, dass in der Lage ist, die Tatsächlichen NMR-Spektren zu maximieren und so eine Abstraktion durch die Wahrscheinlichkeiten zu umgehen. Dabei könnte ein ILP-Model verwendet werden, dass zu einem Spektrum aus den Kandidatenmolekülen das am besten passende Molekül bestimmt. basierend auf diesem Molekül könnte ein weiteres ILP-Model diese Überlagerung für Reaktionen maximieren. Das verwendete ILP-Tool erlaubte jedoch nicht einzelnen Knoten größere Informationen zuzuordnen. Die bisherige Untersuchung eignet sich trotzdem als eine Demonstration der möglichen Kombination beider bisher getrennter Felder.

Danksagung

Ich möchte die Gelegenheit nutzen, mich herzlichst bei dem Arbeitskreis Chemieinformatik zu bedanken. Besonderer Dank gilt Prof. Dr. Daniel Merkle, der mich sowohl tatkräftig bei inhaltlichen Fragen unterstützt hat als auch durch seine große Begeisterung eine konstante Quelle der Motivation war. Ebenso danke ich Dr. Yuri Kolcak, Addie Jordon und Cassandra König für die beratende Expertise zu theoretischen und inhaltlichen Fragen. Auch Dr. Marcel Friedrichs hat mich besonders bei technischen Problemen unterstützt. Zuletzt danke ich Verena Uhlig und Nicole Schellenberg, für die organisatorischen Unterstützung und das sie Prof. Merkle erinnert haben, wenn meine Fragen in den vielen Aufgaben etwas untergegangen sind.

Quellenverzeichnis

[1] J. K. Wegner u. a. „Cheminformatics“. In: *Commun. ACM* 55.11 (Nov. 2012), S. 65–75. ISSN: 0001-0782. DOI: 10.1145/2366316.2366334. URL: https://doi.org/10.1145/2366316.2366334.

[2] H. Chen, T. Kogej und O. Engkvist. „Cheminformatics in Drug Discovery, an Industrial Perspective“. In: *Molecular Informatics* 37.9-10 (2018), S. 1800041. DOI: https://doi.org/10.1002/minf.201800041. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/minf.201800041. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/minf.201800041.

[3] M. J. Willis und M. von Stosch. „Inference of chemical reaction networks using mixed integer linear programming“. In: *Computers and Chemical Engineering* 90 (2016), S. 31–43. ISSN: 0098-1354. DOI: https://doi.org/10.1016/j.compchemeng.2016.04.019. URL: https://www.sciencedirect.com/science/article/pii/S0098135416301156.

[4] J. L. Andersen u. a. „Chemical Transformation Motifs—Modelling Pathways as Integer Hyperflows“. In: *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 16.2 (Apr. 2019), S. 510–523. ISSN: 1545-5963. DOI: 10.1109/tcbb.2017.2781724. URL: https://doi.org/10.1109/TCBB.2017.2781724.

[5] K. Peter, C. Vollhardt und N. E. Schore. *Organische Chemie*. Wiley-VCH, 2020. ISBN: 978-3-527-34582-3.

[6] E. A. Bender und S. G. Williamson. *Foundations of Combinatorics with Applications*. Dover. ISBN-10: 0-486-44603-4. Mineola, NY: Dover Publications, 2006, S. 468. ISBN: 978-0486446035.

[7] J. L. Andersen. „Analysis of Generative Chemistries“. Available at https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa-202723. PhD thesis. Leipzig, DE: Universität Leipzig, Sep. 2015.

[8] J. E. Packer und S. D. Woodgate. „Lewis structures, formal charge, and oxidation numbers: A more user-friendly approach“. In: *Journal of Chemical Education* 68.6 (Juni 1991), S. 456–456. ISSN: 0021-9584. DOI: 10.1021/ed068p456. eprint: https://pubs.acs.org/jceda8/article-pdf/68/6/456/416109/ed068p456.pdf. URL: https://doi.org/10.1021/ed068p456.

[9] R. OC Norman. *Principles of organic synthesis*. Routledge, 2017.

[10] S. Krieger und J. Kececioglu. „Shortest Hyperpaths in Directed Hypergraphs for Reaction Pathway Inference“. In: *Journal of Computational Biology* 30.11 (2023), S. 1198–1225. DOI: 10.1089/cmb.2023.0242. eprint: https://journals.sagepub.com/doi/pdf/10.1089/cmb.2023.0242. URL: https://journals.sagepub.com/doi/abs/10.1089/cmb.2023.0242.

[11] E. Silva Dias u. a. „Efficient Enumeration of Chordless Cycles“. In: *arXiv e-prints*, arXiv:1309.1051 (Sep. 2013), arXiv:1309.1051. DOI: 10.48550/arXiv.1309.1051. arXiv: 1309.1051 [cs.DS].

[12] D.-S. Chen, R. G. Batson und Y. Dang. „Linear Programming—Fundamentals“. In: *Applied Integer Programming*. John Wiley und Sons, Ltd, 2009. ISBN: 9781118166000. DOI: https://doi.org/10.1002/9781118166000. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118166000. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118166000.

[13] C. A. Petri. „Kommunikation mit Automaten“. Dissertation. Darmstadt, Germany: Technische Universität Darmstadt, 1962.

[14] J. L. Peterson. „Petri Nets“. In: *ACM Comput. Surv.* 9.3 (Sep. 1977), S. 223–252. ISSN: 0360-0300. DOI: 10.1145/356698.356702. URL: https://doi.org/10.1145/356698.356702.

[15] P. W. Atkins und R. S. Friedman. *Molecular quantum mechanics*. eng. 5th ed. / Peter Atkins and Ronald Friedman. Oxford: Oxford University Press, 2011. ISBN: 9780199541423.

[16] F. Neese u. a. „The ORCA quantum chemistry program package“. In: *The Journal of Chemical Physics* 152.22 (Juni 2020), S. 224108. ISSN: 0021-9606. DOI: 10.1063/5.0004608. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0004608/16740678/224108_1_online.pdf. URL: https://doi.org/10.1063/5.0004608.

[17] J. L. Andersen u. a. *A Software Package for Chemically Inspired Graph Transformation*. 2016. arXiv: 1603.02481 [cs.FL]. URL: https://arxiv.org/abs/1603.02481.

[18] A. A. Hagberg, D. A. Schult und P. J. Swart. „Exploring Network Structure, Dynamics, and Function using NetworkX“. In: *Proceedings of the 7th Python in Science Conference*. Hrsg. von Gaël Varoquaux, Travis Vaught und Jarrod Millman. Pasadena, CA USA, 2008, S. 11–15. URL: http://conference.scipy.org/proceedings/SciPy2008/paper_2/.

[19] F. Neese. „The ORCA program system“. In: *WIRES Comput. Molec. Sci.* 2.1 (2012), S. 73–78. DOI: 10.1002/wcms.81.

[20] G. Landrum. „RDKit: Open-Source Cheminformatics Software“. In: (2016). URL: https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4.

[21] M. D. Hanwell u. a. „Avogadro: an advanced semantic chemical editor, visualization, and analysis platform“. In: *Journal of Cheminformatics* 4.1 (Aug. 2012), S. 17. ISSN: 1758-2946. DOI: 10.1186/1758-2946-4-17. URL: https://doi.org/10.1186/1758-2946-4-17.

[22] L. Bodis, A. Ross und E. Pretsch. „A novel spectra similarity measure“. In: *Chemometrics and Intelligent Laboratory Systems* 85.1 (2007), S. 1–8. ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2005.10.002. URL: https://www.sciencedirect.com/science/article/pii/S0169743905001620.

[23] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2026. URL: https://www.gurobi.com.

[24] Z. Lin u. a. „Caffeine Synthesis and Its Mechanism and Application by Microbial Degradation, A Review“. In: *Foods* 12.14 (2023). ISSN: 2304-8158. DOI: 10.3390/foods12142721. URL: https://www.mdpi.com/2304-8158/12/14/2721.

[25] A. W. Tazon u. a. „Biotechnological Advances in Vanillin Production: From Natural Vanilla to Metabolic Engineering Platforms“. In: *BioChem* 4.4 (2024), S. 323–349. ISSN: 2673-6411. DOI: 10.3390/biochem4040017. URL: https://www.mdpi.com/2673-6411/4/4/17.

[26] R. L. Benoit und M. Fréchette. „Protonation of hypoxanthine, guanine, xanthine, and caffeine“. In: *Canadian Journal of Chemistry* 63.11 (1985), S. 3053–3056. DOI: 10.1139/v85-506. eprint: https://doi.org/10.1139/v85-506. URL: https://doi.org/10.1139/v85-506.

[27] C. E. Mueller u. a. „Synthesis of paraxanthine analogs (1,7-disubstituted xanthines) and other xanthines unsubstituted at the 3-position: structure-activity relationships at adenosine receptors“. In: *Journal of Medicinal Chemistry* 36.22 (Okt. 1993), S. 3341–3349. ISSN: 0022-2623. DOI: 10.1021/jm00074a015. eprint: https://pubs.acs.org/jmcmar/article-pdf/36/22/3341/7236147/jm00074a015.pdf. URL: https://doi.org/10.1021/jm00074a015.

[28] B. Jayantha u. a. „Engineered Pseudomonas putida monoculture system for green synthesis of 7-methylxanthine“. In: *Green Chemistry* 27.37 (Okt. 2025), S. 11365–11379. ISSN: 1463-9262. DOI: 10.1039/d5gc02883c. eprint: https://pubs.rsc.org/gc/article-pdf/27/37/11365/10400487/d5gc02883c.pdf. URL: https://doi.org/10.1039/d5gc02883c.

[29] K. Bouzková u. a. „Intermolecular Interactions in Crystalline Theobromine as Reflected in Electron Deformation Density and 13C NMR Chemical Shift Tensors“. In: *Journal of Chemical Theory and Computation* 9.6 (Juni 2013), S. 2629–2638. ISSN:

1549-9618. DOI: 10.1021/ct400209b. eprint: https://pubs.acs.org/jctcce/article-pdf/9/6/2629/8808212/ct400209b.pdf. URL: https://doi.org/10.1021/ct400209b.

[30] M. Thevis u. a. „Electrospray ionization mass spectrometric characterization and quantitation of xanthine derivatives using isotopically labelled analogues: an application for equine doping control analysis“. In: *Rapid Communications in Mass Spectrometry* 18.14 (2004), S. 1553–1560. DOI: https://doi.org/10.1002/rcm.1518. eprint: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/rcm.1518. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/rcm.1518.

[31] K. Hanai u. a. „A comparative vibrational and NMR study of cis-cinnamic acid polymorphs and trans-cinnamic acid“. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 57.3 (2001), S. 513–519. ISSN: 1386-1425. DOI: https://doi.org/10.1016/S1386-1425(00)00401-7. URL: https://www.sciencedirect.com/science/article/pii/S1386142500004017.

[32] T. Nakazawa und K. Ohsawa. „Metabolism of Rosmarinic Acid in Rats“. In: *Journal of Natural Products* 61.8 (Aug. 1998), S. 993–996. ISSN: 0163-3864. DOI: 10.1021/np980072s. eprint: https://pubs.acs.org/jnprdf/article-pdf/61/8/993/13045095/np980072s.pdf. URL: https://doi.org/10.1021/np980072s.

[33] R. Świsłocka u. a. „Spectroscopic (FT-IR, FT-Raman, 1H and 13C NMR) and theoretical studies of p-coumaric acid and alkali metal p-coumarates“. In: *Journal of Spectroscopy* 27.1 (2012), S. 546146. DOI: https://doi.org/10.3233/SPE-2012-0568. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3233/SPE-2012-0568. URL: https://onlinelibrary.wiley.com/doi/abs/10.3233/SPE-2012-0568.

[34] F.-H. Kuan, Y. Inoue und R. Chûjô. „Structure and molecular dynamics of solid-state inclusion complexes of cyclodextrin and permethylated cyclodextrin with benzaldehyde studied by high-resolution CP/MAS13C NMR“. In: *Journal of inclusion phenomena* 4.3 (1986), S. 281–290. DOI: https://doi.org/10.1007/BF00658002. eprint: https://link.springer.com/article/10.1007/BF00658002.

[35] J. Yu u. a. „Plant-Derived Caffeic Acid and Its Derivatives: An Overview of Their NMR Data and Biosynthetic Pathways“. In: *Molecules* 29.7 (2024). ISSN: 1420-3049. DOI: 10.3390/molecules29071625. URL: https://www.mdpi.com/1420-3049/29/7/1625.

[36] H. Mart, T. Sökmen und H. Yürük. „The synthesis and thermal properties of oligo-3-hydroxybenzaldehyde synthesized by oxidative polycondensation“. In: *Journal of Applied Polymer Science* 101.2 (2006), S. 892–897. DOI: https://doi.org/10.1002/

app.23537. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/app.23537. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/app.23537.

[37] A. Jha, B.R. Kumar und R. Kumar Ch. „Synthesized Novel Schiff Base of 4-Hydroxybenzaldehyde with Anilines for Optoelectronic Applications“. In: *SSRN Preprint Server* (2024), S. 46. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4934411.

[38] N. Syafni, D. Putra und D. Arbain. „3,4-DIHYDROXYBENZOIC ACID AND 3,4-DIHYDROXYBENZALDEHYDE FROM THE FERN Trichomanes chinense L.; ISOLATION, ANTIMICROBIAL AND ANTIOXIDANT PROPERTIES“. In: *Indonesian Journal of Chemistry* 12.3 (2012), S. 273–278. ISSN: 2460-1578. DOI: 10.22146/ijc.21342. URL: https://journal.ugm.ac.id/ijc/article/view/21342.

[39] M. W. Lodewyk, M. R. Siebert und D. J. Tantillo. „Computational Prediction of 1H and 13C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry“. In: *Chemical Reviews* 112.3 (März 2012), S. 1839–1862. ISSN: 0009-2665. DOI: 10.1021/cr200106v. eprint: https://pubs.acs.org/chreay/article-pdf/112/3/1839/8644981/cr200106v.pdf. URL: https://doi.org/10.1021/cr200106v.

[40] E. Jonas, S. Kuhn und N. Schlörer. „Prediction of chemical shift in NMR: A review“. In: *Magnetic Resonance in Chemistry* 60.11 (2022), S. 1021–1031. DOI: https://doi.org/10.1002/mrc.5234. eprint: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/mrc.5234. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/mrc.5234.

[41] J. Schripsema. „Similarity and differential NMR spectroscopy in metabolomics: application to the analysis of vegetable oils with 1H and 13C NMR“. In: *Metabolomics* 15.39 (2019), S. 1573–3890. DOI: https://doi.org/10.1007/s11306-019-1502-9.

[42] S. J. Connon und S. Blechert. „Recent Developments in Olefin Cross-Metathesis“. In: *Angewandte Chemie International Edition* 42.17 (2003), S. 1900–1923. DOI: https://doi.org/10.1002/anie.200200556. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.200200556. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200200556.

[43] J. Sauer, H. Wiest und A. Mielert. „Eine Studie der DIELS-ALDER-Reaktion, I. Die Reaktivität von Dienophilen gegenüber Cyclopentadien und 9.10-Dimethylanthracen“. In: *Chemische Berichte* 97.11 (1964), S. 3183–3207. DOI: https://doi.org/10.1002/cber.19640971129. eprint: https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/cber.19640971129. URL: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cber.19640971129.

[44] K. B. Wiberg. „The Concept of Strain in Organic Chemistry“. In: *Angewandte Chemie International Edition in English* 25.4 (1986), S. 312–322. DOI: https://doi.org/10.1002/anie.198603121. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.198603121. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.198603121.

[45] J. B. Rowlands u. a. „Towards automatically verifying chemical structures: the powerful combination of 1H NMR and IR spectroscopy“. In: *Chemical Science* 16.45 (Dez. 2025), S. 21590–21599. ISSN: 2041-6520. DOI: 10.1039/d5sc06866e. eprint: https://pubs.rsc.org/sc/article-pdf/16/45/21590/1904936/d5sc06866e.pdf. URL: https://doi.org/10.1039/d5sc06866e.

[46] S. K. Bharti und R. Roy. „Quantitative 1H NMR spectroscopy“. In: *TrAC Trends in Analytical Chemistry* 35 (2012), S. 5–26. ISSN: 0165-9936. DOI: https://doi.org/10.1016/j.trac.2012.02.007. URL: https://www.sciencedirect.com/science/article/pii/S0165993612000702.

Selbstständigkeitserklärung

Nachname: Seng
Vorname: Kilian

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem einzelnen Fall unter genauer Angabe der Quelle (einschließlich des Internets sowie anderer digitaler Informationsquellen) als Entlehnung kenntlich gemacht. Dies gilt auch für eingefügte Zeichnungen, bildliche Darstellungen, Skizzen und Ähnliches. Ich nehme zur Kenntnis, dass die nachgewiesene Unterlassung der Herkunftsangabe als Plagiat bzw. Täuschungsversuch und auch die nachgewiesene Verwendung nicht angegebener KI- gestützter Werkzeuge als versuchte Täuschung gewertet und mit entsprechenden Maßnahmen geahndet wird.

Detmold, 4. November 2026

Ort, Datum

Unterschrift