Konzepte zur Syntheseplanung: Strukturelle Ähnlichkeit und strategische Bindungen

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

zur

Erlangung des Doktorgrades

vorgelegt von

Ralf Fick

aus Eilenburg

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

Tag der mündlichen Prüfung:	25. Juni 1996
Vorsitzender der Promotionskommission:	Prof. Dr. D. Kölzow
Erstberichterstatter:	Prof. Dr. J. Gasteiger
Zweitberichterstatter	Prof. Dr. A. Hirsch

Meinem Doktorvater

Herrn Prof. Dr. J. Gasteiger

danke ich für die vielfältige Unterstützung, ermutigende Diskussionen und nicht zuletzt für die herzliche Aufnahme in seiner Arbeitsgruppe, ohne die die vorliegende Arbeit nicht möglich gewesen wäre.

Weiteren Dank schulde ich:

- Herrn Dr. Wolf-Dietrich Ihlenfeldt für Schaffung des Ausgangspunktes dieser Arbeit und die nie nachlassende Bereitschaft, programminterne Probleme oder Design-Konzepte zu diskutieren,
- Herrn Dr. Markus Wagener, Dr. Wolfgang Witzenbichler und Paul Selzer für die Bereitstellung und Pflege einer stabilen Hard- und Software-Umgebung,
- Frau Dr. Angela Fröhlich für die von ihr entwickelten empirischen Parameter, die eine nicht unwesentliche Grundlage für einen Teil dieser Arbeit darstellen.

Den nicht namentlich erwähnten Mitarbeitern der Arbeitsgruppe danke ich für die ausgezeichnete Arbeitsatmosphäre und vielerlei Hilfestellungen bei Problemen im universitären Alltag.

Für finanzielle Unterstützung gebührt Dank:

- dem Verband der Chemischen Industrie,
- der Deutschen Forschungsgemeinschaft sowie
- der Firma Shell, Schwabenheim

NVNQVAM OTIOSVS

für Hilde, Heinz und Kirsten

Inhaltsverzeichnis

Inhaltsverzeichnis i

Abkürzungen, Symbole und Akronyme iii

1 Motivation 1

2 Computer-gestützte Syntheseplanung 3

2.1 Syntheseplanung und Computer - Konzepte und Methoden 3

2.2 Das WODCA System - Ein Pool von Werkzeugen zur Syntheseplanung 10

2.2.1 Methoden zur Suche von Ausgangsmaterialien 13

2.2.2 Schrittweise Vorstufengenerierung 15

2.2.3 Bewertungsmechanismen 16

2.2.4 Interaktion mit dem Anwender 17

2.2.5 Zielsetzung dieser Arbeit 18

2.3 Ein kleiner Exkurs - Programmtechnische Aspekte 21

3 Ähnlichkeitssuchen mittels Transformationssuche 23

3.1 Prinzip 23

3.2 Technische Realisierung 23

3.2.1 Randbedingungen 23

3.2.2 Skriptsprache *tcl* 25

3.2.3 Chemiebezogene Erweiterungen 28

3.3 Anwendungen in der Syntheseplanung 35

3.3.1 Kataloge 35

3.3.2 Grundlegendes 39

3.3.3 Transformationen zur Erkennung skeletaler Ähnlichkeit 47

3.3.4 Substitutionsmuster als selektivere Ähnlichkeitskriterien 51

3.3.5 Ähnlichkeiten durch generalisierte Reaktionen 68

3.3.6 Hydrolyse-Transformationen 68

3.3.7 Transformationen basierend auf Redox-Prozessen 71

3.3.8 Eliminierungen als Grundlage von Ähnlichkeit 78
3.3.9 Sonderfall Ozonolyse-Transformation 81
3.3.10 Kombinationen von Elementarschritten 84
3.3.11 Benutzerschnittstelle zu den Transformationssuchen 89
3.4 Andere Anwendungen 93
4 Schrittweise Retrosynthese - Suche nach strategischen Bindungen und Erzeugung von Vorstufen 95
4.1 Das Problem 95
4.2 Möglichkeiten zur retrosynthetischen Analyse in WODCA 97
4.3 Implementierte Retrosynthese-Strategien 99
4.3.1 Etablierte Prinzipien 99
4.3.2 Strategien basierend auf CX-Bindungen 100
4.3.3 Strategie für CC-Bindungen 113
4.3.4 Eine formale Strategie für komplexe Ringsysteme 127
4.4 Kommunikation mit dem Benutzer 134
5 Zusammenfassung 137
6 Literaturverzeichnis 139
A Anhang A-1
A.1 Rechenzeiten A-2
A.2 Transformationssuchen A-11

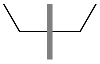
Abkürzungen, Symbole und Akronyme

Graphische Symbole

$\longrightarrow$ .... Pfeil in Reaktionen

$\Longrightarrow$ .... Retrosynthese-Pfeil

 .... Pfeil bei Transformationen oder anderen Suchen

 .... Marker für strategische Bindungen

Alpha-numerische Symbole, physikalische Größen und Abkürzungen

α .... Atompolarisierbarkeit
CC .... Kohlenstoff-Kohlenstoff-Bindung/-Bindungsordnung usw.
χ_π .... π-Elektronegativität
χ_σ .... σ-Elektronegativität
CT .... Konnektivitätstabelle (connection table)
CC .... Kohlenstoff-Kohlenstoff-Bindung, -Bindungsordnung usw.
CX .... Kohlenstoff-Heteroatom-Bindung, -Bindungsordnung usw.
D^+ .... Delokalisierungssatbilisierung positiver Ladungen
D^- .... Delokalisierungsstabilisierung negativer Ladungen
$D^\pm$ .... (Gesamt-)Delokalisierungssatbilisierung von Ladungen
$\Delta\chi_\pi$ .... Differenz der π-Elektronegativität über eine Bindung
$\Delta\chi_\sigma$ .... Differenz der σ-Elektronegativität über eine Bindung
Δq .... Differenz der (Gesamt-)Atompartialladungen zweier Bindungsatome
Δq_π .... Differenz der π-Atompartialladungen zweier Bindungsatome
Δq_σ .... Differenz der σ-Atompartialladungen zweier Bindungsatome
FGA .... Addition funktioneller Gruppen (functional group addition)
FGI .... Austausch funktioneller Gruppen (functional group interconversion)
FGR............... Entfernung funktioneller Gruppen (functional group removal)
GUI .... graphische Benutzerschnittstelle (graphical user interface)
M .... molare Masse
q .... (Gesamt-)Atompartialladung
q_π .... π-Atompartialladung
q_σ .... σ-Atompartialladung
VHLL............ very high level languages

Akronyme

AHMOS	Automatisierte Heuristische Modellierung Organisch-chemischer Synthesen bzw. Automated Heuristic Modeling of Organochemical Syntheses
AIPHOS	Artificial Intelligence for Planning and Handling Organic Synthesis
ALCHEM	ALCHEMy
CAMEO	Computer-Assisted Mechanistic Evaluation of Organic reactions
CASP	Computer-Assistierte SynthesePlanung bzw. Computer Assisted Synthesis Planning
CASREACT	Chemical Abstracts Service REACtions
CHIRON	CHIRal synthON
CHMTRN	CHeMical TRaNslator
CICLOPS	Computers In Chemistry, Logic Oriented Planning of Syntheses
COSYMA	COnception de SYnthèses par une Méthode Analogique
EROS	Erzeugung von Reaktionen zur Organischen Synthesen bzw. Elaboration of Reactions for Organic Synthesis
FLAMINGOES	Formal-Logical Approach to Molecular Interconversions, Nonempirical Generation, Orientation, and Evaluation of Syntheses
GRAMS	Générateur de Réseaux pour l'Apprentissage de Méthodes de Synthèse
HORACE	Hierarchical Organization of Reactions through Attribute and Condition Eduction
IGOR	Interactive Generation of Organic Reactions
LHASA	Logic and Heuristics Applied to Synthesis Analysis
Lilith	kein Acronym, Name einer Feengestalt
MARSEIL/SOS	Multi Approaches for the Research of Synthesis by Efficient and Interactive Logic / Simulated Organic Synthesis
OCSS	Organic Chemical Simulation of Synthesis
PETRA	Parameter Estimation for the Treatment of Reactivity Applications
RAIN	Reactions And Intermediates Network
RDSS	Reaction Design by Synthon Substitution
SECS	Simulation and Evaluation of Chemical Synthesis
SPEK	Synthesis Planning based on Empirical Knowledge
SST	Starting materials STrategies
STN	Scientific and Technical Information Network
SYNCHEM	SYNthetic CHEMistry
SYNGEN	SYNthesis GENerator
TRESOR	TRacing and Evaluation of Synthons in Organic Reactions
VERGIL	Versatile EROS Reactivity functions Generation and Interpretation Language
WODCA	Workbench for the Organization of Data for Chemical Applications

[Corey_1967_19] 1
[Corey_1969_178] 4
[Corey_1971_455] 4
[Fuhrhop_1983_1] 4
[Corey_1985_408] 4
[Corey_69_178] 4
[Corey_1967_19] 4
[Vleduts_1963_117] 5
[Corey_1971_455] 5
[Corey_1972_421] 5
[Corey_1972_431] 5
[Corey_1972_440] 5
[Corey_1972_460] 5
[Corey-1972-440] 5
[Corey_1989_1] 5
[Wipke_74_4825] 6
[Wipke_1977_97] 6
[Gelernter_1973_113] 6
[Bersohn_1972_1897] 6
[Takahashi_1990_436] 6
[Nakayama_1991_495] 6
[Moll_1994_117] 6
[Wipke_1974_4825] 6
[Gelernter_1977_1041] 6
[Braun_1988_43] 6
[Dugundji_1973_19] 6
[Blair_1974_1845] 6
[Bauer_1982_298] 6
[Fontain_1991_469] 6
[Gasteiger_1979_177] 6
[Hendrickson_1985_522] 6
[Hendrickson_1975_5784] 7
[Hendrickson_1992_323] 7
[Funatsu_1988_27] 7
[Hanessian_1990_1887] 8
[Salatin_1980_2043] 8
[Roese_1990_163] 8
[Salatin_1981_5284] 8
[Peishoff_1983_1970] 8
[Peishoff_1985_1056] 8
[Metivier_1987_3724] 8
[Gushurst_1988_3397] 8
[Laird_1990_9] 8
[Bures_1985_4490] 8
[Bures_1988_2504] 8
[Burnier_1983_3923] 8
[Burnier_1984_3001] 8
[Bersohn_1979_137] 9

[Roese_1990_163] 9
[Gelernter_1990_492] 9
[Rose_1994_74] 10
[Jauffret_1990_323] 10
[Ihlenfeldt_1995_2807] 10
[Gasteiger_1979_177] 10
[Gasteiger_1992_270] 13
[Ihlenfeldt_1994_793] 14
[Corey_1971_455] 16
[Ihlenfeldt_Diss.S.105] 16
[Ihlenfeldt_Diss.S.78] 16
[Ihlenfeldt_1994_872] 18
[Ihlenfeldt_1994_872] 19
[Corey_72_421] 21
[Weise_1975_333] 21
[Jorgensen_90_1921] 21
[Hamm_1992_317] 21
[Zefirov_1988_188] 21
[Bauer_1989_269] 21
[Corey_71_455] 21
[Corey_71_455] 21
[Long_94_922] 21
[Sello_94_120] 21
[Moreau_78_187] 21
[Azario_88_720] 21
[PrakInBitterfeld] 21
[Pers.Comm.Goslar] 21
[Wipke_84_71] 21
[Gelernter_77_1041] 21
[Pers.Comm.John] 21
[Hendrickson_85_5228] 21
[Bersohn_72_1897] 21
[Blower_76_1499] 21
[Barone_88_3] 21
[Sussman_1975_1] 25
[Abelson_1985_1] 25
[Clinger_1991_1] 25
[Wall_1991_1] 25
[Wall_1993_1] 25
[Ousterhout_1994_1] 26
[Welch_1995_1] 26
[Ihlenfeldt_1994_793] 29
[Ihlenfeldt_1994_793] 29
[Cahn_1966_413] 29
[Prelog_1982_614] 29
[Ihlenfeldt_1994_793] 29
[Ihlenfeldt_1994_793] 29
[Gasteiger_1984_559] 31

[Gasteiger_1980_3219] 31
[Froehlich_Diss.93] 31
[Froehlich_Diss.93] 31
[Froehlich_Diss.93] 31
[Hutchings_1983_2541] 31
[Hutchings_1983_2541] 31
[Gasteiger_1985_699] 31
[Gasteiger_1980_3219] 31
[Gasteiger_1984_559] 31
[Osterhout_1994_1] 33
[tclForChemistsRefManual] 33
[CACTVS_tcl_Syntax] 33
[CTX_Ref] 36
[Dalby_1992_244] 36
[Bebak_1989_1] 36
[Barnard_1990_81] 36
[Bernstein_1977_535] 36
[CACTVS_Doku] 36
[Wipke_1984_71] 39
[Wipke_1984_71] 39
[CACTVS_SubGenerator] 56
[Gasteiger_1980_3219] 78
[Corey_1967_19] 95
[Hendrickson_1980_103] 99
[Baumer_1990_209] 99
[Baumer_1988_1195] 99
[Corey_1975_6116] 99
[Corey_1974_7724] 99
[Corey_1980_2051] 99
[Corey_1980_221] 99
[Boiten_1993_727] 99
[Corey_1978_2208] 99
[Warren_1982_1] 100
[Fuhrhop_1983_1] 100
[March_1992_1] 100
[Pauwels_1990_470] 107
[Parker_1991_4600] 110
[Michael_1887_113] 113
[Bergman_1959_179] 113
[Oare_1989_227] 113
[Oare_1991_87] 113
[Semmelhack_1991_1] 113
[Parlow_1990_400] 114
[Gasteiger_1992_1114] 114
[Chen_1995_8002] 114
[Kosower_1990_7305] 121
[Kosower_1978_6516] 121
[Kosower_1990_7305] 123

[Narisada_1988_1847] 123
[Morris_1991_3549] 124
[Corey_1989_1] 127
[Corey_1989_1] 127
[Deslongchamps_1984_59] 128
[Corey_1989_1] 129
[CACTVS_Tools] 134
[CACTVS_Tools] 134

AHMOS............ 21
AIPHOS............ 7
ALCHEM 6
CAMEO............ 8
CASP 6
CASREACT 9
CHIRON........... 8
CHMTRN 5
CICLOPS........... 6
COSYMA 21
EROS 8
FLAMINGOES . 21
GRAMS 10
HORACE........... 10
IGOR 6
LHASA............. 5
Lilith 21
MARSEIL/SOS . 21
MASSO 21
OCSS 4
PETRA 28
RAIN 6
RDSS 6
SECS................. 6
SPEK 6
SST 7
STN................... 9
SYNCHEM........ 6
SYNGEN........... 6
TRESOR............ 6
VERGIL 18
WODCA 10

Motivation

Bereits mehr als ein Vierteljahrhundert ist es ein Ziel der chemischen Forschung, die Planung organischer Synthesen zu systematisieren. Basierend auf den Pionierarbeiten von Corey [1] entwickelte sich eine neue Disziplin, die mehrere Wissenschaftszweige zu einem neuen Forschungsgegenstand verschmolz: die computer-gestützte Syntheseplanung. Diese ist primär eine chemische Wissenschaft, im besonderen ein Spezialgebiet der Organischen Chemie. Sie wäre jedoch undenkbar ohne entscheidende Beiträge der informationsverarbeitenden Wissenschaften und der Computertechnologie. Diese Überschneidung einer Reihe von - auch jeder für sich - interessanten und herausfordernden Wissenschaftszweigen macht den besonderen Reiz der computer-gestützten Syntheseplanung aus.

Trotz des beachtlichen Interesses, welches diese Fachrichtung bei Forschergruppen in der ganzen Welt gefunden hat, ist ein Abschluß der Entwicklungen noch nicht absehbar. Diese Arbeit tritt den Beweis an, daß auch gegenwärtig Entwicklungen im Bereich der computer-gestützten Syntheseplanung vonstatten gehen. Die hier vorgestellten Arbeiten fließen in ein eigenes Programmsystem ein, das ebenso wie die Anstrengungen zahlreicher anderer Arbeitsgruppen im folgenden Kapitel vorgestellt wird.

Kapitel 3 befaßt sich mit einer neuen Methode zur Planung organischer Synthesen, die versucht, Relationen zwischen einem Zielmoleküls und möglichen Ausgangsmaterialien herzustellen. Die Einzelheiten der Realisierung, ihre Bedeutung und ihr Potential in der computer-gestützten Syntheseplanung, wie auch mögliche Anwendungen in anderen Bereichen werden diskutiert.

Kapitel 4 schließlich beschäftigt sich mit einem „alten” Problem der computer-gestützten Syntheseplanung: der schrittweisen retrosynthetischen Analyse. Die Bedeutung dieses Problemkreises in Bezug auf unser Programmsystem wird ebenso behandelt wie die Besonderheiten des gewählten Lösungsansatzes und der praktische Wissenserwerb.

Computer-gestützte Syntheseplanung

Dieser Abschnitt stellt die historische Entwicklung des im Titel umrissenen Forschungsbereiches von seinen Anfängen bis in die 90er Jahre dar. Neben den zahlreichen Entwicklungen, die im Laufe der Jahre vollzogen worden sind, wird im zweiten Teil das WODCA Programmsystem vorgestellt.

2.1 Syntheseplanung und Computer - Konzepte und Methoden

Der Versuch, die Planung organischer Synthesen auf eine rationale Grundlage zu stellen, wurde bereits vor mehr als 25 Jahren von Corey [1] unternommen. Er formulierte erstmals Regeln und Schritte, die nach- oder nebeneinander vollzogen werden müssen, um einen Syntheseplan zu entwerfen. Zu dieser Zeit wurde organische Synthese in einer Weise gelehrt, welche die Entwicklung einer Syntheseidee als einen *ad hoc* Prozeß begriff: Einzelbeispiele von eleganten Synthesen wurden dem Studierenden präsentiert, jedes ohne direkte Beziehung zu einem andern. Induktives Ableiten von Lösungen für neue Probleme war erforderlich. Corey war es, der den Chemikern mit der retrosynthetischen Denkweise ein vielversprechendes, deduktives Werkzeug in die Hand gab, mit dem die Planung organischer Synthesen logischer, systematischer und auch einfacher wurde. Die folgenden zwölf Schritte sind aus [1] entnommen. Sie artikulieren das Fundament der retrosynthetischen Analyse:

1. Vereinfache das Syntheseproblem.
2. Erkenne wichtige Struktureinheiten im Syntheseziel.
3. Erzeuge äquivalente Struktureinheiten und modifizierte diese.
4. Addiere funktionelle Gruppen, die reaktionskontrollierend wirken.
5. Trenne systematisch das Molekülskelett an den Struktureinheiten.
6. Leite die notwendigen Reaktionen ab, welche die Vorstufen zur Zielstruktur umsetzen.
7. Wiederhole die Schritte 1 bis 6 für jede erzeugte Vorstufe und jede Sequenz.
8. Erzeuge Vorstufen, bis geeignete Ausgangsmaterialien produziert wurden.
9. Entferne Inkonsistenzen.
10. Identifiziere offene Probleme.
11. Wiederhole die Schritte 1 bis 10 zur Erzeugung alternativer Wege.
12. Bewerte jeden Weg.

Corey war es auch, der eine Reihe von Begriffen etablierte, die im Bereich der Syntheseplanung, insbesondere in ihrem computer-gestützten Umfeld eine zentrale Rolle spielen.

Zunächst muß deutlich zwischen den beiden möglichen Richtungen der Analyse einer Reaktion unterschieden werden: Als *synthetisch* bezeichnet man die Richtung der Umsetzung von Edukten zu Produkten, analog zu einer *Reaktion*, die im Labor durchgeführt wird. *Retrosynthetisch* oder *antithetisch* heißt dagegen die umgekehrte Richtung, wenn Produkte hinsichtlich ihrer möglichen Edukte analysiert werden [2]. Im Unterschied zu einer Reaktion wird der umgekehrte Prozeß als *Transform* bezeichnet [3].

Unter dem Begriff des *Synthons*, ursprünglich relativ weit gefaßt*, wird heute [4], [5] ein Fragment verstanden, welches das (denkbare) reagierende Intermediat in einer Reaktion umfaßt. Synthone sind meistens geladene Spezies. Von ihnen werden *Reagenzien* unterschieden, welche die eigentlich eingesetzten Verbindungen bezeichnen. Demgegenüber beschreibt ein *Retron* [6] diejenige strukturelle Einheit in einem Reaktionsprodukt, welche eine mögliche Reaktion zu ihrem Aufbau signalisiert. Retrone sind damit die Schlüsselelemente zur Entscheidung über die Anwendbarkeit eines bestimmten Transforms. Das Ergebnis einer retrosynthetischen Analyse wird gewöhnlich in Form eines *Synthesebaumes* dargestellt. Das eigentliche Syntheseziel stellt dabei die Wurzel des Baumes dar, der auf dem Kopf stehend (Wurzel oben) dargestellt wird. Zwischenprodukte sind Knoten innerhalb des Baumes, Ausgangsmaterialien die Endpunkte von Ästen. Kanten repräsentieren Reaktionen.

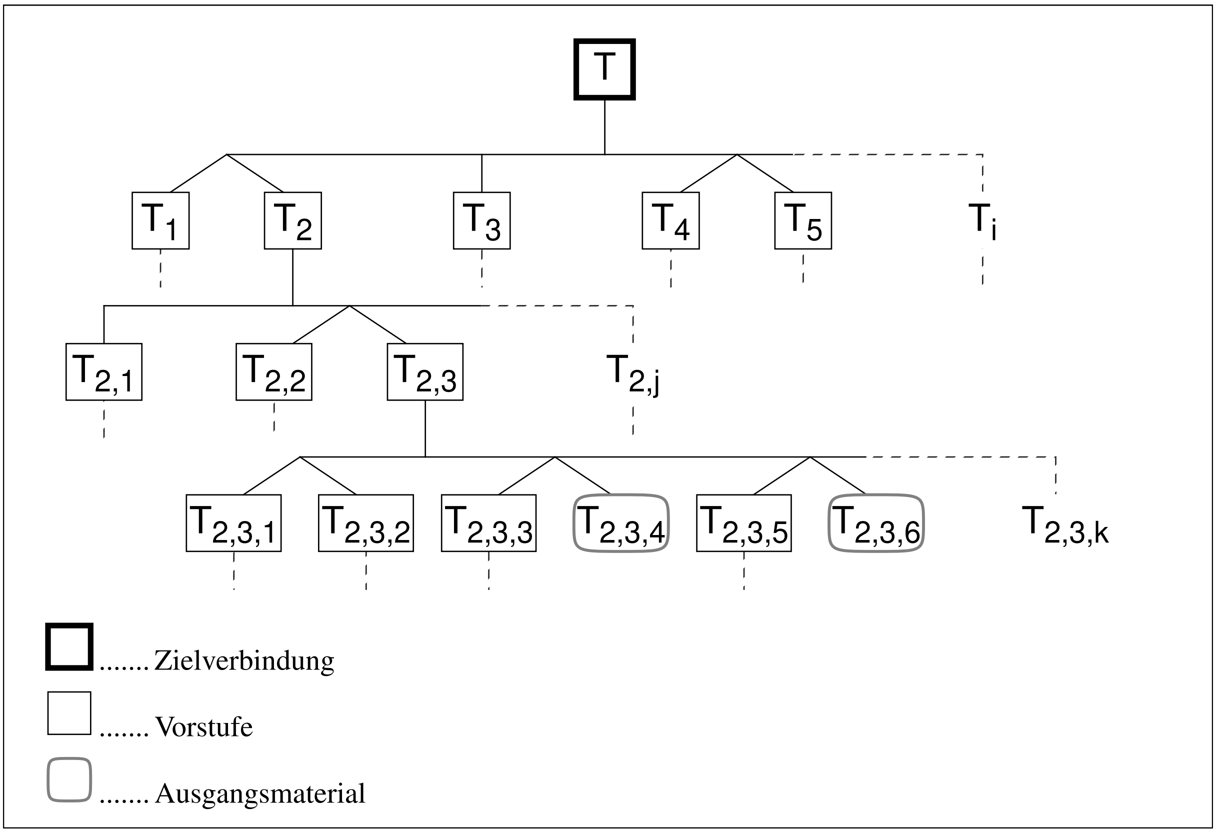


Abb. 2 - 1 Ein (Retro-)Synthesebaum.

Indem man Heuristiken und Regeln zur Planung einer Synthese angeben kann, ergibt sich automatisch die Möglichkeit, einen solchen Algorithmus in Form eines Computerprogramms zu realisieren. Corey und Wipke haben diesen Ansatz mit OCSS† [2], dem ersten Prototypen eines Programms zur computer-gestützten Syntheseplanung, verwirklicht. OCSS ist

* Siehe [1]: „structural units within a molecule which are related to possible synthetic operations"

† Organic Chemical Simulation of Synthesis

der Urvater einer Reihe ähnlich arbeitender Programme, auf die weiter unten noch eingegangen wird.

Es soll hier jedoch zunächst auf eine andere Publikation verwiesen werden, in der ebenfalls wesentliche Grundzüge der später realisierten computer-gestützten Syntheseplanung vorgeschlagen wurden. Vléduts [7] veröffentlichte bereits 1963 einen Artikel, der sich mit einem eindeutigen System zur Indizierung von Reaktionsdaten befaßte. Als eine potentielle Anwendung beschrieb er dabei, wie sich eindeutig indizierte Reaktionen zur Generierung von Synthesevorschlägen einsetzen ließen. Vléduts' *skeleton reaction schemes* entsprechen dabei in etwa den Transforms von Corey. Auch der Begriff des Synthesebaums als Bezeichnung für die Gesamtheit der erzeugten Vorstufen findet sich dort.

Doch zurück zu den ersten realisierten Syntheseplanungsprogrammen. Anfang der 70er Jahre war die erste Version des Nachfolgers von OCSS, LHASA[*] [3], [8], [9], [10], [11] funktionsfähig. Das wesentliche Arbeitsprinzip ist die rationale Anwendung von Transforms auf ein Syntheseziel, um Vorstufen zu generieren. Dieser Prozeß wird fortgesetzt, bis Vorstufen als verfügbar erkannt werden. Kern eines solchen Systems sind die Menge der bekannten Transforms. Bei LHASA können diese in einer eigens entwickelten Sprache (CHMTRN[†] [10]) codiert werden. Dabei wird nicht nur das erforderliche Retron des Transforms codiert, sondern auch Bewertungen spezifiziert, Randbedingungen aufgeführt und Grenzen abgesteckt. Die Transforms sind gruppiert. Funktionalität umwandelnde Transforms werden beispielsweise von solchen zum Aufbau des Kohlenstoff-Skeletts unterschieden. Diese Strukturierung erlaubt die Auswahl von Strategien zur Anwendung bestimmter Transform-Gruppen [12]:

- Strategien basierend auf funktionellen Gruppen. Bestimmte Relationen von funktionellen Gruppen zueinander erlauben die Anwendung von Transforms, welche die Molekülkomplexität drastisch verringern.
- Transform-basierte Strategien: Zu dieser Gruppe gehören besonders mächtige Algorithmen (*long range searches*), die darauf hinarbeiten, die erforderlichen strukturellen Merkmale zur Anwendung eines besonders leistungsfähigen Transforms zu erzeugen. Dazu werden u. U. eine ganze Reihe von taktischen Reaktionsschritten prognostiziert.
- Auf topologischen Merkmalen beruhende Strategien. Hierbei signalisieren generelle strukturelle Merkmale (die nicht nur auf funktionellen Gruppen oder deren Relation zueinander basieren müssen) im Syntheseziel eine Richtung in der Retrosynthese. Beispiele dafür sind bestimmte Relationen zwischen Bindungspaaren oder Lageeigenschaften von Ringbindungen.
- Unmittelbare bzw. opportunistische Strategien. Diese wenden vor allem Transforms zur Umwandlung (*functional group interconversions* - FGI), zur Addition bzw. Entfernung von funktionellen Gruppen (*functional group addition* - FGA bzw. *removal* - FGR) an. In diesem Fall ist deren Anwendung aber ohne lenkenden Hintergrund.
- Strukturbasierte Strategien. Darunter fallen Verfahren, die ein Ausgangsmaterial, einen wesentlichen Baustein (*building block*) oder auch ein zentrales Zwischenprodukt aus Sicht des Syntheseziels erkennen können. Durch Erkennung solcher Vorstufen kann eine bidirektionale Planung zum einen Syntheseziel retrosynthetisch hin zur Vorstufe, zum anderen von der Vorstufe in Syntheserichtung zur Zielverbindung stattfinden.

[*] Logic and Heuristics Applied to Synthesis Analysis

[†] CHeMical TRaNslator

SECS[*] [13], [14], SYNCHEM [15], die Programme aus der Umgebung von Bersohn [16], [17], SPEK[†] [18] und die Familie RDSS[‡]/TRESOR[**] [19] sind ähnlich arbeitende Programme. Die beiden ersten kennen ebenfalls eigene Regelsprachen zur Codierung der Transforms (ALCHEM [13] bzw. eine *PL/1*-ähnliche Sprache [20]). Auf der Grundlage von SECS wurde das CASP[††] Projekt [21] vorangetrieben, das von einem aus schweizer und deutschen Chemieunternehmen bestehendem Konsortium verwirklicht wurde. Hier erreichte die Transform-Bibliothek mehr als 6.000 Einträge. CASP kann als der Endpunkt einer Entwicklung angesehen werden. Offensichtlich macht nicht die Masse in den Transform-Bibliotheken die Fähigkeiten solcher Systeme aus, sondern deren Kombination mit effektiven Methoden der Bewertung von Reaktionspfaden und die Verwendung verschiedener Strategien während der Entwicklungsphase des Synthesebaumes.

Ganz allgemein werden Programme wie die oben beschriebenen als wissensbasierte Systeme bezeichnet. Im Unterschied dazu gibt es eine andere Gruppe von Programmen zur computer-gestützten Syntheseplanung, die auf einer rein mathematisch-logischen Grundlage das Problem in Angriff nehmen. Die Grundlage dafür legten Dugundji und Ugi [22]. Sie benutzten als Strukturabbildung Matrizen, die Bindungen und Elektronen codieren (BE-Matrizen) und statt Transformbibliotheken Reaktions-Matrizen (R-Matrizen), welche die Veränderungen von BE-Matrizen während einer Reaktion beschreiben. Da dieses System stöchiometrieerhaltend arbeitet, wären damit nur Cycloadditionen, Ringöffnungen, Umlagerungen u. ä. möglich. Das wird überwunden, indem einfache Verbindungen wie Wasser, Natronlauge u. dgl. zur BE-Matrix der Zielverbindung hinzugefügt werden. Der große potentielle Vorteil solcher Systeme liegt darin, daß sie - im Gegensatz zu den wissensbasierten Systemen - auch völlig neuartige Reaktionen auf diese Weise generieren *können*. Jedoch ist eine völlig unkontrollierte Generierung von Reaktionen wenig sinnvoll, da der Anwender dann mit einer Unmenge an chemisch zweifelhaften Vorschlägen überschwemmt wird. Das erste derartige Programm war CICLOPS[‡‡] [23]. Andere Programme dieses Typs sind IGOR[***] [24], [25] und RAIN[†††] [26]. Auch die frühen EROS-Versionen (siehe unten) gehörten in diese Kategorie [27]. Man kann diese Systeme als formale Reaktionsgeneratoren auffassen. Mit ihnen wird nicht vordergründig Syntheseplanung betrieben. Vielmehr kann mit ihrer Hilfe gezielt nach noch unbekannten Reaktionen, etwa zur Synthese bestimmter Strukturmerkmale gesucht werden oder sie helfen bei der Formulierung eines Reaktionsmechanismus. Erstere Anwendungen werden auch als Reaktionsplanung [28] oder Reaktionsdesign bezeichnet.

Ein anderer Pionier der computer-gestützten Syntheseplanung ist Hendrickson mit seinem SYNGEN Programm [29]. Sein Ansatz basiert auf der Charakterisierung von Kohlenstoff-Atomen in organischen Verbindungen durch vier Merkmale: die Zahl der σ-Bindungen (σ) und Mehrfachbindungen (Π) zu anderen Kohlenstoff-Atomen, der Bindungen zu Wasserstoff oder anderen elektropositiven Atomen (H) und der Bindungen zu elektronegativen Heteroatomen (Z) [30]. Dieses System eignet sich zur rationalen Beschreibung der Konstitution beliebiger organischer Verbindungen und ihrer Reaktionen. Weiterentwickelt zum Konzept der

[*] Simulation and Evaluation of Chemical Synthesis

[†] Synthesis Planning based on Empirical Knowledge

[‡] Reaction Design by Synthon Substitution

[**] TRacing and Evaluation of Synthons in Organic Reactions

[††] Computer Assisted Synthesis Planning

[‡‡] Computers In Chemistry, Logic Oriented Planning of Syntheses

[***] Interactive Generation of Organic Reactions

[†††] Reactions And Intermediates Network

Halb-Reaktionen [31] ist es die Grundlage einer ganzen Familie* von Programmen zur Syntheseplanung. Ein wesentlicher Nachteil seines Ansatzes besteht darin, daß stereogene Merkmale (chirale Atome, *cis*/*trans*-Relationen an Doppelbindungen) dabei nicht berücksichtigt werden. In [32] weist Hendrickson nach, daß die Ugischen BE- und R-Matrizen und seine Halb-Reaktionen verschiedene Spielarten *eines* rationalen Modells zur Beschreibung organischer Verbindungen und ihrer Reaktionen sind. Es kann daher ebenfalls in die Gruppe der mathematisch-logisch basierten Systeme eingeordnet werden.

Ein hybrides System, das sowohl mittels formaler Reaktionsgeneratoren als auch mit wissensbasierten Komponenten arbeitet, ist AIPHOS† [34]. Der Reaktionsgenerator dient hier dazu, die retrosynthetisch zu brechenden Bindungen auszuwählen (in AIPHOS-Terminologie: *strategic sites*). Ohne eine drastische Einschränkung der formalen Möglichkeiten zur Reaktionsgenerierung wird man bereits in diesem Schritt von Unmengen an Vorschlägen überflutet. Aus dem Bruch solcher Bindungen resultieren zunächst nur Molekülfragmente. Zur Absättigung dieser Fragmente wird auf eine Wissensbasis zurückgegriffen, die automatisch aus Einzelreaktionen ermittelte Abgangsgruppen enthält. Neben der eigentlichen *leaving group* sind auch topologische Eigenschaften ihrer Nachbaratome in der ursprünglichen Reaktion vermerkt. In der Praxis werden für die aus einer strategischen Bindung resultierenden Molekülfragmente in der Regel eine ganze Reihe von Abgangsgruppen für vorgeschlagen. Diese erhalten aber eine Bewertung, so daß der Anwender eine Entscheidungshilfe erhält. Stehen die zu brechende Bindung und deren Synthesevorstufen fest, kann AIPHOS in einer weiteren Wissensbasis nach den geeignetsten Reaktionsbedingungen suchen. Die Lösung dieser Aufgabe basiert auf dem Vergleich der aktuellen Abgangsgruppen mit denen in der Wissensbasis für Reaktionsbedingungen. Es werden so aber nur Prototypen für Bedingungen ermittelt.

Desweiteren wurden einige Programme entwickelt, deren explizite Aufgabe darin besteht, Relationen zwischen Synthesezielen und Ausgangsmaterialien zu erkennen. Ein erster solcher Ansatz stammt von Wipke [35]. Sein SST Programm‡ sucht in einem Katalog von Ausgangsmaterialien (Aldrich-Katalog, ca. 11.000 Verbindungen) ohne Berücksichtigung irgendwelcher Reaktionswege diejenigen Verbindungen, die sich durch skeletale** Ähnlichkeit zum Syntheseziel auszeichnen. Dazu abstrahiert das Programm sowohl die Anfrageverbindung (Syntheseziele) als auch die Katalogverbindungen (Ausgangsmaterialien). SST kennt zwei Abstraktionsstufen: eine vollständige Abstraktion und eine funktionale Abstraktion. Die vollständige Abstraktion reduziert eine Verbindung auf ihr Kohlenstoffskelett, wobei nicht-aromatische Mehrfachbindungen zwischen Kohlenstoffatomen zu Einfachbindungen reduziert werden. Aromatische Bindungen, auch die zu Heteroatomen, bleiben dagegen unverändert. Bei der funktionalen Abstraktion wird das Kohlenstoffskelett der höheren Abstraktionsstufe zusätzlich mit Markern an Positionen versehen, an denen ursprünglich chemische Funktionalität in irgendeiner Form (Mehrfachbindungen, Heteroatome) vorhanden war. Suchfunktionen vergleichen dann diese reduzierten Darstellungen und selektieren so eine Liste von Ausgangs-

---

* SYNGEN - das eigentliche Syntheseplanungsprogramm; SYNOUT [29] - ein Visualisierungsprogramm für die SYNGEN-Ergebnisse; RETRIEVE [32] - ein Programm zur Suche von Literaturreferenzen für vorgeschlagene Syntheseschritte; FORWARD [32] - ein Programm zur Suche nach Synthesewegen ausgehend von Ausgangsmaterialien hin zu einem bestimmten Syntheseziel und schließlich COGNOS [33] - ein Reaktions-Retrievalsystem für speziell vorbereitete Reaktionsdatenbanken.

† Artificial Intelligence for Planning and Handling Organic Synthesis

‡ Benannt nach einem amerikanischen Projekt für ein Überschallverkehrsflugzeug (Super Sonic Traffic), daß große Entfernungen nonstop und schnell überwinden können sollte.

** Skeletale Ähnlichkeit ist über die Verwandtschaft von Molekülskeletten, d. h. meist der C-Skelette, definiert.

materialien. Diese Liste wird dann auf der Ebene der reduzierten Darstellungen bewertet, wodurch sich die Liste weiter eingrenzen läßt. Schließlich werden die korrespondierenden Originalstrukturen zu einer abschließenden Bewertung herangezogen. Das Resultat sind Vorschläge für Ausgangsmaterialien für das gegebene Zielmolekül - ohne irgendwelche Reaktionen zu prognostizieren.

Einen weiteren, aber zusätzlich spezialisierten Ansatz stellt das CHIRON Programm* dar [36]. Wie SST versucht das CHIRON Programm eine synthesebasierte Relation zwischen einem Zielmolekül und möglichen Ausgangsmaterialien herzustellen. Zentraler Schwerpunkt ist dabei die Korrespondenz von stereochemischen Merkmalen. CHIRON steht ein manuell zusammengestellter Katalog von chiralen, racemischen und achiralen Ausgangsmaterialien zur Verfügung. In der gegenwärtigen Version (4.22 vom Februar 1994) umfaßt dieser Katalog 4.567 Verbindungen (davon 1.765 chiral) in13 Dateien. Jedes Katalogmolekül wird auf die Zielverbindung abgebildet, dabei werden Funktionalität, aber besonders auch stereochemische Merkmale verglichen. Lassen sich auf diese Weise keine ausreichenden Übereinstimmungen erzielen, gestattet das Programm begrenzte Transformationen an funktionellen Gruppen der Katalogverbindungen. Dieser Vergleichsprozeß geht einher mit einer Bewertung. Der Wert eines Ausgangsmaterials wird dabei beeinflußt, inwieweit notwendige Refunktionalisierungen machbar sind, dem Maß der Überlappung der Kohlenstoffskelette und inwiefern sich Funktionalität und vor allem die Stereochemie entsprechen. In einem weiteren Operationsmodus können auch CC-Bindungen brechende Transformationen auf der Seite der Katalogverbindungen erlaubt sein, bevor der Vergleichsprozeß abläuft. Als solche brechbaren Bindungen werden nicht-aromatische Doppelbindungen, α-carbonylische oder Bindungen an allylischen Positionen betrachtet. Ist auf diese Weise eine Ähnlichkeit zwischen Syntheseziel und einem chiralen Baustein hergestellt, bleibt es dann Aufgabe des Chemikers zu entscheiden, wie der konkrete Syntheseweg vom Ausgangsmaterial zum Produkt durchgeführt werden kann.

Schließlich müssen grundsätzlich in Vorwärtsrichtung, d. h. in Syntheserichtung arbeitende Programme von den bislang erläuterten unterschieden werden. Solche Systeme sollen Voraussagen über die Produkte einer Reaktion machen, das jedoch nicht nur durch Prognose der Produktstrukturen, sondern darüber hinaus der Produkt- bzw. Isomerenverteilungen. Zu dieser Klasse gehören CAMEO† [37] und EROS6‡ [38]. Beide Systeme benutzen keine Reaktionsmuster. Vielmehr wird die Reaktivität von Atomen und Bindungen unter bestimmten Bedingungen abgeschätzt, um daraus eine Reaktionsablauf zu prognostizieren.

CAMEO besteht aus zahlreichen, fest programmierten Modulen, die viele Teilbereiche der organischen Chemie (basenkatalysierte und nukleophile Reaktionen [39], [40], [41], [42], [43]; Radikalreaktionen [44]; säurekatalysierte und elektrophile Reaktionen [45], [46]; Pericyclenchemie [47], [48]; Redoxreaktionen [49]) behandeln können.

EROS6 verwendet dagegen keine fest codierten Module, die Wissensbasis dieses Programms ist ladbar und wird unabhängig vom eigentlichen Programm erweitert. In den Reaktionsregeln kann auf eine Reihe von physikochemischen Parametern (Atompartialladungen, Polarisierbarkeiten, Bindungsdissoziationsenergien) zugegriffen werden, mit deren Hilfe eine Reaktivitätsfunktion Vorhersagen quantitativer Natur erlaubt. Sind experimentelle Daten in ausreichendem Umfang publiziert, können sehr exakte Reaktivitätsfunktionen durch statistische Analyse automatisch abgeleitet werden. Diese sind wesentlich genauer als die Schätzungen von CAMEO, dafür aber in der Regel auf einen Reaktionstyp begrenzt.

* CHIRal synthON

† Computer-Assisted Mechanistic Evaluation of Organic reactions

‡ Elaboration of Reactions for Organic Synthesis

In neuerer Zeit richtete sich ein Forschungsschwerpunkt aus dem Bereich der computergestützten Modellierung von organischen Reaktionen auf den automatischen Wissenserwerb. Mittlerweile sind umfangreiche Datenbanken an Reaktionen vorhanden, die als In-house-Systeme am Arbeitsplatz des Chemikers verfügbar sind oder über Online-Dienste (z. B. via STN: CASREACT [50]) erreicht werden können. Für alle wissensbasierten Systeme aus dem Bereich Syntheseplanung und Reaktionsvorhersage stellen diese Daten einen immensen Fundus an Information dar. In der Praxis treten jedoch eine Reihe von Schwierigkeiten bei der Verwertung dieses Wissens auf. Einige der in diesem Bereich angesiedelten Systeme seien hier vorgestellt.

Schon 1979 beschrieb Bersohn [51] ein Projekt, welches das für ein Syntheseplanungsprogramm erforderliche Wissen aus englisch-sprachigen Journalen extrahieren sollte. Hierbei wurden relevante Artikel über Schlüsselwörter erkannt, Reaktanden und Produkte durch ihren IUPAC-Namen identifiziert und in eine Bindungstabelle umgewandelt. Nachfolgende Schritte sollten fehlende Reaktionspartner diagnostizieren und das Reaktionszentrum identifizieren. Die so ableitbare Transform-Information muß nachfolgend mit bereits vorhandenen Transforms verglichen werden. Daraus resultierend entsteht entweder ein neues Transform, ein bereits vorhandenes kann verfeinert werden oder aber die Information wird verworfen, weil sie bereits Bestandteil der Wissensbasis ist.

Für das EROS6 System wurden eine Reihe von Programmen entwickelt [38], die Reaktionsinformation aus ChemBase-Reaktionsdatenbanken automatisch zu einer Reaktionsregel verarbeiten. Diese Analyseprogramme erkennen nicht nur das Reaktionszentrum. Darüber hinaus werden für die Atome und Bindungen des Reaktionszentrums auch physikochemische Parameter (Atompartialladungen, Polarisierbarkeiten, Bindungsdissoziationsenergien) errechnet und statistisch validiert. Auf diese Weise werden Fensterbereiche für solche Parameter abgeleitet, die als notwendiges Kriterium in die Reaktionsregel einfließen. Sind auch kinetische Daten verfügbar, kann auch eine multi-variate lineare Modellfunktion für die Kinetik der Reaktion abgeleitet werden. Was diesem System fehlt ist die Möglichkeit, direkt Reaktionsdatenbank-Informationen einzuschleusen. Die ChemBase-Datenbanken, die verwendet werden sollen, müssen bestimmten Restriktionen genügen, damit die Analyseprogramme die Daten auswerten können. Sie werden daher manuell erstellt.

Auch die Gruppe um Gelernter sah sich dem Problem ausgesetzt, daß „es äußerst schwierig ist, Fachleute (Chemiker) zu langfristigem Aufbau und Engagement in der Pflege einer ausgereiften Wissensbasis (für Syntheseplanung) zu begeistern" [52]. Diese hauptsächlich aus Informatikern bestehende Gruppe wollte daher dieses Problem mit Methoden des maschinellen Lernens lösen. Drei Programme entstanden, die - z. T. auf alternativem Weg - aus verfügbaren Reaktionsdatenbanken Komponenten der Wissensbasis für SYNCHEM* ableiten.

- ❏ BRANGÄNE. Aus dem ungeordneten Satz an Einzelreaktionen einer Reaktionsdatenbank ordnet dieses Programm diejenigen mit gleichen Reaktionszentren in Klassen.
- ❏ ISOLDE. Einzelreaktionen mit gleichem Reaktionszentrum (z. B. ermittelt durch BRANGÄNE) werden hinsichtlich der in der Umgebung des Reaktionszentrums anwesenden funktionellen Gruppen untersucht. Das Programm kennt eine Hierarchie von funktionellen Gruppen, die es gestattet, zusammengehörige Gruppen zu generalisieren, d. h. durch generische Gruppen zu verallgemeinern. Auf diese Weise werden aus Einzelreaktionen auch die Einflußfaktoren der Nachbargruppen erkannt und aufbereitet.

* Die zu den LHASA-Transforms analogen Einträge in der Wissensbasis heißen hier *Schemas*.

- TRISTAN. Bereitet zunächst eine Einzelreaktion aus einer Reaktionsdatenbank für CAMEO auf. CAMEO gibt seine Reaktivitätsprognose für die Edukte der Reaktion an TRISTAN zurück. Favorisiert CAMEO das Produkt der Einzelreaktion, leitet TRISTAN aus den Entscheidungskriterien von CAMEO einen Reaktionskontext ab. Das resultierend Reaktionsschema wird mit den in der Wissensbasis vorhandenen verglichen. Ist es noch nicht vorhanden, wird ein neues Reaktionsschema hinzugefügt. Andernfalls wird die Kontextinformation mit der bereits vorhandenen gemischt.

Eine Weiterentwicklung von BRANGÄNE und ISOLDE ist das Programm HORACE* [53]. Es umfaßt die Funktionalität beider Programme. Darüber hinaus kann es empirisch berechnete physikochemische Parameter zur Charakterisierung des Reaktionszentrums heranziehen. Diese physikochemischen Parameter (π- und σ-Elektronegativität sowie Werte zur quantitativen Abschätzung der Stabilisierung positiver oder negativer Ladungen durch Delokalisierung) sind die gleichen, die auch EROS6 bzw. dem WODCA System zur Verfügung stehen.

Das von einer französischen Gruppe betriebene GRAMS-Projekt† verfolgt ein ähnliches Ziel, realisiert die klassifizierte und generalisierte Reaktionsinformation aber in Form eines hierarchischen Netzes [54]. Während andere Wissensbasen nur noch die in den Transforms codierte Information beinhalten, bleiben die Originalreaktionen im Reaktionsnetz von GRAMS immer erhalten. Das Reaktionsnetz ist baumartig aufgebaut, wobei die Blätter die Einzelreaktionen zum Inhalt haben. Verzweigungsstellen sind Abstraktionen von Einzelreaktionen oder höhere Verallgemeinerungen von bereits generalisierten Reaktionen. Für die Syntheseplanung wird für ein gegebenes Zielmolekül die ähnlichste Produktstruktur im Netz der klassifizierten Reaktionen gesucht. Dies Suche beginnt an der Wurzel des Baumes, das ist die allgemeinste Reaktion. Wird ein Mapping zwischen diesem Knoten und dem Syntheseziel gefunden, wird mit den Vaterknoten fortgesetzt. Die Suche auf einem Ast des Baumes wird abgebrochen, wenn ein Vaterknoten nicht mehr im Syntheseziel gefunden wird. Eine so gefundene Abbildung eines Reaktionsknotens auf das Syntheseziel wird dann bewertet. In diese Bewertung gehen ein: das Niveau der Abstraktion des noch passenden Knotens, die Zahl der zugrundeliegenden Originalreaktionen, die Ausbeuten der Originalreaktionen und deren Konsistenz‡. Im Extremfall wird auf diese Weise eine Originalreaktion prognostiziert.

2.2 Das WODCA System - Ein Pool von Werkzeugen zur Syntheseplanung

Dieser Abschnitt widmet sich ausschließlich der Vorstellung des im Arbeitskreis Gasteiger von Dr. W.-D. Ihlenfeldt entwickelten Programmsystems WODCA** [55], [56]. Dieses entstand im Zeitraum von etwa 1988 bis 1991. Es baut zu wesentlichen Teilen auf der Datenstruktur und den Algorithmen des EROS Systems auf. Dieses konnte zu dieser Zeit bereits auf mehr als 15-jährige Entwicklungen zurückblicken [27].

Das WODCA System gibt dem Chemiker eine Vielzahl von Methoden in die Hand, die ihn bei der Planung von organischen Synthesen unterstützen - nicht ihn für diesen Schritt überflüssig machen. Zunächst muß man sich vergegenwärtigen, daß der Chemiker selbst eine Syn-

* <u>H</u>ierarchical <u>O</u>rganization of <u>R</u>eactions through <u>A</u>ttribute and <u>C</u>ondition <u>E</u>duction

† <u>G</u>énérateur de <u>R</u>éseaux pour l'<u>A</u>pprentissage de <u>M</u>éthodes de <u>S</u>ynthèse

‡ Konsistenz bedeutet hier, daß für die Originalreaktionen möglichst ähnliche Ausbeuten publiziert sind.

** <u>W</u>orkbench for the <u>O</u>rganization of <u>D</u>ata for <u>C</u>hemical <u>A</u>pplications these für eine organische Zielverbindung nicht in der konsequent rationalen Form entwickelt, wie das in Abbildung 2 - 1 dargestellt wird. Vielmehr vollzieht ein Chemiker den Designprozeß durch Anwendung verschiedenster Methoden, manche davon schwer rationalisierbar. Möglicherweise „sieht“ er unmittelbar im Syntheseziel eine Schlüsselreaktion, oder aber ein Ausgangsmaterial. Er wird auch ganz rational einige konkrete Syntheseschritte auf dem Papier vornehmen - sowohl retrosynthetisch ausgehend vom Syntheseziel, als auch in Syntheserichtung beginnend bei Ausgangsmaterialien oder Zwischenstufen. Vielleicht versucht er seinen Plan in Richtung einer ihm besonders gut bekannten Reaktion zu entwickeln. Dieses ganze Arsenal von Ansätzen veranschaulicht die folgende Abbildung.

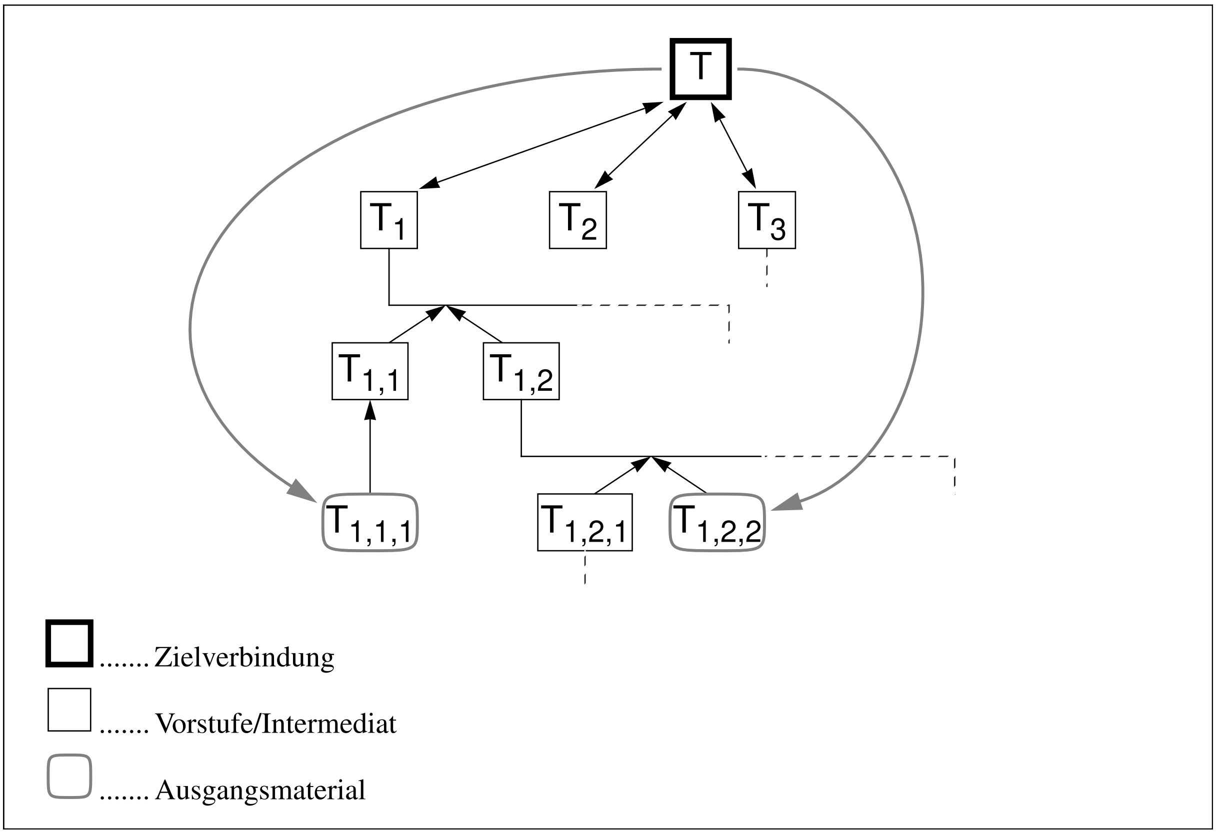


Abb. 2 - 2 Ein durch eine Vielzahl an Methoden eingeschränkter Synthesebaum. Direkte Suchen nach Ausgangsmaterialien, Auswahl und Bewertung strategischer Bindungen, Reaktionsabschätzungen in Syntheserichtung sind die wichtigsten dieser Verfahren. (Vgl. Abb. 2 - 1)

WODCA will den Chemiker bei dieser Art des Denkens in allen Phasen unterstützen. Es bietet daher Methoden zur Suche geeigneter Ausgangsmaterialien, stellt Funktionen zur schrittweisen Retrosynthese bereit und schließt Verfahren zur Bewertung, etwa von Listen von Ausgangsmaterialien, ein. Darüber hinaus besteht das WODCA System nicht aus nur einem monolithischen Programm. Vielmehr steht eine ganze Kollektion an Werkzeugen für dedizierte Aufgaben zur Verfügung. Die folgende Übersicht veranschaulicht diesen Methodenpool.

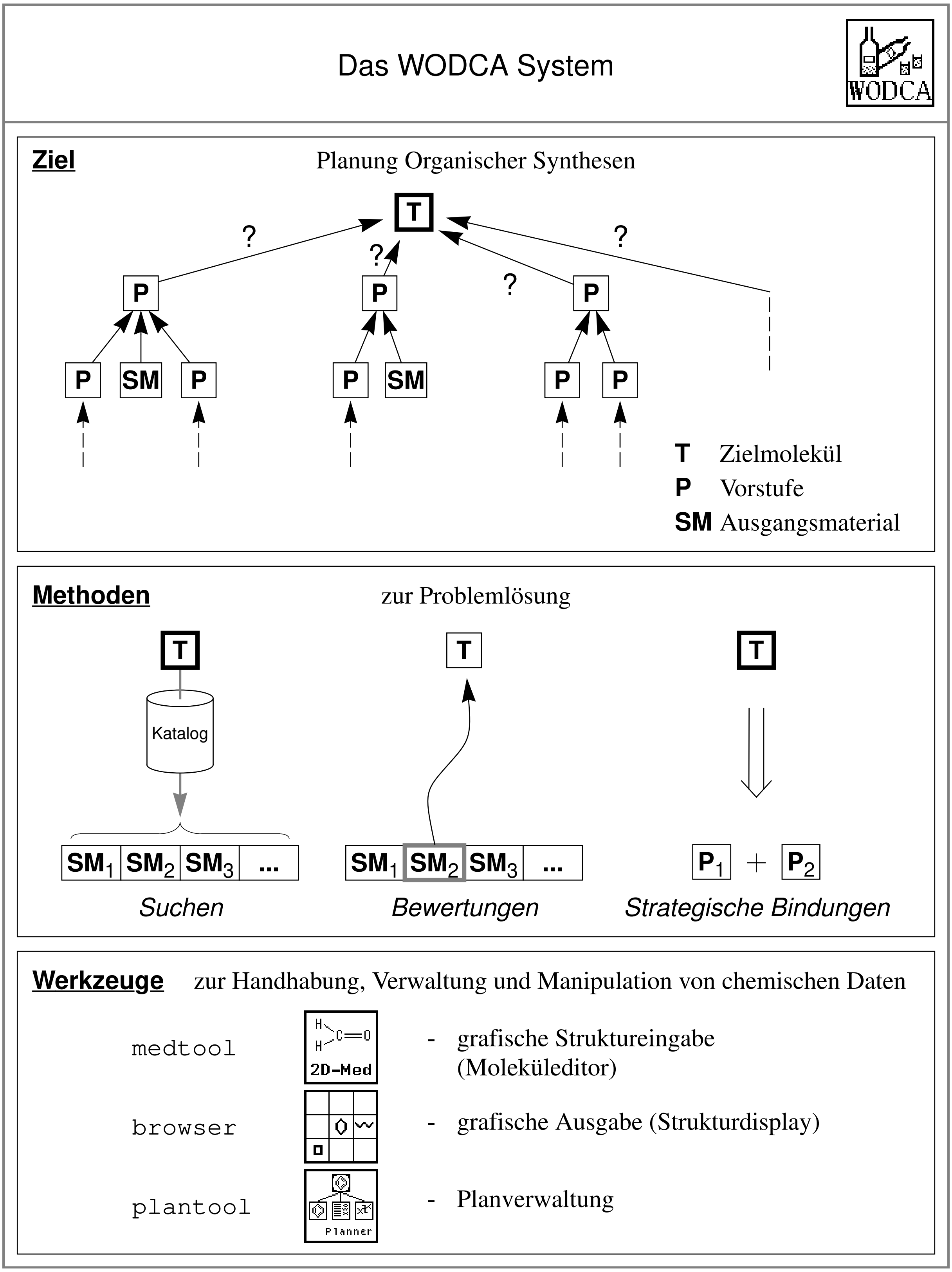


Abb. 2 - 3 Übersicht über das Programmsystem. Methoden (inhärente Funktionalität) und Werkzeuge (externe Programme) können unterschieden werden.

Die klare Verteilung von Aufgaben auf einzelne Programme hat zahlreiche Vorteile. Die wichtigsten sind die leichtere Wartbarkeit und die Verteilung von Entwicklungsarbeit an Einzelprogrammen sowohl zeitlich als auch personell, die Möglichkeit des flexiblen Anpassens der Arbeitsumgebung eines Anwenders an die zu lösende Aufgabe und die Verwendung einzelner Tools losgelöst vom Gesamtsystem.

Im folgenden sollen einige prinzipielle Methoden des Kernsystems vorgestellt werden.

2.2.1 Methoden zur Suche von Ausgangsmaterialien

Eine erste wichtige Methodengruppe sind die Verfahren zur Suche nach Ausgangsmaterialien für ein Syntheseproblem. Dazu stehen dem WODCA System Strukturdatenbanken zur Verfügung. Zu Beginn der Arbeiten an dieser Dissertation waren das der Katalog der Firma Janssen Chimica, Belgien mit 8.464 Verbindungen, eine Sammlung von Chemikalien der Firma Merck-Suchardt mit 3.561 Verbindungen, desweiteren die auch im Programm CHIRON Version 4.21 verwendete Datenbank mit 2.211 Einträgen und zwei kleinere Datenbanken mit 114 bzw. 274 Verbindungen.[*] Es existieren eine Reihe klassischer, stark an chemischen Datenbank-Systemen orientierte Methoden zur Suche von zu einem gegebenen Molekül verwandten Verbindungen:

- Namenssuche: Ist die untersuchte Verbindung beilspielsweise ein Benzol-Aromat kann eine Suche nach Verbindungen mit dem Wortstamm `benzol` oder `benzen` sinnvoll sein.
- Summenformeln oder Elementzähler: Ein einfaches Ähnlichkeitskriterium ist die Summenformel einer Verbindung. So gefundene Isomere können auch für Anwendungen außerhalb der Syntheseplanung, etwa bei der Strukturaufklärung hilfreich sein. Bei der Suche mit Elementzählern können Bereiche von Atomsorten definiert werden. Auch können weitere, nicht vorgegebene Elemente im Ergebnis vorkommen. Auf diese Weise lassen sich z. B. einfach alle Verbindungen mit wenigstens zehn C-Atomen auswählen. Zusätzlich können auch Pseudo-Atome, wie z. B. chirale Zentren, Ringe bestimmter Größe oder die Zahl der aromatischen Ringe gesucht werden.
- Vollstruktursuche: Diese Methode resultiert in Treffern mit vollständiger Übereinstimmung zu einer Anfragestruktur.[†] In der Syntheseplanung terminiert eine solche erfolgreiche Suche für eine Vorstufe definitiv einen Ast des Synthesebaumes.
- Substruktursuchen: Diese ist ein klassisches Verfahren zur Erkennung molekularer Ähnlichkeit. Man beachte jedoch, das es Aufgabe des Anwenders ist, die Substruktur - also das Kriterium der Ähnlichkeit - aus seinem Problem abzuleiten. Mitunter gibt es auch Fälle, in der eine Substruktur nicht genügend auf ein Ziel hin ausgerichtet werden kann (siehe Abschnitt 3.3.4).

Eine für das WODCA System neu entwickeltes Verfahren zur Ähnlichkeitssuche ist die Transformationssuche [56]: Wird eine Regel zur Manipulation von Atomen und Bindungen auf zwei (oder mehrere) Verbindungen angewendet, und sind die beiden *resultierenden* Verbindungen danach *identisch*, kann gefolgert werden, daß die beiden *Original*verbindungen hinsichtlich der verwendeten Regel *ähnlich* sind.

Ein Beispiel soll das verdeutlichen. Die dreifach funktionalisierte, acyclische Verbindung (**1**) aus Abbildung 2 - 4 und das Cyclopenten-Derivat (**2**) werden einer Ozonolyse unterworfen. An (**1**) ändert sich dadurch nichts, der Cyclopentenring dagegen wird aufgebrochen. Nachfolgend werden Substituentenpositionen generalisiert.[‡] Die beiden so *transformierten*

[*] Die Begriffe Katalog, Strukturdatenbank, Datenbank von Ausgangsmaterialien werden im weiteren synonym verwendet. Auf abweichende Bedeutung wird an entsprechender Stelle ausdrücklich verwiesen.

[†] Wie im Abschnitt 3.3.2 erläutert wird, muß diese Übereinstimmung nicht unbedingt 100%ig sein. Man denke etwa an Tautomere.

Verbindungen werden miteinander verglichen. In diesem Fall sind sie *identisch*, damit sind (**1**) und (**2**) bezüglich einer Ozonolyse/Substitutionsoperation einander *ähnlich*. Zweifellos wäre (**2**) ein geeignetes Ausgangsmaterial zur Synthese von (**1**). Um diese Methode effizient zu gestalten, wird auf ein aufwendiges und zeitraubendes Atom-by-Atom-Mapping beim Vergleichen der transformierten Verbindungen verzichtet und statt dessen ein Hashcodevergleich durchgeführt. Als Hashcode soll hier eine eindeutige Zahl verstanden werden, die nur für eine bestimmte chemische Struktur charakteristisch ist. Ein solcher Hashcode kann um Größenordnungen schneller berechnet werden, als ein Mapping-Verfahren arbeiten kann. Der im WODCA System verwendete Hashcode-Algorithmus [57] hat sich im Routinebetrieb als sehr zuverlässig und leistungsfähig bewährt. Durch die kompakte Codierung einer chemischen Struktur in einer einzigen Zahl ist es auch möglich, die Transformationsresultate für einen ganzen Katalog im voraus zu berechnen, in Dateien bereitzuhalten und zur Programmlaufzeit vollständig im Kernspeicher zu halten.*

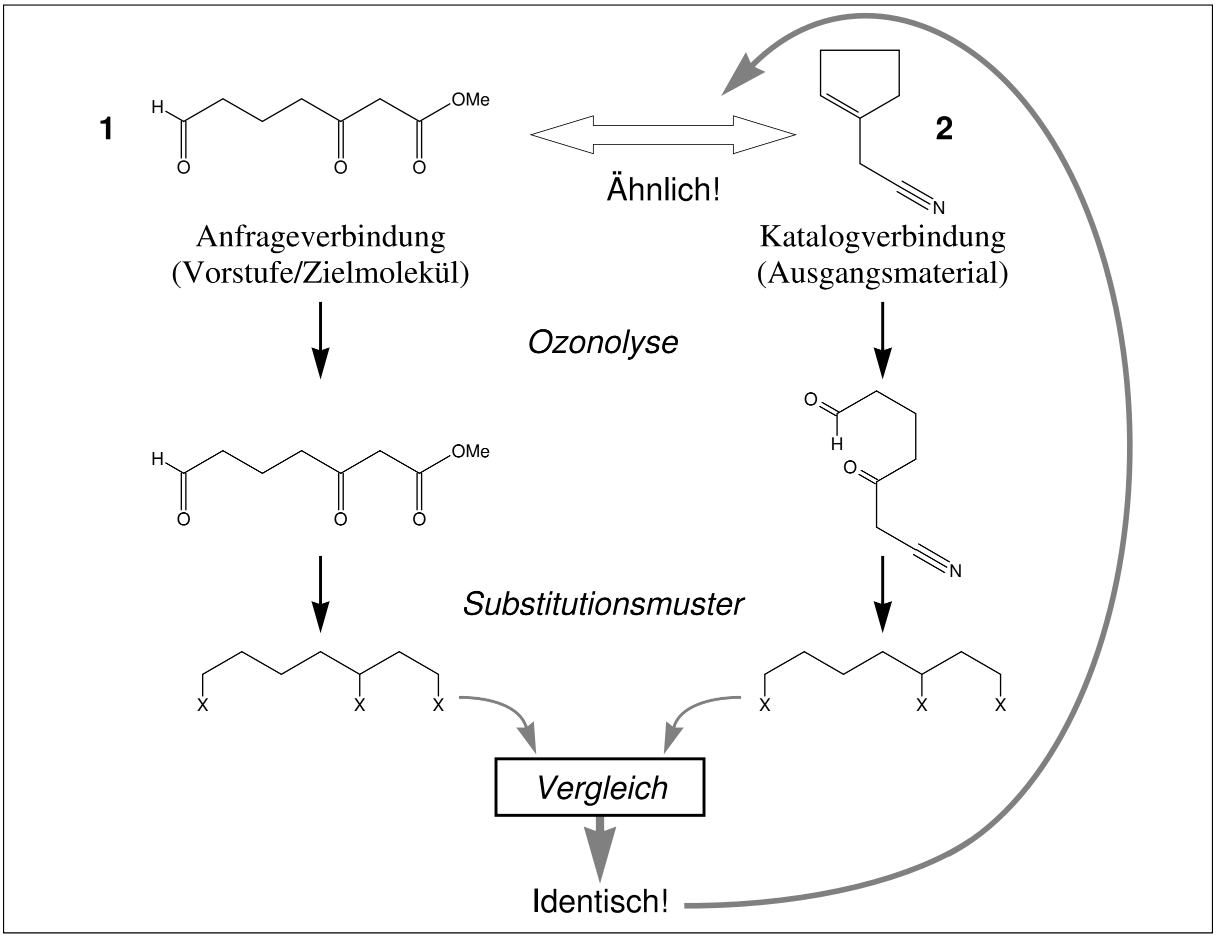


Abb. 2 - 4 Prinzip der Ähnlichkeitssuche mittels Transformation. Die Transformationsregel (hier Ozonolyse und Substitutionsmuster) wird auf die zu vergleichenden Verbindungen angewendet. Das Ergebnis wird über die resultierenden Hashcodes verglichen.

‡ Auf welch vielfältige Weise das geschehen kann stellt Abschnitt 3.3.4 ausführlich dar.

* Die Hashcodes des WODCA Systems sind als 32-Bit-Integer-Zahlen realisiert. Für 10.000 Verbindungen werden im Kernspeicher gerade einmal 40 kByte belegt.
Prinzipiell kann nicht ausgeschlossen werden, daß zwei verschiedene Verbindungen den gleichen Hashcode erhalten. Die Länge des Hashcodes (Zahl der Stellen) begrenzt letztendlich die Menge der kollisionsfrei codierbaren Verbindungen.

Die technische Seite der Transformationssuche erläutert Abbildung 2 - 5. Weitere Einzelheiten zu den Transformationssuchen in der aktuellen WODCA Version finden sich in Abschnitt 3.3.

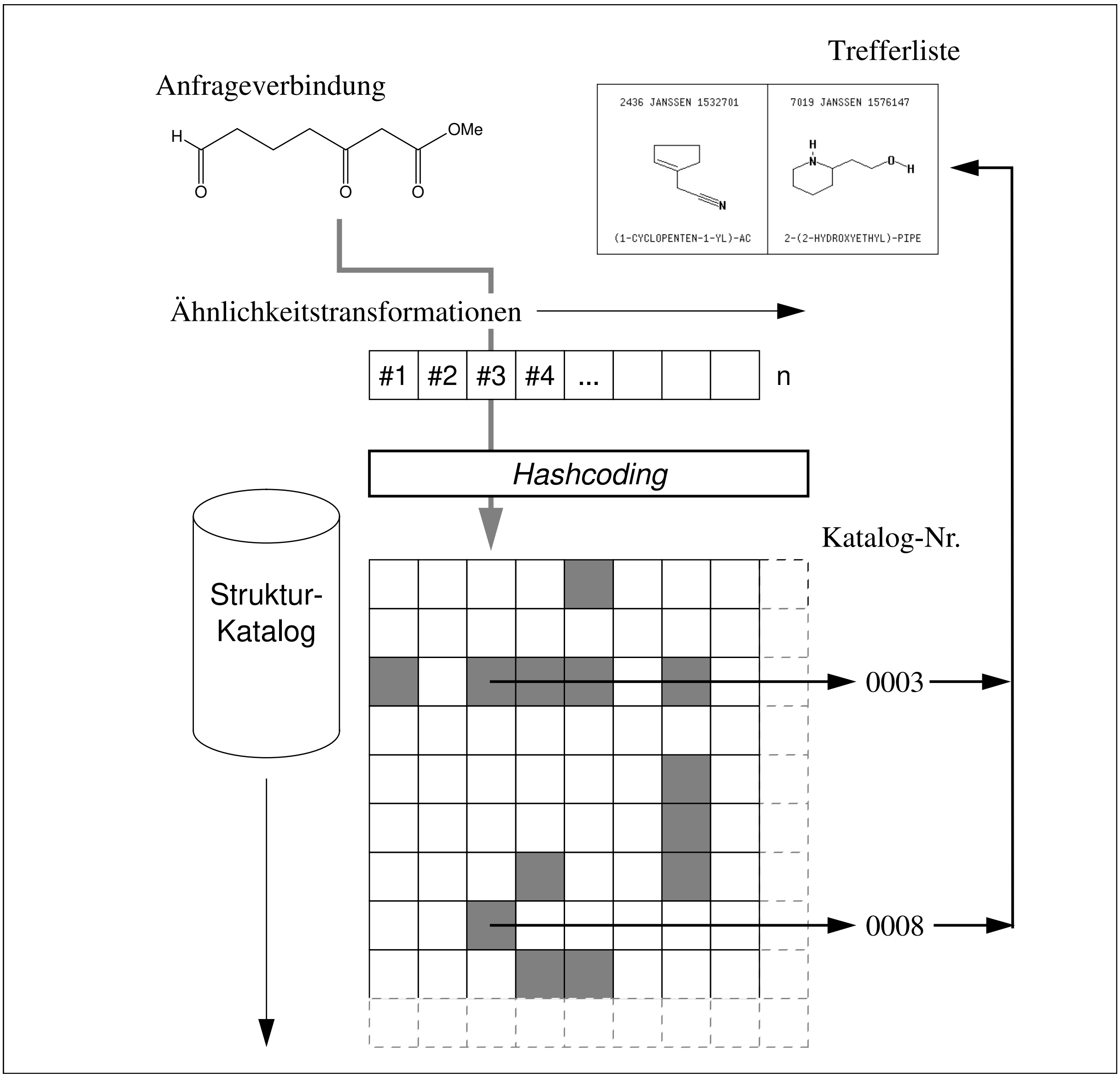


Abb. 2 - 5 Implementation der Transformationssuche. Die Ähnlichkeitsregeln wurden vorab auf alle Datenbankmoleküle angewendet. Die resultierenden Hashcodes werden parallel zur Datenbank in Dateien bereitgehalten. Zum Zeitpunkt der Durchführung einer Transformationssuchen muß nur noch die Anfrageverbindung transformiert, der Hashcode berechnet und dieser gegen einen aus einer Datei geladenen Vektor abgeglichen werden. Übereinstimmende Hashcodes führen zur Selektion der betreffenden Datenbankmoleküls in der Trefferliste.

2.2.2 Schrittweise Vorstufengenerierung

Nicht immer wird es möglich sein, direkt von einem Zielmolekül mittels der Methoden des Abschnitts 2.2.1 zu Ausgangsmaterialien zu gelangen. Der Chemiker würde dann Coreys schrittweisem Retrosyntheseansatz folgen. Das zentrale Problem bei dieser Vorgehensweise ist die Auswahl von Bindungen, die retrosynthetisch gebrochen werden sollen. Nicht jede Bindung ist gleich gut für diesen Zweck geeignet. Eine Zerlegung sollte chemisch sinnvoll sein (durch synthetisch ausnutzbare Reaktivität begründet) und zu einer Vereinfachung der Struktur führen. Man sucht nach sog. *strategischen Bindungen*. Der Begriff wird schon in [3] verwendet. In dieser Arbeit werden darunter diejenigen Bindungen in einem Syntheseziel (oder einer Vorstufe) verstanden, deren retrosynthetischer Bruch zur Generierung geeigneter Vorstufen sinnvoll ist.* In der Regel werden solche Bindungsbrüche zur Reduktion der Komplexität der Zielstruktur führen.

Mit der eigentlichen Deklaration einer Bindung als strategisch, ist ein weiteres Problem verknüpft. Durch den retrosynthetischen Bruch entstehen primär Molekülfragmente mit freien Valenzen. Die Absättigung dieser Positionen muß auf geeignete Weise erfolgen, so daß man zu echten Reagenzien kommt.

WODCA stellt für diese Aufgabe zwei prinzipielle Lösungen zur Verfügung: die manuelle Spezifikation von strategischen Bindungen bzw. ihre automatische Ermittlung durch ladbare Funktionen. Die manuelle Spezifikation eröffnet dem Chemiker die unmittelbare Möglichkeit, steuernd in die Planungsrichtung einzugreifen. Die auf seiner Erfahrung basierende Auswahl einer strategischen Bindung lenkt das Synthesedesign. Zum einen betrachtet der Chemiker dadurch das mit WODCA erzielbare Ergebnis eher als „seine” Synthese, zum anderen ist diese Möglichkeit wichtig bei selteneren topologischen Merkmalen (spezielle Ringsysteme, ausgefallene Substitutionsmuster) oder auch als Korrekturmöglichkeit von automatisch vorgeschlagenen Bindungsbrüchen. Diese automatischen Verfahren zur Ermittlung strategischer Bindungen bestanden aus ladbaren Funktionen, die in einer speziellen Programmiersprache codiert waren und außerhalb des eigentlichen Programms entwickelt werden konnten.† Sie waren allerdings in ihrer Mächtigkeit eher begrenzt und auf sehr spezielle Situationen ausgerichtet. Beispiele sind die Erkennung von möglichen Diels-Alder-Reaktionen in Cyclohexenringen, Carbanionenchemie an carbonylischen Zentren oder die Erkennung der Spaltbarkeit von Peptidbindungen [58]. Dennoch konnten damit Vorstufen generiert werden, die dann erneut den Ähnlichkeitssuchen aus Abschnitt 2.2.1 unterworfen wurden. Es war eine Aufgabe dieser Arbeit, die Fähigkeiten des WODCA Systems hinsichtlich der Auswahl strategischer Bindungen durch verbesserte automatische Funktionen zu erweitern (siehe Kapitel 1).

2.2.3 Bewertungsmechanismen

Bewertungen sind an verschiedenen Stellen der Planung einer organischen Synthese wünschenswert und notwendig. Beispielsweise sollten aus eine Liste von möglichen Ausgangsmaterialien diejenigen ausgewählt werden können, die der Referenzverbindung *am ähnlichsten* sind. Nach *vollzogener* Generierung von Vorstufen müssen Prüfungen realisierbar sein, welche die Reaktivität der zu verbindenden Atome abschätzen oder Selektivitätsprobleme erkennen können. Derartige Verfahren werden von WODCA in den meisten Fällen wie bei strategischen Bindungen durch ladbare Funktionen vollzogen.

Eine sehr komplexe Bewertungsfunktion existiert zur Beantwortung der Frage: Welche Verbindung aus einer Liste von Ausgangsmaterialien ist die geeignetste zur Synthese eines Zielmoleküls [59]. Diese Bewertungsfunktion überlagert die Zielstruktur mit jeder Verbindung aus der Liste. Es wird die maximale gemeinsame Substruktur der beiden Verbindungen ermittelt. Typischerweise gibt es mehrere solcher maximaler Überlagerungen. Für jede davon werden von der Bewertung dann Bonus- und Maluspunkte für verschiedene Kriterien vergeben:

* Man beachte, daß diese Definition nicht notwendiger Weise eine Vereinfachung der Zielstruktur hinsichtlich Funktionalität, skeletaler Topologie oder stereochemischer Merkmale *fordert*.

† Zu technischen Einzelheiten siehe Abschnitt 2.2.5.

- Übereinstimmung von chiralen Zentren
- Zahl der im Ausgangsmaterial zu brechenden bzw. zu modifizierenden Bindungen
- Abschätzung der Reaktivität an Atomen und Bindungen, die im Ausgangsmaterial zu verändern sind.
- relative Position von Funktionalität zueinander

Bei mehreren gemeinsamen Substrukturen erhält das Ausgangsmaterial die höchste errechnete Bewertung. Mittels der so zugeordneten Werte kann eine Liste einfach geordnet oder gar besonders schlecht bewertete Fälle ganz entfernt werden.

Das WODCA System hat auch eine Anbindung an ein Reaktionsvorhersageprogramm (EROS6). Damit kann für vorgeschlagene Vorstufen getestet werden, ob die zugrundeliegende Reaktion in Syntheserichtung anwendbar ist. Das setzt im Falle von EROS6 allerdings eine entsprechende Reaktionsregel voraus.

Ein anderes denkbares Verfahren zur Prüfung der Einsetzbarkeit der vorgeschlagenen Vorstufen wäre die Recherche in einer Reaktionsdatenbank. Diese Möglichkeit war für das zu Beginn dieser Arbeit vorhandene WODCA System nicht gegeben. Dafür ist es nicht notwendig, in WODCA selbst ein solches Recherchesystem einzubauen. Vielmehr sollte sich das Programm in einer Umgebung (d. h. hier: Betriebssystem mit fenster-basierter Benutzeroberfläche) verwenden lassen, wo solche Reaktionsdatenbanken bereits benutzt werden können.

2.2.4 Interaktion mit dem Anwender

Das WODCA System ist interaktiv, d. h. der Chemiker, der das Programm zur Planung einer Synthese einsetzt, greift an vielen Stellen als Entscheidungsträger in den Ablauf ein. Nach dem Start des Programms, bei dem der Anwender über die für diese Sitzung notwendigen Tools entscheidet (Moleküleditor, Displayprogramme u. dgl.), erfolgt die Eingabe eines Zielmoleküls mittels graphischem Editor oder durch Einlesen einer Datei. Diese Zielverbindung wird dann an die Spitze des zu entwickelnden Syntheseplanes eingetragen. Der Chemiker entscheidet nun, welche Methoden er aus dem WODCA Arsenal einsetzten will. Er wird also Ähnlichkeitssuchen benutzen, strategische Bindungen ermitteln, Vorstufen generieren und Bewertungen vornehmen. Dazu wählt er aus Menüpunkten aus, aktiviert Kommandos über Schalter oder setzt Kommandos per Hand ab. Die Ergebnisse dieser Verfahren können sämtlich im Syntheseplan verwaltet werden. Das gibt dem Anwender die Gelegenheit, die Planung zu einem späteren Zeitpunkt wieder aufzunehmen oder aber an verschiedenen Punkten des Planes fortzufahren. Für alle diese Tätigkeiten stehen dem Chemiker mit graphischen Benutzerschnittstellen ausgestattete Programme zur Verfügung. Er arbeitet in einer Fenster-basierten Umgebung namens SunView. Abbildung 2 - 6 gibt den Bildschirm wieder, wie er sich einem Benutzer während einer typischen WODCA Sitzung präsentiert.

Die direkte Einbindung des Chemikers in den Programmablauf kann als Vorteil verstanden werden. Er kommuniziert mit dem System, um von diesem Anregungen zu erhalten. Diese können unmittelbar zu neuen Fragestellungen führen. Der Anwender steht in echtem Dialog mit dem Programm.

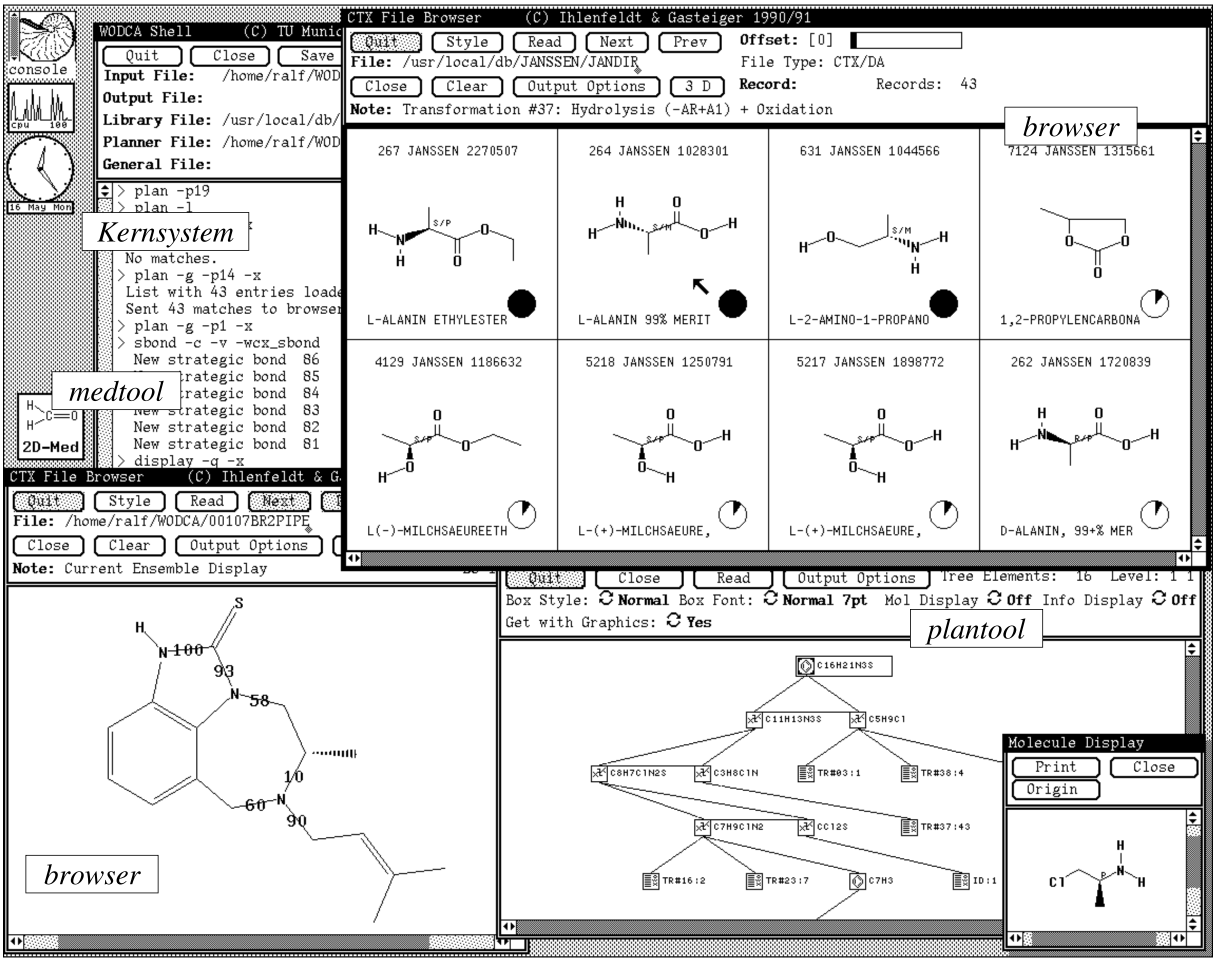


Abb. 2 - 6 Typischer Bildschirm einer WODCA Sitzung. Mehrere Programme arbeiten gleichzeitig und präsentieren dem Chemiker Informationen in graphischer Form. Das eigentliche *Kernsystem* stellt die Funktionalitäten zur Syntheseplanung zur Verfügung. Der *browser* rechts oben stellt eine Liste von Ausgangsmaterialien einschließlich der vergebenen Bewertung dar. Der zweite *browser* links unten visualisiert das aktuelle Zielmolekül mit Bewertungen über Bindungen, die als strategisch erkannt wurden. Das *plantool* rechts unten verwaltet den gesamten Syntheseplan. Der Moleküleditor *medtool* ist zum Icon geschlossen (Mitte links).

2.2.5 Zielsetzung dieser Arbeit

Zum Verständnis der Zielsetzung ist es erforderlich, an dieser Stelle den Systemaufbau von WODCA zu erläutern. Die Separation bestimmter Funktionen und deren Bereitstellung durch separate Programme wurde bereits auf Seite 12 erwähnt. Aber auch das Kernsystem selbst kann in verschiedene Bereiche unterteilt werden. Es enthält Funktionen zum Lesen und Schreiben chemischer Informationen, es verwaltet Kataloginformationen und erlaubt den Zugriff darauf, es organisiert Synthesepläne, stellt die Substruktur-, Transformations- und alle anderen Suchen bereit. Der Zugriff auf Methoden zur Suche strategischer Bindungen und der Generierung von Vorstufen oder die Bewertungsmechanismen erfolgt über Programmschnittstellen. Dadurch sind die Funktionen, welche die eigentliche Arbeit etwa bei einer Bewertung leisten, *nicht* Bestandteil des Kernsystems. Stattdessen werden diese in externen Dateien codiert in einer speziellen Sprache (*VERGIL+++* [60]) bereitgestellt. Die Vorteile liegen erneut auf der Hand:

- ❏ Dynamisch verwendbar, d. h. neue oder veränderte Funktionen können ohne Recompilierung des Kernsystems sogar zur Laufzeit eingeschleust werden.
- ❏ Einfache Austauschbarkeit. Die Unabhänigkeit solcher Definitionen befreit einen potentiellen Entwickler von sonst erforderlichem Wissen über das Management das gesamten Quellcodes des Kernsystems.
- ❏ Separate Verwaltung und Entwicklung von Methoden wird hierdurch möglich.

Dem Kernsystem werden diese Methodendefinitionen jedoch in einer anderen Form zur Verfügung gestellt: als *Forth*-Module. Der Grund dafür liegt in der einfachen Implementation eines Forth-Systems und dem schnellen Laufzeitverhalten von *Forth*-Code. Diesem Vorteil steht jedoch die ungewöhnliche Syntax und andere spezielle Eigenheiten von *Forth* gegenüber, welche die Programmierung von *Forth*-Funktionen selbst für in anderen Sprachen erfahrene Experten erschwert. *VERGIL*[*] sollte daher eine einfachere Sprachdefinition bereitstellen, die auch für Chemiker als Programmier-Laien verständlich ist und so die Akzeptanzbarriere von *Forth* durchbricht. Zur einfachen Überführung von *VERGIL*-Code in *Forth* wurde ein spezieller Crosscompiler geschaffen [60].

Auch die graphische Benutzeroberfläche, die Schnittstelle der Kommunikation zwischen Programm und dem Chemiker ist separat vom Kernsystem implementiert. Sie basiert auf dem Fenstersystem SunView von Sun Microsystems.

Die folgende Abbildung veranschaulicht diesen Systemaufbau.

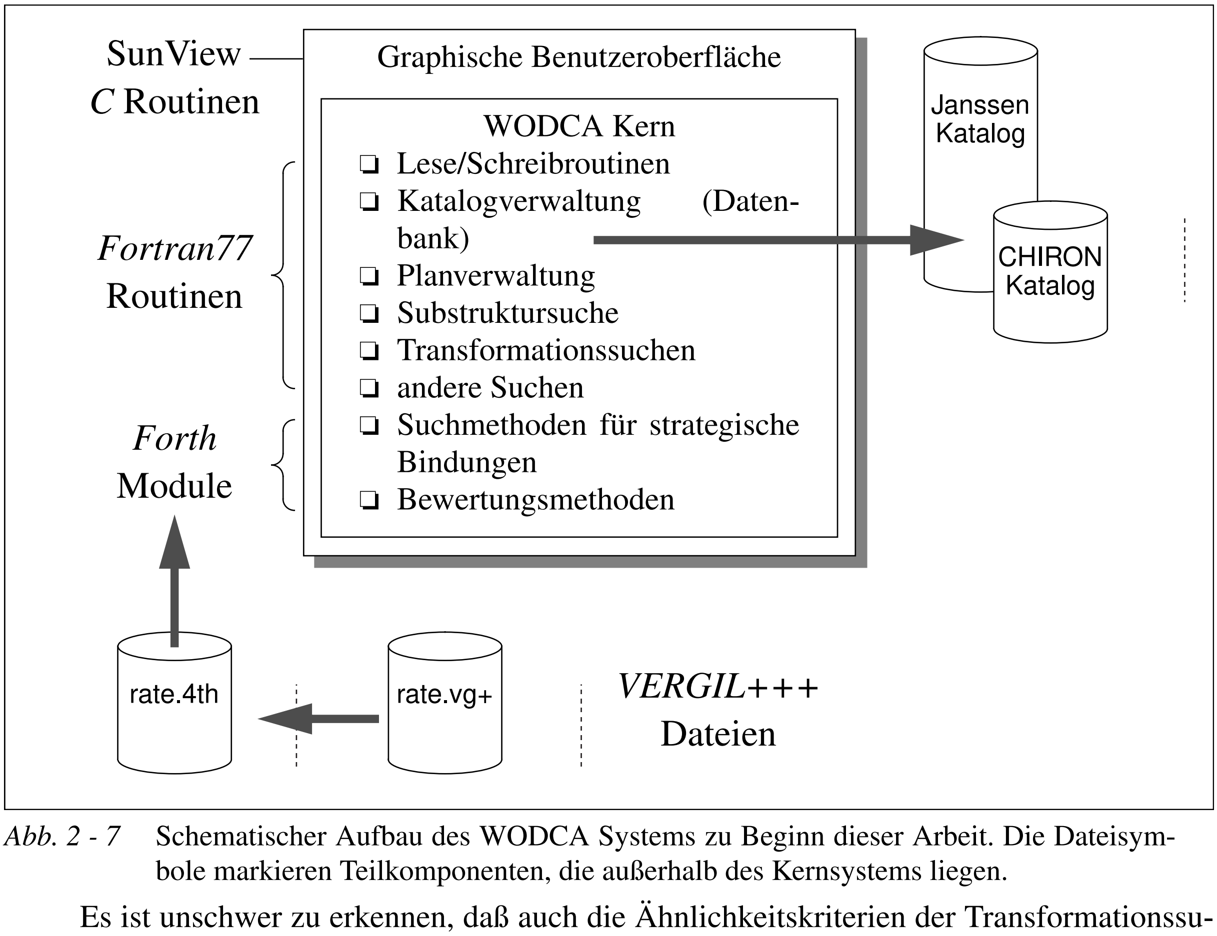


Abb. 2 - 7 Schematischer Aufbau des WODCA Systems zu Beginn dieser Arbeit. Die Dateisymbole markieren Teilkomponenten, die außerhalb des Kernsystems liegen.

Es ist unschwer zu erkennen, daß auch die Ähnlichkeitskriterien der Transformationssuchen chemisches Wissen darstellen. Daher ist es wünschenswert, daß diese genau wie die

[*] Versatile EROS Reactivity functions Generation and Interpretation Language

Bewertungsmethoden oder Suchmethoden für strategische Bindungen in externen Dateien definiert werden können.

Der akzeptanzbedingte Umweg über eine eigene chemische Programmiersprache zur Definition von chemischem Wissen für das WODCA System verkompliziert vor allem die Wartung und Entwicklung des Programms. Es wäre demnach sinnvoll, im Zuge der Erweiterung des Subsystems, das für die Verarbeitung von externem Wissen zuständig ist, dieses auf einfachere, aber dennoch leistungsfähige Weise zu reimplementieren.

Die Verwendung von SunView als Basis der graphischen Benutzeroberfläche fesselt das WODCA System an Sun Workstations. Mittlerweile ist dieser Standard auch von der Software-Entwicklung überholt worden. Die heutige Standardgrundlage von graphischen Oberflächen im Workstationbereich ist das X Window System des MIT. Daraus ergab sich der Zwang, die Benutzerschnittstelle für das WODCA System neu zu implementieren.

Abschnitt 3.2 zeigt, daß diese Anforderungen mittels einer neuen Standardsprache verwirklichen lassen. Dieser programmtechnische Aspekt ist ein wesentlicher Bestandteil dieser Arbeit.

Darüber hinaus muß dieses sehr flexible und mächtige System mit chemischem Leben erfüllt werden, um den Nachweis antreten zu können, daß das Programm ein sinnvolles Werkzeug für den Chemiker darstellt. Dieser Aufgabe stellt sich diese Arbeit, indem die Ähnlichkeitskriterien für die Transformationssuchen gründlich überarbeitet wurden (Abschnitt 3.3). Darüber hinaus konnten leistungsfähige Funktionen zur Suche nach strategischen Bindungen (Kapitel 4) entwickelt werden.

2.3 Ein kleiner Exkurs - Programmtechnische Aspekte

Die computer-gestützte Syntheseplanung ging einher mit einer gewaltigen Entwicklung ihrer materiellen Grundlage: der Computertechnik. Die ersten Programme mußten mit wenigen Kilobyte Speicher auskommen und benutzten - aus heutiger Sicht - geradezu archaische Schnittstellen wie Zeichentableau und mehrere Terminalschirme [8]. Heutige Workstations stellen dagegen Hauptspeicher im Bereich von Dutzenden von Megabyte bereit, verwalten Gigabytes an Plattenplatz und präsentieren sich mit Multitasking- und Multiuser-fähigen Betriebssystemen und graphischen Fensteroberflächen. Auch die Arbeitsplatzrechner, ob IBM-kompatibler PC oder Macintosh-Rechner, liegen in ihrem Angebot an Hauptspeicher und Plattenplatz um Größenordnungen über den Maschinen der Pionierzeit der computer-gestützten Syntheseplanung, ganz zu schweigen von den außerordentlich gewachsenen Prozessorleistungen.

Tab. 2 - 1 Zusammenstellung von Syntheseplanungssystemen. Die Einträge sind alphabetisch nach den Programmnamen geordnet.

Programm	Sprache(n)	Referenz
AHMOS	*PL/1*	[61]
CAMEO	*Fortran*	[62]
COSYMA	*C/Fortran*	[63]
EROS	ursprünglich *PL/1*, z. Z. *Fortran* Migration nach *C++* in Gang	
FLAMINGOES	*Basic*	[64]
IGOR	*PL/1* *Fortran/Assembler*	[24] [25]
LHASA-1 / OCSS	*Assembler* (*DECAL*)	[3]
LHASA-10	*Fortran*	[3]
LHASA	*Fortran/C/Pascal* (*CHMTRN*)	[65]
Lilith	*Fortran*	[66]
MASSO	*Fortran*	[67]
MARSEIL/SOS	*Forth*	[68]
RDSS	*C*	[69]
SECS	*Fortran* (*ALCHEM*)	[70]
SST	Fortran	[35]
SYNCHEM1/2	*PL/1*, *Assembler* (in Version 1), Migration nach *C* in Gang	[20] [71]
SYNGEN	*Fortran*	[29]
TRESOR	*C*	siehe RDSS
WODCA	*Fortran/C* (*Forth/VERGIL+++*)	
-[a]	*Lisp*	[16]
-[a]	*Lisp*	[72]
-[a]	*Basic*	[73]

a. Name des Programms in der Referenz nicht erwähnt.

Ein anderer technischer Aspekt betrifft die einem Programmsystem zugrundeliegende Programmiersprache. Wie die Tabelle 2 - 1 zeigt, existiert dabei eine große Vielfalt, was die verwendeten Sprachen anbetrifft. Es gibt auch Systeme, die eine ganze Reihe verschiedener Sprachen in sich vereinen. Grundsätzlich läßt sich aber erkennen, daß der Schwerpunkt auf wenigen Hochsprachen (*Fortran*, *C*) liegt, wobei sich in jüngerer Zeit ein Trend zu objektorientierten Sprachkonzepten (vor allem *C++*) abzeichnet.

Ähnlichkeitssuchen mittels Transformationssuche

Das Kapitel befaßt sich mit einer speziellen Form der Ähnlichkeitssuche, ihrer programmtechnischen Realisierung sowie ihrer Anwendung im Bereich der computer-gestützten Syntheseplanung.

3.1 Prinzip

Die Ähnlichkeitssuche soll in der computer-gestützten Syntheseplanung zur Erkennung von Verwandtschaften zwischen Synthesezielen bzw. Vorstufen und bekannten Ausgangsmaterialien dienen. Das Prinzip der Transformationssuche wurde bereits im Abschnitt 2.2.1 erläutert. Es umfaßt im Wesentlichen die Anwendung einer Regel, die sowohl auf ein Syntheseziel, als auch auf ein Ausgangsmaterial angewendet wird. Transformiert eine solche Regel beide Verbindungen in das gleiche Resultat, folgt daraus Ähnlichkeit der beiden Verbindungen. Die Definition der Ähnlichkeit geschieht demnach durch die Transformationsregel. Diese ist gleichsam das Ähnlichkeitskriterium. Ein Ziel dieser Arbeit war es, diese Transformationsregeln in externen Dateien, unabhängig vom eigentlichen WODCA System zu halten und auf diese Weise eine flexible Entwicklung von Ähnlichkeitskriterien zu ermöglichen.

3.2 Technische Realisierung

Dieser Abschnitt beschreibt, wie die Transformationssuche realisiert wurde, d. h. es werden programmiertechnische Details besprochen. Neben den Randbedingungen die zu erfüllen waren (3.2.1), wird auf die benutzte Skriptsprache (3.2.2) und deren notwendige Erweiterungen (3.2.3) eingegangen.

3.2.1 Randbedingungen

Mit dem zu Beginn dieser Arbeiten existierenden Programmsystem WODCA konnte die Nützlichkeit der im Abschnitt 2.2.1 beschriebenen Suchmethode im Besonderen für die Syntheseplanung gezeigt werden. Allerdings traten bei der routinemäßigen Anwendung eine Reihe von Kritikpunken zu Tage:

- ❑ Während andere Techniken des WODCA Systems, wie die Suchmethoden nach strategischen Bindungen oder Bewertungsmethoden für Trefferlisten, in einer Form realisiert waren, die es gestattet, das für den Computer aufgearbeitete chemische Wissen unabhängig vom eigentlichen Programmkern zu halten, war das für die Vielzahl an Ähnlichkeitskriterien nicht so. Diese Kriterien wurden in das Gesamtsystem hineincompiliert. Das bedeutet, Änderungen an diesen Kriterien waren nur möglich, wenn man das Gesamtsystem recompilieren *konnte*.

- Der unerfahrene Anwender erhält keinerlei Hilfestellungen bei der Auswahl eines geeigneten Ähnlichkeitskriteriums. Er wird durch die Art der Präsentation der Transformationssuchen als eine komplette Liste aller verfügbaren Ähnlichkeitsdefinitionen eher irritiert. Das schränkt die Akzeptanz des Gesamtsystems beim Endbenutzer ein.
- Die bestehende Lösung zur externen Verwaltung und Bereitstellung von chemischem Wissen (Bewertungsfunktionen, Suchmethoden nach strategischen Bindungen - siehe oben) besteht aus einem komplexen, zweistufigen System. Dieses stellt ein zusätzliches Problem bei einer Portierung des Gesamtsystems oder von Teilen, die diese Fähigkeiten benutzen, dar. Außerdem verkompliziert es die Handhabung des Systems unnötig.

Diese Kritikpunkte sollten bei einer umfassenden Reimplementierung überwunden werden. Daraus ergeben sich die folgenden Anforderungen an eine neue Lösung:

- Die verschiedenen Ähnlichkeitskriterien sollen in externen Dateien verwaltet werden. Sie stehen dem Kernsystem zur Laufzeit zur Verfügung. Sie sollen aber unabhängig vom Kernsystem entwickelt und getestet werden können. Idealerweise sollten diese externen Dateien in einem über alle Standard-Computerarchitekturen (UNIX, PC, Macintosh) verwendbaren Format realisiert werden, die dadurch ebenso leicht modifizierbar wie auch problemlos portierbar wären. Weiterhin sollte die Codierung in einer leicht verständlichen, höheren Programmiersprache erfolgen.
- Das Management aller Dateien mit externem chemischen Wissen soll vereinheitlicht werden. Das erleichtert enorm die Wartbarkeit des Gesamtsystems wie auch die Einarbeitung neuer Benutzer und Entwickler. Betroffen davon sind die Ähnlichkeitssuchen, Suchmethoden nach strategischen Bindungen oder Bewertungsfunktionen.
- Reorganisation der Benutzerschnittstelle. Das gilt für das Gesamtsystem bezüglich einer Anpassung an aktuelle Grafikstandards wie auch hinsichtlich der Absicherung einer Portierung. Für die Ähnlichkeitssuchen sind insbesondere eine einfache Auswahl zu ermöglichen und Hilfefunktionen zur Entscheidungsfindung, welches Ähnlichkeitskriterium vorteilhaft ist, anzubieten. Dazu können die verschiedenen Ähnlichkeitskriterien nach ihrer Anwendungsbreite und nach ihrer Selektivität geordnet werden (siehe 3.3.11).

Was also primär benötigt wird, ist eine geeignete Programmiersprache, die ...

- leistungsfähig ist. Die Interaktion des Benutzers mit dem Programm darf nicht durch lange Antwortzeiten behindert werden. Deshalb muß die gewählte Lösung möglichst schnell arbeiten.
- integrierbar und erweiterbar ist. Es muß möglich sein, die gewählte Programmiersprache als Subsystem in das Kernsystem einzubetten. Falls der eigentliche Sprachstandard nicht ausreicht, muß es einfach möglich sein, die Sprache durch neue Konstrukte zu erweitern.
- portabel ist. Insellösungen, die nur für ein Betriebssystem oder gar nur eine Computerarchitektur existieren, sind ungeeignet. Daher sollten Implementationen der gewählten Sprache wenigstens für IBM-kompatible PCs *und* UNIX Workstations existieren.
- eine einfache Syntax besitzt. Eine solche Sprache ist leicht erlernbar und erleichtert daher den Einstieg neuer Benutzer enorm.
- dokumentiert ist und für die Support verfügbar ist. Das sind zwei nicht zu unterschätzende Punkte. Eine Sprache, mag sie noch so leistungsfähig und dabei einfach sein, ist ungeeignet, wenn das Wissen über ihre Anwendungsrichtlinien nicht weitergegeben werden kann. Auch wollen wir eine solche Sprache nicht in ihrem Kern weiter-

entwickeln. Stattdessen sind Ansprechpartner erwünscht, die systemimmanente Fehler -sofern vorhanden - beseitigen und bei Problemen zur Verfügung stehen.

- geringe Kosten verursacht. Damit sind sowohl Anschaffungs- wie Unterhaltungskosten oder Lizenzgebühren gemeint. Das ist ein nicht unwesentliches Kriterium gerade in akademischen Institutionen.

Man wird unschwer erkennen, das dieser Anforderungskatalog sehr umfassend ist und mitunter einander gegenläufig wirkende Kriterien umfaßt. Es hängt dann von der konkreten Anwendung ab, ob z. B. ein Mehr an Leistungsfähigkeit mit einer komplizierteren Sprachsyntax erkauft werden muß.

3.2.2 Skriptsprache *tcl*

Es existieren eine Reihe von Programmiersprachen, die mehr oder weniger die im vorangehenden Abschnitt aufgezählten Anforderungen erfüllen. Diese Sprachen werden als *very high level languages* (VHLL) bezeichnet:

- *Scheme*:
  Bei dieser Sprache handelt es sich um einen *Lisp*-Dialekt. Die erste Beschreibung von *Scheme* datiert aus dem Jahre 1975 [74]. Ein erstes Textbuch erschien 1984 [75]. Die Sprache verbreitete sich an US-amerikanischen Hochschulen, womit auch eine Diversifizierung des Sprachumfangs einherging. Um letzterem entgegenzuwirken und einen verbindlichen Sprachstandard zu definieren, werden Berichte herausgegeben. Der aktuelle datiert vom 2. November 1991 [76]. Obwohl es sich bei *Scheme* um eine klar definierte, überschaubare Sprache handelt, hat sie kaum Verbreitung außerhalb der Informatikwelt gefunden. Die Gruppe der Nutzer umfaßt schätzungsweise einige Hundert weltweit.
- *Perl*[*]
  Perl ist eine interpretierte Sprache, die vor allem auf das Bearbeiten von Textdateien im weitesten Sinne optimiert wurde. Sie wird ebenso für Aufgaben aus dem Bereich des UNIX Systemmanagements eingesetzt. Ihr Design ist vor allem von praktischen Gesichtspunkten bestimmt (leichte Verwendbarkeit, Effizienz, Vollständigkeit). *Perl* gilt als leicht erlernbar für Programmierer mit Erfahrung im Umgang mit *C*, *sed*, *awk*, *sh* - alles Werkzeuge hauptsächlich aus dem UNIX-Bereich. Die Sprache wurde Ende 1994 grundlegend überarbeitet. Seit dem 16. Oktober 1994 steht die aktuelle Version 5 für die wichtigsten Architekturen und Betriebssysteme zur Verfügung. Als Referenz gilt [77], [78]. Pflege und Weiterentwicklung erfährt die Sprache einzig und allein durch ihren Autor L. Wall. Gleichzeitig steht jedoch eine Usenet Diskussionsgruppe mit vielen tausend *Perl*-Benutzern bei Problemen zur Verfügung.
- *Python*
  Diese Sprache kann als die jüngste und modernste der hier vorgestellten Sprachen angesehen werden. *Python* wurde von G. van Rossum am niederländischen Centrum voor Wiskunde en Informatica Amsterdam entwickelt. Auch diese Sprache wird interpretiert; sie kann interaktiv benutzt werden; sie ist von Haus aus objekt-orientiert und gilt als leistungsstark bei einfacher Syntax. *Python* kann durch *C* oder *C++*-Funktionen erweitert werden. Ihr fehlte jedoch zum Zeitpunkt der Entscheidungsfindung für diese Arbeit wie *Scheme* oder *Perl* eine Anbindung an GUI-Grafik (siehe unten).[†]

[*] Practical Extraction and Report Language

[†] Mittlerweile bietet *Python* eine sehr gelungene Anbindung an *tk* oder die *Motif*-Bibliothek.

Außerdem scheint diese Sprache die gleiche Nische zu besetzen wie *Tcl*, jedoch mit einer zeitlichen Verzögerung.

- *Tcl**
  Für diese Arbeit fiel die Entscheidung auf *tcl*, weil diese Sprache nach ersten Projektstudien am geeignetsten erschien. Deshalb folgen an dieser Stelle ein paar ausführlichere Bemerkungen zu *tcl*.
  Tcl entstand an der University of California at Berkeley unter der Leitung von John Osterhout. Im Frühjahr 1988 begannen die Arbeiten an einer wiederverwendbaren Kommandosprache. Diese Sprache wurde als eine *C*-Bibliothek implementiert, wodurch die Sprache leicht in beliebige Anwendungsprogramme integriert werden kann. Weiterhin wurde von Anfang an sehr viel Wert darauf gelegt, daß die Sprache durch Erweiterungen in ihren Fähigkeiten an das Anwendungsproblem angepaßt werden kann. Zusätzliche Anforderungen, wie die Möglichkeit eine graphische Benutzerschnittstelle (Graphical User Interface - GUI) ebenso flexibel und wiederverwendbar zu implementieren, führten zur Entwicklung von *tk*, dem graphischen ToolKit auf der Basis der unter UNIX standardisierten X Bibliothek. Die Arbeiten an *tk* begannen 1989. Diese Erweiterung von *tcl* um graphische Fähigkeiten stellt einen der wesentlichen Vorteile gegenüber den anderen Sprachen dar.†
  Tcl ist zweifellos eine leistungsfähige und attraktive Skriptsprache. Sie fand Anfang der 90er Jahre schnelle Verbreitung in der UNIX-Welt und wurde dort mittlerweile ein Quasi-Standard. Sie ist eine interpretierte Sprache. *Tcl* ist heute für zahlreiche Computerarchitekturen und Betriebssysteme verfügbar, darunter UNIX-Rechner und IBM-kompatible PCs. Inzwischen wird das Paket von Mitarbeitern von Sun Microsystems Laboratories, Inc. unter der Leitung des Originalautors gewartet und weiterentwickelt. Diese Gruppe treibt gegenwärtig vor allem die Portierung des Paketes auf IBM-kompatible PCs und Macintosh-Rechner voran. Auch die Entwicklung eines *tcl*-Compilers ist dort geplant während ein Drittanbieter bereits mit einen kommerziellen *tcl*-Compiler auf dem Markt ist. *Tcl* selbst ist nach wie vor völlig kostenfrei erhältlich und kann auch in kommerziellen Paketen verwendet werden. Es sind keinerlei Lizenzgebühren fällig. Als Dokumentation stehen mehrere Medien zur Verfügung. Zum einen erhält man mit der *tcl*-Distribution Online-Manualseiten, die dem Programmierer jederzeit zur Verfügung stehen. Desweiteren sind zwei Textbücher kommerziell erhältlich [79], [80]. *Tcl* stellt daher ein gerade auch für universitäre Umgebungen hervorragend geeignetes Werkzeug dar. Die nachfolgenden Abschnitte werden die Eignung dieser Sprache für chemierelevante Problemstellungen unter Beweis stellen.

Mittels der gewählten Skriptsprache steht ein Basissprachumfang zur Verfügung zu dem die folgenden wichtigen Elemente zählen: Variablen, Listen und Prozeduren; Kontrollstrukturen (Schleifen, `if-else`-Verzweigungen, `switches`); Dateizugriffsfunktionen (für allgemeinen File I/O); arithmetische, relationale, logische und bitweise Operatoren; Manipulationsfunktionen für Listen und Zeichenketten (strings).

Die folgenden Zeilen geben ein kleines Beispiel:

* Tool Command Language, gesprochen *tickel*

† Es existieren *Perl*- und *Scheme*-Varianten, die die *tk*-Bibliothek benutzen. Im Fall von *Perl* ist diese Anbindung jedoch äußerst umständlich und dadurch schwer verwendbar.

```
1 # Procedure min
2 #---------------
3 # Takes a list of arguments and returns the smallest one.
4 # Args:
5 #   args - A variable number of values from which the
6 #          smallest one has to be found.
7 # Returns:
8 #   The smallest list element.
9 #-----------------------------------------------------------
10 proc min { args } {
11     set minimum [lindex $args 0]
12     foreach arg $args {
13         if { $arg < $minimum } {
14             set minimum $arg
15         }
16     }
17     return $minimum
18 }
```

Mittels Kommentar (`#`) werden Aufgabe, Argumente und Rückgabewerte einer Prozedur[*] (`min`) dokumentiert (Zeilen 1 bis 9). Dann folgt die Definition der Prozedur. Einer Variablen (`minimum`) wird als Initialisierungswert das erste Argument zugewiesen (Zeile 11). In der darauf folgenden Schleife (`foreach`, Zeilen 12 bis 16) werden alle Argumente mit der Variable `minimum` verglichen. Ist ein Argument kleiner, so nimmt `minimum` dessen Wert an. Der Rückgabewert (Zeile 17) ist dann das kleinste Element der Argumentliste. Zweifellos sind diese Zeilen auch für Nicht-Programmierer verhältnismäßig leicht verständlich.

Der folgende Abschnitt wird verdeutlichen, warum diese Merkmale allein für eine chemieorientierte Anwendung nicht ausreichen und die Erweiterbarkeit einer Skriptsprache daher ein wesentliches Auswahlkriterium darstellt.

[*] Prozedur ist der für *tcl* typische Begriff. Er kann synonym für Funktion oder Routine verstanden werden.

3.2.3 Chemiebezogene Erweiterungen

Aus dem vorangegangenen Abschnitt wurde ersichtlich, warum *tcl* eine geeignete Sprache für die hier zu behandelnden Probleme darstellt. Es ist jedoch unschwer nachvollziehbar, daß eine solche, für Anwendungen beliebiger Natur entwickelte Sprache, nicht direkt diejenigen Objekte und Informationseinheiten in ihrer Semantik abbildet, die innerhalb der chemischen Informationsverarbeitung (im engeren Sinne in der Syntheseplanung) benötigt werden. Objekte wie „Atome", „Bindungen" oder „Moleküle" existieren für *tcl* nicht. Gleichwohl will man natürlich auf diese Objekte zugreifen und sie handhaben können. Eine *Erweiterung* der Skriptsprache, um „Chemie" handhaben zu können, ist daher erforderlich. Nachfolgend soll auf die Einzelheiten der erforderlichen Informationen (Daten) eingegangen werden.

Die für eine flexible Programmierschnittstelle zugänglichen und für die organische Synthese relevanten Informationen lassen sich in mehrere Gruppen einteilen.

Als Grundlegendes braucht eine Sprache, die molekulare Strukturen handhaben soll, natürlich genau wie eine konventionelle, compilierte Sprache Zugriff auf die Konnektivitätstabelle (connection table - CT) der chemischen Verbindung. In Standard-Hochsprachen wie *C*, *Fortran* und dgl. besteht diese Zugriffsmöglichkeit automatisch, da die CT als ein Datenkonstrukt in der jeweiligen Hochsprache implementiert wird. Für eine Interpretersprache wie *tcl*, die in ein Programm eingebettet ist, existiert dieser Zugriff nicht *per se*. *Tcl* muß über seine Sprachkonstrukte auf die Datenstruktur* des einbettenden Programms zugreifen können. Das bedeutet z. B., ein Molekül wird vom Basisprogramm eingelesen und verwaltet, *tcl* erhält aber die *Möglichkeit*, auf diese Informationen zuzugreifen und - wie man noch sehen wird - u. U. diese sogar zu verändern.

Zu diesen <u>grundlegenden</u> Informationen gehören demnach solche Basisdaten wie die Zahl der Atome, Bindungen oder Einzelmoleküle; Identifikatoren (typischerweise Indices oder Labels) für einzelne Atome, Bindungen oder Moleküle, um gezielt Einzelinformationen für diese Objekte erhalten oder manipulieren zu können; Informationen zu Bindungsordnungen, der Atomart (chemisches Symbol, Ordnungszahl), formale Ladungen oder die Zahl freier Elektronen von Atomen. Eine vollständige Liste dieser Basisdaten umfaßt Tabelle 3 - 1.

Davon ableitbar sind eine ganze Reihe weiterer Informationen. Diese <u>abgeleiteten</u> Daten sollen *tcl* ebenfalls zur Verfügung stehen, jedoch vom zugrundeliegenden Programm (*C*/*Fortran*) ermittelt werden. Dazu gehören beispielsweise Informationen über Ringmitgliedschaften von Atomen und Bindungen oder die Aromatizitätsinformation. Diese Gruppe von Informationen wird in Tabelle 3 - 2 zusammengefaßt.

Desweiteren ist es erforderlich, daß die Interpretersprache (*tcl*) auch auf <u>physikochemische</u> Parameter zugreifen kann. Natürlich gilt hier aus Gründen der Geschwindigkeit ihrer Berechnung - wie für die vorangehenden abgeleiteten Informationen - daß diese Parameter nicht in *tcl*, sondern durch compilierte Funktionen des Kernsystems ermittelt werden. Unter diesen physikochemischen Parametern werden all jene Werte verstanden, die auch vom Programm PETRA† ermittelt werden, insbesondere jedoch diejenigen, die in Tabelle 3 - 2 beschrieben werden. Tatsächlich werden beim Zugriff von *tcl* auf solche Informationen, die gleichen Routinen wie im PETRA Programm verwendet.

---

* Datenstruktur meint an dieser Stelle beliebige Variablen, Vektoren, Strukturen usw. die für das Basisprogramm in *C* oder *Fortran* definiert wurden.

† <u>P</u>arameter <u>E</u>stimation for the <u>T</u>reatment of <u>R</u>eactivity <u>A</u>pplications

Tab. 3 - 1 Grundlegende chemische Informationseinheiten, die für Erweiterungen von *tcl* transparent sein müssen.

Information	*tcl*-Schlüsselwort	Zugriffsart[a]
ensemblezentrierte Informationen		
Name des Ensembles	`E_NAME`	L/S
Identifikator des Ensembles	`E_IDENT`	L/S
Ensemblehashcode [57]	`E_HASH`	L
Ensemblehashcode mit Berücksichtigung von Stereochemie und Isotopen	`E_HASHS`	L
Textzeilen in Form von Kopf- und Fußzeilen, die für graphische Darstellungen relevant sind	`E_GRAPHICS_TEXT`	L/S
molekülzentrierte Daten		
Molekülidentifikatoren (Indices[b])	`M_LABEL`	L
Molekülhashcode [57]	`M_HASH`	L
bindungszentrierte Daten		
Bindungsidentifikatoren (Indices[b])	`B_LABEL`	L
Bindungsordnung	`B_ORDER`	L/S
Stereodeskriptoren der Bindung[c]	`B_CIPSTEREO` `B_CTSTEREO` `B_LABSTEREO`	L
Bindungshashcodes [57]	`B_HASH`	L
Flags für grafische Darstellungen[d]	`B_FLAGS`	L/S
atomzentrierte Daten		
Ordnungszahlen	`A_ELEMENT`	L/S
Symbole	`A_SYMBOL`	L/S
Atomidentifikatoren (Labels)	`A_LABEL`	L/S
Stereodeskriptoren der Atome[e]	`A_CIPSTEREO` `A_DLSTEREO` `A_LABSTEREO`	L
formale Ladungen	`A_NOM_CHARGE`	L/S
Zahl freier (nicht an Bindungen beteiligter) Elektronen	`A_FREE_ELECTRONS`	L/S
Atomhashcodes [57]	`A_HASH`	L
Flags für graphische Darstellungen[f]	`A_FLAGS`	L/S

a. L - Lesezugriff, S - Schreibzugriff
b. Derzeit sind das Indices. Es ist aber auch die Benutzung von Labels an Stelle oder zusätzlich zu Indices denkbar.
c. `CIP`: nach Cahn, Ingold und Prelog [81], [82]; `CT`: *cis/trans*-Nomenklatur; `LAB`: basierend auf der lokalen Atomumgebung
d. Beispiele sind invertierte Darstellung, Kennzeichnung durch *, Umrahmung durch einen Kasten usw.
e. Siehe c; `DL`: nach D/L-Nomenklatur
f. Das sind Attribute wie Fettdruck, gestrichelte Darstellung, Keilbindung etc.

Tab. 3 - 2 Abgeleitete chemische Informationseinheiten, die für Erweiterungen von *tcl* greifbar sein müssen.

Information	*tcl*-Schlüsselwort	Zugriffs-art[a]
ensemblezentrierte Daten		
Zahl der Ringe im SSSR[b]	E_NSSSR	L
Zahl der Ringe im ESSR[c]	E_NESSR	L
molekülzentrierte Daten		
Molekülgewicht	M_WEIGHT	L
ringbezogene Daten		
Liste der Atome je Ring	R_ATOMS	L
Liste der Bindungen je Ring	R_BONDS	L
Ringidentifikatoren[d]	R_LABEL	L
Aromatizitätskennzeichnung	R_AROMATIC	L
Ringgröße	R_SIZE	L
Ringsystemzugehörigkeit	R_SYSTEM	L
bindungszentrierte Daten		
Aromatizitätskennzeichnung	B_ISAROMATIC	L
Kennzeichnung aromatischer *und* zu aromatischen Systemen α-ständiger Bindungen	B_ISALPHAARO	L
Kennzeichnung hetero-aromatischer Bindungen	B_ISHETARO	L
Zähler für Ringmitgliedschaft[e]	B_RING_COUNT	L
Flags für Zugehörigkeit zu Ringen der Größen 3-7	B_[3-7]RING	L
Flags für Zugehörigkeit zu bestimmten Ringgrößen: kleine (3/4), mittlere (5-7) und große (über 8) Ringe	B_SMALL_RING B_MEDIUM_RING B_LARGE_RING	L
atomzentrierte Daten		
Aromatizitätskennzeichnung	A_ISAROMATIC	L
Kennzeichnung zu aromatischen Systemen α-ständiger Atome	A_ISALPHAARO	L
Zähler für Ringmitgliedschaft[e]	A_RING_COUNT	L
Kennzeichnung von Brückenkopfatomen	A_BRIDGEHEAD	L
Kennzeichnung von Spirozentren	A_SPIRO	L
Zahl der Wasserstoff-Nachbarn	A_HCOUNT	L
Atomgewichte	A_WEIGHT	L
IUPAC Gruppennummer	A_IUPAC_GROUP	L
freie Valenzen	A_FREE_VALENCE	L
Valenzen	A_VALENCE	L

a. L - Lesezugriff, S - Schreibzugriff
b. SSSR: Smallest Set of Smallest Rings. Siehe Abschnitt 4.3.4.
c. ESSR: Extended Set of Smallest Rings. Dieser Satz an Ringen ist um solche Ringe erweitert, die Hüllringe zweier SSSR-Ringe sind (wobei Ringpaare ausgeschlossen sind, die genau eine gemeinsame Bindung haben.)
d. Indiziert werden die primären und sekundären Ringe (siehe Abschnitt 4.3.4).
e. In Ringen des SSSR.

Tab. 3 - 3 Physikochemische Daten, die für *tcl* zur Verfügung stehen müssen.

Information	*tcl*-Schlüsselwort	Zugriffs-art[a]
bindungszentrierte Daten		
Differenz der π-Elektronegativität der Bindungsatome je Bindung ($\Delta\chi_\pi$)	`B_DENPI`	L
Differenz der σ-Elektronegativität der Bindungsatome je Bindung ($\Delta\chi_\sigma$)	`B_DENSIG`	L
Differenz der π-Atompartialladungen der Bindungsatome je Bindung (Δq_π)	`B_DQPI`	L
Differenz der σ-Atompartialladungen der Bindungsatome je Bindung (Δq_σ)	`B_DQSIG`	L
Differenz der (Gesamt-)Atompartialladungen der Bindungsatome je Bindung (Δq)	`B_DQTOT`	L
Bindungspolarisierbarkeit [83]	`B_POLARIZ`	L
die durch den PEOE-Algorithmus über die Bindung verschobene Ladungsmenge [84]	`B_SQIT`	L
(Gesamt-)Delokalisierungsstabilisierung[b] von Ladungen bei heterolytischen Bindungsbruch $D^{\pm}$ [86]	`B_SDELOC`	L
Delokalisierungsstabilisierung einer positiven Ladung bei heterolytischem Bindungsbruch D^{+} [86]	`B_PDELOC`	L
Delokalisierungsstabilisierung einer negativen Ladung bei heterolytischem Bindungsbruch D^{-} [86]	`B_NDELOC`	L
atomzentrierte Daten		
π-Elektronegativität (χ_π) [87]	`A_ENPI`	L
σ-Elektronegativität (χ_σ) [87]	`A_ENSIG`	L
π-Atompartialladung (q_π) [88]	`A_QPI`	L
σ-Atompartialladung (q_σ) [84]	`A_QSIG`	L
(Gesamt-)Atompartialladung (q)	`A_QTOT`	L
Atompolarisierbarkeit (α) [83]	`A_POLARIZ`	L

a. L - Lesezugriff, S - Schreibzugriff

Eine letzte Gruppe an Informationen umfaßt Angaben zur zweidimensionalen Darstellung sowie Angaben zur räumlichen Gestalt von chemischen Strukturen. Es liegt auf der Hand, daß ein offenes Interpretersystem Zugang zu dreidimensionaler Molekülinformation haben muß, wenn es potentiell in der Lage sein soll, stereochemische Problemstellungen zu bearbeiten. Derartige Informationen benennt Tabelle 3 - 2.

Tab. 3 - 4 Informationen zur 2D-Darstellung und zur räumlichen Gestalt, die für *tcl* transparent sein müssen.

Information	*tcl*-Schlüsselwort	Zugriffs-art[a]
ensemblezentrierte Daten		
Standardbindungslänge im Molekülplot im Maßstab der 2D-Koordinaten	`E_STDBLE`	L
Das begrenzende Rechteck, das den Molekülplot vollständig umschließt.	`E_2DEXTENT`	L
atomzentrierte Daten		
zweidimensionale Atomkoordinaten	`A_XY`	L/S
dreidimensionale Atomkoordinaten	`A_XYZ`	L

a. L - Lesezugriff, S - Schreibzugriff

Neben den voranstehend beschriebenen Informationseinheiten muß es *tcl* auch ermöglicht werden, grundsätzliche Operationen auf den chemischen Objekten durchzuführen. Einige Forderungen, bestimmte Eigenschaften von Atomen, Bindungen oder Molekülen ändern zu können (Schreibzugriff) finden sich bereits in den obenstehenden Tabellen. Das bedeutet, diese Informationen einstellen, sie setzten zu können. Demgegenüber steht die reine Abfrage von Informationen (Lesezugriff). Notwendig ist auch, bestimmte Objekte völlig zu entfernen bzw. neu zu kreieren.

Will man beispielsweise aus einem molekularen Ensemble, d. h. einem Satz von mehreren Einzelmolekülen, nur das größte erhalten, so müssen alle anderen Moleküle gelöscht werden. Oder man will eine Addition von Atomen vornehmen. Dann muß es möglich sein, diese Atome zu definieren und neue Bindungen zwischen diesen und möglicherweise bestehenden Atomen zu bilden. Dabei kann es geboten sein, zunächst andere Bindungen zu brechen.

Diese Operationen faßt Tabelle 3 - 5 zusammen.

Tab. 3 - 5 Operationen, die *tcl* auf chemischen Objekten ausführen können muß.

Information	*tcl*-Schlüsselwort	Zugriffs-art[a]
Abfragen bestehender Information	`get`	L
Manipulieren bestehender Information	`set`	S
Erzeugen von Atomen, Bindungen	`create`	S
Entfernen von Atomen, Bindungen, Molekülen	`delete`	S
Kopieren ganzer Informationsblöcke (z. B. Ensemble)	`copy`	L/S
Neuberechnung von Informationen (vollständig oder teilweise)	`need`	S

a. L - Lesezugriff, S - Schreibzugriff

Darüber hinaus sind auch weitere, höhere Operationen denkbar, die zum Sprachumfang einer chemietauglichen Skriptsprache gehören müssen, etwa die Suche nach einer Substruktur auf dem Molekülskelett. Natürlich ist man mit einer Sprache, die obige Elemente enthält, bereits in der Lage, diese Substruktursuche als Funktion zu implementieren. Jedoch kann es aus Geschwindigkeitsgründen geboten sein, derartige Funktionalität in den Sprachumfang aufzunehmen. Ein Beispiel ist das für *tcl* implementierte `match` Kommando. Dieses wird in den WODCA Bewertungsfunktionen für gefundene Ausgangsmaterialien benutzt, um maximale gemeinsame Substrukturen zwischen der Anfrageverbindung und den Treffern der Liste zu finden, die dann eingehender untersucht werden.

Als letzte Anmerkung bleibt nun nur noch die Forderung, das die gewünschten Erweiterungen sich nahtlos in die Syntax der Standardsprache einfügen sollten. *Tcl* ist eine Sprache, die komplexe Information listenbasiert behandelt [79]. Dementsprechend werden auch bei den neu geschaffenen Erweiterungskommandos Rückgabewerte in Listenform erzeugt.

Fünf neue Kommandos sind ausreichend, um die in den voranstehenden Tabellen aufgeführten Funktionalitäten zu realisieren. Diese fünf Kommandos unterscheiden sich durch den Typ des Objektes, auf das sie einen Zugriff ermöglichen:

- ❏ `atom` Atome
- ❏ `bond` Bindungen
- ❏ `ring` Ringe
- ❏ `mol` (Einzel-)Moleküle. Das können auch Ionen sein.
- ❏ `ens` Molekulare Ensembles. Ensembles können mehrere Moleküle/Ionen umfassen.

Jedes dieser Kommandos kennt mehrere, maximal vier Unterkommandos und zwar `create|delete|get|set` für `atom` und `bond` Kommando, `delete|get` für das `mol` und `copy|get|set|need` für das `ens` Kommando. In der Syntax folgt danach eine Spezifizierung des Datenblocks, auf dem das Kommando arbeiten soll. Dazu eine Anmerkung: In der auf *Fortran77* basierenden EROS- und WODCA-Umgebung existieren drei wesentliche Speicherbereiche für vollständige Molekülinformation. An diese Speicherbereiche sind die Bezeichner `VORHER`, `MOLINF`, `NACHER` angelehnt. Äquivalent dazu sind die Bezeichner `ens0`, `ens1`, `ens2`. In einer anderen zugrundeliegenden Programmierumgebung, die andere, u. U. auch mehr als drei Speicherbereiche für Molekülinformationen kennt, kann das gleiche Syntaxmodell verwendet werden. Die Bezeichner `ens0`, `ens1` ... `ens`*n* sind dann als Verweise auf die jeweiligen Datenblöcke dieser Umgebung zu verstehen. Für *tcl* ändert sich dadurch *nichts*.*

Eine dokumentarische Beschreibung der Funktionalität aller erlaubten Kombinationen findet sich in [90]. Die folgenden Zeilen geben einige instruktive Beispiele wieder.

Beispiel 1:
Im Datenblock *dblock* befinde sich Toluol (siehe Abb. 3 - 1). Das folgende Kommando informiert über die Zugehörigkeit der einzelnen Atome zu einem aromatischen System. Die Reihenfolge der Ausgabe entspricht der Reihenfolge der Atomnumerierung. Aromatische Atome erhalten eine Eins für die Gültigkeit, nicht-aromatische eine Null für die Ungültigkeit der Information.

```
  ens get dblock A_ISAROMATIC
→ 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
```

Beispiel 2:
Für das gleiche Beispielmolekül wird diesmal ermittelt, wieviel Wasserstoffatome dem Atom 7 (Label) benachbart sind.

```
  atom get dblock 7 A_HCOUNT
→ 3
```

* Die Syntax der chemischen Erweiterungen für *tcl* wurde an jener des CACTVS Systems [89] ausgerichtet, um weitgehende Kompatibilität auf der Ebene der *tcl*-Skripts zwischen diesen beiden Programmen zu gewährleisten. Die eigentliche Programm-Datenstruktur unterscheidet sich jedoch erheblich.

Beispiel 3:
Nun soll eine Bindung zwischen dem aliphatischen C-Atom (`7`) und einem seiner Wasserstoffnachbarn (z. B. `13`) gebrochen und ein Sauerstoffatom eingefügt werden. Das Resultat soll Benzylalkohol sein. Dazu wird zunächst die betreffende Bindung gebrochen (`20` sei der Index der Bindung zwischen Atom `7` und Atom `13`) ...

```
  bond delete dblock 20
→
```

dann ein Sauerstoffatom erzeugt ...

```
  atom create dblock O
→ 16
```

und schließlich die beiden neuen Bindungen geknüpft.

```
  bond create dblock {7 16}
→ 22
  bond create dblock {13 16}
→ 31
```

Das Kommando zur Addition eines neuen Atoms verlangt als Parameter das chemische Symbol des betreffenden Elements*. Der Rückgabewert ist das Label (die Nummer) des neuen Atoms. Zur Knüpfung einer Bindung müssen die Nummern der beteiligten Atome (Labels) angegeben werden. Da keine Bindungsordnung angegeben wird, entsteht eine Einfachbindung (Vorgabewert). Rückgabewert ist hier der Index der neuen Bindung.

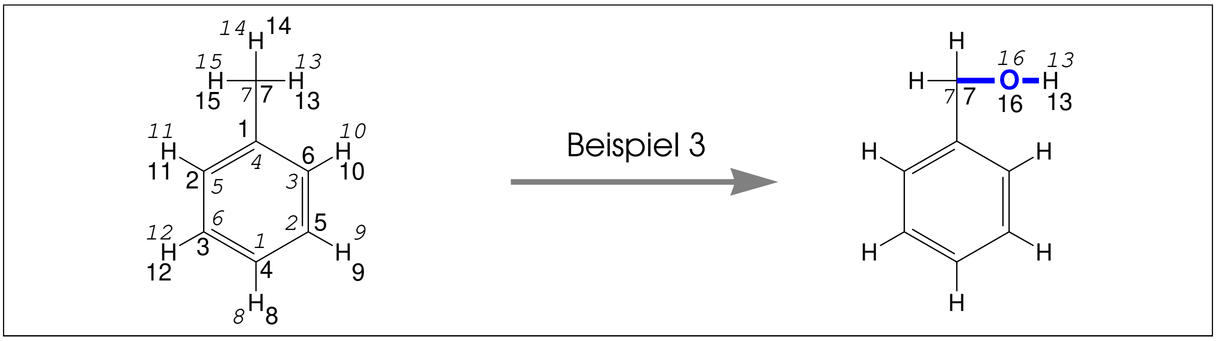


Abb. 3 - 1 Beispielmolekül Toluol (links). Nummern in kursiver Schrift (*1*) entsprechen der Reihenfolge der Eingabe der Atome (Indices), fette Nummern (**1**) können dagegen manuell zugeordnet sein und identifizieren ein Atom eindeutig (Labels).
Rechts: Resultat der Anwendung der Kommandosequenz aus Beispiel 3 (siehe Text).

Mittels der hier vorgestellten Erweiterungen ist man in der Lage, sowohl die Transformationsregeln dieses Abschnitts als auch die Funktionen zur Suche strategischer Bindungen (Kapitel 4) zur codieren. *Tcl* ist mit diesen Erweiterungen ausgezeichnet als Sprache der WODCA Wissensbasis geeignet. Genaue Zahlen, die das belegen, finden sich in den Abschnitten 3.3.1 und im Anhang A.

* Alternativ ist auch die Angabe der Ordnungszahl möglich.

3.3 Anwendungen in der Syntheseplanung

Dieser Abschnitt beschreibt Einzelheiten der für die Syntheseplanung implementierten Transformationssuchen. Nach einigen allgemeinen Bemerkungen, die für alle Transformationsregeln gelten, werden letztere im Detail vorgestellt.

3.3.1 Kataloge

Um geeignete Ausgangsmaterialien für ein Syntheseziel auswählen zu können, sind Sammlungen solcher Verbindungen in computer-lesbarer Form (Datenbanken) notwendig. Darauf wurde bereits im Abschnitt 2.2.1 eingegangen. An dieser Stelle sollen diejenigen Kataloge mit Ausgangsmaterialien benannt werden, die innerhalb dieser Arbeit Verwendung fanden (siehe Tab. 3 - 6).

Übernommen wurde der Katalog der Firma Janssen Chimica, Belgien (Stand 1987) aus dem alten WODCA System. Allerdings mußten die Hashcode-Tabellen für die neuen, *tcl*-basierten Transformationssuchen neu berechnet werden. Obwohl dieser Katalog nicht mehr dem aktuellen Stand entspricht*, dient er weiterhin als wertvolles Reservoir für Synthesebausteine, da davon ausgegangen werden kann, daß die verzeichneten Verbindungen bei dem einen oder anderen Feinchemikalienanbieter erhältlich sein werden.

Neu ist der Katalog der Fluka Chemie AG, der seit Mitte 1995 korrespondierend zur gedruckten Ausgabe 1995/96 als flukalog© 1.0 in Form einer Strukturdatenbank für MDL‘s ISIS System erworben werden kann. Dieser Katalog wurde in das für WODCA erforderliche Dateiformat konvertiert.

Ein für Anwendungen in der Syntheseplanung weniger geeigneter, nichtsdestoweniger aber sehr interessanter Katalog ist der VCH Trivialnamenkatalog (Band 1 und 2), der uns in elektronischer Form vom FIZ Chemie Berlin zur Verfügung gestellt wurde. Enthalten sind mehr als 20.000 Verbindungen. Das sind hauptsächlich Naturstoffe ohne weitere Informationen zu kommerzieller Verfügbarkeit. Daher ist dieser Katalog weniger zur ausgangsmaterial-orientierten Syntheseplanung geeignet. Auch dieser Katalog mußte konvertiert werden.

Tab. 3 - 6 Wichtige Kataloge des WODCA Systems

Anbieter	Originalgröße	Größe als WODCA Katalog	davon leere[a] Einträge	Bemerkungen
Janssen Chimica	8.464[b]	8.464	0	Stand 1987
Fluka Chemie	16.781	16.769	2.512	Entspricht flukalog 1.0 (1995).
FIZ Chemie Berlin (VCH)	21.812	20.441	0	Stand Band 1 und 2 der 1. Auflage des gedruckten Werkes (1993)

a. Leer bedeutet, daß ein solcher Katalogeintrag keine Strukturformel enthält. Andere Informationen wie Namen, Reinheit, Packungsgrößen etc. können aber vorhanden sein.
b. Von der vorangehenden WODCA Version übernommen.

* Janssen Chimica firmiert inzwischen unter dem Namen Acros Chimica, der neueste Katalog hat ca. 2.000 Einträge mehr.

Bei der Konvertierung der Kataloge in das CTX[*]-Format [91] gehen einige Einträge verloren. Das sind Verbindungen, die aufgrund des Größenlimits (mehr als 300 Atome, 600 Bindungen oder 75 Ringe) oder nicht codierbarer chemischer Merkmale (z. B. 3-Zentren-Bindungen wie in Boranen oder Metallionen mit mehr als acht Liganden wie in Metallocenen) vom WODCA System nicht behandelt werden können. Fehlcodierungen in den Originaldaten wurden soweit wie möglich korrigiert. Prinzipiell können beliebige Sammlungen chemischer Verbindungen als WODCA Katalog aufbereitet werden. Für einige verbreitete Dateiformate (Molfile, SDFile von MDL [92]; SMD[†] [93], [94]; PDB[‡] [95]) stehen Konvertierungsprogramme bereit [89]. Meist kann für jedes definierte Dateiformat mit begrenztem Aufwand ein Konvertierungstool programmiert werden.

Tab. 3 - 7 Durchschnittliche Zahlen für alle nicht leeren Einträge der verschiedenen Kataloge des WODCA Systems.

Katalog	Zahl der			
	Atome	Schweratome	C-Atome	Ringe
Janssen	22	12	8	1
Fluka	28	14	9	1
Trivialnamen	43	21	16	2

Aus der konvertierten CTX-Datei wird eine index-sequentiell lesbare Master-Datei (ebenfalls in CTX) angelegt. Parallel dazu werden eine ganze Reihe weiterer Dateien erzeugt, die einen schnellen (und portablen) Zugriff auf die Einträge gestatten. Von großer Bedeutung sind die für die Transformationssuchen notwendigen Hashcode-Dateien. Je Ähnlichkeitskriterium wird eine binäre Datei erzeugt, die das betreffende Transformationsresultat je Katalogverbindung als 32-Bit-Integer-Zahl codiert. Der Zeitaufwand für die verschiedenen Schritte der Katalogerstellung ist recht unterschiedlich. Während die Dateiformat-Konvertierung in der Größenordnung von etwas mehr als einer Zehntel CPU-Sekunde[**] je Verbindung liegt, beträgt der Aufwand für die Erstellung der Master-Datei etwa 2-3 CPU-Millisekunden[††] je Verbindung. Deutlich rechenintensiver ist dagegen die Generierung der Transformations-Hashcodes. Diese Aufgabe wird *tcl*-basiert durchgeführt, d. h. von einem Interpreter-Programm. Abbildung 3 - 2 stellt den Zeitaufwand (CPU-Zeit[‡‡]) je Katalogeintrag für einige Transformationen grafisch dar. Vollständiges Zahlenmaterial findet sich im Anhang A in Tabelle A - 1. Soweit vorhanden, sind Vergleichswerte für die analoge Berechnung mittels eines compilierten *Fortran*-Programms angegeben. Die *tcl*-basierte Berechnung dauert um einen Faktor von 14 (Hydrolyse ohne Aromatensubstituenten und Oxidation) bis 5 (Ozonolyse) länger. Der Leistungsunterschied ist deutlich, jedoch durchaus akzeptabel. Dazu muß man bedenken, daß diese Berechnung nur einmalig bei *Vorbereitung* des Kataloges oder bei Neudefinition einer Transformationsregel durchgeführt werden muß. Bei *Anwendung* der Transformation werden

* Clear Text File

† Standard Molecular Data

‡ Brookhaven Protein Data Bank

** Konvertierung des Fluka-Kataloges aus dem SDFile-Format nach CTX mittels eines *tcl*-Skripts. Die eigentliche Arbeit machen dabei allerdings *C*-Routinen. Zeit gültig für Sun Sparc 10/512 bzw. 10/40 oder SGI Indigo^2 R4000.

†† Gültig für Sun Sparc 10/512 bzw. 10/40. Sun Sparc 10/20 liegen bei etwa 5 ms.

‡‡ Gemeint ist damit die Summe aus der eigentlichen Prozeßzeit und der Zeit, die dieser Prozeß für Systemaufrufe benötigte.

Zeiten in der Größenordnung der Spalte „Zeit pro Eintrag“ beansprucht, d. h. maximal einige Sekunden - unabhängig von der Größe des durchsuchten Katalogs.* Dem Gewinn der flexiblen Definition von Ähnlichkeitssuchen steht also ein annehmbarer höherer Zeitaufwand bei der Vorbereitung eines Kataloges gegenüber. Man beachte auch, daß die Größe der in einem Katalog enthaltenen Verbindungen und deren Topologie (Ringe) nicht unerheblich die Transformationen beeinflussen (vgl. dazu Tabelle 3 - 7). Die durchschnittlich größten Einträge mit den meisten Atomen und Ringen finden sich im Trivialnamen-Katalog. Für diesen Katalog werden auch etwa doppelt so lange Zeiten je Eintrag beobachtet wie für den Janssen-Katalog, Extremfälle brauchen gar dreimal soviel Zeit. Bei den eher vergleichbaren Katalogen Janssen und Fluka findet man noch um ca. $1^1/_2$ mal längere Zeiten für den etwas größere Verbindungen enthaltenden Fluka-Katalog. Diese Werte können als Anhaltspunkte dienen, um abzuschätzen, mit welchen Aufwand für einen beliebigen Katalog gerechnet werden muß. Für einen Katalog mit einem Inhalt ähnlich dem des Fluka-Kataloges, 10.000 Verbindungen, 38 Transformationen und einem durchschnittlichen Aufwand von 1,4 CPU-Sekunden je Eintrag resultieren 532.000 Sekunden, das ist eine Rechenzeit von sechs Tagen, drei Stunden, 46 Minuten und 40 Sekunden - ein durchaus vertretbarer Aufwand.

* Das mag zunächst verblüffend klingen. Die eigentliche Suchzeit, d. h. der Vergleich des Hashcodes der transformierten Anfrageverbindung mit den im Voraus berechneten Hashcodes der Katalogverbindungen, liegt im Bereich von bestenfalls *Milli*sekunden.

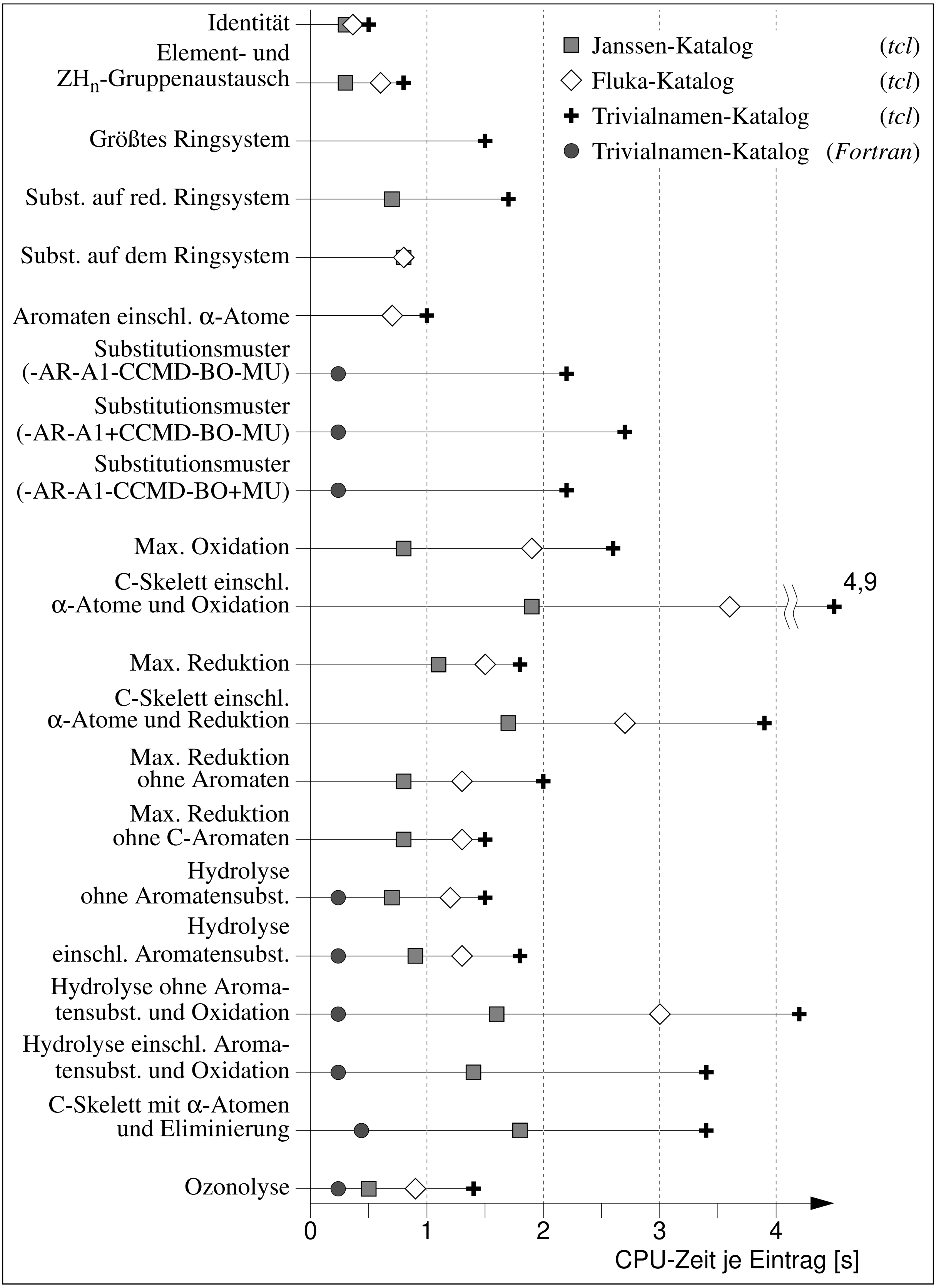


Abb. 3 - 2 Zeitaufwand je Katalogeintrag (Verbindung) für verschiedene Transformationssuchen. Alle dargestellten Werte wurden für eine Sun-Workstation (Sparc 10/512) und eine SunOS-Programmversion ermittelt.

3.3.2 Grundlegendes

Die Transformationssuchen sollen primär dazu dienen, für gegebene Verbindungen ähnliche Ausgangsmaterialien auszuwählen. Es sind also Verbindungen gefragt, die sich zur Synthese der Anfrageverbindungen einsetzen lassen Die zu diesem Zweck notwendige Definition von Ähnlichkeit basiert vor allen auf zwei Grundlagen: dem (Kohlenstoff-)Skelett und der auf dem Skelett verteilten Funktionalität.

Das Skelett einer Verbindung kann als die erste Basis von für Synthesen relevanter Ähnlichkeit dienen. Oft kann es das reine Kohlenstoffskelett sein. CC-Mehrfachbindungen können dabei meist ignoriert werden.

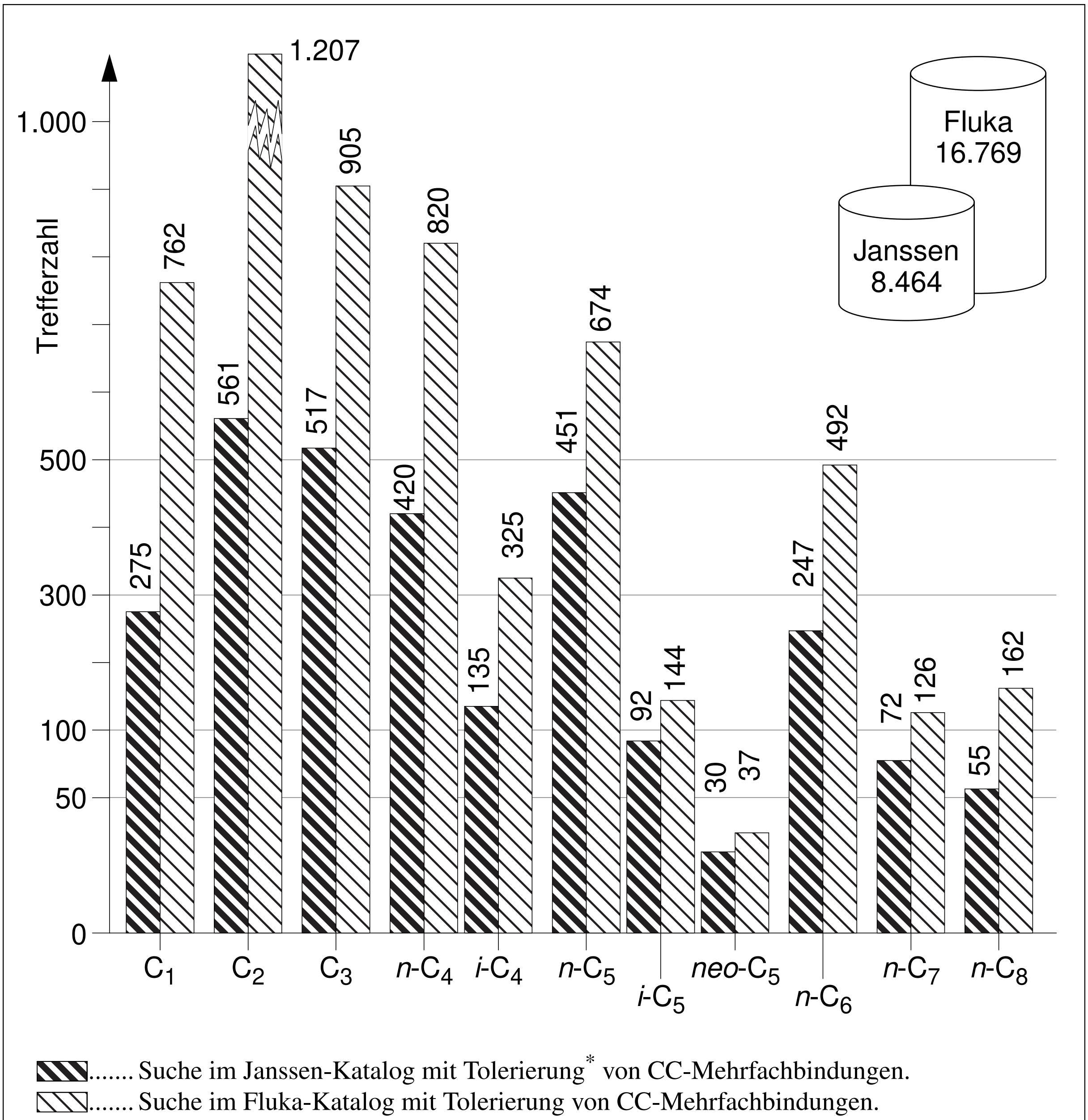
Wipke [35] schätzte, daß nur etwa 20 % aller CC-Bindungen eines Syntheseziels nicht schon im Ausgangsmaterial vorhanden waren. Dagegen wurden ca. 40 % aller C-Heteroatom-Bindungen (im folgenden CX-Bindungen) im Syntheseziel erst im Laufe der Synthese geknüpft. Will man beispielsweise ein bestimmtes C_5-Skelett synthetisieren, ist es naheliegend, dieses Skelett bereits aus einem Ausgangsmaterial zu übernehmen. Dieses sehr einfache Kriterium kann die aus einem Katalog von Ausgangsmaterialien in Frage kommenden Verbindungen bereits drastisch einschränken. Abbildung 3 - 3 unterstreicht das. Die Zahl an Verbindungen, die einem bestimmten (acyclischen) Skelett entsprechen, liegt im Extremfall (C_2-Einheit) in der Größenordnung von 1.000. Größere Skelette sind sehr viel rarer.

Dieser Art der skeletalen Ähnlichkeit bedienen sich die WODCA-Transformationssuchen „C-Skelett“ oder „Reduziertes C-Skelett“.

Einen Spezialfall stellen Ringsysteme dar. Handelt es sich um reine Kohlenwasserstoffe gilt auch hier, daß CC-Bindungen viel seltener als CX-Bindungen im Verlauf der Synthese aufgebaut werden. Aromatische Systeme stellen eine weitere Besonderheit dar. Für diese läßt sich eine noch höhere Tendenz zur Übernahme aus einem Ausgangsmaterial feststellen [35], d. h. aromatische Bindungen werden noch *seltener* im Verlauf der Synthese verändert. Gehören zu einem Ringsystem jedoch Heteroatome (cyclische Ether, Amine, Lactone, Lactame etc., Pyridine, Furane, Thiophen usw.), stellen diese Heteroatome einen zentralen Bestandteil des Skeletts dar, sind also als Ähnlichkeitskriterium einzubeziehen. Vor diesem Hintergrund wurden die WODCA-Transformationssuchen „Größtes Ringsystem“, „Ringe und C-Skelett“ und „C-Skelett und Aromaten einschließlich α-Atome“ entwickelt.

Offensichtlich ist, daß mittels dieser Kriterien auch die *allgemeinste* Art an skeletaler Ähnlichkeit definiert wird. Das bedeutet, wenn diese Kriterien für eine gegebene Anfrageverbindung keinerlei Treffer liefern, daß spätestens dann die Methoden zur Suche nach strategischen Bindungen angewandt werden müssen. Werden aber zu viele Verbindungen gefunden (in der Praxis einige Hundert oder mehr Treffer), müssen selektivere Ähnlichkeitskriterien angewendet werden.

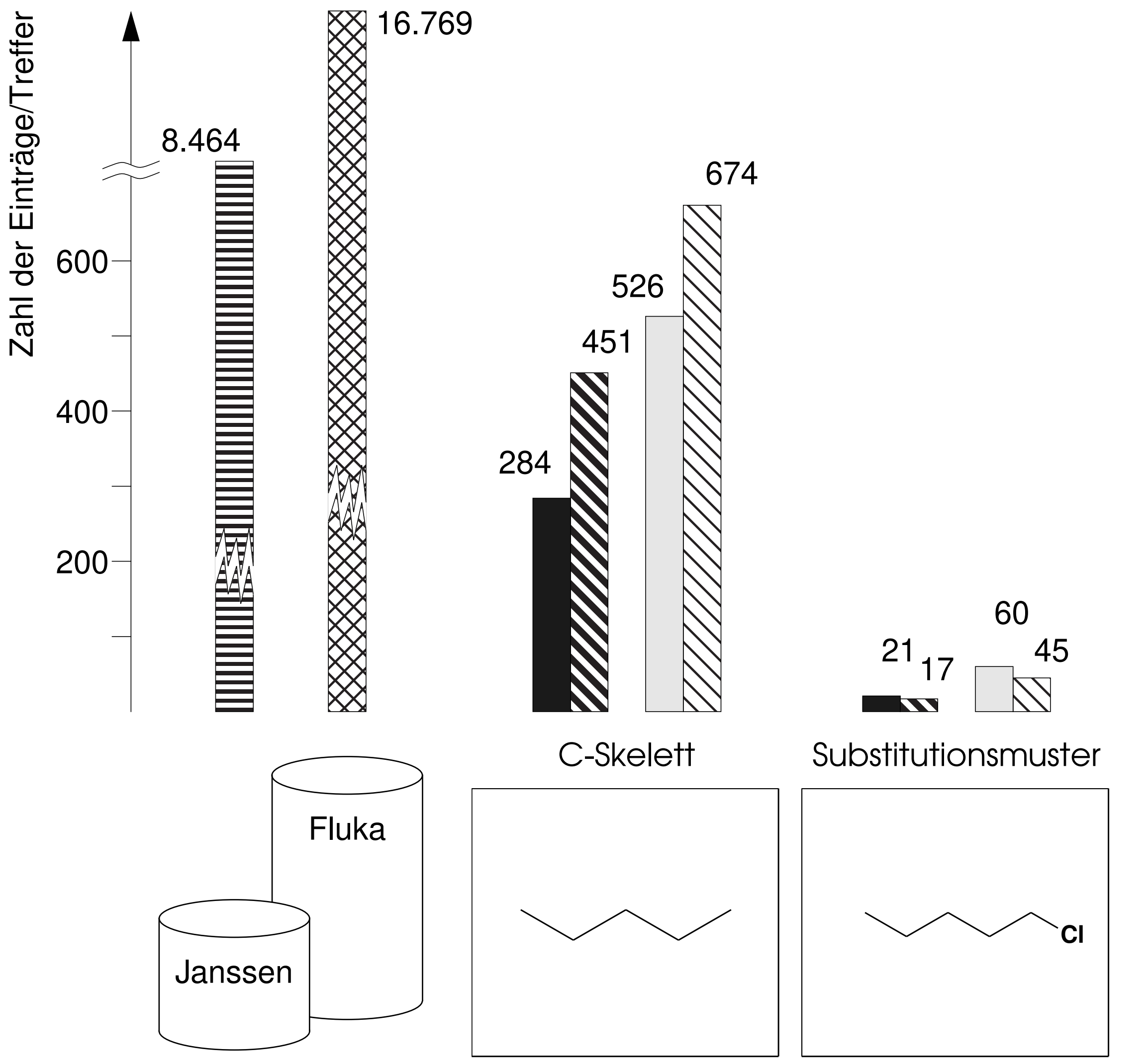
Selektivere Ähnlichkeit läßt sich problemlos auf der Basis der Funktionalität einer Verbindung definieren. Ein bestimmtes Skelett allein ist ein guter Filter zur Auswahl von Verbindungen aus einem Katalog. Findet auch die Funktionalität Beachtung, so reduzieren sich die in Frage kommenden Verbindungen noch drastischer (siehe Abb. 3 - 4).



* Tolerierung von CC-Mehrfachbindungen heißt, daß diese keine Rolle bei der Suche spielen. Butadien wird danach wie *n*-Butan als ein C_4-Treffer erfaßt.

Abb. 3 - 3 Zahl an Katalogverbindungen, die bestimmten C-Skeletten entsprechen (Suche mit Transformation „Reduziertes C-Skelett“). Angegeben sind die Trefferzahlen für C-Skelette von C_1 bis C_8. Man beachte, daß hier wirklich Skelette gesucht wurden und nicht Elementzählungen.

Mit den verschiedenen Möglichkeiten, wie Substituenten behandelt werden können, d. h. wie sie die Ähnlichkeit von Verbindungen beeinflussen können, befaßt sich Abschnitt 3.3.4 eingehend.



.... Suche im Janssen bzw. Fluka Katalog ohne Tolerierung von CC-Mehrfachbindungen

.... Suche im Janssen bzw. Fluka Katalog mit Tolerierung von CC-Mehrfachbindungen. Man beachte daß solche Bindungen bei einer Substitutionsmuster-Suche als Substituenten zählen.

Abb. 3 - 4 Beispiele für die Anwendung einfacher Ähnlichkeitskriterien zur Auswahl von Ausgangsmaterialien. Ganz links die Gesamtzahl an Verbindungen zweier durchsuchter Kataloge. Daneben die Zahl der Verbindungen, die ein *n*-Pentan-Skelett enthalten. Rechts eine Ähnlichkeitssuche nach beliebig in 1-Position substituierten *n*-Pentanen.

Für alle Transformationen des WODCA Systems gelten einige grundlegende Konventionen, die immer zur Anwendung kommen.

Viele Transformationsregeln fragmentieren eine Originalverbindung. Resultat sind also mehr oder weniger große transformierte Verbindungen. In solchen Fällen wird nur das größte Fragment zur Hashcodeberechnung herangezogen, die anderen Fragmente werden verworfen. Dem liegt die Erfahrung zugrunde, daß Ähnlichkeit zwischen den *wichtigsten* Teilen zweier Verbindungen erkannt werden soll und sich diese Wichtigkeit am ehesten über die *Größe* dieser Teile messen läßt.

Im alten WODCA System lag der Größenberechnung Gleichung (3-1) zugrunde.

$$\text{Größe(Mol)} = 1.000.000 \cdot n_{atom}(\text{Mol}) - 10 \cdot M(\text{Mol}) + \text{Mol} \qquad (3\text{-}1)$$

mit
Mol......... Molekülindex
n_{atom}....... Zahl der Atome je Molekül
M........... molare Masse des Moleküls

Man beachte, daß die Molmasse abwertend wirkt. Das bedeutet, daß leichtere Fragmente als *größer* betrachtet werden als schwerere Fragmente. Das wurde eingeführt, um Phenylringe abzuwerten, und von der Zahl der Atome gleichgroße, aber leichtere Fragmente höherzubewerten. Das führt aber zu unerwünschten Nebeneffekten. Mit obigem Ansatz wird aus den in Abbildung 3 - 5 dargestellten Molekülen *n*-Propanol als das *größte* ausgewählt.

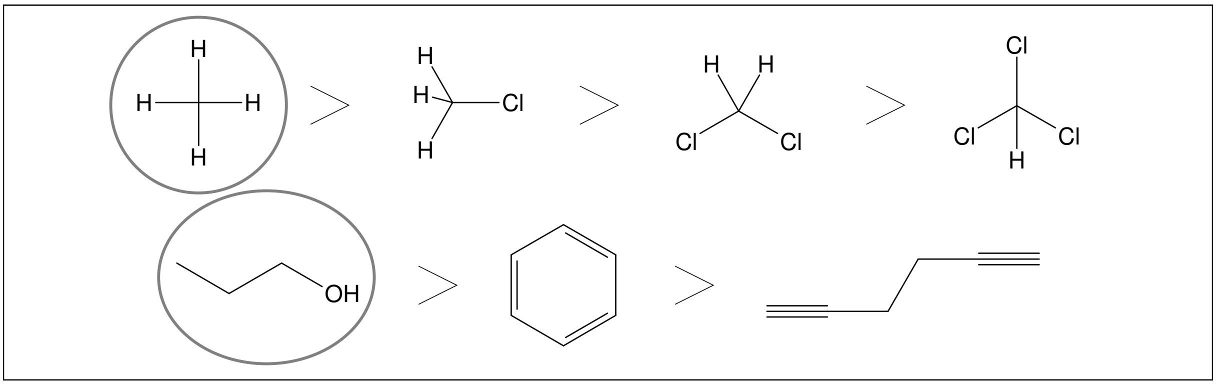


Abb. 3 - 5 Auswahl des „größten" Moleküls nach Gleichung (3-1). Oben: Methan ist *größer* als die Chlorderivate. Unten: Propanol wird als *größer* betrachtet als Benzol, aber auch als Hexadiin(1,5). Dabei hängt die Plazierung Benzol/Hexadiin von der Reihenfolge der Moleküleingabe ab.

Im WODCA System Version 3 wird dieses Verfahren daher durch den folgenden Ansatz ersetzt:

$$\text{Größe(Mol)} = 100.000 \cdot n_{atom}(\text{Mol}) + 10 \cdot M(\text{Mol}) \qquad (3\text{-}2)$$

Hier wird bei gleicher Anzahl an Atomen das schwerere Molekül als *größer* betrachtet. Sollte auch dann noch keine Entscheidung möglich sein, werden die Hashcodes zur Entscheidung herangezogen: Das Molekül mit dem größten Hashcode wird ausgewählt. Für die Beispiele in Abbildung 3 - 5 heißt das: Chloroform bzw. Hexadiin sind die *größten* Moleküle.

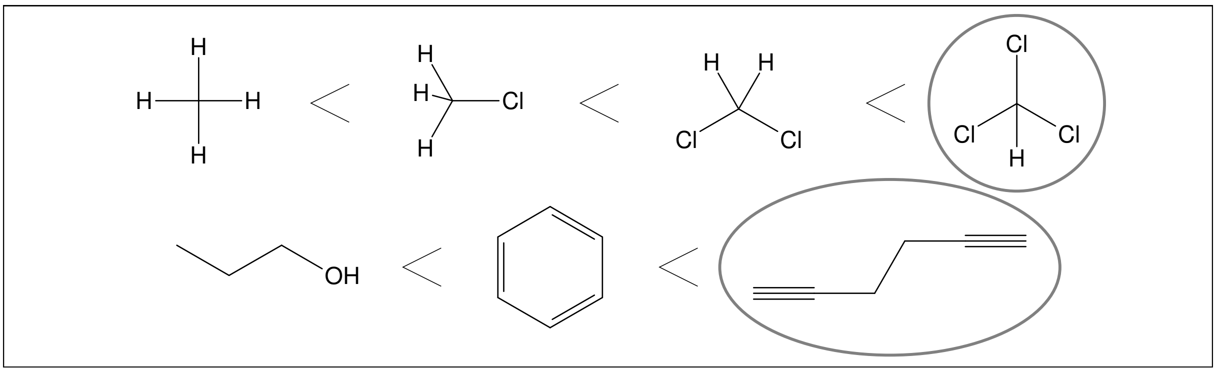


Abb. 3 - 6 Auswahl des größten Moleküls im WODCA System Version 3. Diese Entscheidung entspricht mehr der Intuition des Chemikers. Man beachte, daß die Entscheidung Benzol/Hexadiin erst durch die Hashcodes der beiden Moleküle bestimmt wird.

Nach wie vor werden Phenylringe damit tendenziell abgewertet. Allerdings werden Phenylringe jetzt erst durch aliphatische C_4-Einheiten als *größte* Moleküle verdrängt, mit Gleichung (3-1) waren dazu bereits C_3-Einheiten ausreichend.

Weiterhin werden Verbindungen, wenn sie einer Transformationsregel unterworfen werden, normalisiert. Das bedeutet, daß Tautomere in *ein* definiertes Isomer überführt werden, Ladungsstellen beseitigt werden. Dieses Vorgehen soll absichern, daß verschieden codierte, aber offensichtlich identische Verbindungen, auch als identisch erkannt werden (siehe Abb. 3 - 7).

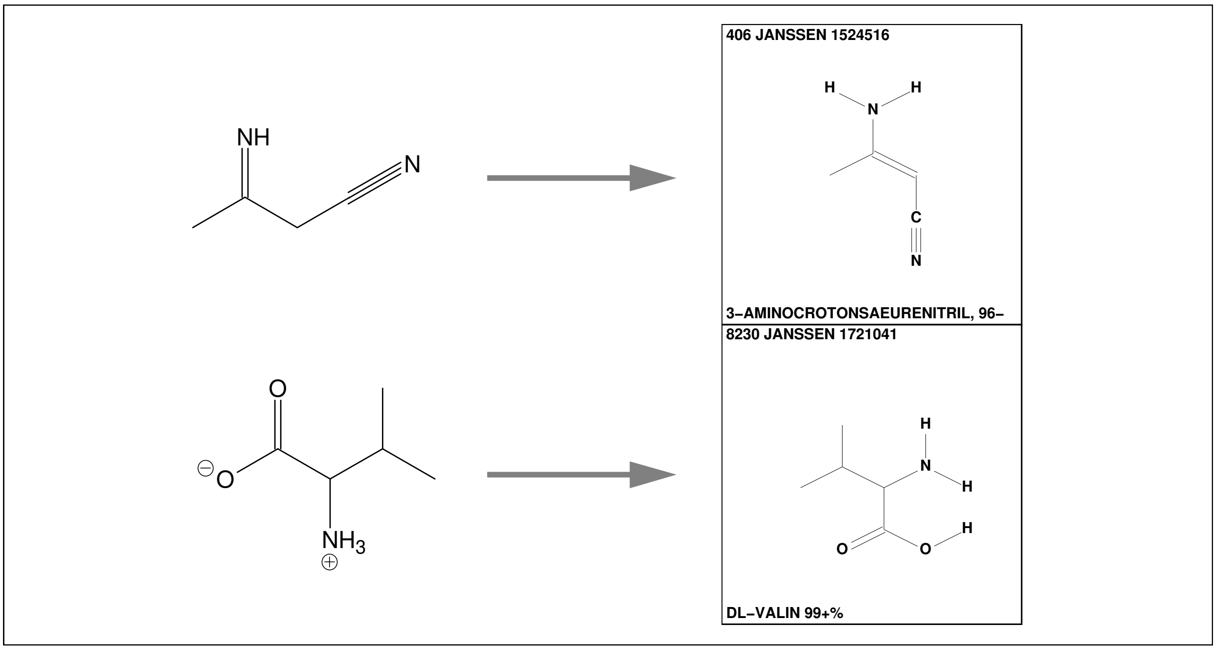


Abb. 3 - 7 Beispiele für die Nützlichkeit der Normalisierung: Oben: Die als Imid formulierte Anfrage findet die als Aminocrotonsäurenitril codierte tautomere Katalogverbindung. Unten: Die mit Ladungen codierte Aminosäure findet ihr ladungsfreies Pendant im Katalog.

Die für das WODCA System Version 3 erarbeiteten Transformationsregeln, die nunmehr als *tcl*-Skripts codiert sind, orientieren sich an den Erfahrungen, die mit Transformationen des alten WODCA Systems gemacht wurden. Einige Transformationsregeln wurden ersatzlos gestrichen, weil deren Nützlichkeit zweifelhaft oder aber ihr Unterschied zu anderen Regeln marginal ist. Dafür wurden einige bisher nicht verfügbare Transformationsregeln codiert, die sich im routinemäßigen Gebrauch als sehr effektiv erweisen. Die Algorithmen der übernommenen Transformationsregeln wurden gründlich überarbeitet. Sie unterscheiden sich im Einzelfall deutlich von ihren *Fortran77*-Vorgängern. Das folgt schon allein aus der Tatsache, daß sich die Möglichkeiten der codierenden Sprachen (*Fortran77* bzw. *tcl*) wesentlich unterscheiden. Die folgende Tabelle stellt alle im alten WODCA System vorhandenen sowie die nun in der Version 3 verfügbaren Transformationsregeln zusammen.

Tab. 3 - 8 Gegenüberstellung der *Fortran77*- und *tcl*-basierten Transformationsregeln

Name	*F77*	*tcl*	Klasse [a]
Identität	✘	✘	S
Identität unter Beachtung von Stereochemie und Isotopenmustern	✘	✘	S
Größtes Molekül	✘	✘	S
Elementaustausch innerhalb einer Spalte des Periodensystems	✘	—	S
Elementaustausch (wie vorher) und Austausch von ZH_n-Gruppen	✘	✘	S/R
C-Skelett und Elementaustausch	✘	—	S/R
Größtes Ringsystem	✘	✘	S
Substitutionspositionen auf dem reduzierten Ringsystem	✘	✘	S
Substitutionspositionen auf dem Ringsystem	—	✘	S
Ringe und C-Skelett	✘	✘	S
C-Skelett	✘	✘	S
Reduziertes C-Skelett	✘	✘	S
Aromatensystem einschließlich α-Atome	✘	✘	S
C-Skelett einschließlich α-Atome	✘	✘	S
C-Skelett einschließlich α-Atome einschließlich Stereochemie	✘	—	S
C-Skelett und Aromaten einschließlich α-Atome	✘	✘	S
C-Skelett und Aromaten einschließlich α-Atome einschließlich Stereochemie	✘	—	S
Entfernung terminaler Alkylgruppen	✘	—	S
Entfernung von homologierenden CH_2-Gruppen	✘	—	S
Entfernung symmetrischer Substituenten	✘	—	S
Substitutionsmuster (-AR+A1-CCMB-BO-MU)[b]	—	✘	S/R
Substitutionsmuster (-AR+A1+CCMB-BO-MU)[b]	—	✘	S/R
Substitutionsmuster (-AR+A1-CCMB-BO+MU)[b]	—	✘	S/R
Substitutionsmuster (-AR+A1+CCMB-BO+MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR+A1-CCMB+BO+MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR+A1+CCMB+BO+MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR-A1-CCMB-BO-MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR-A1+CCMB-BO-MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR-A1-CCMB-BO+MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR-A1+CCMB-BO+MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR-A1-CCMB+BO+MU)[b]	✘	✘	S/R
Substitutionsmuster (-AR-A1+CCMB+BO+MU)[b]	✘	✘	S/R
Maximale Oxidation	✘	✘	R
C-Skelett inklusive α-Atome und maximale Oxidation	✘	✘	R
Maximale Reduktion	✘	✘	R

Tab. 3 - 8 (Forts.) Gegenüberstellung der *Fortran77*- und *tcl*-basierten Transformationsregeln

Name	*F77*	*tcl*	Klasse [a]
C-Skelett inklusive α-Atome und maximale Reduktion	✘	✘	R
Maximale Reduktion (ohne Aromaten)	✘	✘	R
Maximale Reduktion (ohne C-Aromaten)	✘	✘	R
Hydrolyse ohne Aromatensubstituenten	✘	✘	R
Hydrolyse einschließlich Aromatensubstituenten	✘	✘	R
Hydrolyse ohne Aromatensubstituenten und Oxidation	✘	✘	R
Hydrolyse einschließlich Aromatensubstituenten und Oxidation	✘	✘	R
Eliminierung	✘	✘	R
Eliminierung und Substitutionsmuster (-AR-A1+CCMB-BO-MU)[b]	✘	—	R
Eliminierung und Substitutionsmuster (-AR+A1+CCMB-BO-MU)[b]	—	✘	R
C-Skelett einschließlich Aromaten mit α-Atomen, dann Eliminierung	✘	✘	R
C-Skelett einschließlich Aromaten mit α-Atomen, dann Reduktion, dann Eliminierung	✘	—	R
Eliminierung, dann Reduktion	✘	—	R
Eliminierung, dann Reduktion ohne Aromaten	✘	—	R
Eliminierung, dann Reduktion ohne C-Aromaten	✘	—	R
Eliminierung, dann Reduktion ohne Aromaten, dann Substitutionsmuster (-AR-A1-CCMB-BO-MU)	✘	—	R
Ozonolyse	✘	✘	R
Ozonolyse und Substitutionsmuster (-AR-A1-CCMB-BO-MU)[b]	✘	—	R
Ozonolyse und Substitutionsmuster (-AR+A1-CCMB-BO-MU)[b]	—	✘	R
Ozonolyse, dann Hydrolyse, dann Oxidation	✘	—	R
Ozonolyse, dann C-Skelett, dann Reduktion	✘	—	R

a. S: substruktur-orientiert; R: reaktionsklassen-orientiert
b. Zur Bedeutung der Kürzel siehe Abschnitt 3.3.4.

Zwei Klassen von Transformationen können unterschieden werden:

- substruktur-orientierte Transformationen
- reaktionsklassen-orientierte Transformationen

Unter substruktur-orientierten Transformationen kann man all jene Ähnlichkeitskriterien verstehen, die mittels klassischer Substruktursuchen ebenfalls erzielt werden können. Jedoch befreit eine substruktur-orientierte Transformation den Anwender von der Spezifikation einer Substruktur. Eine Originalstruktur (d. h. eine chemisch korrekt codierte Verbindung ohne freie Valenzen oder Listen von Atomarten an bestimmten Positionen) wird direkt für die Suche verwendet. Die Transformationsregel definiert die Substruktur. Außerdem ist die Transformationssuche in der Regel* schneller als eine Substruktursuche: Im ersten Fall werden nach der automatischen Manipulation der Anfrageverbindung nur Hashcodes verglichen. Im letzten Fall

* Bestenfalls liegen Substruktursuche und Transformationssuche gleichauf. Das ist besonders dann zu erwarten, wenn das Screening vor der eigentlichen Substruktursuche die in Frage kommende Zahl möglicher Treffer stark reduziert.

muß ein intelligentes Screening die tatsächlich Atom für Atom auf die Anfrage abzubildenden Katalogverbindungen drastisch reduzieren. Dieses Mapping ist bereits für einige Dutzend Paare aufwendiger als der Vergleich von Tausenden von Hashcodes.

Die an Reaktionsklassen orientierten Transformationen erweisen sich in der Praxis als noch leistungsfähiger für die computer-gestützte Syntheseplanung.* Es werden mit diesen Transformationen Ergebnisse erzielt, die mit herkömmlichen Substruktursuchen in dieser Selektivität und Vollständigkeit nicht oder aber nur mittels mehrerer Substruktursuchen erreichbar sind. Desweiteren werden gerade für die Syntheseplanung interessante Relationen zu Ausgangsverbindungen gefunden, die dann auch für eine Synthese einsetzbar sind.

Es sei an dieser Stelle daran erinnert, daß durch die neue Implementation mittels einer Skriptsprache, jederzeit auf einfache Art und Weise weitere Transformationsregeln codiert oder aber vorhandene ausgetauscht werden können. Auch sind diese Transformationsregeln mit der Ähnlichkeit die sie definieren, an der Syntheseplanung orientiert. Andere Sätze von Transformationsregeln, für Ähnlichkeitssuchen in anderen Anwendungsbereichen (z. B. Pharmakologie) sind leicht vorstellbar.

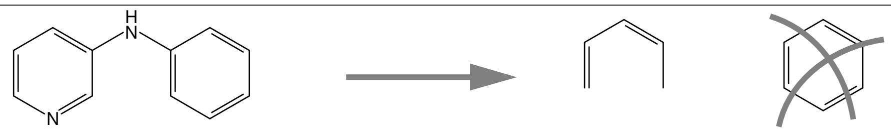
* Im Einzelfall kann die Einordnung einer Transformation in die eine oder andere Gruppe strittig sein.

3.3.3 Transformationen zur Erkennung skeletaler Ähnlichkeit

Dieser Abschnitt beschreibt diejenigen Ähnlichkeitskriterien, die Relationen zwischen Anfrageverbindungen und Katalogverbindungen auf der Basis ihrer Kohlenstoffskelette oder anderer genereller topologischer Merkmale erkennen (Übersicht 3 - 1). Die Verteilung von Substituenten bzw. die Anordnung von funktionellen Gruppen spielt hierbei - wenn überhaupt - eine untergeordnete Rolle.

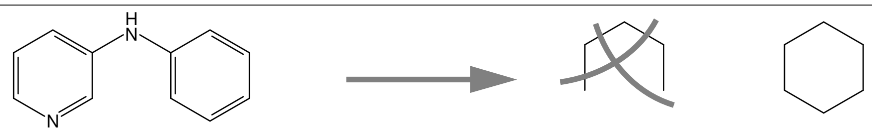
Übersicht 3 - 1 Ähnlichkeitskriterien zur Erkennung skeletaler Verwandtschaften

C-Skelett



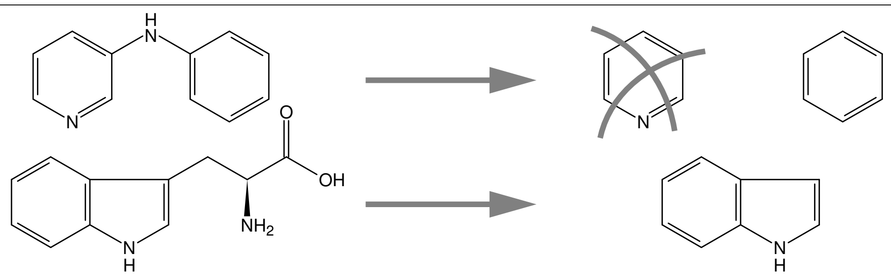
Alle Kohlenstoff-Heteroatom-Bindungen[a] werden gespalten, unabhängig vom Bindungsgrad (Heterocyclen werden also geöffnet!), Kohlenstoff-Kohlenstoff-Bindungsordnungen[b] werden nicht verändert, freie Valenzen mit Wasserstoff gesättigt. Nur das größte Fragment bleibt erhalten.

Reduziertes C-Skelett



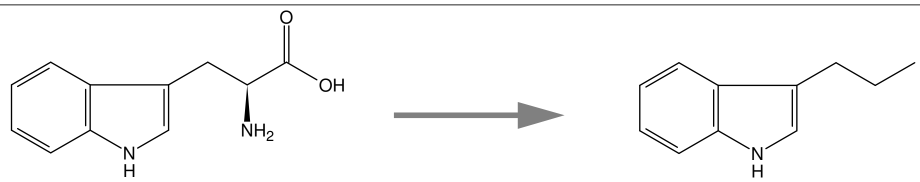
Alle CX-Bindungen werden gespalten, unabhängig vom Bindungsgrad (Heterocyclen werden also geöffnet!), CC-Bindungsordnungen werden zu Einfachbindungen reduziert, freie Valenzen mit Wasserstoff gesättigt. Nur das größte Fragment bleibt erhalten.

Größtes Ringsystem



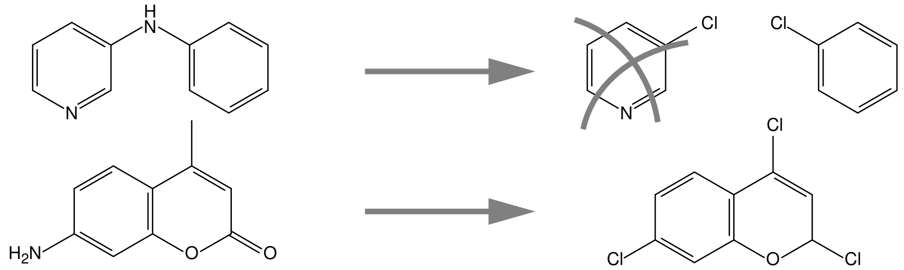
Alle Bindungen von Ringatomen zu acyclischen Zentren werden gespalten, unabhängig von Bindungsgrad oder chemischer Natur. Fragmente die dann keine Ringatome mehr enthalten werden verworfen. Für die verbliebenen Fragmente werden nun auch alle Bindungen, die keine Ringbindungen sind, gespalten. Freie Valenzen werden dann mit Wasserstoff gesättigt und nur das größte Fragment bleibt erhalten.

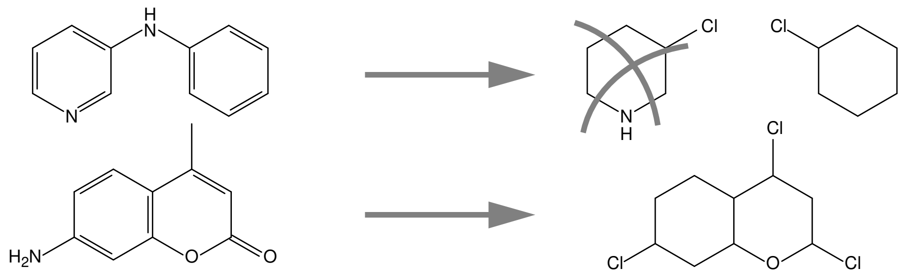
Ringe und C-Skelett

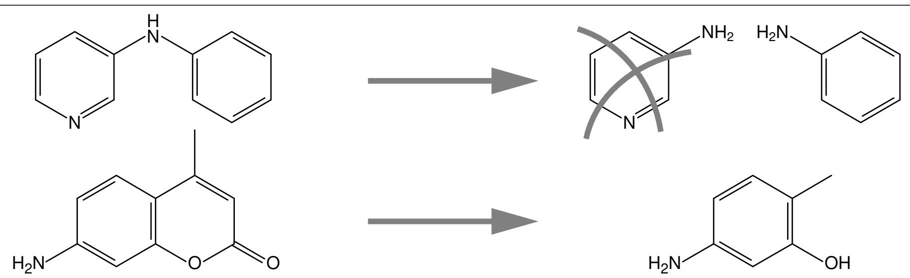


CX-Bindungen werden gespalten. Ausgenommen werden jene, bei denen mindestens das Heteroatom ein Ringatom ist. Freie Valenzen werden mit Wasserstoff gesättigt. Nur das größte Fragment bleibt erhalten.

Übersicht 3 - 1 (Forts.) Ähnlichkeitskriterien zur Erkennung skeletaler Verwandtschaften

Substitutionsmuster auf dem Ringsystem

Für alle Ringatome wird ermittelt, ob eine Bindung zu einem Nicht-Ringatom existiert, das ein Schweratom ist. Danach werden alle exocyclischen Bindungen gebrochen. Fragmente die dann keine Ringatome mehr enthalten werden verworfen. Anschließend wird an jedem Ringatom, das ursprünglich mindestens einen Schweratom-Substituenten hatte, genau ein Chloratom als Markeratom addiert. Nur das größte Fragment bleibt erhalten.

Substitutionsmuster auf dem reduzierten Ringsystem

Wie vorangehendes Kriterium. Allerdings werden hier Mehrfachbindungen in Ringen zu Einfachbindungen reduziert.

Aromatensystem einschließlich α-Atome

Zunächst werden alle aromatischen Atome ermittelt. Für jedes dieser Atome wird dann geprüft, ob es einen exocyclischen Schweratom-Nachbarn hat. Dann werden alle Bindungen von aromatischen Atomen zu nicht-aromatischen Zentren gebrochen. Anschließend all jene Fragmente verworfen, die kein Aromatenatom mehr enthalten. Schließlich werden die ursprünglich vorhandenen exocyclischen Schweratom-Nachbarn wieder addiert. Noch freie Valenzen werden dann mit Wasserstoff gesättigt und nur das größte Fragment bleibt erhalten.

a. Im folgenden immer CX-Bindung usw.
b. Im folgenden immer CC-Bindung/Bindungsordnung usw.

Die voranstehende Tabelle veranschaulicht die Wirkungsweise der Ähnlichkeitskriterien. Die letzten drei Transformationen sind Grenzfälle: Neben dem eigentlichen topologischen Merkmal (aromatisches System/Ringsystem) werden *auch* Substituenteninformationen (in sehr allgemeiner Form bzw. beschränkt auf das aromatische System) einbezogen. Die Suche nach Verbindungen mit einem bestimmten Ringsystem, u. U. ohne Berücksichtigung von Bindungsordnungen in Ringen, mit bestimmten Substitutions*positionen** ist zwar ein wesentlich anspruchsvolleres Ähnlichkeitskriterium als das Ringsystem allein, jedoch immer noch wesentlich allgemeiner als die Kriterien des Abschnitts 3.3.4. In der Tendenz liegt „Aromatensystem einschließlich α-Atome“ dagegen schon sehr nahe im Bereich des Ähnlichkeitsbegriffs der substitutions-orientierten Transformationssuchen.

Es wäre ohne Schwierigkeiten möglich, ein Ähnlichkeitskriterium „Aromatensystem“ zu definieren. Für die praktische Anwendung im Bereich der Syntheseplanung ist es aber in der Regel nicht interessant, etwa für ein Cumarin alle einen Phenylring als größtes Aromatensystem enthaltende Ausgangsverbindungen zu finden. Vielmehr ist das gleichzeitige Vorhandensein des heterocyclischen Ringes oder des entsprechenden C-Skelettes in einem Ausgangsmaterial erwünscht. Diese Bedingungen werden aber durch „Größtes Ringsystem“ bzw. „Ringe und C-Skelett“ erfaßt.

Inwieweit allein die Kriterien „C-Skelett“ bzw. „Reduziertes C-Skelett“ die in Frage kommende Zahl an Ausgangsmaterialien einschränken, zeigen folgende Zahlen†:

Es gibt drei acyclische Pentan-Gerüste.

- 284 bzw. 526 Ausgangsmaterialien in den Katalogen Janssen bzw. Fluka haben das gleiche „C-Skelett“ wie *n*-Pentan. Werden auch CC-Mehrfachbindungen erlaubt („Reduziertes C-Skelett“) findet man 451 bzw. 674 Verbindungen. Das sind 3-5 % der Gesamtheit aller Katalogeinträge.
- Für 2-Methyl-butan (*i*-Pentan) sind es nur noch 65 bzw. 105 Ausgangsmaterialien mit gleichem „C-Skelett“ und 92 bzw. 144 mit Tolerierung von CC-Mehrfachbindungen. Das sind um 1 % der Katalogeinträge.
- Noch seltener sind Gerüste von 2,2-Dimethyl-propan (*neo*-Pentan). Nur 32 bzw. 42 Ausgangsmaterialien haben das gleiche „C-Skelett“ im Janssen- bzw. Fluka-Katalog. 30 bzw. 37 sind es im Falle der Tolerierung von CC-Mehrfachbindungen‡. Das sind nur weniger als ein halbes Prozent aller Katalogeinträge.

Im Fall des nächst größeren, acyclischen C-Skeletts hat man fünf Hexane zu betrachten:

- Für *n*-Hexan findet man 176 bzw. 454 Verbindungen im Janssen- bzw. Fluka-Katalog. Läßt man auch CC-Mehrfachbindungen zu sind es 247 bzw. 492 Ausgangsmaterialien. (unter 3 %)
- Bei 2-Methyp-pentan (*i*-Hexan) sind es 47 bzw. 75 mit gleichem „C-Skelett“ oder 89 bzw. 125 mit Tolerierung von CC-Mehrfachbindungen. (unter 1 %)
- Das „C-Skelett“ von 3-Methylpentan findet sich in 31 bzw. 86 Katalogverbindungen. 63 bzw. 118 sind es, wenn CC-Mehrfachbindungen erlaubt werden. (weniger als 0,75 %)
- Beim 2,2-Dimethyl-butan (*neo*-Hexan) sinken die Zahlen auf 10 bzw. 48 („C-Skelett“) und 12 bzw. 53 („Reduziertes C-Skelett“). (weniger als 0,33 %)
- 2,3-Dimethylbutan schließlich ist am seltensten: 8 bzw. 11 mal ist dieses „C-Skelett“ im Janssen- bzw. Fluka-Katalog vertreten, 18 bzw. 19 Verbindungen sind es, wenn CC-Mehrfachbindungen gestattet werden. (um 0,1 %)

* Man beachte, daß Hetero- und C-Substituenten hier völlig gleichwertig behandelt werden. Synthetisch gesehen wird man jedoch kaum eine C-Seitenkette gegen ein Heteroatom austauschen.

† Für eine tabellarische Aufbereitung siehe Tab. 5 - 2 bzw. 5 - 4 im Abschnitt A.

‡ Im *neo*-Pentan selbst kann es keine CC-Mehrfachbindungen geben. Die unterschiedlichen Zahlen resultieren daher, daß am Ende der Transformation andere Skelettfragmente als größer oder kleiner als ein *neo*-Pentan-Gerüst bewertet werden.

Daraus folgen diese Anwendungshinweise:

⇨ Allgemein wächst die Selektivität bezüglich des C-Skelettes in folgender Reihenfolge: „Reduziertes C-Skelett“ < „C-Skelett“ < „Ringe und C-Skelett“ (acyclische Skelette) bzw. „Größtes Ringsystem“ < „Ringe und C-Skelett“ (Skelette mit Ringen).

⇨ Acyclische Kohlenwasserstoffskelette: Je größer die C-Einheit und je stärker verzweigt das Skelett ist, desto selektiver werden schon einfache, toplogisch orientierte, allgemeine Ähnlichkeitskriterien. Gleichzeitig wächst damit ihre Bedeutung als Suchmethode.

3.3.4 Substitutionsmuster als selektivere Ähnlichkeitskriterien

Wie bereits im Abschnitt 3.3.2 (vgl. Abb. 3 - 4) erwähnt, ist das Substitutionsmuster eines Syntheseziels oder -bausteins ein bestens geeignetes Ähnlichkeitskriterium, um verwandte Ausgangsverbindungen aufzufinden. Jedoch muß man sich zunächst vergegenwärtigen, was unter dem Begriff Substitutionsmuster verstanden werden soll. Bindungen zwischen Skelett- und Heteroatomen werden zweifellos ein Substitutionsmuster bestimmen. Aber *ein* Skelettatom (C-Atom) kann *ein oder mehrere* Heterosubstituenten haben. Solche Heteroatome können über Einfach-, Doppel- oder Dreifachbindungen mit dem Skelett verbunden sein. Schließlich stellt sich die Frage, ob ein Heteroatom in einem aromatischen Heterocyclus einen Substituent für die benachbarten C-Atome darstellt oder es selbst besser als Skelettbestandteil betrachtet werden soll. Eine weitere Besonderheit sind (nichtaromatische) CC-Mehrfachbindungen. Diese könnten als funktionelle Gruppe (d. h. Substituentenäquivalent) oder ausschließlich als Skelettkomponente verstanden werden. Für diese Arbeit wird <u>Substitutionsmuster</u> daher wie folgt definiert:

Unter dem Substitutionsmuster einer Verbindung wird deren Skelett verstanden, auf dem die vorhandene chemische Funktionalität durch Markeratome* generalisiert wurde.

Was im Einzelnen unter chemischer Funktionalität verstanden werden soll, kann stark variieren (siehe oben) und ist daher im WODCA System durch Schalter steuerbar. Diese Schalter faßt die folgende Tabelle zusammen.

Tab. 3 - 9 Möglichkeiten der Erweiterung oder Einschränkung chemischer Funktionalität, die mit einem Substitutionsmuster erfaßt werden soll.

Merkmal	zutreffend	nicht zutreffend	Abkürzung
Heteroatome in aromatischen Systemen zählen als Substituenten	*+AR*	*-AR*	AR: -AR ← → +AR
zu Aromaten α-ständige Heteroatome zählen als Substituenten	*+A1*	*-A1*	A1: -A1 ← → +A1
(nichtaromatische) Mehrfachbindungen zwischen C-Atomen zählen als Substituenten	*+CCMB*	*-CCMB*	CCMB: +CCMB ← → -CCMB
Bindungsordnungen zu Substituenten zählen	*+BO*	*-BO*	BO: +BO ← → -BO
Mehrfachsubstitution zählt	*+MU*	*-MU*	MU: +MU ← → -MU
Ähnlichkeitsbereich			eng — breit
zu erwartende Trefferzahl			kleiner — größer
zu erwartende Zahl der erforderlichen Reaktionen vom gefundenen Ausgangsmaterial zur Anfrageverbindung			kleiner — größer

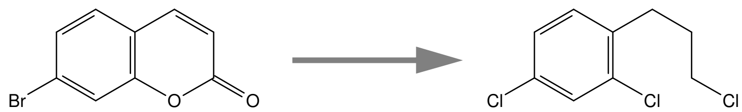
* Als Markeratome werden von WODCA Chloratome verwendet.

Während die ersten drei Schalter (aromatische Heteroatome zählen als Substituenten oder nicht [*AR*], α-Heteroatome werden generalisiert oder nicht [*A1*], CC-Mehrfachbindungen sind Substituenten [*CCMB*]) beeinflussen, *wo* Substituenten auf einem gegebenen Skelett zu finden sind, bewirken die letzten beiden Schalter nur *wie* vorhandene Substituenten im Substitutionsmuster erfaßt werden. Weiter oben in der Tabelle stehende Schalter haben dabei Auswirkungen auf weiter unten aufgeführte Bedingungen (siehe Übersicht 3 - 2). Auf diese Weise entsteht eine Hierarchie von Substitutionsmustern.

Die verschiedenen Schalter werden benutzt, um unterschiedliche Ähnlichkeitskriterien zu definieren, die alle auf das Substitutionsmuster abzielen. Diese Kriterien beschreibt Übersicht 3 - 2.

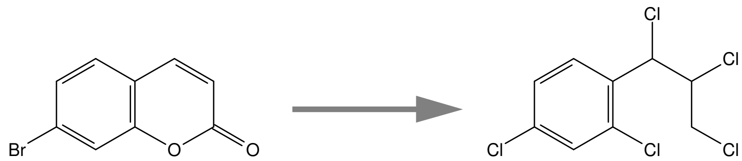
Übersicht 3 - 2 Ähnlichkeitskriterien basierend auf dem Substitutionsmuster von Verbindungen

Substitutionsmuster -AR+A1-CCMB-BO-MU (1)



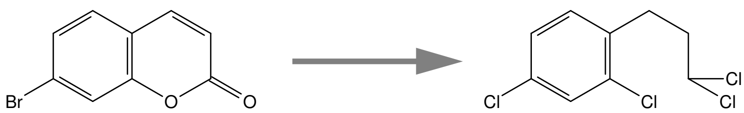
Alle CX-Bindungen werden gebrochen, ausgenommen sind aromatische CX-Bindungen. An dem so extrahierten Skelett wird danach überall dort, wo ursprünglich mindestens ein Substituent (ganz gleich welcher Bindungsordnung) vorhanden war, ein Chloratom addiert. Nichtaromatische CC-Mehrfachbindungen werden zu Einfachbindungen reduziert. Noch freie Valenzen werden mit Wasserstoffen abgesättigt. Nur das größte Fragment bleibt erhalten.

Substitutionsmuster -AR+A1**+CCMB**-BO-MU (2)



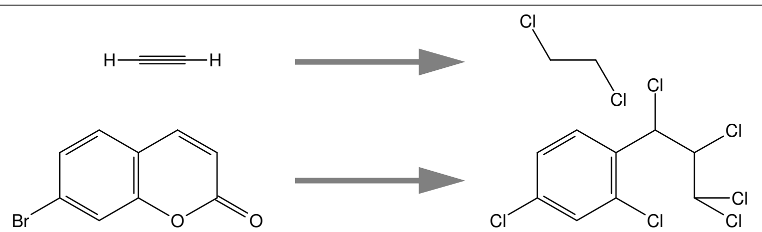
Wie (1). Nichtaromatische CC-Mehrfachbindungen werden jedoch durch Addition je eines Chloratoms je Bindungsatom aufgelöst.

Substitutionsmuster -AR+A1-CCMB-BO**+MU** (3)



Wie (1), jedoch werden je Skelettatom so viele Chloratome addiert, wie ursprünglich Substituenten vorhanden waren.

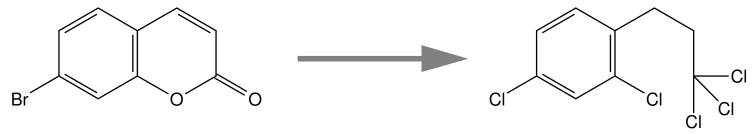
Substitutionsmuster -AR+A1**+CCMB**-BO**+MU** (4)



Wie (2), nichtaromatische CC-Mehrfachbindungen werden durch Addition von Chloratomen (je eines pro Bindungsatom und Bindungsgrad über Eins) reduziert. An Skelettatomen werden so viele Chloratome addiert, wie ursprünglich Substituenten vorhanden waren.

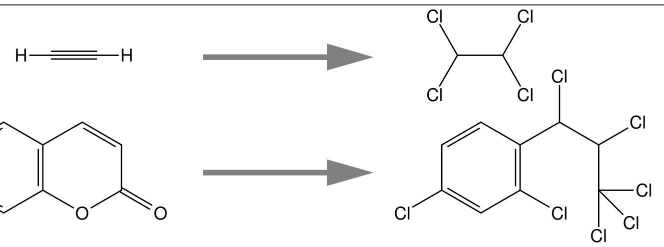
Übersicht 3 - 2 (Forts.) Ähnlichkeitskriterien basierend auf dem Substitutionsmuster von Verbindungen

Substitutionsmuster -AR+A1-CCMB**+BO+MU** (5)



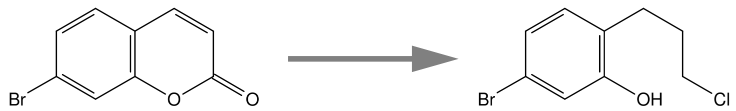
Wie (3). Für mehrfach gebundene Substituenten werden jedoch eine dem Bindungsgrad entsprechende Anzahl Chloratome vermerkt.

Substitutionsmuster -AR+A1**+CCMB+BO+MU** (6)



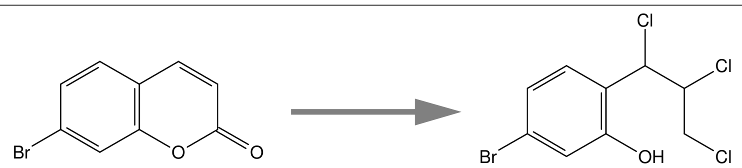
Wie (5). Für mehrfach gebundene Substituenten werden jedoch eine dem Bindungsgrad entsprechende Anzahl Chloratome vermerkt.

Substitutionsmuster -AR**-A1**-CCMB-BO-MU (7)



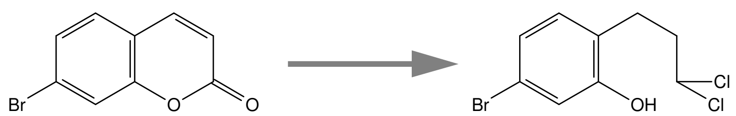
Wie (1). Im Unterschied werden jedoch Bindungen von einem aromatischen Atom zu einem nichtaromatischen Heteroatom unverändert übernommen.

Substitutionsmuster -AR**-A1**+CCMB-BO-MU (8)



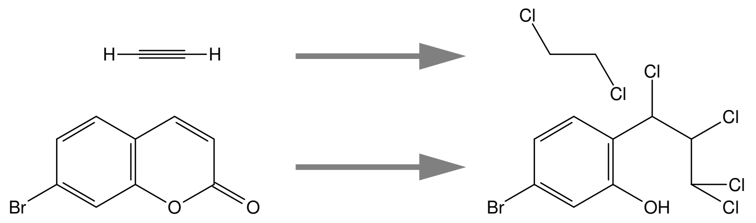
Wie (2). Im Unterschied werden jedoch Bindungen von einem aromatischen Atom zu einem nichtaromatischen Heteroatom unverändert übernommen.

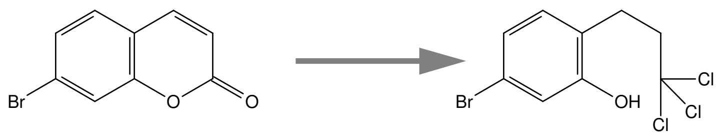
Substitutionsmuster -AR**-A1**-CCMB-BO+MU (9)

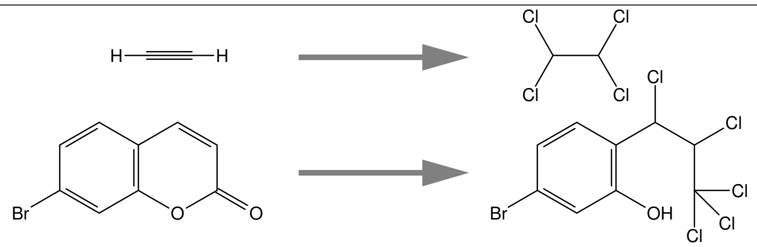


Wie (3). Im Unterschied werden jedoch Bindungen von einem aromatischen Atom zu einem nichtaromatischen Heteroatom unverändert übernommen.

Übersicht 3 - 2 (Forts.) Ähnlichkeitskriterien basierend auf dem Substitutionsmuster von Verbindungen

Substitutionsmuster -AR**<u>-A1</u>**+CCMB-BO+MU	(10)
	
Wie (4). Im Unterschied werden jedoch Bindungen von einem aromatischen Atom zu einem nichtaromatischen Heteroatom unverändert übernommen.	

Substitutionsmuster -AR**<u>-A1</u>**-CCMB+BO+MU	(11)
	
Wie (5). Im Unterschied werden jedoch Bindungen von einem aromatischen Atom zu einem nichtaromatischen Heteroatom unverändert übernommen.	

Substitutionsmuster -AR**<u>-A1</u>**+CCMB+BO+MU	(12)
	
Wie (2). Im Unterschied werden jedoch Bindungen von einem aromatischen Atom zu einem nichtaromatischen Heteroatom unverändert übernommen.	

Der untere Teil von Tabelle 3 - 9 veranschaulicht, wie es sich auf eine Ähnlichkeitssuche auswirkt, wenn ein Schalter benutzt (+) oder ignoriert (–) wird: Müssen mehr oder spezifischere Kriterien erfüllt werden, wird ein restriktiverer Ähnlichkeitsbereich definiert. Sind die Kriterien zwangloser, wird der Ähnlichkeitsbereich umfassender. Mit dem Ähnlichkeitsbereich parallel laufen die zu erwartenden Trefferzahlen bzw. die mögliche Zahl an Reaktionen, die eine gefundene Verbindung in die Anfrageverbindung umwandeln: großzügigerer Begriff von Ähnlichkeit ≡ höhere mögliche Trefferzahl ≡ möglicherweise höhere Zahl an Umsetzungen vom Ausgangsmaterial zur Anfrageverbindung.

Eine Anmerkung zu dem Merkmal *Mehrfachbindungen zwischen C-Atomen* (*CCMB*). In der Regel führen Substitutionstransformationen ohne Zählung der CC-Mehrfachbindungen als Substituenten (bei sonst gleichen Parametern) zu *höheren* Trefferzahlen. Das ist einleuchtend, wenn man beispielsweise ein langkettiges Alken sucht, das eine Doppelbindung in 1-Position und einen Heterosubstituenten in ω-Position besitzt. Ohne Beachtung der Doppelbindung wird man eine Reihe von einfach-endständig substituierten Alkanen finden können , mit Beachtung der Doppelbindung u. U. (das Substitutionsmuster entspricht dann einem 1,2,ω-substituierten Alkan) gar keine Verbindung. 6-Hydroxy-hex-1-en (**3**) als Beispiel (siehe Abb. 3 - 8) führt zu 39 Treffern ohne Dubletten im Fluka-Katalog *ohne* Beachtung von CC-Mehrfachbindungen. Gefordert wurde außerdem gleiche Bindungsordnung (+*BO*) und gleiche Anzahl an Substituenten (+*MU*) wie in der Anfrage. *Mit* Beachtung von CC-Mehrfachbindungen resultieren nur vier Treffer. Ist die Kettenlänge aber gering, wie beispielsweise im Allylchlorid (**4**), findet man mit Beachtung der Doppelbindung zusätzlich zu allylisch substituierten Verbindungen *auch* 1,2,3-substituierte Verbindungen wie Glycerin. Dadurch *können* mehr Treffer zustandekommen als ohne Beachtung der Doppelbindung. Mit Allylchlorid resultieren 81 Treffer *mit*, hingegen 89 Treffer *ohne* Beachtung von CC-Mehrfachbindungen.

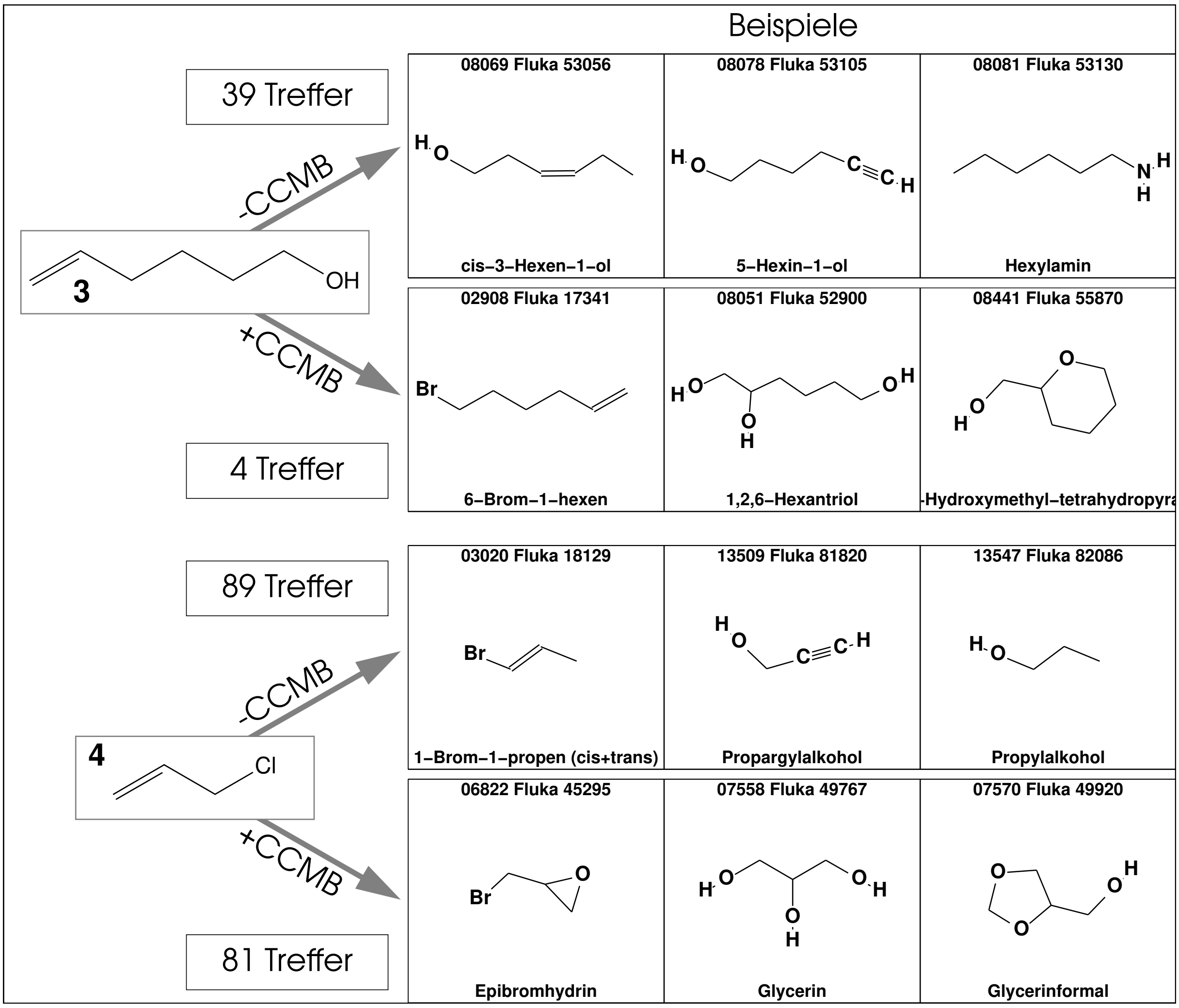


Abb. 3 - 8 Substitutionsmuster-orientierte Transformationssuchen (*-AR+A1±CCMB+BO+MU*) einmal mit (*+CCMB*) und einmal ohne (*-CCMB*) Beachtung von CC-Mehrfachbindungen. Im oberen Fall resultieren deutlich mehr Treffer ohne Beachtung von CC-Mehrfachbindungen. Mit Allylchlorid sind die Trefferzahlen praktisch gleichauf.

Schwerer abschätzbar ist die zu erwartende Zahl der Treffer bei Vergleich von zwei Substitutionstransformationen, bei denen mehr als ein Parameter variiert. In der Regel verhalten sich die Variationen additiv, d. h. wenn mit dem Schalter *-AR* weniger Treffer erwartet werden können, als mit *+AR* und mit *-A1* weniger als mit *+A1*, dann können mit *-AR-A1* weniger Treffer erwartet werden als mit allen anderen Kombinationen von *AR* und *A1*. Nicht vorhersagbar ist allerdings, ob mit *-AR+A1* oder mit *+AR-A1* mehr Treffer zu erwarten sind. Das hängt ausschließlich vom Datenmaterial in den durchsuchten Katalogen ab.

Für die Abschätzung der am meisten spezifisch suchenden bzw. der großzügigsten Substitutionstransformation folgt aus dem Vorangehenden, daß die Substitutionstransformation mit den Schaltern *-AR-A1+CCMB+BO+MU* das engste Ähnlichkeitskriterium umfaßt, die Schalter *+AR+A1-CCMB-BO-MU* das allgemeinste.*

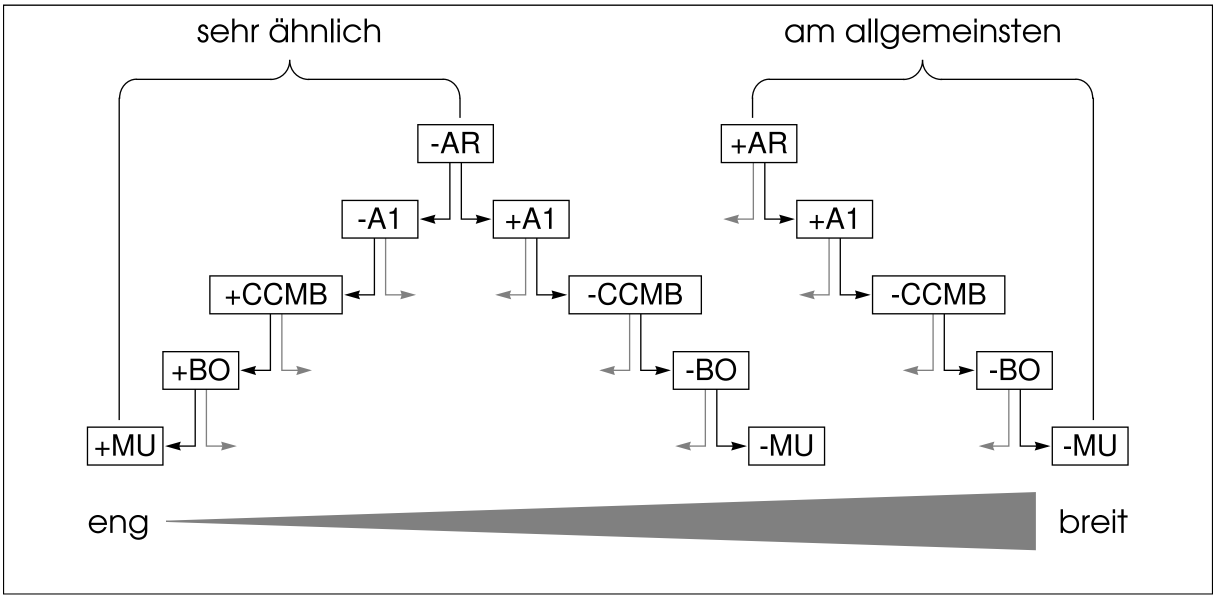
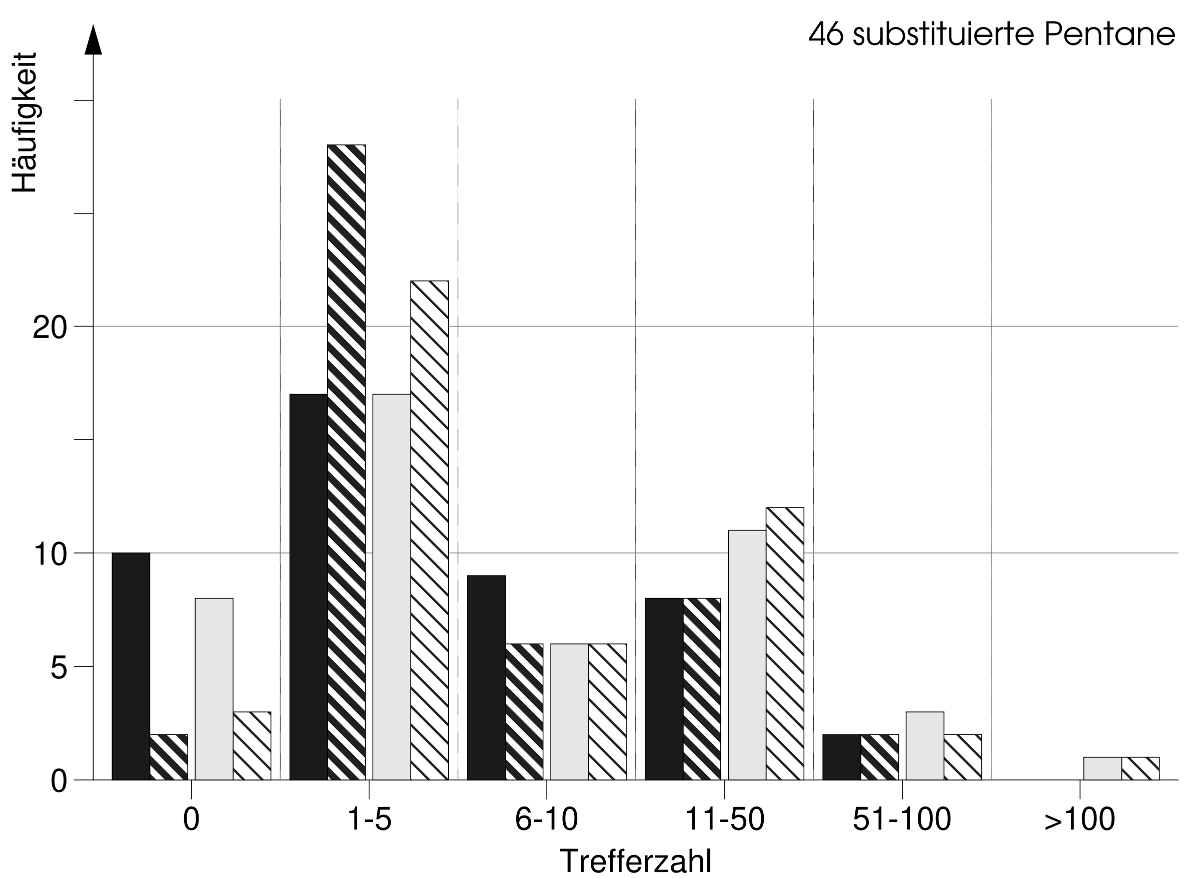


Abb. 3 - 9 Hierarchie von Substitutionstransformationen: Additive Wirkung mehrerer Schalter, welche die Art des interessierenden Substitutionsmusters beeinflussen. Grau gezeichnete Verzweigungen deuten an, daß auch eine Einstellung in dieser Richtung möglich ist. Aus Platzgründen sind jedoch nur drei Kombinationen dargestellt.

Die mit den verschiedenen Substitutionsmustern erzielbaren Abstufungen bei Ähnlichkeitssuchen soll im Folgenden an einigen Beispielen verdeutlicht werden.

Zunächst sei daran erinnert, daß auf dem Substitutionsmuster basierende Ähnlichkeitssuchen deshalb von Interesse sind, weil nur skelett-basierte Transformationen u. U. zu große Trefferlisten ergeben (siehe Abschnitt 3.3.3, Seite 49f). Diese haben eine typische Größe von einigen Hundert bis einigen Duzend Treffern. Als Beispiel dienen die acyclischen Pentan- und Hexangerüste, für die alle möglichen Substitutionsmuster generiert wurden [96]. Für das Substitutionsmuster werden hier Bindungs*ordnung*en zu Heteroatomen ignoriert, auch *Mehrfach*substitution wird nicht unterschieden. Das entspricht den Schaltern *-AR±A1±CCMB-BO-MU* des WODCA Systems. Bindungen zu α-aromatischen Heteroatomen sind hier uninteressant. CC-Mehrfachbindungen wurden jedoch einmal beachtet und einmal ignoriert. Daraus ergeben sich leicht unterschiedliche Trefferlisten. Die folgende Abbildung veranschaulicht die Größenordnung der Trefferlisten für Ähnlichkeitssuchen mit allen möglichen substituierten Pentanen (46 Verbindungen) und deren Häufigkeit bezogen auf dieses allgemeinste Substitutionsmuster. Man erkennt, daß nur in sehr seltenen Fällen Trefferlisten zwischen 50 und 100 oder gar darüber auftreten. Allerdings gibt es auch Substitutionsmuster, die überhaupt nicht in Ausgangsmaterial-Katalogen enthalten sind. Das einzige Substitutionsmuster auf C_5-Einheiten, das im Janssen-Katalog mit beiden Substitutionsmuster-Definitionen keinerlei Treffer ergibt, ist die Prototyp-Verbindung 2,4-Dichlor-2-methyl-butan. Im Falle des Fluka-Kataloges gilt entsprechendes für 1,3-Dichlor-2-methyl-butan und 1,1‘,2,3-Tetrachlor-2-methyl-butan.

* Da derzeit keine Substitutionstransformation implementiert ist, die Heteroatome in Aromaten als Substituenten betrachtet (*+AR*), ist die allgemeinste *verfügbare* Substitutionstransformation *-AR+A1-CCMB -BO-MU*.



.... Werte für die Transformationssuche „Substitutionsmuster (-AR+A1-CCMB-BO-MU)“ im Janssen-Katalog.

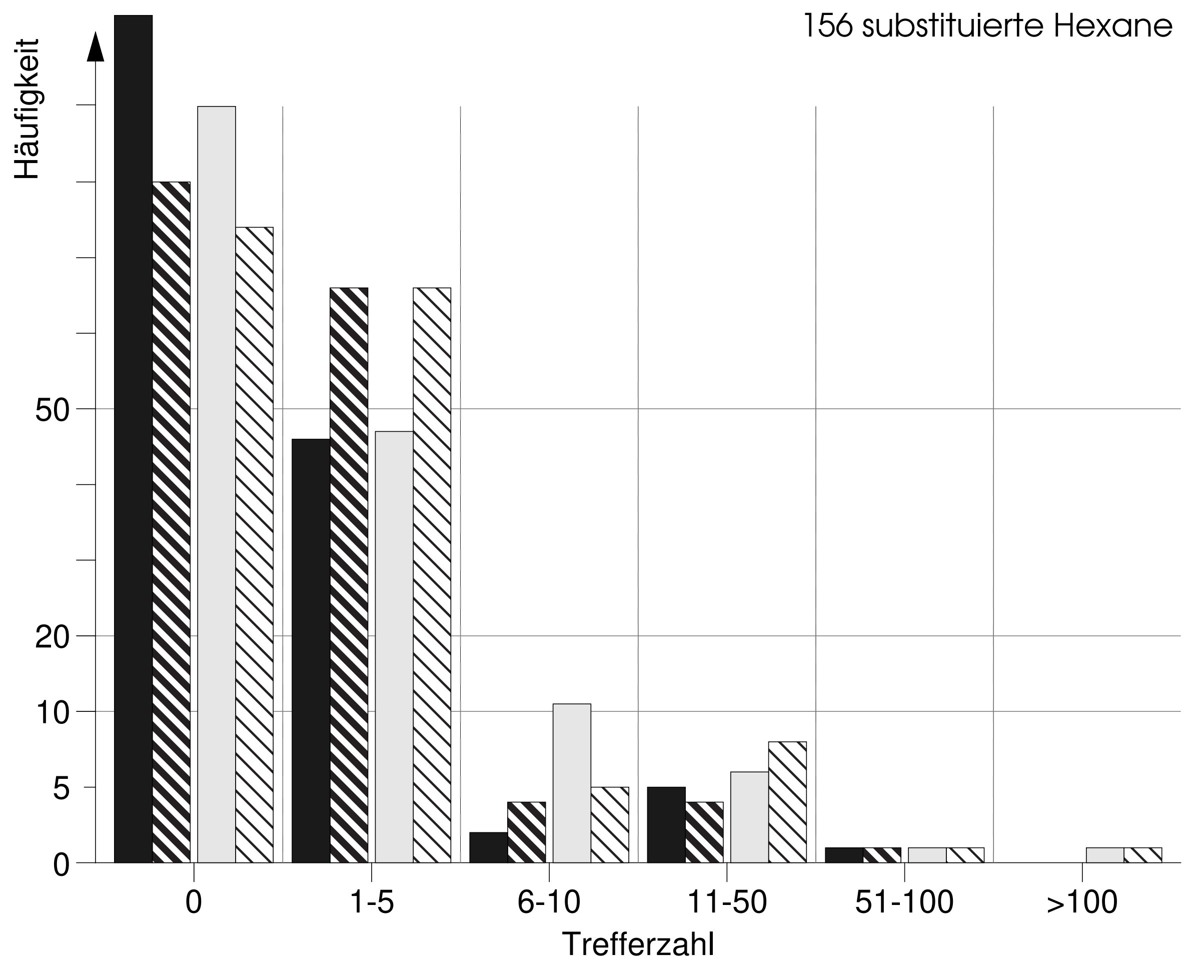
.... Werte für die Transformationssuche „Substitutionsmuster (-AR+A1+CCMB-BO-MU)“ im Janssen-Katalog.

.... Werte für die Transformationssuche „Substitutionsmuster (-AR+A1-CCMB-BO-MU)“ im Fluka-Katalog

.... Werte für die Transformationssuche „Substitutionsmuster (-AR+A1+CCMB-BO-MU)“ im Fluka-Katalog

Abb. 3 - 10 Histogramm der ermittelten Treffer für Substitutionsmustersuchen mit allen möglichen Substitutionspositionen auf acyclischen C_5-Skeletten. Bei bis zu zehn Substitutionsmustern werden keinerlei ähnliche Verbindungen gefunden (ganz links). Sehr selten (bei bis zu drei Substitutionsmustern) werden mehr als 50 oder gar 100 Treffer gefunden (rechte Spalten).

Die analoge Untersuchung für alle möglichen substituierten acyclischen Hexane (siehe Abb. 3 - 11): 156 verschiedene Substitutionsmuster gibt es hier. Deutlich wird, daß der Anteil an Substitutionsmustern, für die keinerlei Analogon im Katalog gefunden werden kann, stark angestiegen ist. Im Janssen-Katalog werden 77 Substitutionsmuster von keiner der beiden Substitutionsmuster-Definitionen gefunden, im Fluka-Katalog sind das immerhin 70. Hohe Trefferraten sind dementsprechend deutlich zurückgegangen.



Legende siehe Abb. 3 - 10

Abb. 3 - 11 Histogramm der ermittelten Treffer für Substitutionsmustersuchen mit allen möglichen Substitutionspositionen auf acyclischen C_6-Skeletten. Bei bis zu 102 und wenigstens 74 Substitutionsmustern (von 156 möglichen) werden keinerlei ähnliche Verbindungen gefunden (ganz links). Das bedeutet, daß praktisch die Hälfte der denkbaren Substitutionsmuster nicht als Ausgangsmaterial zur Verfügung steht. Extrem selten (je ein Substitutionsmuster) werden mehr als 50 oder gar 100 Treffer gefunden (rechte Spalten). Man beachte, daß die Skalierung der Häufigkeitsachse anders ist als in Abb. 3 - 10.

Fazit: Das Substitutionsmuster in seiner allgemeinsten Form ist das geeignete Mittel, um mit zu großzügigen (skelettbasierten) Ähnlichkeitskriterien ermittelte Trefferzahlen zu reduzieren. Manchmal ist die auf dem Substitutionsmuster basierende Ähnlichkeit bereits zu eng gefaßt, so daß keine verwandten Ausgangsmaterialien gefunden werden können. Im Rahmen einer Syntheseplanung wird man dann entweder auf die mit skeletalen Ähnlichkeiten gefundenen Katalogverbindungen zurückgreifen oder aber die Methoden aus Kapitel 4 anwenden.

Nachdem bislang nur mit abstrakten Trefferzahlen argumentiert wurde, soll nun anhand geeigneter Substitutionsmuster gezeigt werden, inwieweit durch Variation des Substitutions*begriffs* unterschiedliche Trefferlisten erzielt werden. In Frage kommen dazu Verbindungen mit Substitutionsmustern, für die im allgemeinsten Falle hohe Trefferzahlen resultierten.

Auf ein (oder mehrere) Beispielmolekül(e) werden alle auf dem Substitutionsmuster basierenden Transformationssuchen angewendet und in einem möglichst repräsentativen Katalog gesucht. Die so ermittelten Zahlen an Treffern dienen zur Ordnung der einzelnen Transformationssuchen nach ihrem Ähnlichkeitsbereich.

Als erstes Beispiel wird ein Aminosäurederivat gewählt: 2-Amino-2-methyl-bernsteinsäurediamid (**5**). Das C-Skelett ist eine *i*-Pentaneinheit, dreifach funktionalisiert. Eine Ähnlichkeitssuche im Fluka-Katalog basierend auf dem allgemeinen Substitutionsmuster unter Einschluß von Doppelbindungen (*-AR+A1+CCMB-BO-MU*) findet die in Abbildung 3 - 12 dargestellten fünf Treffer. (Zum Vergleich: 105 Katalogverbindungen entsprechen exakt diesem C-Skelett) Werden CC-Mehrfachbindungen nicht als Funktionalität gewertet (*-AR+A1***-CCMB***-BO-MU*), erzielt man nur vier Treffer: 3-Methyl-3-buten-1-ol (**6**) wird dann nicht gefunden. Zweifellos sind alle Katalogverbindungen in Abbildung 3 - 12 geeignete Ausgangsmaterialien für die gewählte Anfrageverbindung. Hervorzuheben ist das gefundene Oxetan-Derivat (**7**). Würde der Oxetan-Sauerstoff als Skelettatom betrachtet, wäre diese Verbindung nicht gefunden worden. Damit wird unterstrichen, daß die Definition des Substitutionsmusters ausgehend vom C-Skelett wesentlich für Anwendungen in der Syntheseplanung ist.

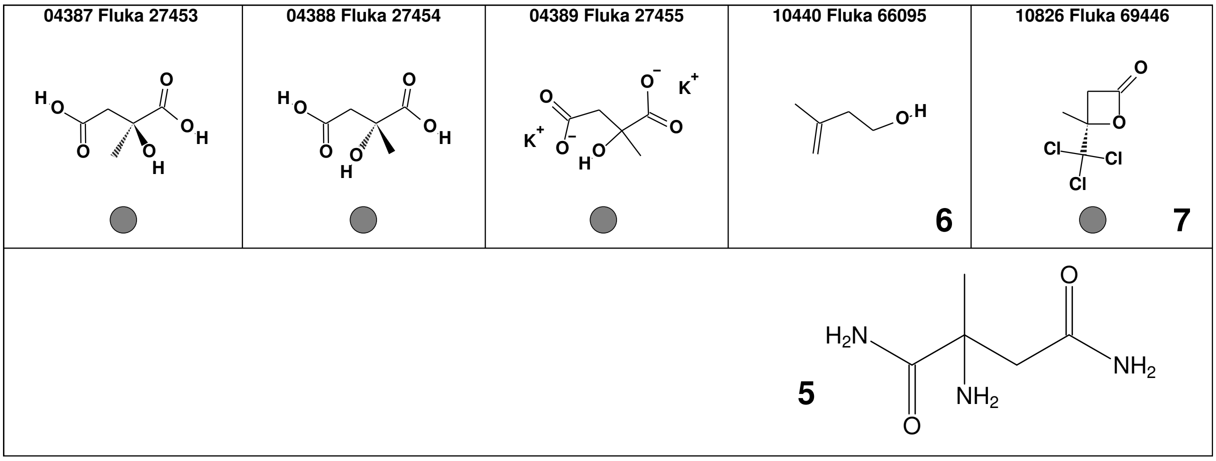


Abb. 3 - 12 Suche nach geeigneten Ausgangsmaterialien für 2-Amino-2-methyl-bernsteinsäurediamid (rechts unten) mittels substitutionsmuster-orientierter Transformationssuchen. Fünf Verbindungen findet man mit dem Kriterium „Substitutionsmuster (-AR+A1+CCMB-BO-MU)“ im Fluka-Katalog, nur vier Treffer (markiert mit ●) mit „Substitutionsmuster (-AR+A1-CCMB-BO-MU)“.

Als nächstes Beispiel dient ein häufiger zu findendes Substitutionsmuster: Isopentylchlorid (**8**). Wiederum wird im Fluka-Katalog zuerst mittels der allgemeineren Substitutionskriterien *-AR+A1-CCMB-BO-MU* bzw. *-AR+A1+CCMB-BO-MU* gesucht. Gefunden werden 21 (siehe Abb. 3 - 13) bzw. 14 Treffer (letztere markiert mit ●). Will man diese Trefferzahlen einschränken, kann man zunächst fordern, daß die Zahl der Substituenten in Anfrage und Treffer übereinstimmen soll (*+MU*). Dann werden 15 (*-AR+A1-CCMB-BO+MU* ⊘) bzw. zehn (*-AR+A1+CCMB-BO+MU* ⊘) Treffer erzielt. Damit werden Carbonsäureanaloga ausgeschlossen. Schließlich kann auch verlangt werden, daß die Bindungsordnung von Substituenten dem der Anfrage entspricht (*+BO*). In diesem Fall reduzieren sich die Trefferzahlen auf 13 (*-AR+A1-CCMB+BO+MU* ⊜) bzw. neun (*-AR+A1+CCMB+BO+MU* ⦶). Jetzt werden auch Aldehydanaloga ausgeschlossen. Die Trefferreduktion ist deutlich: Aus 16.769 Katalogverbindungen haben 105 das gewünschte C-Skelett, 21 das entsprechende Substitutionsmuster im allgemeinsten Sinn und nur neun das Substitutionsmuster der Anfrageverbindung, bei dem sich die Oxidationsstufen exakt entsprechen. Ohne Frage kann das Isopentylbromid (**9**),

-amin (**10**) oder der Isopentylalkohol (**11**) mittels einer Substitutionsreaktion in das gewünschte Chlorid überführt werden.

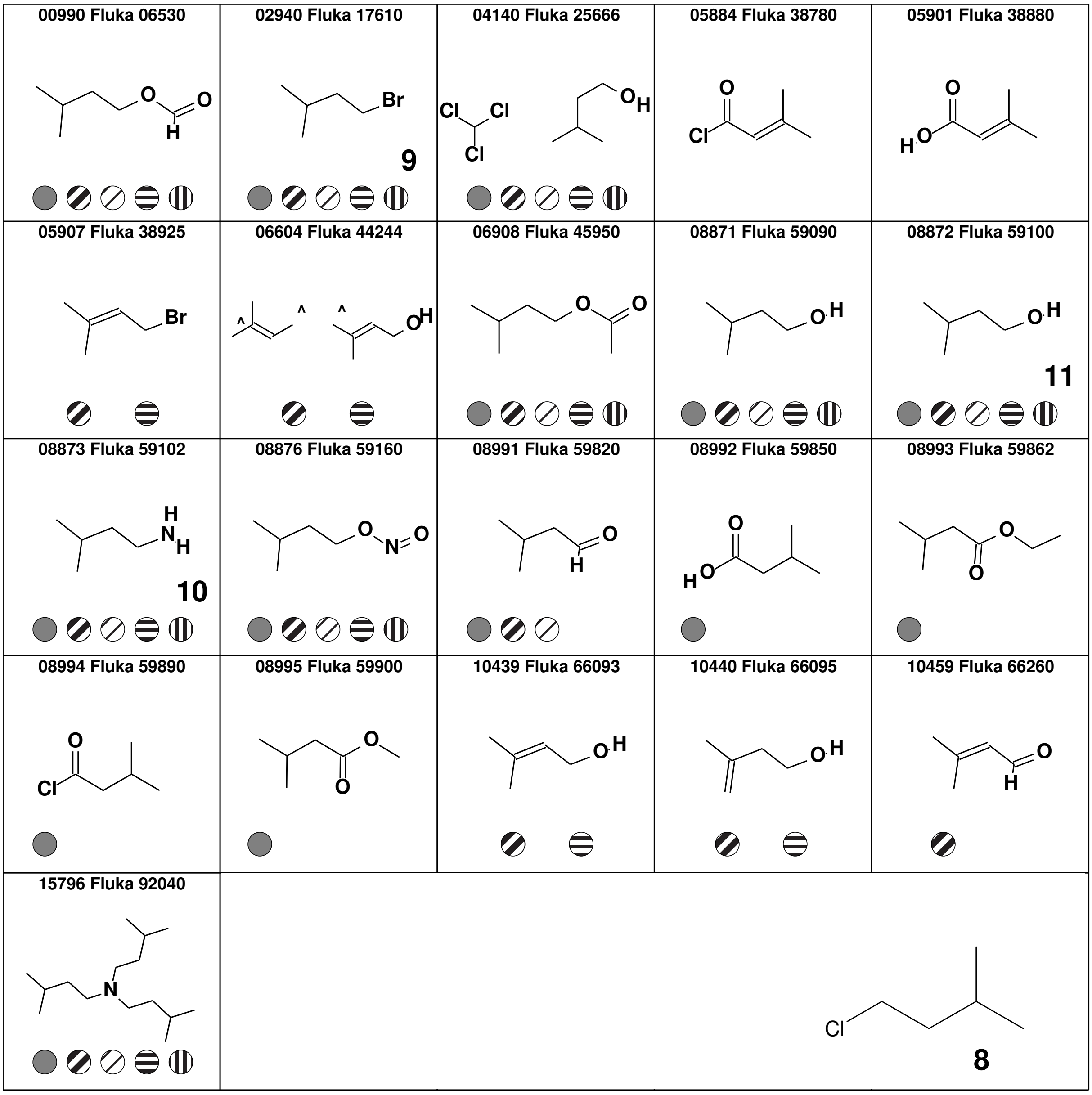


Abb. 3 - 13 Ähnlichkeitssuchen mit Isopentylchlorid. 21 Treffer resultieren mit dem Kriterium „Substitutionsmuster (-AR+A1-CCMB-BO-MU)“. Andere substitutionsmuster-orientierte Transformationssuchen führen zu reduzierten Trefferzahlen. (Erklärung der Markierungen im Text.)

Gesucht seien nun all jene Katalogverbindungen, deren größtes zusammenhängendes C-Skelett aus einer Cyclohexan-Einheit besteht, die an allen Skelettatomen mindestens einfach funktionalisiert ist. CC-Mehrfachbindungen sollen als Funktionalität betrachtet werden (*-AR+A1+CCMB-BO-MU*). Eine solche Suche liefert für den Janssen-Katalog 17 Treffer (siehe Abb. 3 - 14). Die dazu notwendige Anfragestruktur wäre beispielsweise 1,2,3,4,5,6-Hexachlor-cyclohexan (**12**). Wenn man CC-Mehrfachbindungen ausschließt (*-AR+A1* ***-CCMB****-BO-MU*), bleiben nur noch zwölf Treffer (in Abb. 3 - 14 mit ● markiert). Wird verlangt, daß genauso viele Substituenten je Skelettatom in den Treffern vorhanden sind wie in der Anfrage, resultieren acht bzw. elf Treffer (*-AR+A1+CCMB-BO****+MU*** bzw. *-AR+A1*

-CCMB*-BO****+MU*** ( ), je nachdem ob man wiederum CC-Mehrfachbindungen als Funktionalität betrachtet oder nicht. Schließlich kann man auch noch fordern, daß mögliche Substituenten nur über Einfachbindungen - wie in der Anfrage - gebunden sind (*-AR+A1 ±CCMB****-BO+MU***). Dann bleiben nur noch zwei Treffer ( ), auch bei Betrachtung von CC-Mehrfachbindungen als Substituenten. Es sei angemerkt, daß durchaus unterschiedliche Trefferzahlen im letzten Fall resultieren könnten: 3,5-Cyclohexadien-1,2-diol würde beispielsweise bei Ignorierung der Doppelbindungen nicht als Treffer erscheinen, im anderen Fall aber gefunden werden.

Die verschiedenen Trefferzahlen verdeutlichen es: Von der allgemeinsten Definition eines Substitutionsmusters (17 Treffer) kann durch geeignete Wahl von Zusatzkriterien deutlich eingeschränkt werden (bis auf zwei Treffer). Zweifellos sind das *meso*-Inosit (**13**) und dessen Derivat (**14**) die zu 1,2,3,4,5,6-Hexachlor-cyclohexan ähnlichsten Verbindungen was ihr Substitutionsmuster betrifft.

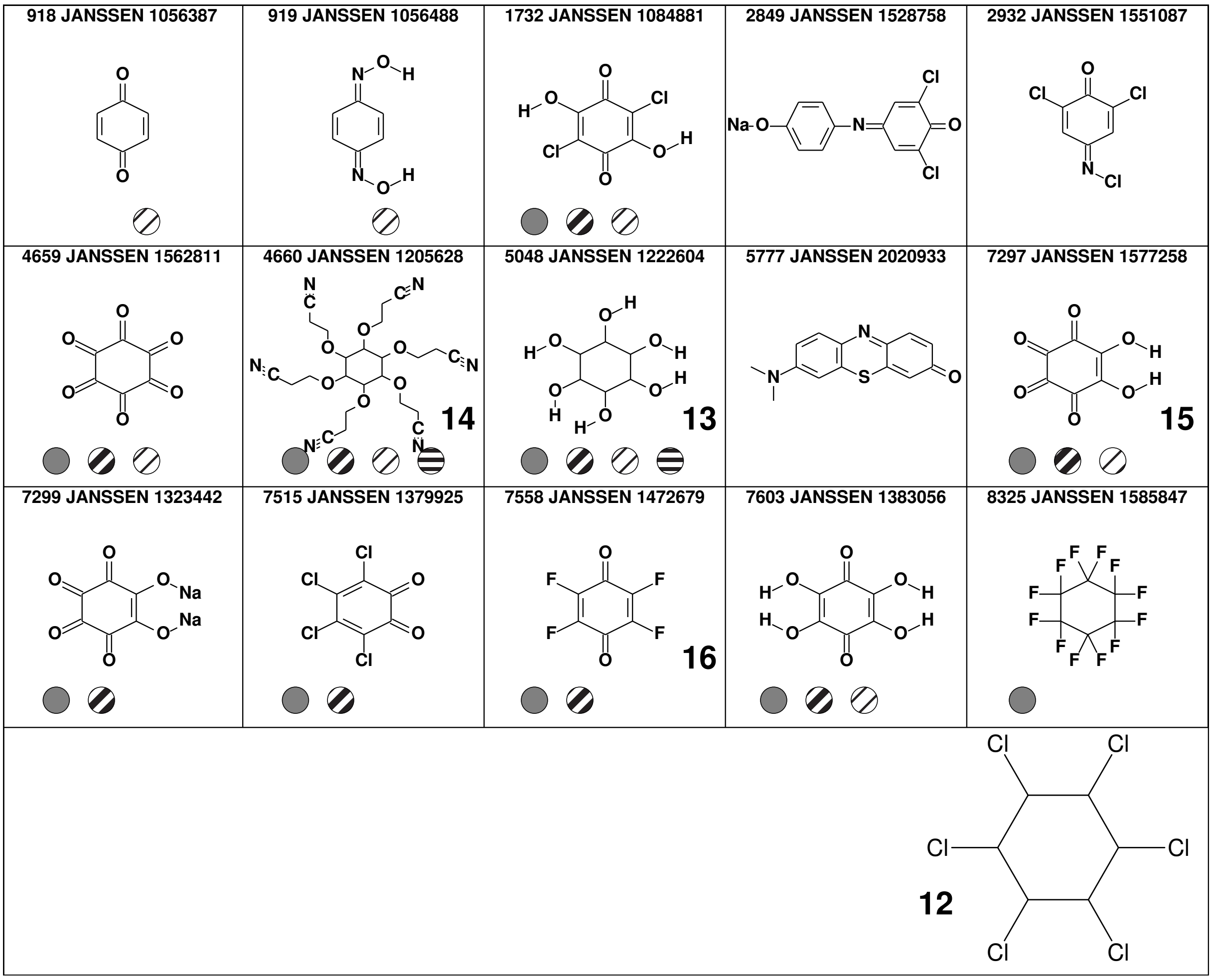


Abb. 3 - 14 15 von 17* Treffern, die im Janssen-Katalog für die Anfrageverbindung 1,2,3,4,5,6-Hexachlor-cyclohexan (**12**, unten rechts) mit der Transformation „Substitutionsmuster (-AR+A1+CCMB-BO-MU)“ ermittelt werden. Die Bedeutung der markierten Treffer ist im Text erläutert.

* Die nicht dargestellten Treffer sind das Kalium-Salz der Rhodizonsäure (**15**) und das Brom-Analogon zu Tetrafluoro-*p*-Benzochinon (**16**).

Als letztes Beispiel dient ein Benzoesäure-Derivat: 3-Amino-4-chlor-benzotrifluorid (**17**). Dessen allgemeinstes Substitutionsmuster (*-AR+A1±CCMB*[*]*-BO-MU*) findet sich in 41 Verbindungen des Fluka-Kataloges. 25 davon sind in Abbildung 3 - 15 dargestellt. In diesem Fall führt die Forderung nach gleicher Anzahl an Substituenten je Skelettatom in Anfrage und Katalogverbindung ohne Beachtung von Mehrfachbindungen zu den Substituenten (*-AR+A1±CCMB*<u>***-BO+MU***</u>) zu einer drastischen Verminderung der Trefferzahl (vier, markiert in Abb. 3 - 13 mit): Nur Verbindungen, die wie die Anfrage dreifach an der Methylseitenkette heterosubstituiert sind, werden gefunden. Werden aber CX-Mehrfachbindungen als Äquivalente für Mehrfachsubstitution beachtet (*-AR+A1±CCMB*<u>***+BO***</u>*+MU*), steigt die Trefferzahl wieder an (18,). Dann sind auch Säure- oder Nitrilgruppen als Äquivalente zulässig. Eine wichtige Besonderheit bei Aromaten sind unmittelbar am Kern vorhandene Heterosubstituenten. Während bei acyclischen Skelettatomen ein Austausch von vorhandenen Heteroatomen durch andere oft problemlos möglich ist, gilt diese Regel für den Aromatenkern nicht. Zwar existieren auch hier Methoden zur Umfunktionalisierung (etwa die Umwandlung von Nitrogruppen in Phenole), jedoch erfordern solche Methoden meist recht drastische Bedingungen. Es ist daher wünschenswert, Ähnlichkeitssuchen basierend auf dem Substitutionsmuster so zu gestalten, daß Heterosubstituenten am Aromatenkern nicht wie andere Substituenten generalisiert werden. Diesem Zweck dient der Schalter *A1*. Ähnlichkeitssuchen mit dem Schalter *-A1* fordern daher für 3-Amino-4-chlor-benzotrifluorid (**17**), daß mögliche Treffer neben einem Benzyl-Skelett in *meta*-Position einen Stickstoff- und in *para*-Position einen Chlor-Substituenten tragen. Damit (*-AR-A1±CCMB-BO-MU*) werden statt 41 (siehe oben) nur noch fünf Treffer erzielt ( ).

[*] Der Schalter für CC-Mehrfachbindungen spielt in diesem Beispiel keine Rolle. Das gilt aber nicht grundsätzlich für aromatische Verbindungen. Sind die Seitenketten am Aromatenkern nur lang genug und entsprechend funktionalisiert, spielt auch in diesen Fällen der Schalter *CCMB* eine Rolle.

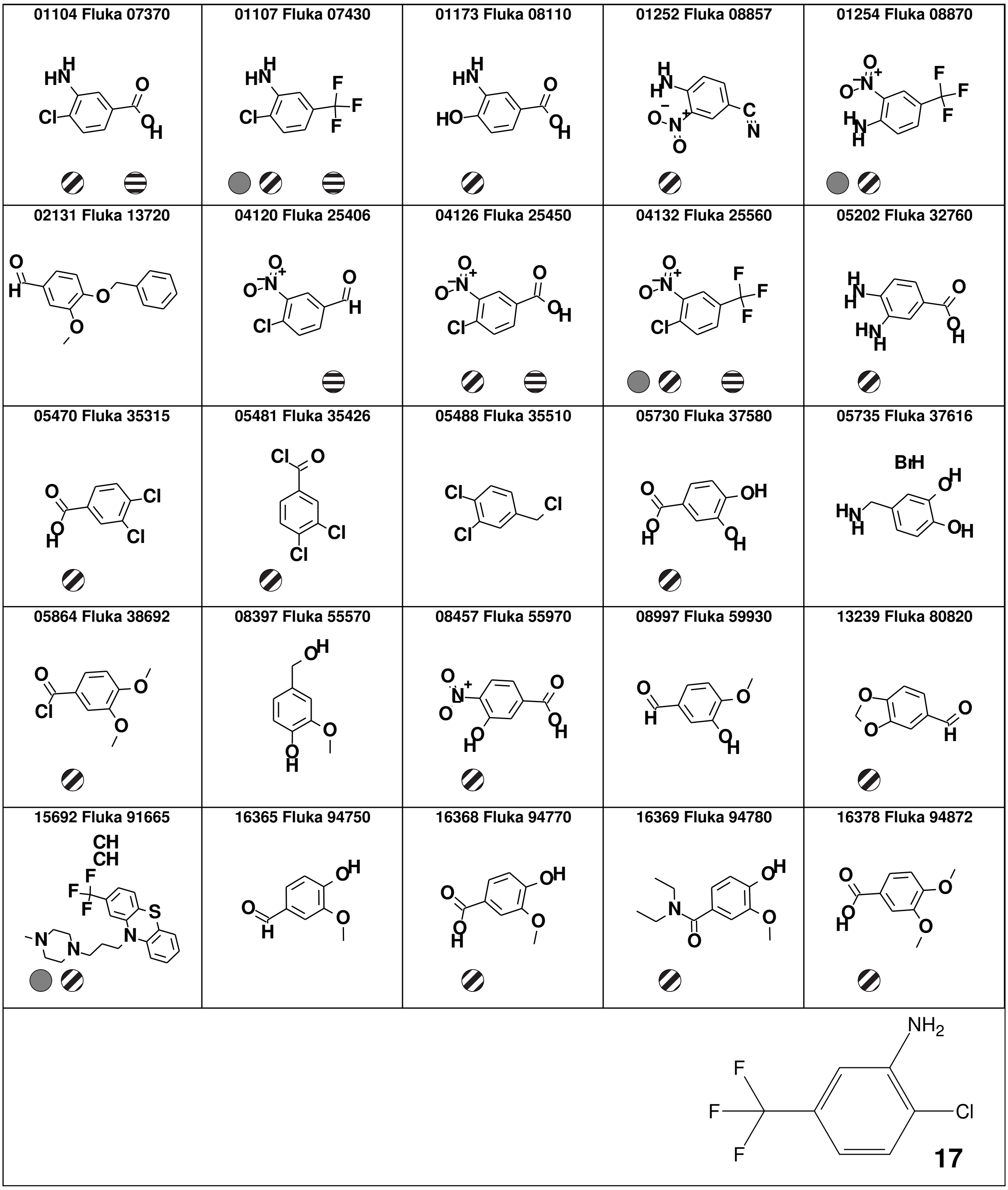


Abb. 3 - 15 Substitutionsmuster-orientierte Ähnlichkeitssuchen mit 3-Amino-4-chlor-benzotrifluorid (**17**, unten rechts). Dargestellt sind 25 von 41 Treffern erzielt mit dem Kriterium „Substitutionsmuster (-AR+A1±CCMB-BO-MU)“.

Es sei an dieser Stelle angemerkt, daß sich die hier dargelegten Ergebnisse auch mittels klassischer Substruktursuchen erzielen lassen. Es gibt jedoch Fälle, bei denen die gewünschte Spezifität der Transformationsregel entweder gar nicht oder nur durch mehrere Substruktursuchen erreicht wird. Ein Beispiel zeigen die folgenden Abbildungen. Gesucht werden sollen Verbindungen, die folgende Kriterien erfüllen:

- ❏ Ein endständig funktionalisierter C_4-Baustein mit gestreckter Kette.
- ❏ Das C_1-Atom in Oxidationsstufe 3, d. h. funktionalisiert als Carbonsäure, Ester, Trihalogenverbindung, Nitril oder dgl.; sonst ohne Heterosubstituenten.
- ❏ Mehrfachbindungen auf dem C-Skelett sind erlaubt, etwa weil davon ausgegangen werden kann, daß sie selektiv reduziert werden könnten.

Diese Kriterien lassen sich einfach mit einer einzigen Transformationssuche erfüllen: „Substitutionsmuster -AR±A1-CCMB+BO+MU“. Man beachte, daß als Anfrage beliebige Repräsentanten der obigen Kriterien dienen können. Das Suchergebnis unterscheidet sich *nicht*, wenn mit Buttersäure (**18**), *Ortho*-Buttersäuremethylester (**19**) oder Crotonsäurenitril (**20**) als Anfrage gearbeitet wird (siehe Abb. 3 - 16). Dagegen sind allein drei einzelne Substrukturen vonnöten, um überhaupt alle möglichen Varianten für eine endständige Oxidationsstufe 3 auf einem *n*-Butangerüst zu formulieren.* Um auch CC-Mehrfachbindungen zu erlauben, müssen noch einmal je vier weitere Substrukturen spezifiziert werden. Allene werden dann immer noch nicht erfaßt.

Als Resultat erhält man 32 (Transformationssuche, siehe Abb. 3 - 16) bzw. 42 (Substruktursuchen, siehe Abb. 3 - 17) Treffer. Die 32 Treffer der Transformationssuche sind komplett Bestandteil des Ergebnisses der Substruktursuchen. Die zusätzlich gefundenen Treffer zeigt Abbildung 3 - 18. Die Ursache, warum diese Verbindungen mittels der Transformationssuche nicht gefunden werden, liegt an der Größe des substituierten Butyl-Derivats im Vergleich zu anderen Molekülfragmenten (siehe Abschnitt 3.3.2). Da für das Ergebnis einer Transformationssuche nur das größte resultierende Molekül relevant ist, scheiden sämtliche Verbindungen in Abbildung 3 - 18 aus. Ganz besonders überzeugend ist der Treffer Cholesterylbutyrat (**21**). Die Butyrat-Gruppe (die geforderte Substruktur) ist ein relativ bedeutungsloser Rest am Steroidgerüst. Richtigerweise werden solche Einträge bei den für die Syntheseplanung entwickelten Transformationssuchen ausgeschlossen. Liegt die Betonung bei einer solchen Suche jedoch darauf, tatsächlich alle Verbindungen aufzuspüren, die eine gegebene Substruktur enthalten, gleich in welchem molekularem Umfeld, dann ist die klassische Substruktursuche die Methode der Wahl.

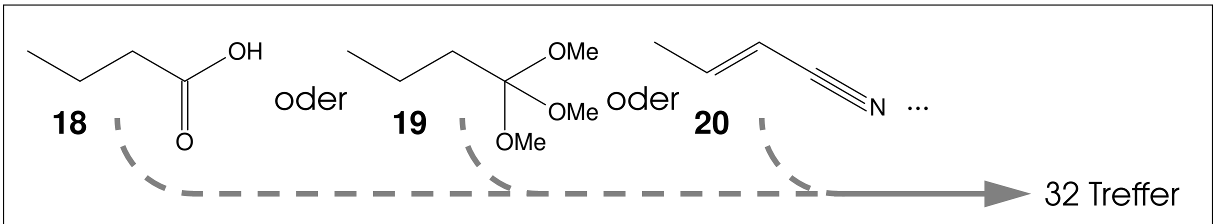


Abb. 3 - 16 Transformationssuche nach einer endständig funktionalisierten C_4-Einheit mit Tolerierung von CC-Mehrfachbindungen („Substitutionsmuster -AR±A1-CCMB+BO+MU“). Als Anfrage kann jede der dargestellten Verbindungen direkt verwendet werden.

* Man könnte einwenden, daß mit freien Valenzen gearbeitet werden könnte. Ein solches Vorgehen würde aber dazu führen, daß neben den gewünschten Verbindungen auch andere C-Skelette als Treffer gefunden werden.

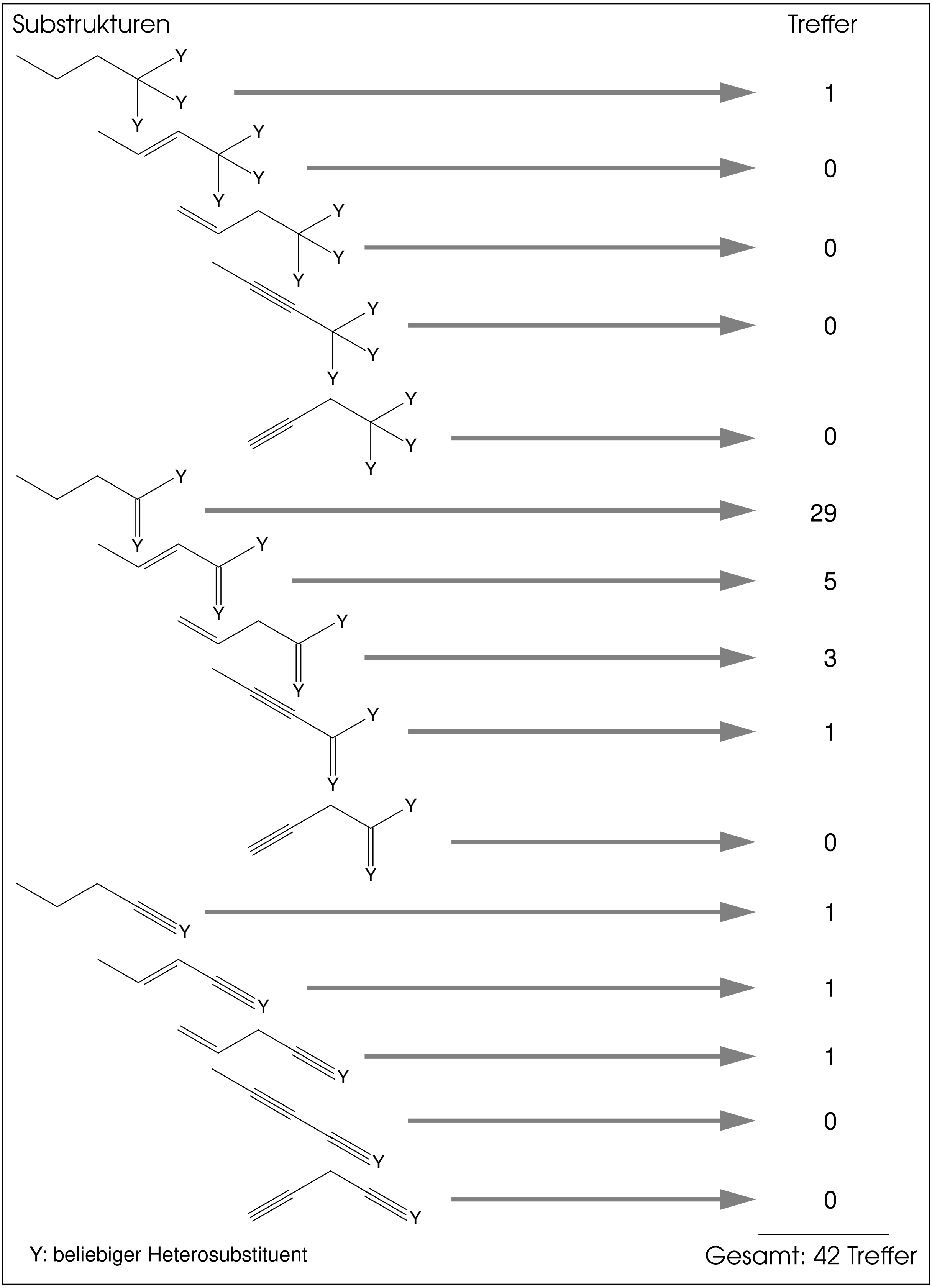


Abb. 3 - 17 Klassische Substruktursuche mit der gleichen Aufgabe wie in Abb. 3 - 16. Hier müssen eine ganze Reihe von Substrukturen formuliert werden, um das entsprechende Ergebnis zu erzielen.

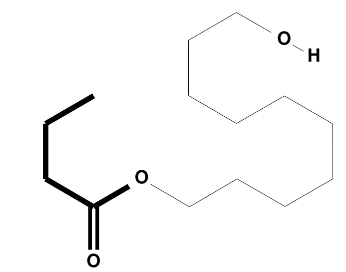
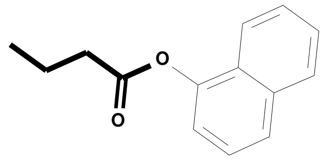
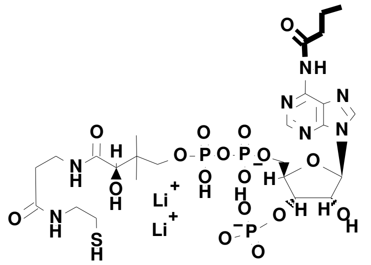
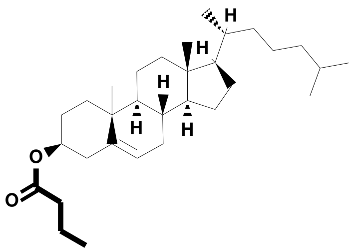
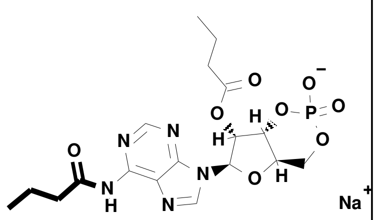
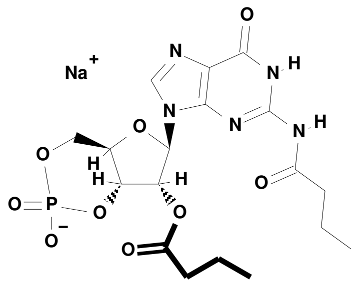
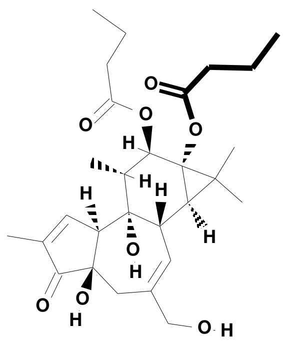
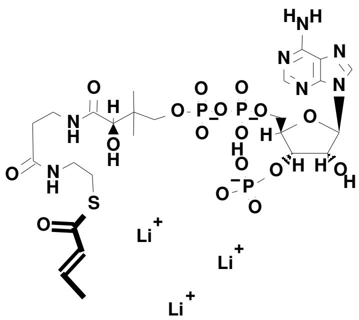
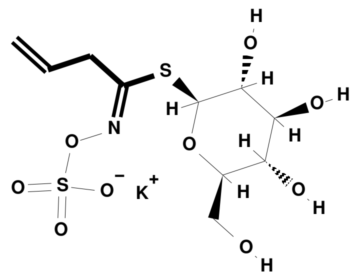
00009 Fluka 19355	00012 Fluka 19363	00021 Fluka 20785
		
Buttersäure–(10–hydroxydecyl)–ester	Buttersäure–1–naphthylester	Butyryl–Coenzym A Dilithiumsalz Hydrat
00023 Fluka 26770	00024 Fluka 34988	00026 Fluka 34995
 **21**	 **22**	
Cholesteryl–butyrat	N(6),2'–O–Dibutyryl–adenosin–3',5'–cyclo	N(2),2'–O–Dibutyryl–guanosin–3',5'–cyclo
00028 Fluka 79345	00032 Fluka 28007	00037 Fluka 85440
		
Phorbol–12,13–dibutyrat	Crotonoyl–Coenzym A Trilithiumsalz	Sinigrin Monohydrat

Abb. 3 - 18 Neun* von zehn Treffern, die mit Substruktursuchen zusätzlich gefunden werden. Die Substrukturen sind in den Katalogverbindungen fett markiert.†

* N(6),2'-O-Dibutyryl-adenosin-3',5'-cyclophosphat Natriumsalz Monohydrat (**22**, Fluka 34988) ist noch ein weiteres Mal im Katalog enthalten (Fluka 34990).

† In manchen Katalogverbindungen findet sich die Substruktur mehrfach. Auf deren Kennzeichnung wurde verzichtet.

Aus den gewonnenen Erfahrungen im Umgang mit substitutionsmuster-basierten Ähnlichkeitssuchen ergeben sich folgende Anwendungshinweise:

⇨ Die Suche sollte zunächst mit einem möglichst allgemein definierten Substitutionsmuster durchgeführt werden. Werden sehr hohe Trefferzahlen erzielt, kann durch Variation des Substituentenbegriffs später eingeschränkt werden.

⇨ Besitzt die Anfrage keine CC-Mehrfachbindungen und auch keine Heterosubstituenten in vicinaler Relation wird das allgemeinste Substitutionsmuster - und damit die größte mögliche Trefferzahl - mit *-AR+A1-CCMB-BO-MU* gefunden, andernfalls mit *-AR+A1+CCMB -BO-MU*.

⇨ Die Menge der mit der Anfrage am ähnlichsten Verbindungen findet man meistens mit den Schaltern *-AR+A1+CCMB+BO+MU*.

⇨ Eine bedeutende Verminderung der Trefferzahlen für aromatische Anfrageverbindungen läßt sich durch den Schalter *-A1* erreichen.

3.3.5 Ähnlichkeiten durch generalisierte Reaktionen

Die bisher beschriebenen Ähnlichkeitskriterien gehören zur Gruppe der substruktur-orientierten Transformationen (vgl. Abschnitt 3.3.2, Seite 45ff). Nachfolgend werden die reaktionsklassen-orientierten Ähnlichkeitskriterien behandelt. Zunächst werden die Kriterien einzeln nacheinander vorgestellt. Schließlich lassen diese sich (z. T. auch mit Transformationen der anderen Gruppe) zu noch leistungsfähigeren Kriterien kombinieren. Das wird im Anschluß belegt.

3.3.6 Hydrolyse-Transformationen

Die Hydrolyse ist eine wichtige, sehr generelle Reaktion in der Chemie. Mit ihrer Hilfe lassen sich elementare Umfunktionalisierungen durchführen. Größere Verbindungen (z. B. Ester) zerfallen dabei oft in wesentlich kleinere, elementarere Komponenten. Es wird deshalb hier nochmals daran erinnert, daß am Ende alle Fragmente bis auf das größte verworfen werden.

Hingewiesen sei an dieser Stelle auch daran, daß die Hydrolyse als Transformationsregel sowohl auf die Anfrageverbindung, als auch auf jedes Datenbankmolekül angewendet wird bzw. wurde. Dadurch können im Sinne einer synthetischen Umsetzung eines so gefundenen Datenbankmoleküls zur Anfrageverbindung auch andere Reaktionen als nur Hydrolyseschritte notwendig sein, etwa Substitutionen. Es wird weiter unten noch gezeigt, welche Konsequenzen das hat.

Grundsätzlich werden *formale* Hydrolyseschritte zur Definition des Ähnlichkeitskriteriums an substituierten Kohlenstoff-Atomen, aber auch an den Atomen der Stickstoff-Gruppe und Chalkogenen (mit Ausnahme von Stickstoff und Sauerstoff), durchgeführt. Unterschieden werden kann, ob das betreffende Atom mit hydrolysierbaren Substituenten aromatisch ist oder nicht.* Das trägt der Tatsache Rechnung, daß aromatische Substituenten mitunter nur unter recht drastischen Bedingungen (oder gar nicht) hydrolysierbar sind.

Welche Bindungen hydrolysiert werden und welche (Transformations-)Produkte dabei entstehen, zeigen die folgenden Abbildungen. Die wesentlichen funktionellen Gruppen, die der Hydrolyse unterworfen werden sind

- Ester (sowohl Carbonsäure- als auch Ester von Schwefel-Sauerstoffsäuren oder Phosphor-Sauerstoffsäuren) und verwandte Verbindungen wie Amide, Säurehalogenide, Lactone und Lactame usw.,
- Acetale, Ketale, Aminale, Iminale und weitere Analoga,
- Enolether, Enamine und Analoga,
- Epoxide, Episulfide und Oxetane sowie deren höhere Analoga,
- carbonylanaloge Verbindungen wie Imine, Oxime, Hydrazone, Semicarbazone,
- freie Substituenten wie Halogene (jedoch *nicht* bei ausschließlich fluorierten Atomen) oder Aminogruppen,
- Nitrile, Isocyanate und Analoga,
- Nitrogruppen, Diazoverbindungen,
- Organo-Metallverbindungen wie Alkyl-Lithium-, Grignard-Verbindungen oder Borane.

* Die Schalter werden in der Benutzerschnittstelle von WODCA mit „no alphas“ bzw. „with alphas“ bezeichnet.

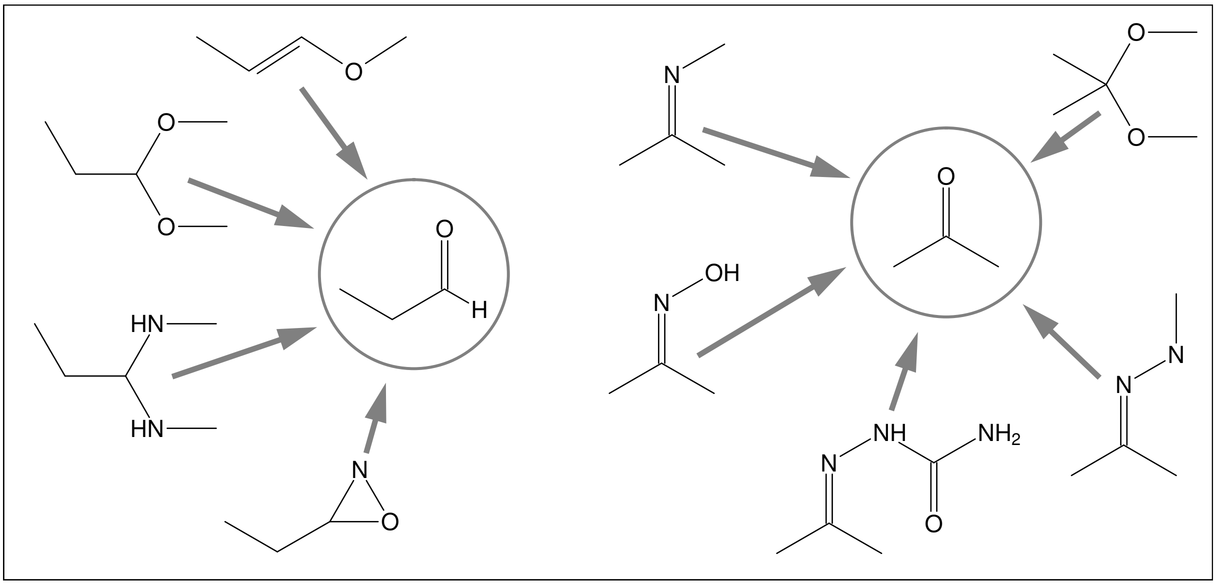


Abb. 3 - 19 Wirkung der Hydrolysefunktion auf einige Carbonyl- und carbonylanaloge Verbindungen. Es wird immer die korrespondierende Carbonylstammverbindung (Aldehyd bzw. Keton) gebildet.

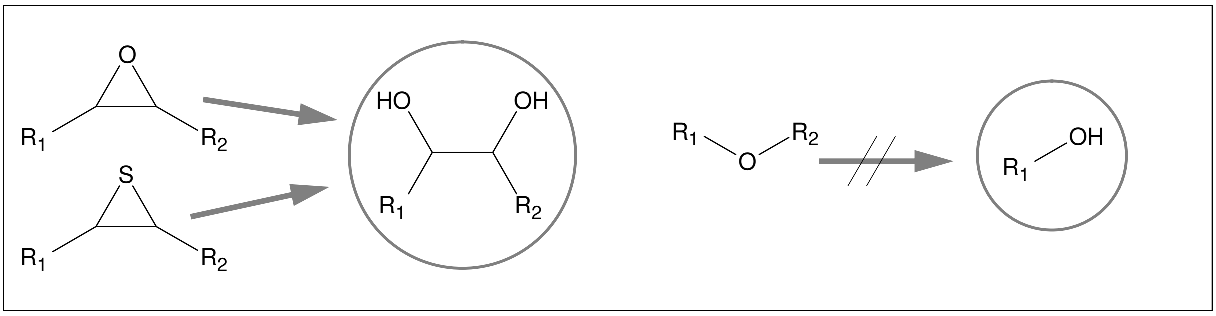


Abb. 3 - 20 Wirkung der Hydrolysefunktion auf Epoxide, Episulfide etc. Andere Ether werden *nicht* verändert.

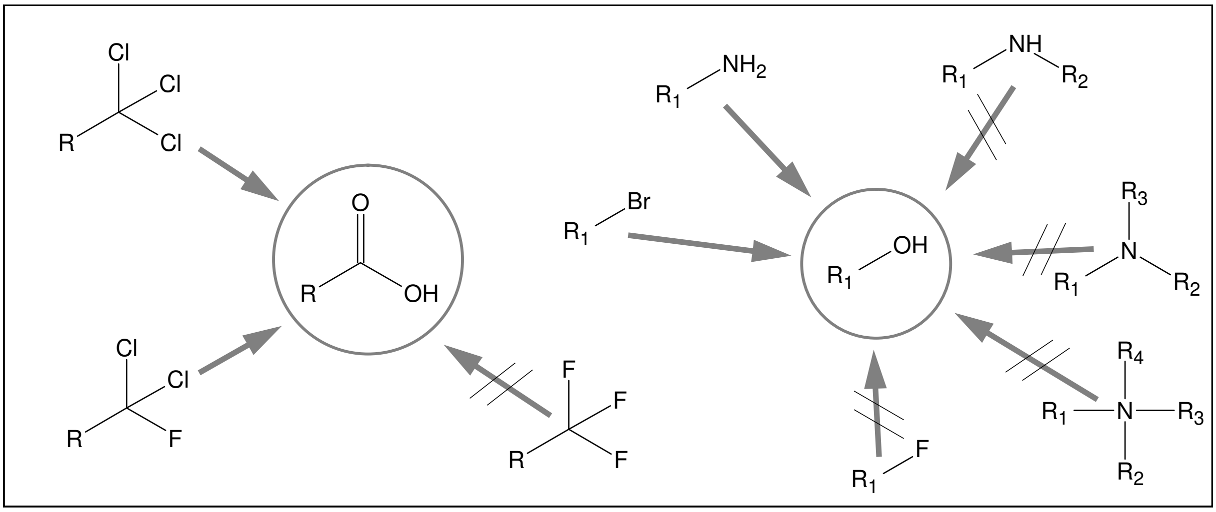


Abb. 3 - 21 Hydrolyse von Halogensubstituenten und Aminen. Freie Aminogruppen werden immer hydrolysiert, sekundäre und höhere Amine dagegen *nicht* (rechts). Fluorsubstitutenten werden nur dann hydrolysiert, wenn andere hydrolysierbare Substituenten gleichzeitig vorhanden sind (links).

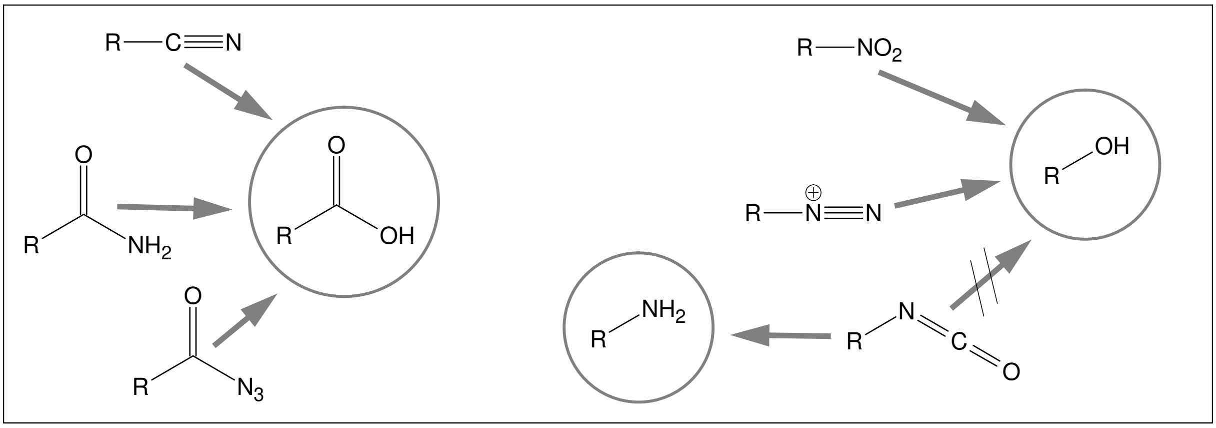


Abb. 3 - 22 Transformation einiger Stickstoff-haltiger Gruppen. Man beachte, daß aus Isocyanaten nicht die entsprechenden Alkohole, sondern die korrespondierenden Amine entstehen.

Unverändert bleiben dagegen Ether, Thioether usw., höhere Amine, Phosphine usw.* sowie Bindungen zwischen den Elementen der IV. Hauptgruppe (Silane, Stannane). Auch Heteroaromaten (Triazine beispielsweise), die durchaus hydrolytisch spaltbar sind, werden von diesen Hydrolysekriterium *nicht* verändert.

In ihrer Wirkung ähneln Hydrolyse-Transformationen den Substitutions-Transformationen. Während dort Substituenten durch Chloratome verallgemeinert werden, entstehen hier Sauerstoff-haltige Substituenten. Jedoch gibt es deutliche Unterschiede: Die Substitutions-Transformationen fragmentieren eine Verbindung *immer* an CX-Bindungen. Hydrolysen belassen dagegen eine Reihe solcher Bindungen (siehe oben). Auch die Behandlung von Fluor-Substituenten birgt einen Unterschied. Die Konsequenz daraus ist, daß die Hydrolysen ein wesentlich *engeren* Begriff von synthetisch nutzbarer Ähnlichkeit definieren als Substitutions-Transformationen. Die Zahl der notwendigen Umsetzungen von einem gefundenen Ausgangsmaterial zur gewählten Anfrageverbindung ist dadurch *geringer.*

Anwendungshinweise:

⇨ Hydrolyse-Transformationen gehören zu den Suchen, die einen sehr engen Ähnlichkeitsbegriff definieren. Sie sind daher dort am besten geeignet, wo andere, allgemeinere Transformationssuchen unakzeptabel große Trefferlisten erzielen.

⇨ Wie bei den Transformationen basierend auf dem Substitutionsmuster (vgl. Abschnitt 3.3.4) können Substituenten am Aromatenkern von der Hydrolyse ausgenommen werden. Für den Begriff der Ähnlichkeit bedeutet das eine nähere Verwandtschaft. Synthetisch gesehen sind dann weniger Umsetzungen von einem gefundenem Ausgangsmaterial hin zur Anfrageverbindung zu erwarten.

⇨ Besonders leistungsstark ist der Hydrolyse-Begriff als Ähnlichkeitskriterium in Kombination mit Redox-Operationen (siehe Abschnitt 3.3.10).

* Sekundäre Amine können beispielsweise nicht *generell* durch Hydrolyse wieder in primäre Amine gespalten werden.

3.3.7 Transformationen basierend auf Redox-Prozessen

Eine weiterer Typ von Reaktionen, die in weiten Bereichen der Synthese Anwendung finden, sind Oxidationen und Reduktionen. Letztere haben bereits bei topologisch orientierten Transformationssuchen (vgl. Abschnitt 3.3.3) eine Rolle gespielt, dort allerdings begrenzt auf Mehrfachbindungen des C-Skeletts. Hier sollen Redoxreaktionen dazu dienen, Atome in einen vergleichbaren Oxidationszustand zu versetzen.

Für das Kriterium „Maximale Oxidation“ werden folgende formale Schritte unternommen: Betroffen werden von dieser Transformation nur Atome der Stickstoff-Gruppe, Chalkogene und substituierte Kohlenstoffatome. Alle anderen Atome bleiben unbeeinflußt.

- ❑ Primäre Alkohole, Aldehyde, Säurehalogenide, Trihalogenverbindungen werden zu den korrespondierenden Carbonsäuren oxidiert.
- ❑ Sekundäre Alkohole und Dihalogenverbindungen werden zu den korrespondierenden Ketonen oxidiert.
- ❑ Geminale Aminoalkohole werden wie Säureamide behandelt.
- ❑ Primäre Amine, Nitrosoverbindungen, Hydroxylamine und Oxime werden zu den korrespondierenden Nitroverbindungen oxidiert.
- ❑ Aus sekundären Aminen entstehen die korrespondierenden Hydroxylamine; aus tertiären Aminen die N-Oxide. Nitrile ergeben ebenfalls N-Oxide.
- ❑ Nitrosamine und Hydrazine werden analog behandelt, d. h. endständige einfach gebundene N-Atome werden zu Nitrogruppen, sekundäre N-Atome zu Hydroxylamin oxidiert.
- ❑ Gleiches gilt für Isonitrile und Isocyanate, d. h. die N-Atome werden zu N-Oxid-Analoga.
- ❑ Primäre Phosphine und P-Oxide, phosphinige Säuren und deren Säurehalogenide sowie Halogenide von Phosphonsäuren werden zu den korrespondierenden Phosphonsäuren oxidiert.
- ❑ Sekundäre Phosphine, P-Oxide und Halogenide von Phosphinsäuren werden zu den korrespondierenden Phosphinsäuren.
- ❑ Tertiäre Phosphine werden zu den analogen P-Oxiden umgesetzt.
- ❑ Höhere Elemente der Stickstoff-Gruppe werden wie Phosphor behandelt.
- ❑ Mercaptane, Sulfensäuren und deren Halogenanaloga, Sulfinsäuren und deren Halogenide sowie Sulfonsäurehalogenide werden zu den korrespondierenden Sulfonsäuren oxidiert.
- ❑ Sulfide und Sulfoxide werden zu den korrespondierenden Sulfonen oxidiert.
- ❑ Höhere Chalkogene werden ebenso behandelt.

Die folgenden Abbildungen veranschaulichen diese Transformationen.

Eine spezielle Behandlung erfahren wiederum Fluor-Substituenten an C-Atomen. Sind sie Substituent an einem sp^2-Kohlenstoff, werden sie zu Hydroxylgruppen hydrolysiert. Als Substituenten an sp^3-Kohlenstoffen bleiben sie dagegen unverändert.

Unverändert, weil bereits im höchsten Oxidationszustand oder weil durch Oxidation nicht angreifbar, bleiben auch z. B. einfache Halogenide, quartäre Ammonium- oder Phosphoniumverbindungen, Carbonsäuren, Phosphin- und Phosphonsäuren, tertiäre N- und P-Oxide, P-Ylide bzw. Alkylidenphosphorane, Thion-, Dithio- und Thiolsäuren sowie S-Ylide bzw. Alkylidensulfane.

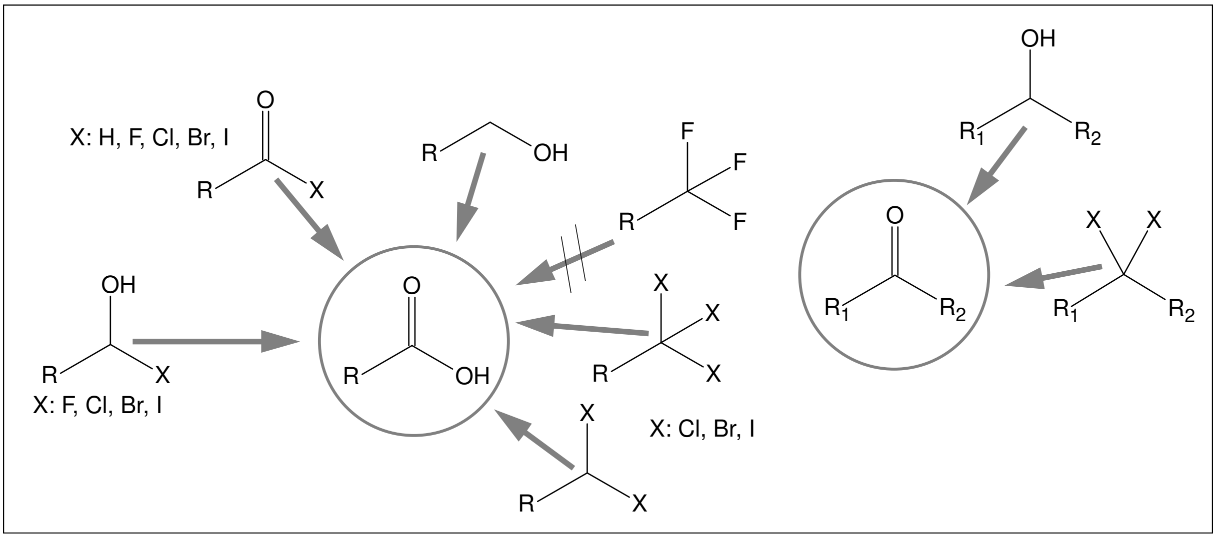


Abb. 3 - 23 Wirkung der Oxidationsregel auf Kohlenstoffverbindungen. Wo möglich, wird die korrespondierende Säure bzw. das Keton gebildet.

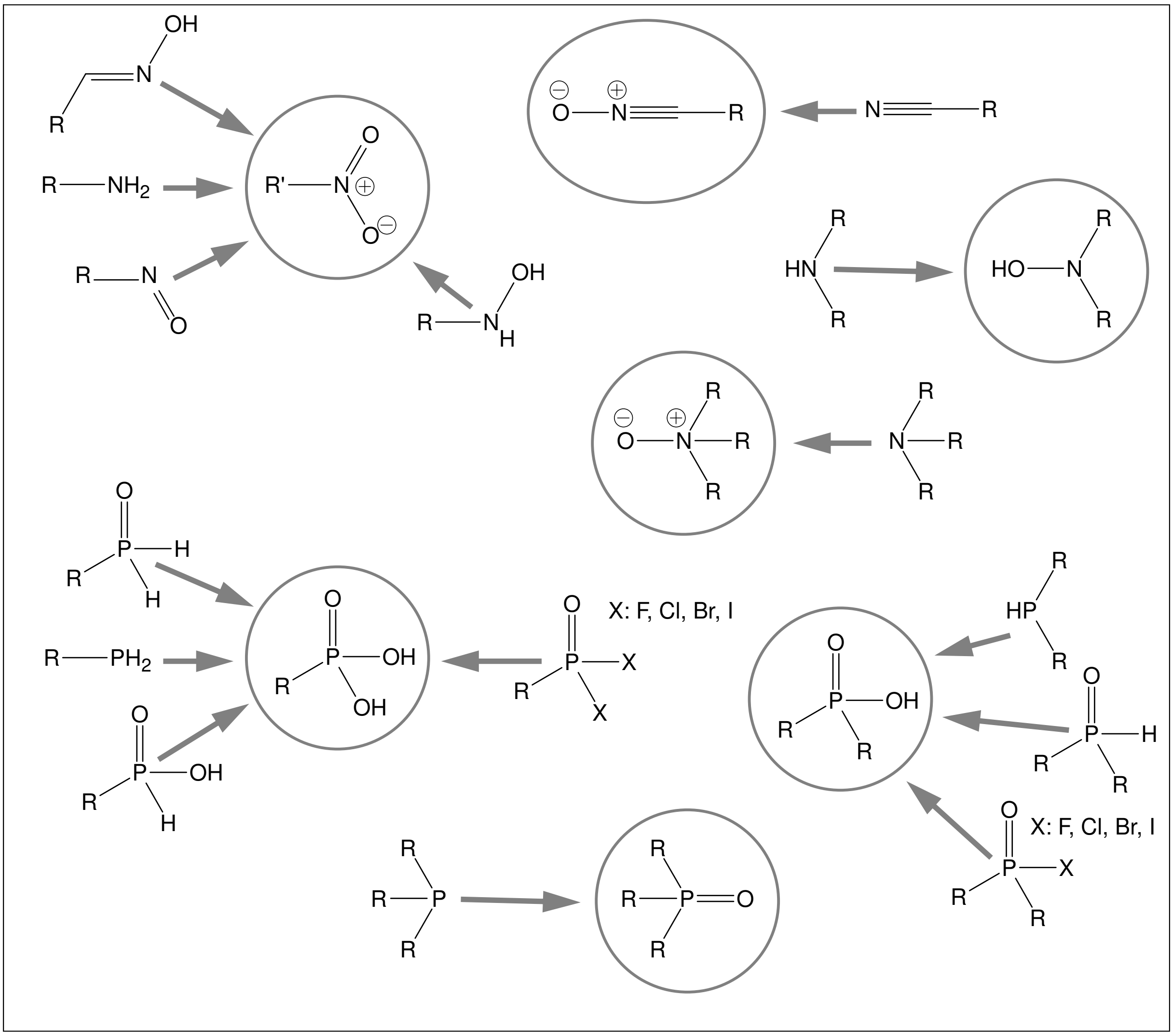


Abb. 3 - 24 Oxidation von Stickstoff- und Phosphor-haltigen Gruppen. Die Transformation führt zu den höchsten Oxidationsstufen.

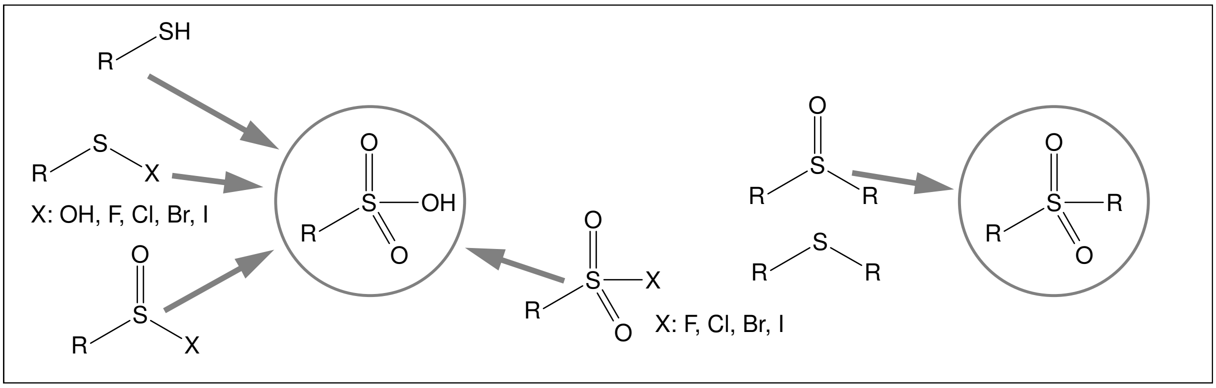


Abb. 3 - 25 Wirkung der Oxidationstransformation auf Schwefelverbindungen. Es entstehen die höchsten Oxidationsstufen.

Das Kriterium „Maximale Reduktion“ reduziert Bindungen zwischen Heteroatomen und beliebige Mehrfachbindungen, d. h. sowohl CC- als auch CX-Mehrfachbindungen. Im einzelnen werden folgende Transformationen vorgenommen (siehe auch Abb. 3 - 26):

- ❏ Bindungen zwischen zwei Heteroatomen werden grundsätzlich gespalten.
- ❏ Thionsäuren, Carbonsäuren und deren Halogenide sowie Aldehyde werden zu den korrespondierenden Alkoholen reduziert.
- ❏ Thiolsäuren und Dithiosäuren, Sulfonsäuren und deren Halogenide, Sulfinsäuren und deren Halogenide sowie Disulfide werden zu den korrespondierenden Mercaptanen reduziert.
- ❏ Carbonsäureamide und die tautomeren Imidosäuren, Carbonsäureazide, Amidine, Nitrile, Nitro- und Nitrosoverbindungen, Hydrazine, Hydroxylamine, Cyanate und Thiocyanate werden zu den korrespondierenden primären Aminen reduziert.

Mehrfachbindungen in Aromaten allgemein* („Maximale Reduktion (ohne Aromaten)“) bzw. nur Kohlenstoff-Aromaten (Benzole, Naphthole usw.; „Maximale Reduktion (ohne C-Aromaten)“) können von einer Reduktion ausgenommen werden. Das trägt der Tatsache Rechnung, daß Aromaten oft von einem Ausgangsmaterial übernommen werden sollen. Die Ähnlichkeitssuche kann auf diese Weise eingeschränkt werden.

Oxidations- und Reduktionstransformationen können oft alternativ eingesetzt werden. Sie führen beide z. B. alle Sauerstoff-substituierten Derivate eines Kohlenwasserstoffes vom Alkohol über den Aldehyd bis zur Carbonsäure in einer Trefferliste zusammen. Unterschiede ergeben sich aber, wenn CC-Mehrfachbindungen in der Anfrage vorkommen wie auch bei bestimmten Sonderfällen wie z. B. Disulfiden (siehe Abb. 3 - 27). Unterschiedlich ist auch die Behandlung von vollständig halogenierten Skelettpositionen: Bei der Oxidation werden solche Positionen in die entsprechenden Ketone bzw. Carbonsäuren überführt; bei der Reduktion tritt dort keine Veränderung auf.

Alle bislang vorgestellten Redox-Kriterien definieren eine sehr begrenzte Ähnlichkeit. Das resultiert vor allem daher, daß eine Verbindung an Heteroatomen, die mehrere C-Skelette verbinden, *nicht* fragmentiert wird†. Um dem Begriff der Ähnlichkeit basierend auf Redoxreaktionen eine etwas erweiterte Basis zu geben, wurden zwei weitere Transformationen implementiert: „C-Skelett inklusive α-Atome und maximale Oxidation“ und „C-Skelett inklusive

* Allgemein bedeutet hier: Es wird keine Unterscheidung zwischen Hetero- und reinen C-Aromaten vorgenommen.

† Man vergleiche dieses Verhalten mit den Hydrolyse- oder Substitutionstransformationen.

α-Atome und maximale Reduktion“. Diese Transformationen stellen bereits Kombinationen einzelner Ähnlichkeitsoperationen dar.

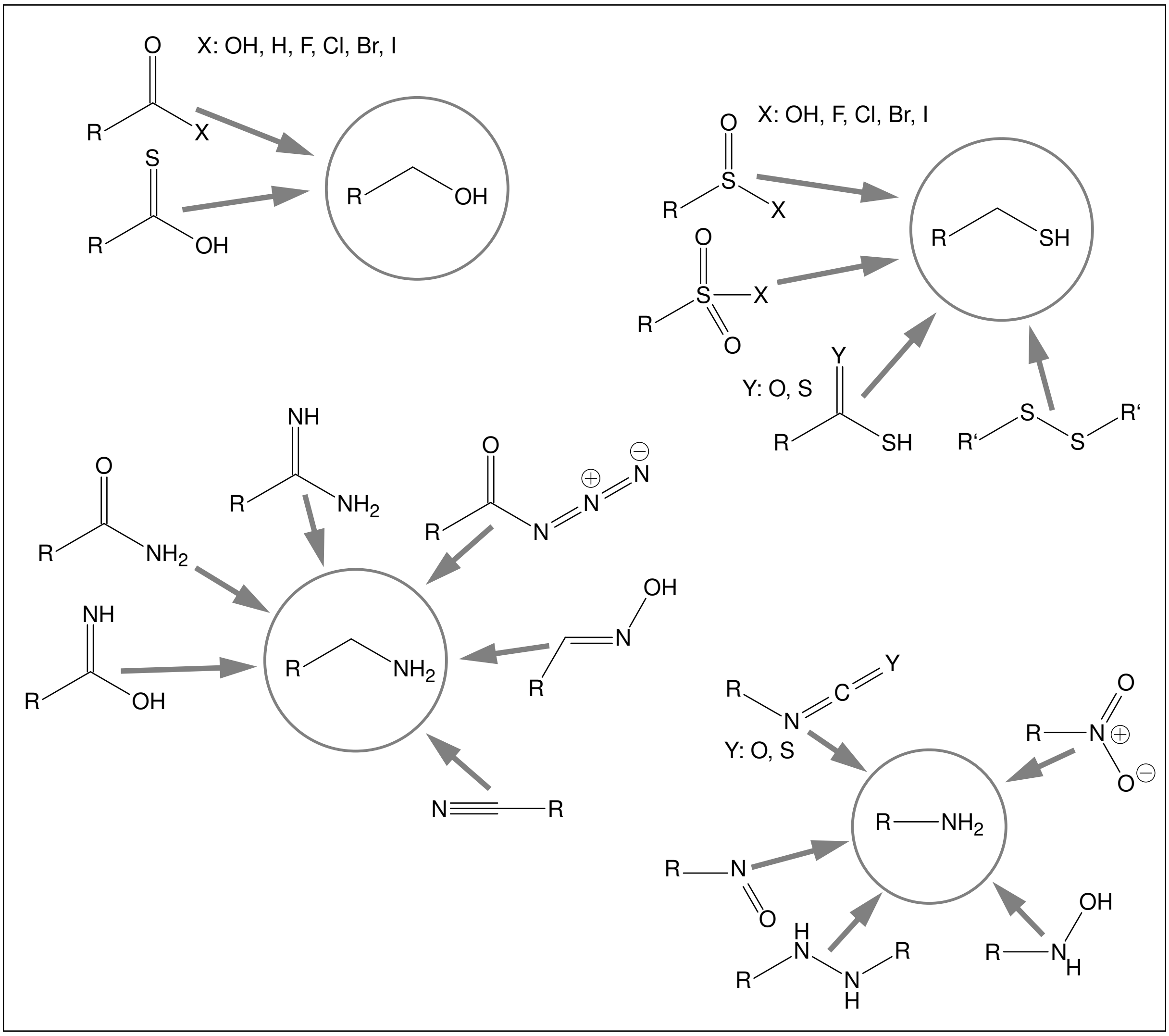


Abb. 3 - 26 Wirkung der Reduktionstransformationen. Links oben resultieren Alkohole. Rechts oben schwefelhaltige Verbindungen, deren Reduktion zu Mercaptanen führt. Unten Verbindungen, die in Amine überführt werden.

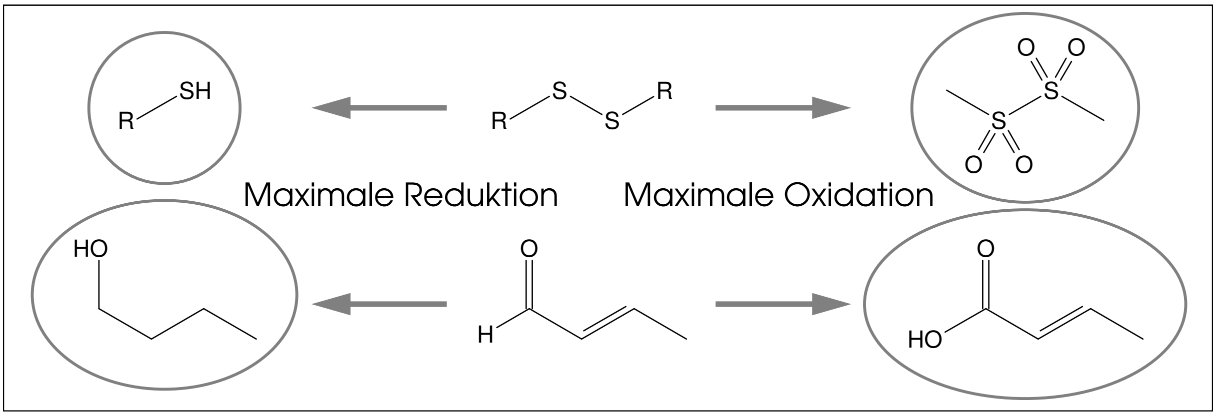
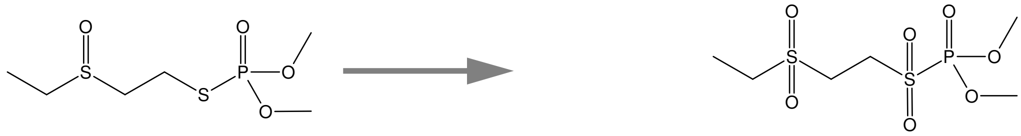
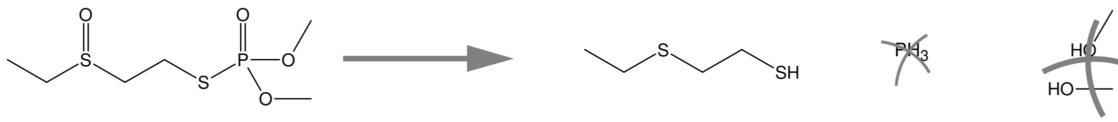
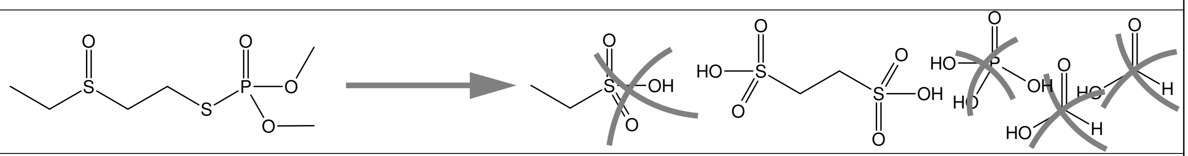


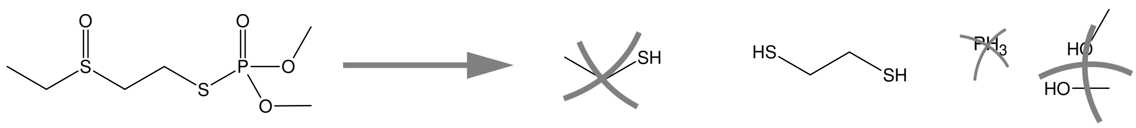
Abb. 3 - 27 Unterschiedliche Wirkung der Reduktions- und Oxidations-Transformationen bei CC-Mehrfachbindungen und Disulfiden.

Übersicht 3 - 3 Ähnlichkeitskriterien basierend auf Redoxreaktionen

Maximale Oxidation

Oxidierbare Atome (substituierte Kohlenstoffe, Chalkogene, Elemente der Stickstoff-Gruppe) werden in ihre jeweils höchsten Oxidationszustand überführt.

Maximale Reduktion

Reduzierbare Atome (substituierte Kohlenstoffe, Chalkogene, Elemente der Stickstoff-Gruppe) und Gruppen (Bindungen zwischen Heteroatomen, Mehrfachbindungen) werden auf die niedrigste Oxidationsstufe reduziert.

C-Skelett inklusive α-Atome und maximale Oxidation

Zunächst werden die C-Skelette extrahiert, wobei unmittelbar daran gebundene Heteroatome (α-Atome) ebenfalls erhalten bleiben. Verbrückt ein Heteroatom mehrere C-Skelette wird dieses dupliziert. Offene Valenzen werden dann gesättigt und dann die Oxidationsregel angewandt. Nur das am Ende größte resultierende Fragment bleibt erhalten.

C-Skelett inklusive α-Atome und maximale Reduktion

Wie oben wird bei der Extraktion der C-Skelette und der Behandlung der α-Atome verfahren. Statt der Oxidationsregel wird hier die maximale Reduktion angewandt. Nur das am Ende größte Fragment bleibt erhalten.

Als ein Anwendungsbeispiel dient Äpfelsäure (**23**). Eine Suche im Fluka-Katalog mit der Transformation „Maximale Oxidation“ ergibt neun Treffer (siehe Abb. 3 - 28, Treffer markiert mit ●). Die selbe Trefferliste wird auch mit dem Kriterium „Maximale Reduktion“ erhalten. Erweitert man den Ähnlichkeitsbegriff mit dem Kriterium „C-Skelett inklusive α-Atome und maximale Oxidation“ werden 18 Treffer gefunden. In der mittels „C-Skelett inklusive α-Atome und maximale Reduktion“ erzielbaren Liste fehlt 4-Trichlor-acetessigsäure-ethylester (**24**). Zum Vergleich: Das allgemeinste substitutions-basierte Ähnlichkeitskriterium (-AR+A1-CCMB-BO-MU) findet 107 Treffer, das mit der engsten Definition des Substitutionsmusters (-AR+A1+CCMB+BO+MU) immerhin 43.

In einem zweiten Beispiel sollen geeignete Ausgangsmaterialien für Pimelindiimidosäurediethylester (**25**) gefunden werden. Eine Suche im Fluka-Katalog mit dem Kriterium „Maximale Reduktion“ findet einen Treffer: den entsprechenden Pimelinsäureester (**26**, siehe

Abb. 3 - 29). Keinerlei Treffer findet sich mit „Maximale Oxidation“. Die Ähnlichkeitssuche „C-Skelett inklusive α-Atome und maximale Reduktion“ fördert sieben Treffer zu Tage. Ebenfalls einen Treffer findet man mit „C-Skelett inklusive α-Atome und maximale Oxidation“: den zur Anfrage analogen Dimethylester (**27**). Die für die Synthese interessante Verbindung ist jedoch Pimelinsäuredinitril (**28**), die nur mit einem nicht zu eng begrenztem Ähnlichkeitskriterium („C-Skelett inklusive α-Atome und maximale Reduktion“) gefunden wird.

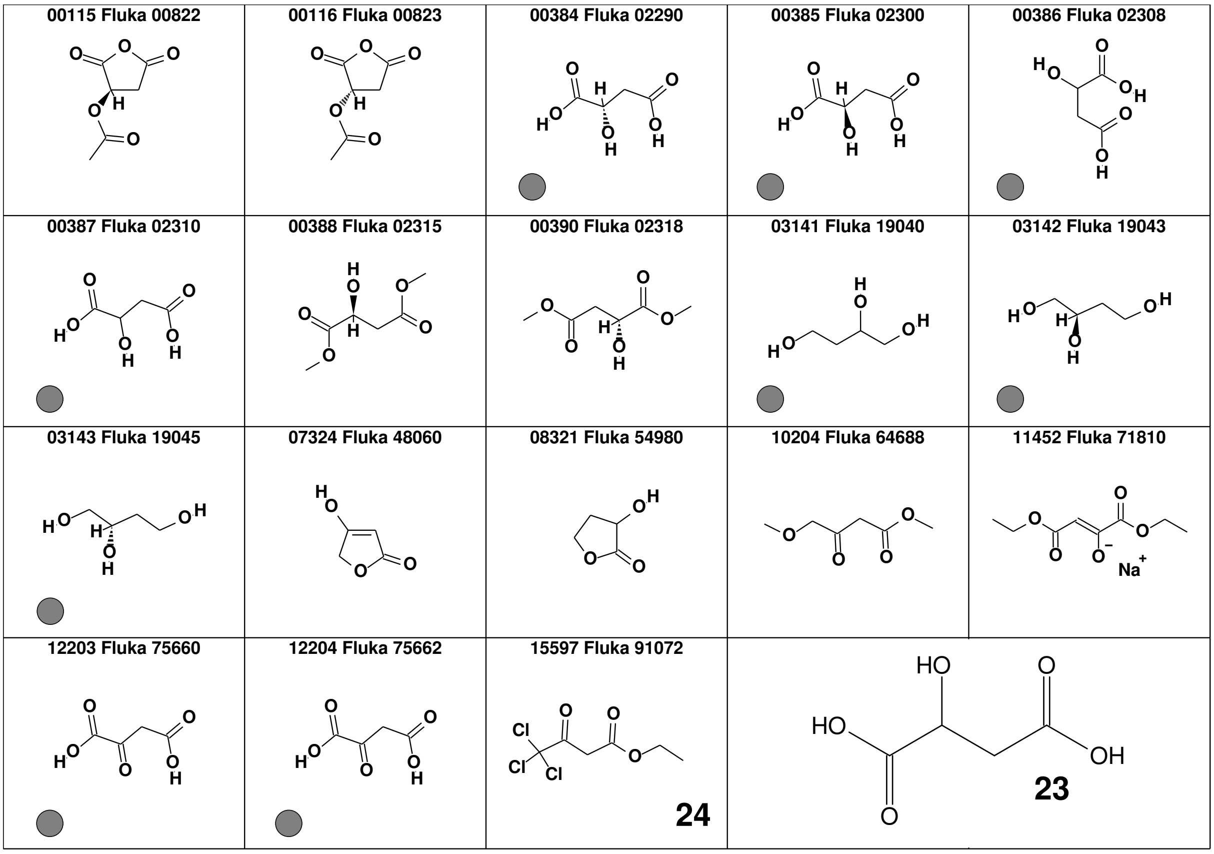


Abb. 3 - 28 Ähnlichkeitssuchen basierend auf Redoxreaktionen für Äpfelsäure (**23**, rechts unten). Die dargestellte Liste von 18 Verbindungen wird mit „C-Skelett inklusive α-Atome und maximale Oxidation“ gefunden. Markierungen sind im Text erläutert.

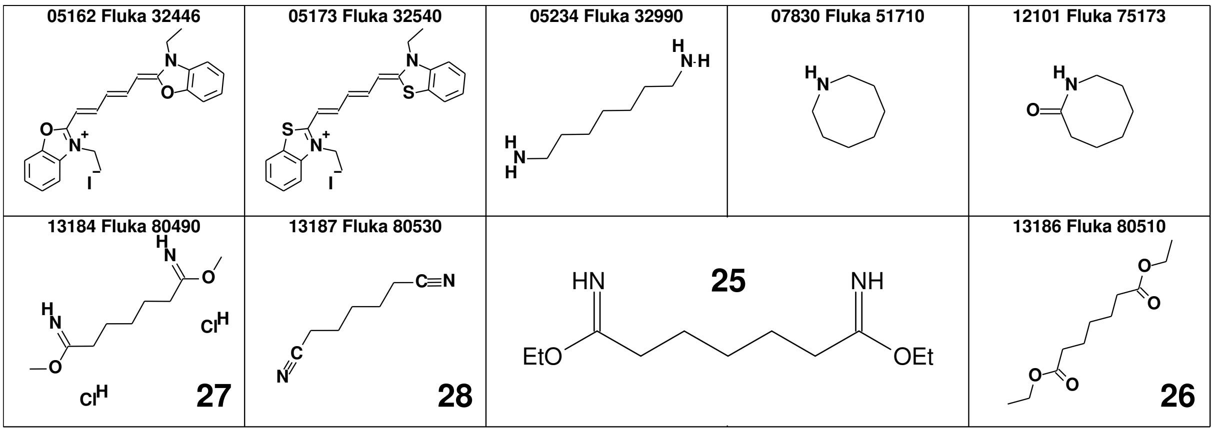


Abb. 3 - 29 Suche nach Ausgangsmaterialien für Pimelindiimidsäurediethylester (**25**, Mitte unten) mittels redox-basierter Transformationen. Sieben Treffer ergibt die Transformationssuche „C-Skelett inklusive α-Atome und maximale Reduktion“ (oben und links). „Maximale Reduktion“ führt zu genau einem Treffer (**26**, rechts unten).

Anwendungshinweise:

⇨ Die redox-basierten Transformationssuchen Oxidation/Reduktion definieren eine chemisch wichtige, wenngleich aber sehr eng begrenzte Ähnlichkeit. Sie sollten daher erst dann angewendet werden, wenn andere, allgemeinere Ähnlichkeitskriterien (etwa die substitutionsbasierten Ähnlichkeitssuchen) unbefriedigend große Trefferlisten ergeben.

⇨ In der Kombination C-Skelett und Oxidation bzw. Reduktion nehmen diese beiden Transformationssuchen eine Mittelstellung zwischen den sehr allgemeinen und den sehr spezifischen Kriterien ein, wobei die Tendenz hin zu den spezifischeren geht. Sie sollten daher noch vor den „reinen“ Oxidations-/Reduktionstransformationen verwendet werden.

3.3.8 Eliminierungen als Grundlage von Ähnlichkeit

Auch Eliminierungen sind ein sehr genereller Reaktionstyp in der organischen Chemie. Das wesentliche an Eliminierungen hinsichtlich ihres Einsatzes als Kriterium von Ähnlichkeit zwischen Verbindungen besteht darin, daß freie Heterosubstituenten unter Bildung von CC-Mehrfachbindungen von einem Skelett entfernt werden. Unter *freien* Heterosubstituenten werden solche Heteroatome verstanden, die genau eine Bindung zu einem anderen Schweratom haben, als weitere Nachbarn also höchstens Wasserstoff-Atome. In der Begriffswelt der Syntheseplanung ist ein solcher Prozeß eine Entfernung funktioneller Gruppen (FGR*). Gleichwohl soll eine Relation zwischen Verbindungen mit CC-Mehrfachbindungen und mit analog einfach oder vicinal substituierten Verbindungen hergestellt werden.

Die wesentlichen Umwandlungen, welche diese Transformationsregel vornimmt, sind (in dieser Reihenfolge):

- ❏ 1,4-Eliminierungen. Als eliminierbar gelten Halogene sowie freie Chalkogene und Elemente der Stickstoff-Gruppe.
- ❏ 1,2-Eliminierungen. Ist keine 1,4-Eliminierung möglich, werden 1,2-Eliminierungen vorgenommen. Für beide gilt: Die Bredtsche Regel wird immer eingehalten, d. h. Doppelbindungen zu Brückenkopfatomen werden nicht gebildet. Bildung von Allenen findet ebenfalls nicht statt. Kommen mehrere C-Atome in Frage, zu denen die Doppelbindung gebildet werden könnte, erhält das C-Atom den Vorzug, welches die höchste σ-Atompartialladung† hat. Dadurch entspricht die vorgenommenen Eliminierung einer Sajtsev-Orientierung.
- ❏ Decarboxylierungen. Freie Carbonsäuregruppen werden eliminiert, wenn sie nicht an ein Brückenkopfatom gebunden sind und in γ-Position ein Donoratom vorhanden ist. Als Donoratome zählen beliebige Schweratome, die über eine (nicht-aromatische) Mehrfachbindung zum Atom in β-Position gebunden sind. Diese Kriterien sichern, daß ein cyclischer Übergangszustand möglich ist, wodurch die Decarboxylierung vonstatten geht.

Hervorhebenswert ist für diese Transformationsregel die Verwendung eines physikochemischen Parameters (σ-Atompartialladung). Üblicherweise basieren die Manipulationen der Transformationsregeln auf topologischen und anderen qualitativen Kriterien. Dem WODCA System stehen jedoch eine ganze Reihe von Berechnungsmethoden zur Verfügung, welche die quantitative Berechnung von physikochemischen Eigenschaften bei gleichzeitiger Sicherstellung der Interaktionsfähigkeit mit dem Benutzer erlauben.

Wie die Redox-Transformationen Oxidation und Reduktion erfordert die Eliminierungs-Transformation eine deutlich enge Verwandtschaft zwischen zwei Verbindungen, damit sie als ähnlich erkannt werden. Das liegt vor allem darin begründet, daß *freie* Substituenten notwendig sind. Jedoch kann auch die Eliminierung als Elementarschritt einer Extraktion der C-Skelette inklusive der α-Atome (unter Erhalt von Aromaten) nachgeschaltet werden (siehe Übersicht 3 - 4). Diese Transformationsregel „C-Skelett inklusive α-Atome und Eliminierung“ führt auch die Ester von decarboxylierbaren Carbonsäuren, Ether und höhere Amine in einer Trefferliste zusammen, definiert also eine großzügigere Ähnlichkeit.

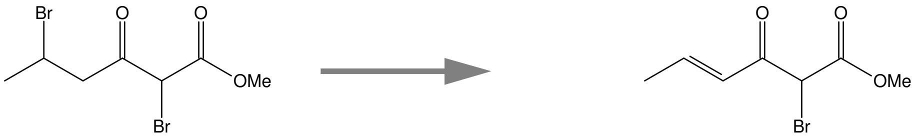
---

* Functional Group Removal

† Die σ-Atompartialladungen werden nach dem Algorithmus von Gasteiger und Marsili [84] während der Transformation berechnet.

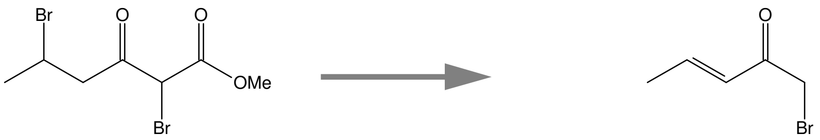
Übersicht 3 - 4 Ähnlichkeitssuchen basierend auf Eliminierungen

Eliminierung



1,4- und 1,2-Eliminierungen freier Substituenten (Halogene, Chalkogene, Elemente der Stickstoff-Gruppe) sowie Decarboxylierung freier Carbonsäuregruppen in Nachbarschaft zu Donoratomen (siehe Text). Nur das größte resultierende Fragment bleibt erhalten.

C-Skelett inklusive α-Atome und Eliminierung



Zunächst werden die C-Skelette extrahiert, wobei unmittelbar daran gebundene Heteroatome (α-Atome) ebenfalls erhalten bleiben. Verbrückt ein Heteroatom mehrere C-Skelette wird dieses dupliziert. Offene Valenzen werden dann gesättigt und dann die Eliminierungsregel angewandt. Nur das am Ende größte resultierende Fragment bleibt erhalten.

Ein Beispiel für die Anwendung der Eliminierungs-Transformation ist die Suche nach zu 2-Methyl-2-penten (**29**) ähnlichen Verbindungen. Die Suche ergibt drei Treffer im Fluka-Katalog. Neben der Anfrageverbindung werden zwei Alkohole gefunden, die sich durch Eliminierung unter Sajtsev-Orientierung genau zu dem gewünschten Olefin umsetzen lassen (siehe Abb. 3 - 30).

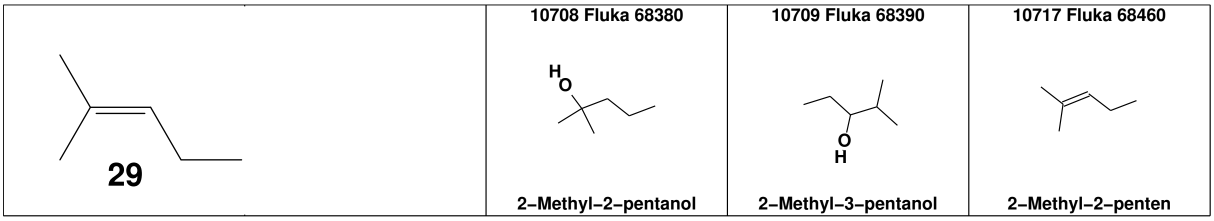


Abb. 3 - 30 Zueinander Ähnliche Verbindungen nach dem Kriterium „Eliminierung“. Ein solch spezifisches Ergebnis ist mittels klassischer Substruktursuchen unmöglich. Links die Anfrageverbindung.

Die Arbeitsweise der Transformation „C-Skelett inklusive α-Atome und Eliminierung“ kann anhand einer Suche nach Lävulinsäure-Bausteinen (**30**) verdeutlicht werden. Sechs Verbindungen werden für Lävulinsäure im Fluka-Katalog gefunden. Die in Abbildung 3 - 31 dargestellten Verbindungen werden durch die Transformation alle in die Anfrageverbindung umgewandelt. α-Angelicalacton (**31**) ist das innere Anhydrid von Lävulinsäure, auch die Methylen-Verbindung (**32**) entspricht nach Hydrolyse und Ketonisierung der enolischen Hydroxylgruppe genau der Anfrage. Bemerkenswert ist der letzte Treffer. Soll ein Lävulinsäure-Baustein etwa an der endständigen Methylgruppe elektrophil angegriffen werden, dann ist 3-Oxo-adipinsäuredimethylester das Ausgangsmaterial der Wahl, da hier diese Position durch die (später decarboxylierbare) Estergruppe doppelt aktiviert ist.

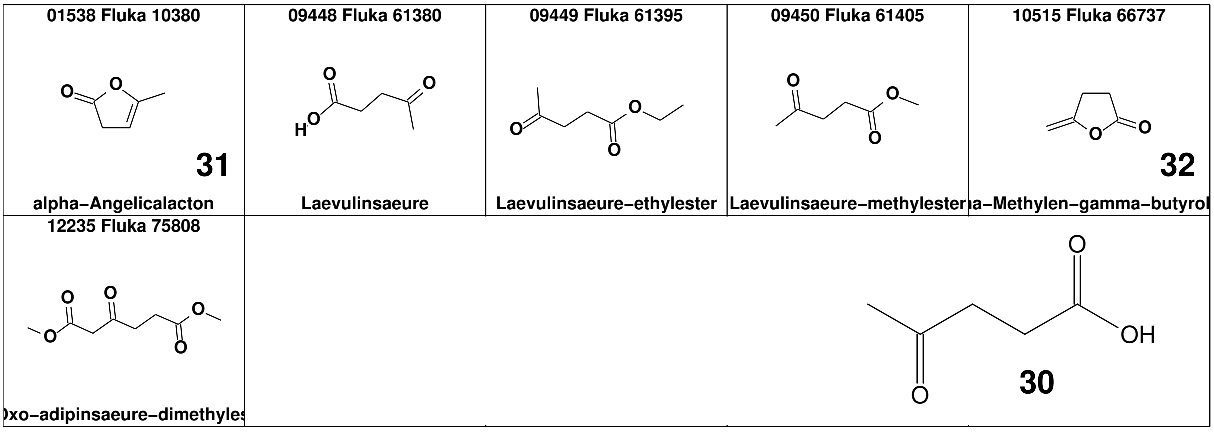


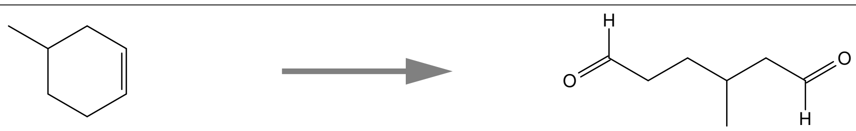
Abb. 3 - 31 Suche nach Lävulinsäure-Analoga mit der Transformation „C-Skelett inklusive α-Atome und Eliminierung“. Im Fluka-Katalog erhält man sechs Treffer.

3.3.9 Sonderfall Ozonolyse-Transformation

Alle bislang vorgestellten reaktionsklassen-orientierten Transformationssuchen manipulieren ausschließlich Bindungen zwischen Kohlenstoff- und Heteroatomen, mitunter wird auch die Bindungsordnung von CC-Bindungen verändert, jedoch werden skeletale Bindungen zwischen Kohlenstoffatomen *nicht* verändert. Die einzige Ausnahme von dieser Regel sind die Eliminierungs-Transformationen (siehe Abschnitt 3.3.8), weil dort freie Carbonsäuregruppen abgespalten werden, wodurch sich das C-Skelett verändert.

Es gibt jedoch Anwendungsfälle in der computer-gestützten Syntheseplanung, wo es wünschenswert ist, daß Verbindungen, deren C-Skelett nicht übereinstimmt, aber aus anderen Gründen in enger Relation steht, mittels eines Ähnlichkeitskriteriums gefunden werden sollten. Beispiele dafür sind 1,6-funktionalisierte Verbindungen, die aus Cyclohexenen erhalten werden können oder Aldehydfunktionen in Verbindungen mit Carbonsäuregruppen, die am einfachsten aus den entsprechenden Methylen-Verbindungen synthetisiert werden. Der diese Relation herstellende Reaktionstyp ist die Ozonolyse.

Übersicht 3 - 5 Ozonolyse als Ähnlichkeitskriterium

Ozonolyse

Nicht-aromatische CC-Doppelbindungen werden gespalten und als Aldehydfunktion gesättigt. Nur das größte Fragment bleibt erhalten.

Ein recht eindrucksvolles Beispiel für den Wert dieser Transformationsregel ist die Suche nach geeigneten Ausgangsmaterialien für den Semialdehyd (**33**) in Abbildung 3 - 32. Zieht man zunächst eine andere naheliegende Transformation heran (Maximale Reduktion), findet man im Janssen-Katalog fünf Asparaginsäure-Derivate. Mittels einer milden Oxidation ließe sich aus *L*-Homoserin (**34**) die Anfrageverbindung synthetisieren. Eine gute Alternative ist jedoch 2-Amino-pent-4-ensäure (**35**), das mit der Transformation „Ozonolyse“ als einzige Verbindung aus dem Katalog selektiert wird. Durch die Titelreaktion entsteht aus dem Ausgangsmaterial glatt der gewünschte Semialdehyd.

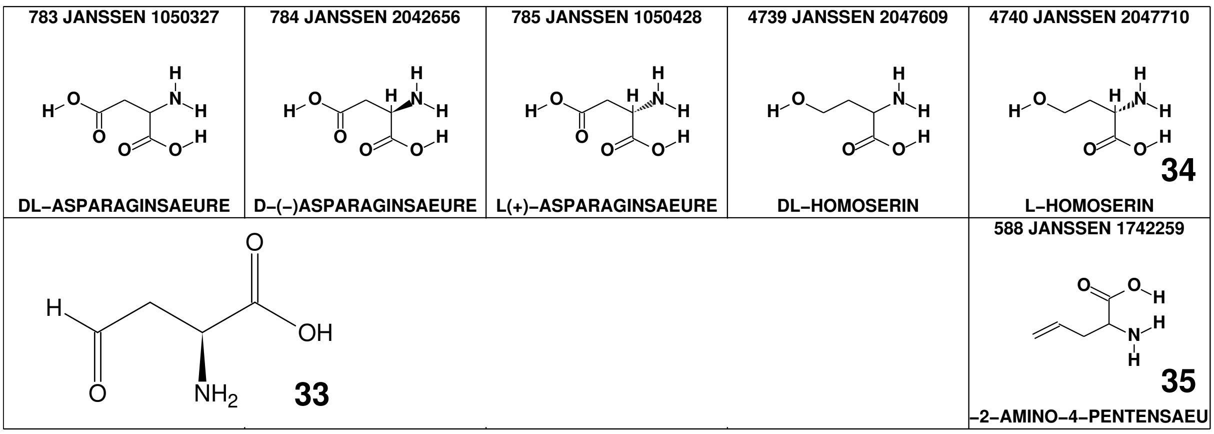


Abb. 3 - 32 Ähnlichkeitssuchen mit Asparaginsäure-semialdehyd (**33**, unten links) im Janssen-Katalog. Eine Suche mit dem Kriterium „Maximale Reduktion“ findet die oberen fünf Verbindungen. Die Transformation „Ozonolyse“ findet als einzige Verbindung das Penten-Derivat (**35**, unten rechts).

Folgende allgemeine Anwendungshinweise lassen sich ableiten:

⇨ Die Ozonolyse definiert eine eng begrenzte Art von Ähnlichkeit. Sie ist besonders für einige Spezialfälle von Interesse, dort aber sehr leistungsfähig.

⇨ Für ausschließlich α,ω-Dialdehyde werden mit dieser Transformation die verwandten Cycloalkene gefunden, sofern sie als Ausgangsmaterial vorhanden sind.

⇨ Für Aldehyde, die z. B. durch Redoxreaktionen aufgrund störender anderer Gruppen nicht hergestellt werden können, werden die um eine C-Einheit längeren Methylen-Verbindungen gefunden, sofern im Katalog enthalten.

Da sich diese Transformationsregel in *tcl* in sehr kompakter Form codieren läßt, sei die Hauptprozedur an dieser Stelle abgedruckt.

```
1  proc ozonolyze { dblock } {
2     set bonds2delete {}
3     foreach bond [ens get $dblock B_LABEL] {
4        if [bond get $dblock $bond B_ISAROMATIC] continue
5        if [bond get $dblock $bond B_ORDER]!=2 continue
6        if ![string match \
7           „C C“ [bond get $dblock $bond A_SYMBOL]] continue
8        lassign [bond get $dblock $bond A_LABEL] at1 at2
9        if $at1>$at2 continue
10       lappend bonds2delete [list $at1 $at2]
11    }
12    foreach pair $bonds2delete {
13       lassign $pair at1 at2
14       bond delete $dblock [findBond $dblock $at1 $at2]
15       bond create $dblock „$at1 [atom create $dblock 8]“ 2
16       bond create $dblock „$at2 [atom create $dblock 8]“ 2
17    }
18    return
19 }
```

Die Zahlen am rechten Rand sind kein Bestandteil des Skripts. Sie dienen nur der Erläuterung der Zeilen. Die Prozedur hat den Namen `ozonolyze` (Zeile 1) und ein Argument (`dblock`), das auf die zu behandelnde Verbindung verweist. Als erstes wird eine Variable initialisiert (Zeile 2). In einer Schleife über alle Bindungen der gerade untersuchten Verbindung (Zeile 3 bis 11) wird für jede Bindung geprüft: Ist die Bindung aromatisch? Wenn ja, wird sie ignoriert (Zeile 4). Ist die Bindung keine Doppelbindung? Wenn ja, wird sie ignoriert (Zeile 5). Für Bindungen, die bislang noch nicht verworfen wurden, wird nun geprüft, ob es sich um CC-Bindungen handelt. Das `bond`-Kommando in Zeile 7 liefert im Fall einer CC-Bindung die Zeichenkette „`C C`“ als Resultat. Wurde eine CC-Doppelbindung gefunden, werden für diese die Labels der Bindungsatome ermittelt (Zeile 8). Da im WODCA System Bindungen doppelt verwaltet werden (Richtung Atom1→Atom2 und Atom1→Atom2), wird mit der `if`-Anweisung in Zeile 9 gesichert, das eine bestimmte CC-Doppelbindung nur einmal prozessiert wird. Nach Abarbeitung der Schleife enthält die Variable `bonds2delete` die Atompaare aller interessierenden CC-Doppelbindungen. In einer zweiten Schleife (Zeile 12 bis 17) werden die Atomlabels ausgelesen (Zeile 13), die Bindung zwischen diesen Atomen gebrochen (Zeile 14) und für jedes Atom eine Doppelbindung zu einem neu erzeugten Sauerstoff (`atom`-Kommando) definiert (Zeilen 15 und 16). Danach kehrt die Prozedur in die aufrufende Ebene zurück. Dort findet noch eine Standard-Nachbehandlung statt (Normalisierung, Entfernung aller außer dem größten Molekül). Beendet wird die Arbeit von *tcl* mit einer Abfrage des Hashcodes für die dann mit `dblock` referenzierte Verbindung. Der Vergleich dieses Hashcodes mit denen der Katalogverbindung geschieht dann außerhalb von *tcl*.

Sicher ist das oben abgedruckte Skript kein Klartext. Jedoch wird jeder ein wenig mit Programmierung vertraute Chemiker relativ leicht den beschriebenen Algorithmus der Ozonolyseregel erkennen können. Ebenso einfach ist es dann natürlich auch, weitere solche Regeln zu programmieren.

3.3.10 Kombinationen von Elementarschritten

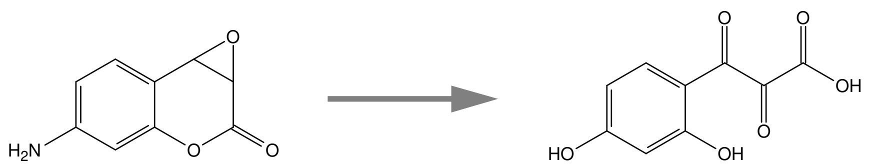
In einigen der vorangegangenen Abschnitten wurden bereits elementare Ähnlichkeitskriterien durch Kombinationen zu neuen Transformationen verknüpft. Hier sollen nun weitere solche Kombinationen vorgestellt werden.

Es wurde gezeigt, daß Hydrolyse- und Oxidations-Transformationen (siehe Abschnitte 3.3.6 bzw. 3.3.7) jeweils für sich wertvolle Ähnlichkeitskriterien für die Syntheseplanung darstellen. Als Kombinationen erhält man jedoch eine besonders leistungsfähige Transformationssuche, die in ihrer Mächtigkeit mit den allgemeineren Substitutionssuchen (siehe Abschnitt 3.3.4) vergleichbar ist, bei gerade für den Einsatz im Bereich der computer-assistierten Syntheseplanung höherer chemisch-synthetischer Relevanz.

Zwei andere vom Charakter her sehr allgemeine Ähnlichkeitskriterien definieren auch die Kombination der Eliminierung bzw. Ozonolyse mit einer Substitutionsmuster-basierten Transformation. Diese beschreiben Ähnlichkeiten, die im Bereich der allgemeinsten Substitutions-Transformation (*-AR+A1-CCMB-BO-MU*) liegen oder gar darüber hinausgehen.

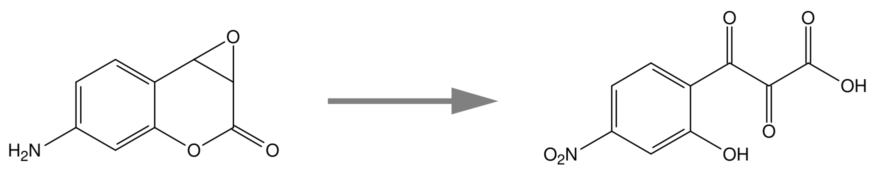
Übersicht 3 - 6 Kombinationen von Elementaroperationen zur Definition leistungsfähigerer Ähnlichkeitskriterien

Hydrolyse (einschließlich Aromatensubstituenten[a]) und Oxidation



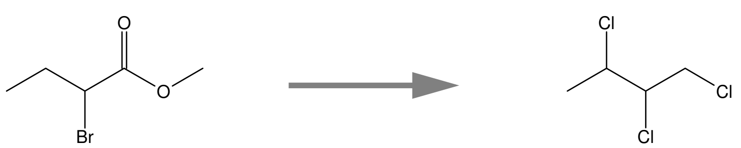
Zuerst werden Hydrolyseschritte, wie in Abschnitt 3.3.6 beschrieben, durchgeführt. Anschließend werden oxidierbare Atome (substituierte Kohlenstoffe, Chalkogene, Pnictogene) in ihre jeweils höchsten Oxidationszustand überführt. Nur das am Ende größte Fragment bleibt erhalten.

Hydrolyse (ohne Aromatensubstituenten[b]) und Oxidation



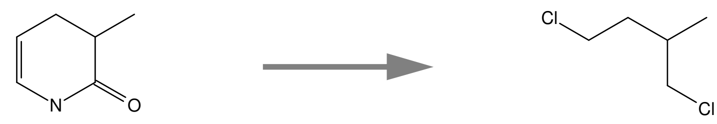
Wie voranstehend, jedoch werden zu Aromatensystemen α-ständige Heterosubstituenten von der Hydrolyse ausgenommen.

Eliminierung und Substitutionsmuster (-AR+A1+CCMB-BO-MU)



Zunächst Durchführung von Eliminierungen gemäß Abschnitt 3.3.8. Anschließend Bestimmung des Substitutionsmusters unter Beachtung von CC-Mehrfachbindungen aber Ignorierung der Zahl der Substituenten je Skelettatom und der Bindungsordnung zwischen Skelett- und Heteroatom. Nur das am Ende größte resultierende Fragment bleibt erhalten.

Übersicht 3 - 6 (Forts.) Kombinationen von Elementaroperationen zur Definition leistungsfähigerer Ähnlichkeitskriterien

Ozonolyse und Substitutionsmuster (-AR+A1-CCMB-BO-MU)

Zunächst Anwendung der Ozonolyseregel nach Abschnitt 3.3.9. Danach Ermittlung des Substitutionsmuster wobei CC-Mehrfachbindungen nicht als Substituenten gezählt werden und auch der Zahl der Substituenten je Skelettatom und der Bindungsordnung zwischen Skelett- und Heteroatom keine Bedeutung zukommt. Nur das am Ende größte Fragment bleibt erhalten.

a. In der Benutzeroberfläche von WODCA mit „with alphas“ bezeichnet.
b. In der Benutzeroberfläche von WODCA mit „no alphas“ bezeichnet.

Als instruktives Beispiel dient 2,3-Dibrombernsteinsäure (**36**). Suchen im Janssen-Katalog ergeben für das allgemeinste Substitutionsmuster (*-AR+A1-CCMB-BO-MU*) 38 Treffer. Das Substitutionsmuster, bei dem die Oxidationsstufen sich entsprechen müssen (*-AR+A1-CCMB-BO-MU*), resultiert in 13 Treffern. Diese 13 Treffer werden auch mit dem Kriterium „Hydrolyse“ erzielt. Mit dem kombinierten Kriterium „Hydrolyse und Oxidation“ erhält man jedoch nur 30 Treffer. Diese sind sämtlich in den 38 Treffern der ersten Substitutionsmuster-Suche enthalten. Die fehlenden acht Treffer sind in Abbildung 3 - 33 dargestellt.

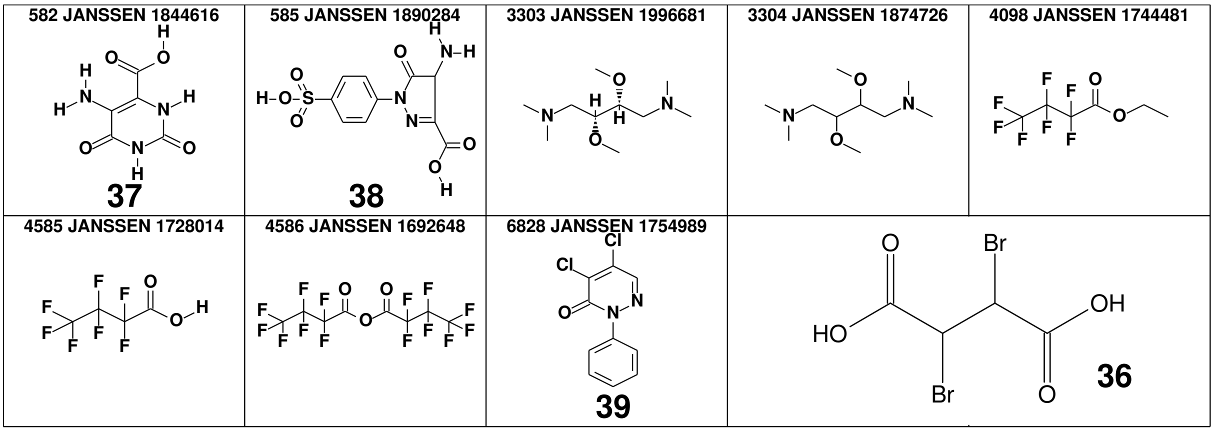


Abb. 3 - 33 Acht Treffer, die für 2,3-Dibrombernsteinsäure (unten rechts) mit der Transformationssuche „Substitutionsmuster (-AR+A1-CCMB-BO-MU)“ gefunden werden, nicht aber mit „Hydrolyse und Oxidation“.

Die fluorhaltigen Verbindungen werden bei Hydrolyse-Transformationen nicht gefunden, da Fluorsubstituenten als nicht hydrolysierbar gelten. Auch die Methylamin- bzw. Methylether-Verbindungen fallen deshalb heraus, weil solche Substituenten nicht hydrolysiert werden. Etwas schwieriger erkennbar ist, warum das Pyrimidin- (**37**) und die Hydrazin-Derivate (**38**, **39**) von der kombinierten Transformation nicht gefunden werden. Das verdeutlicht Abbildung 3 - 34. Im Falle des Pyrimidin-Derivats wird nur die *freie* Aminogruppe hydrolysiert. In den anderen beiden Fällen sind andere Molekülteile nach Anwendung der Transformation größer.

Ein Beispiel dafür, wie nützlich eine Kombination mit einem Substitutionsmuster-Kriterium sein kann, ist die Suche nach Ausgangsmaterialien für 2,4-Dibrom-2-methyl-butan (**40**).

Zur Erinnerung: Für diese Verbindung wird mit dem Substitutionsmuster allein (-*AR+A1±CCMB-BO-MU*), keine Verbindung im Janssen-Katalog gefunden. Die kombinierte Transformation „Eliminierung und Substitutionsmuster“ findet dagegen 13 Verbindungen (siehe Abb. 3 - 35). Durch die Transformation treffen sich alle diese Verbindungen bei Isopren. Das wahrscheinlich beste Ausgangsmaterial für die Anfrageverbindung 2,4-Dibrom-2-methyl-butan dürfte 3-Hydroxy-3-methyl-but-1-en sein (**41** in Abb. 3 - 35). Bei dessen Umsetzung mit HBr durch elektrophilen Angriff entsteht das gewünschte Markovnikov-Produkt. Der Austausch der Hydroxylgruppe durch eine Substitution führt zur gewünschten Verbindung 2,4-Dibrom-2-methyl-butan (**40**).

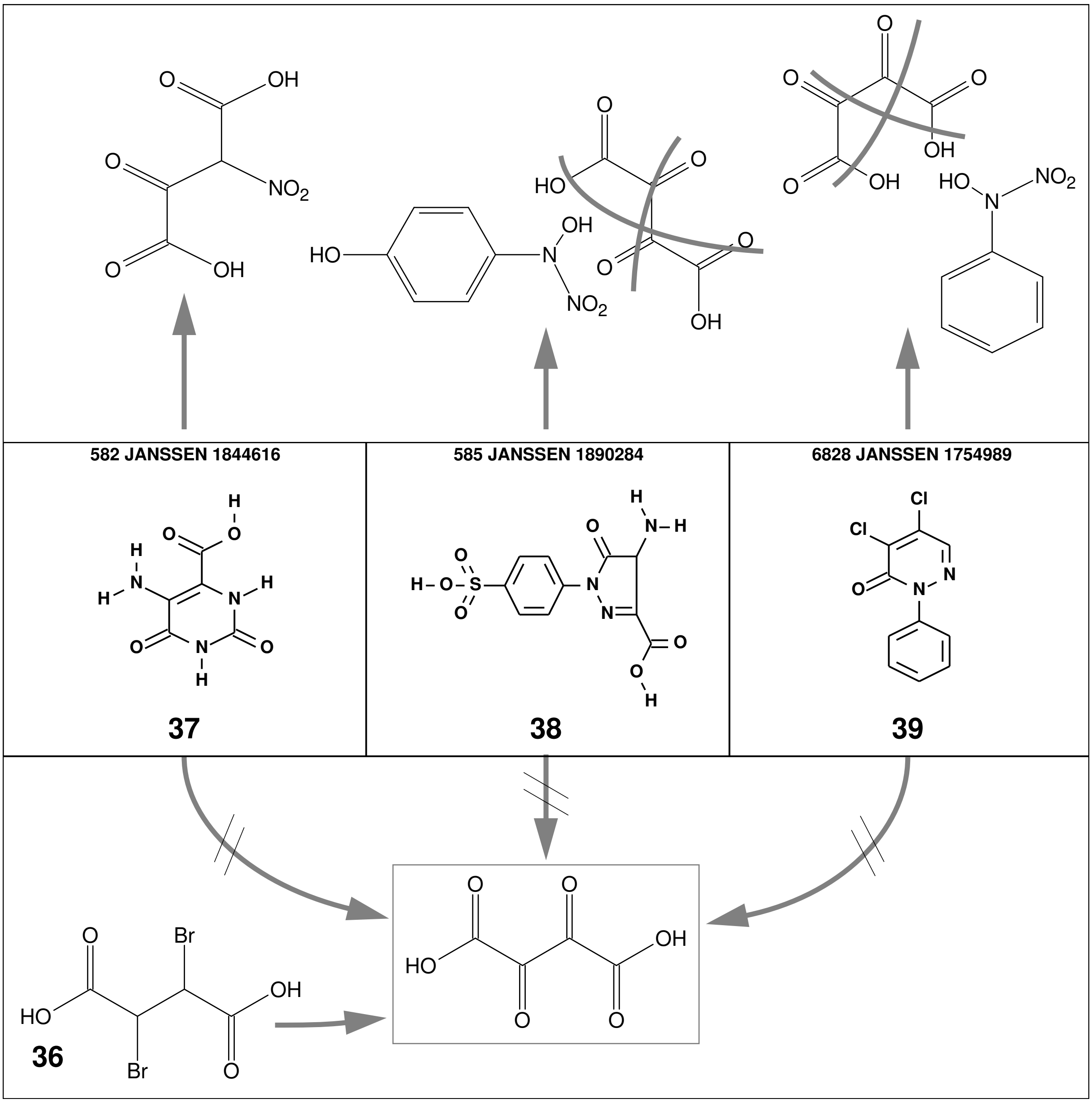


Abb. 3 - 34 Resultierende Strukturen nach Anwendung der Transformation „Hydrolyse und Oxidation“ (oben). Unten in der Mitte die Struktur, die durch die Transformation erhalten werden müßte, damit obige Katalogverbindungen als ähnlich zu 2,3-Dibrombernsteinsäure (**36**, unten links) erkannt würden.

Die kombinierte Transformation aus „Eliminierung und Substitutionsmuster“ kann also besonders dann eine wertvolle Ergänzung sein, wenn die substitutionsmuster-basierten Transformationen allein kein oder kein befriedigendes Ergebnis erzielen. Dasselbe gilt auch für die Kombination „Ozonolyse und Substitutionsmuster". Als Beispiel dient das Triol (**42**) aus Abbildung 3 - 36. Im Janssen-Katalog mit dem Substitutionsmuster allein (*-AR+A1 -CCMB-BO-MU*) werden keine ähnlichen Verbindungen gefunden. Das gilt bei Zählung von CC-Mehrfachbindungen als Substituenten ebenfalls (*-AR+A1***+CCMB***-BO-MU*). Die kombinierte Transformation „Eliminierung und Substitutionsmuster" findet dagegen den Treffer 2-Methyl-pentan-2,4-diol (**43** in Abb. 3 - 36). An dieser Stelle viel besser geeignete Ausgangsmaterialien ermittelt jedoch die Kombination „Ozonolyse und Substitutionsmuster". Die freie 3,4-Dihydro-2,2-dimethyl-γ-pyron-6-carbonsäure (**44**) bzw. ihr Butylester (**45**) werden von der Ozonolyse genau in die 1,2,4-sauerstoff-substituierte C_6-Einheit gespalten, die mit ihrem Substitutionsmuster der Anfrageverbindung entspricht. Eine Reduktion der Carbonylgruppen führt zu dem entsprechenden Triol.

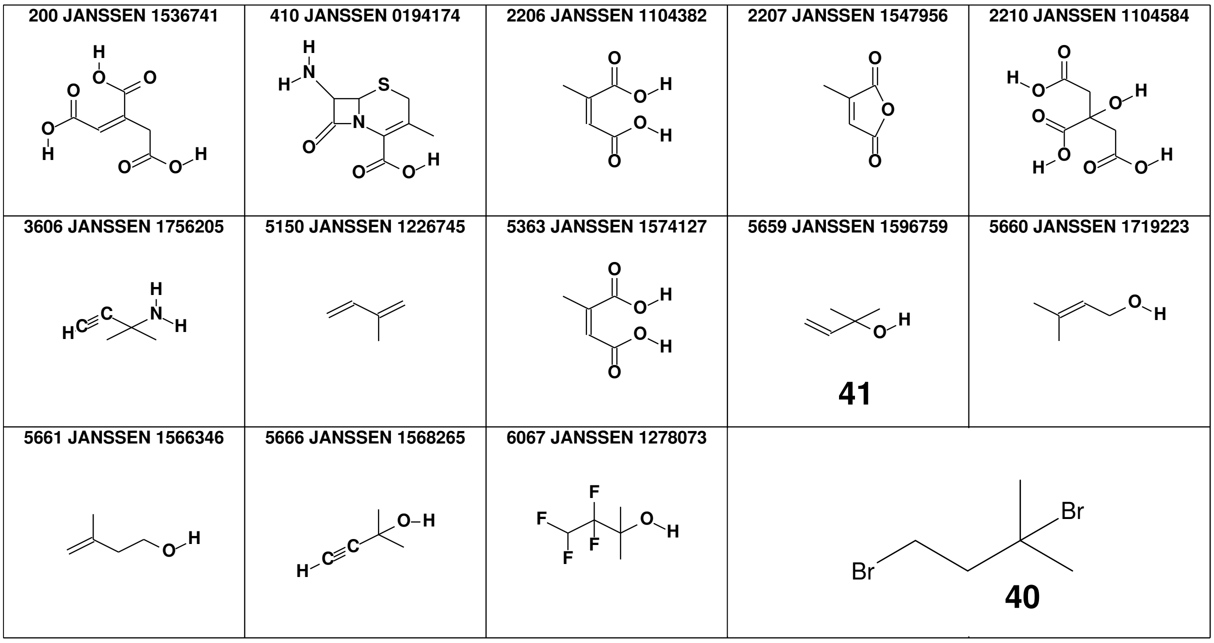


Abb. 3 - 35 Ergebnis einer Suche mit der Transformation „Eliminierung und Substitutionsmuster" für die Anfrageverbindung 2,4-Dibrom-2-methyl-butan (**40**, rechts unten): 13 Verbindungen werden im Janssen-Katalog gefunden.

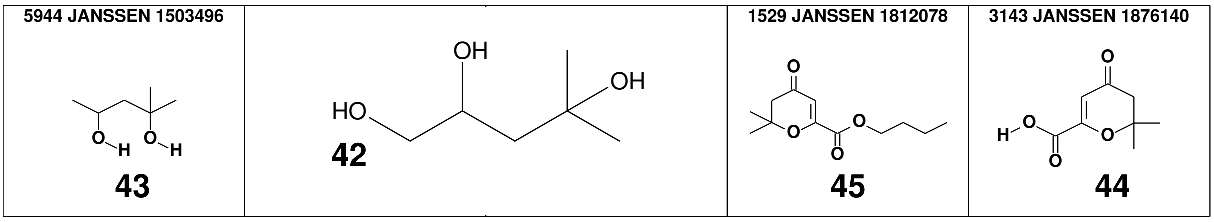


Abb. 3 - 36 Suchergebnisse für 2-Methyl-pentan-2,4,5-triol (**42**, links) im Janssen-Katalog. Die Verbindung (**43**) wird mit der Transformation „Eliminierung und Substitutionsmuster" gefunden, die beiden anderen mit „Ozonolyse und Substitutionsmuster".

Aus den Erfahrungen mit den kombinierten Transformationen resultieren die folgenden Anwendungshinweise:

⇨ Die Kombination aus Hydrolyse und Oxidation stellt eine Alternative zu den allgemeinsten Substitutions-Transformationen dar. Ihre Ähnlichkeitsdefinition ist jedoch restriktiver und stärker an tatsächlich durchführbaren synthetischen Umsetzungen orientiert.

⇨ Werden Aromatensubstituenten von der Hydrolyse ausgenommen, wird der Ähnlichkeitsbereich eingeengt. Die Verwendung dieser Transformation ist angezeigt, wenn die analoge Transformation mit Hydrolyse von Aromatensubstituenten eine extrem große Trefferliste erzeugt.

⇨ Die Kombination von Eliminierung bzw. Ozonolyse mit einem Substitutionsmuster stellen eine exzellente Ergänzung zu den allgemeinen Substitutions-Transformationen dar. Sie sollten besonders dann zum Einsatz kommen, wenn die Substitutions-Transformationen allein keine befriedigenden Resultate erzielen.

3.3.11 Benutzerschnittstelle zu den Transformationssuchen

Wie in den Zielsetzungen für diese Arbeit (siehe Abschnitt 2.2.5) dargelegt, mußte auch die grafische Benutzeroberfläche des WODCA Systems aus programmtechnischen (überholter Software-Standard) wie aus ergonomischen Gründen (nicht ausreichend benutzerfreundlich und intuitiv) umgestaltet werden. An dieser Stelle soll dargelegt werden, wie sich die Transformationssuchen heute dem Anwender präsentieren.

Der Zugriff auf die Transformationssuchen im alten WODCA System veranschaulicht Abbildung 3 - 37. Über ein Menü wird auf die komplette Liste aller verfügbaren Transformationsregeln zugegriffen. Zweifellos wird ein gelegentlicher bzw. unerfahrener Nutzer des Programms durch die Fülle der Möglichkeiten überfordert. Was fehlt, ist eine Hilfestellung bei der Auswahl der für das aktuelle Syntheseproblem passendsten Transformationsregel.

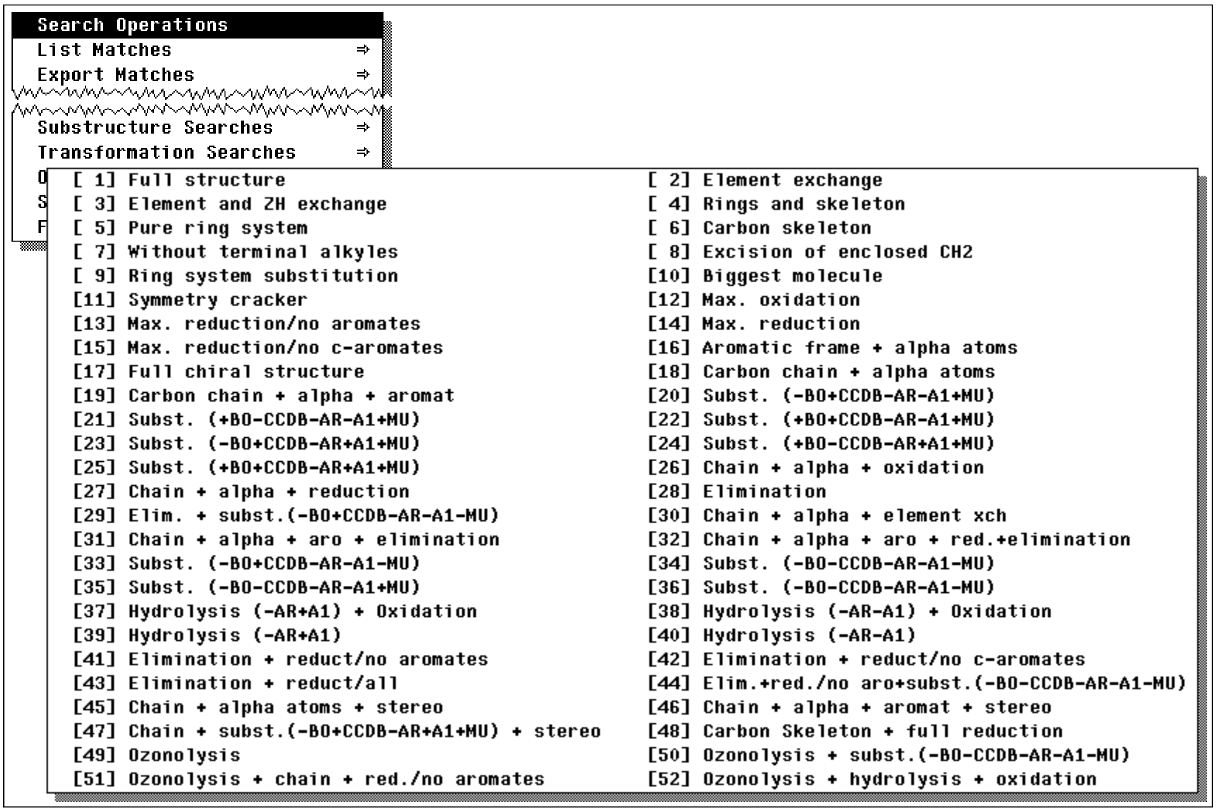


Abb. 3 - 37 Zugriffsweg auf die Transformationssuchen im alten WODCA System.

In der Version 3 des WODCA Systems präsentieren sich die Transformationssuchen in Form eines Pop-up-Dialoges, der aus dem Hauptmenü erreichbar ist (siehe Abb. 3 - 38). Das Dialogfenster ist mehrfach unterteilt. Im rechten Bereich finden sich die textuellen Titel der Transformationsregeln, jetzt jedoch als eine Abrollliste. Darüber befinden sich Auswahlmenüs, welche die Menge der angebotenen Transformationsregeln beeinflussen.

Über den Schalter Ähnlichkeitsbereich (*Similarity Range*) legt der Anwender fest, ob er ein mehr oder minder großzügiges Ähnlichkeitskriterium verwenden möchte. Anders ausgedrückt: Der Benutzer gibt an, daß er eher wenig Treffer erhalten möchte bzw. zunächst einmal viel Treffer verlangt. Er formuliert damit Randbedingungen für den Ähnlichkeitsbereich einer Transformationssuche. Als Wahlmöglichkeiten stehen zur Verfügung:

- ❑ *Narrow*
  Ein enger Ähnlichkeitsbereich wird gefordert. Das bedeutet, daß nur solche Transformationen interessieren, die eine sehr limitierte Variabilität in den verwandten Strukturen zulassen. Beispiele wären die Transformationen „Maximale Reduktion“, Maximale Oxidation“, „Hydrolyse“, „Eliminierung“.
- ❑ *Broad*
  Das andere Extrem: Möglichst großzügige Kriterien für Ähnlichkeit (und damit mehr Treffer) werden angefordert. Beispiele hierfür sind „C-Skelett“, „Größtes Ringsystem“ oder „Substitutionspositionen auf dem reduzierten Ringsystem“.
- ❑ *Intermediate*
  Der Zwischenbereich umfaßt alle Transformationsregeln, die nicht zu den Extrembereichen zählen.
- ❑ *Unspecified*
  Diese Option bedeutet, daß die angezeigten Transformationsregeln ohne Angabe eines Ähnlichkeitsbereiches aus der Gesamtzahl aller verfügbaren Transformationen selektiert werden. Diese Option wird für einen routinierten Benutzer von Vorteil sein.

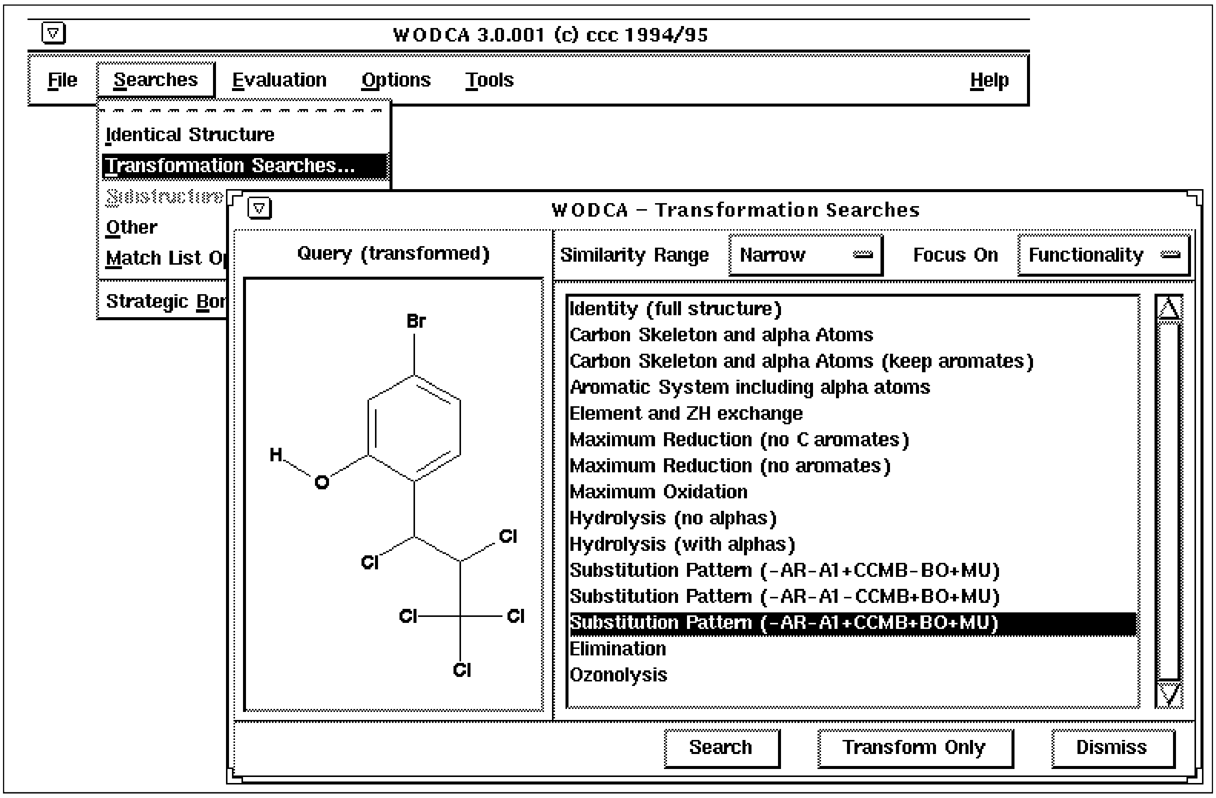


Abb. 3 - 38 Zugriff auf die Transformationssuchen im WODCA System Version 3. Oben das Hauptmenü des Programms mit dem aktivierten Suchen-Menü. Rechts unten das Dialog-Fenster zur Auswahl der Transformationen.

Der Schalter Schwerpunkt (*Focus On*) stellt dem Chemiker eine weitere interessante Möglichkeit bereit, Randbedingungen für die Auswahl der Transformationsregeln festzulegen. Hier dient die chemische Natur, dienen spezielle Merkmale der gerade untersuchten Anfrageverbindung als Auswahlkriterium. Besitzt eine Verbindung ein komplexes Ringsystem, so wird der Anwender vor allem an Ähnlichkeitskriterien interessiert sein, die dieses molekulare Merkmal berücksichtigen. Soll eine offenkettige Verbindung gesucht werden, ist wahrscheinlich die Verteilung von funktionellen Gruppen das interessanteste Kriterium dafür, was ähnlich sein soll. Dazu dienen die folgenden Optionen:

- *Functionality*
  Es werden Transformationsregeln ausgewählt, die die Funktionalitätsverteilung auf dem Molekülskelett berücksichtigen. Typische Beispiele sind die Substitutions-Transformationen, aber auch die Redox-Transformationen u. a.
- *Skeleton*
  Hier werden Ähnlichkeitskriterien angefordert, die sich auf das Skelett einer Verbindung konzentrieren. Funktionalität spielt nur eine untergeordnete Rolle. Dazu gehören „Reduziertes C-Skelett“, „Größtes Ringsystem“ oder „Ringe und C-Skelett“.
- *Rings*
  Diese Option ist wie die folgende (*Aromaticity*) ein Spezialfall der Option Skeleton. Sie kann gleichermaßen wichtig sein bei Verbindungen mit Ringen als zentrale Bestandteile wie auch dann, wenn nur Teile eines größeren Moleküls Ringe enthalten. Die Ähnlichkeitssuchen finden dann selektiv für diese Molekülteile interessante Ausgangsmaterialien.
- *Aromaticity*
  Sehr eng verwandt mit der voranstehenden Option (*Rings*). Stellt jedoch eine nützliche Randbedingung bei der Suche nach Ausgangsmaterialien für Verbindungen mit Aromatensystemen dar.
- *Unspecified*
  Wie bei dem anderen Auswahlmenü kann die Auswahl der Transformationsregeln unbeeinflußt von molekularen Merkmalen bleiben.

Auf der linken Seite, überschrieben mit *Query (transformed)*, befindet sich eine Zeichenfläche, auf der bei Anwendung einer Transformation das aus der eingesetzten Anfrageverbindung resultierende Transformationsprodukt dargestellt wird. Dem Benutzer wird so ein Instrument bereitgestellt, das ihm erlaubt, die Wirkung einer gewählten Transformationsregel visuell zu verfolgen. Die transformierte Anfrageverbindung ist diejenige Verbindung, in die alle Verbindungen einer Trefferliste überführt werden, quasi das Gemeinsame zwischen Anfrage und Treffern.

Am unteren Rand des Dialogfensters befinden sich außerdem eine Reihe von Knöpfen, welche die eigentlichen Aktionen (z. B. den Suchvorgang) auslösen. Während mit *Search* die Transformationssuche gestartet wird, wodurch in jedem Falle eine Trefferliste, u. U. eine leere Liste, resultiert, kann mittels *Transform Only* die Ähnlichkeitsdefinition nur auf die Anfrage angewendet werden, etwa um vor der eigentlichen Suche zu prüfen, ob die beabsichtigte Transformation stattfindet. Praktisch dauern aber beide Aktionen annähernd gleich lang. Geschwindigkeitsbestimmend sind zum einen die Manipulation der Anfrageverbindung und die grafische Darstellung des Transformationsresultates, die in beiden Fällen stattfinden. Die tatsächliche Suche in einem Katalog (der Vergleich von Hashcodes, siehe Abschnitt 2.2.1) fällt dagegen kaum ins Gewicht.*

Die substitutions-basierten Transformationen wurden besonders berücksichtigt. Um dem Anwender das Verständnis dafür zu erleichtern, wie die verschiedenen Schalter die Definition eines Substitutionsmusters beeinflussen (siehe Abschnitt 3.3.4), wurde ein zusätzlicher Dialog implementiert (siehe Abb. 3 - 39). Dort ist verbal umschrieben, welche Wirkung etwa der Schalter *A1* hat. Der Anwender selektiert Kombinationen nach seinen Wünschen und das Pro-

* Das gilt für Kataloge bis zu einer Größenordnung von etwa 20.000 Einträgen (Größter, dem WODCA System bei Abschluß dieser Arbeit zur Verfügung stehendender Katalog: 20.441 Einträge).

gramm aktiviert die betreffende Transformation.* Dieser Dialog kann für erfahrene Anwender deaktiviert werden.

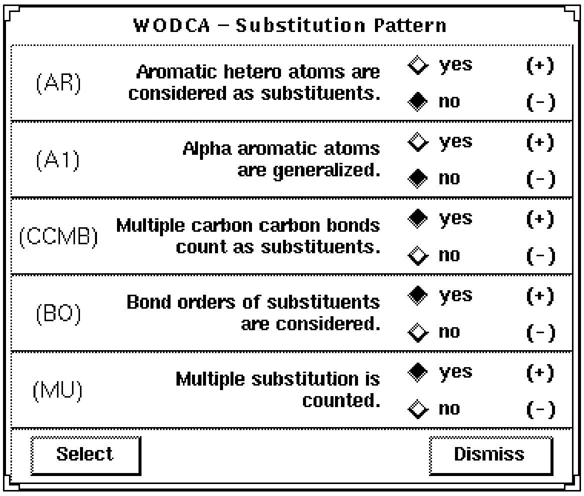


Abb. 3 - 39 Zusätzliches Dialog-Fenster zur Auswahl von Substitutions-Transformationen.

Auf diese Weise wird es auch dem unerfahrenen Nutzer möglich, auf einfachem, intuitivem Weg eine geeignete Transformation, zugeschnitten auf sein Syntheseproblem, aus der doch recht umfangreichen Menge vorhandener Transformationssuchen (39 bei Abschluß dieser Arbeit) auszuwählen und das Prinzip ihrer Ähnlichkeitsdefinition zu erfassen.

* Es ist durchaus möglich, das Kombinationen gewählt werden, die nicht zur Verfügung stehen. Darüber informiert ggf. das Programm den Benutzer.

3.4 Andere Anwendungen

An dieser Stelle sollen einige Anwendungen der Transformationssuchen erläutert werden, die nicht unmittelbar zur Syntheseplanung dienen, aber dennoch wertvolle Informationen liefern.

Da für einen Katalog Tabellen von Hashcodes gemäß den Transformationsregeln ermittelt werden, kann man diese Tabellen benutzen, um folgende Fragestellungen zu beantworten:

- ❏ Wie unterscheiden sich zwei Kataloge?
- ❏ Welche Vielfalt an chemischen Strukturen birgt ein Katalog?

Um die erste Frage zu konkretisieren: Man ist an der Zahl der Verbindungen interessiert, die exklusiv nur in einem Katalog vorkommen und an der Zahl der Verbindungen, die in beiden Katalogen vorhanden sind. Aus diesen Zahlen läßt sich abschätzen, inwieweit ein weiterer Katalog zusätzliche chemische Information in sich birgt. Denn ein weiterer Katalog, der kaum neue Verbindungen bietet, stellt kaum einen zusätzlichen Wert dar. Auch die Frage, ob eine bestimmte Verbindung schon in einem Katalog vorhanden ist, läßt sich leicht beantworten. Tabelle 3 - 10 berichtet über eine Analyse der WODCA-Kataloge.

Tab. 3 - 10 Katalog-Analyse: Zahl gemeinsamer Einträge (Überlappungen) und Zahl der nur in einem Katalog vorhandenen Verbindungen.

Katalog	verschiedene[a] Einträge	gemeinsam mit Katalog			Einträge in Katalog	... und nicht in Katalog		
		Janssen	Fluka	Trivial namen		Janssen	Fluka	Trivial-namen
Janssen	7.847		4.147	1.449	Janssen		3.700	6.398
Fluka	10.941			1.677	Fluka	6.794		9.264
Trivial-namen	17.084				Trivial-namen	15.635	15.407	

a. Verschiedene Einträge ist die Zahl der nicht identischen Verbindungen. Manche Verbindungen erscheinen mehrfach in Katalogen, etwa weil sie in verschiedenen Reinheitsgraden angeboten werden. Stereochemische Unterschiede werden nicht berücksichtigt.

Die Überschneidung des Janssen- mit dem Fluka-Katalog ist zwar deutlich (über 4.000 gemeinsame Verbindungen, das sind mehr als 50 % bezogen auf den Janssen-Katalog), andererseits auch nicht überraschend, da ja beide Kataloge von Feinchemikalien-Anbietern stammen. Daß dagegen der Trivialnamen-Katalog nur 1.449 bzw. 1.677 Verbindungen mit den Katalogen Janssen bzw. Fluka gemeinsam hat, konnte erwartet werden. Eine entsprechende Analyse ermittelt eine Gesamtzahl von 29.742 verschiedenen Verbindungen aus der Gesamtzahl von 45.674 Einträgen der drei Kataloge Janssen, Fluka und Trivialnamen.

Gefragt nach der chemischen Vielfalt eines Kataloges, wird hier wieder im Hinblick auf die Syntheseplanung. Vielfalt bedeutet dann: Wie viele verschiedene C-Skelette? Wieviel verschiedene Ringsysteme? Wie viele verschiedene Substitutionsmuster? etc. Diese Fragen können anhand der durch die entsprechenden Transformationsregeln resultierenden Hashcode-Tabellen leicht beantwortet werden. Die Zahlen enthält Tabelle 3 - 11. Man beachte beispielsweise, daß die Zahl der verschiedenen C-Skelette in den Katalogen Janssen und Fluka nur ein Bruchteil der tatsächlichen Anzahl an Verbindungen ausmacht (11,5 bzw. 9,3 % der Zahl verschiedener Einträge). Noch niedriger liegen diese Werte, wenn man auch CC-Mehrfachbindungen ignoriert (Transformation „Reduziertes C-Skelett“). Nur 516 bzw. 603 verschiedene C-

Skelette enthalten dann die Kataloge Janssen und Fluka. Betrachtet man die reinen Ringsysteme gibt es davon 461 bzw. 629 verschiedene bei Janssen bzw. Fluka. Deutlich höher liegt diese Zahl natürlich für den Trivialnamen-Katalog (3.302). Hier sind ja auch statt Bausteinen für Synthesen Naturstoffe zu finden.

Tab. 3 - 11 Variationen einzelner struktureller Merkmale in verschiedenen WODCA-Katalogen. Die Zahlen geben die Anzahl der nach der entsprechenden Transformation verschiedenen Verbindungen je Katalog an.

Kriterium (Transformation)	Janssen	Fluka	Trivial-namen
Gesamtzahl verschiedener Einträge (Identität)	7.847	10.941	17.084
Gesamtzahl verschiedener Einträge unter Berücksichtigung von Stereochemie und Isotopen (Stereochemische Identität)	7.940	11.533	
C-Skelett	905	1.020	4.949
Reduziertes C-Skelett	516	603	2.650
Ringe und C-Skelett	1.591	1.926	7.913
Ringsystem	461	629	3.302
Substitutionspositionen auf dem Ringsystem	1.145	1.326	6.474
Substitutionspositionen auf dem reduzierten Ringsystem	855	1.049	5.107
Substitutionsmuster (-AR+A1-CCMB-BO-MU)	2.113	2.098	8.098
Substitutionsmuster (-AR+A1+CCMB-BO-MU)	2.234	2.255	8.876
Substitutionsmuster (-AR+A1-CCMB+BO+MU)	2.656	2.617	9.614
Substitutionsmuster (-AR+A1+CCMB+BO+MU)	2.864	2.878	10.500
Eliminierung und Substitutionsmuster	2.295	2.335	9.386
Ozonolyse und Substitutionsmuster	2.030	2.049	8.183

Schrittweise Retrosynthese - Suche nach strategischen Bindungen und Erzeugung von Vorstufen

In diesem Kapitel werden die Mechanismen vorgestellt, mit denen im WODCA System schrittweise Vorstufen abgeleitet werden können. Mehrere Algorithmen sind verwirklicht worden, die verschiedene Strategien zur Lösung dieses Problems anbieten.

4.1 Das Problem

Es kann verschiedenste Gründe dafür geben, daß die im Kapitel 3 beschriebenen Methoden zur Suche geeigneter Ausgangsmaterialien nicht zu befriedigenden Ergebnissen führen. Die häufigsten sind:

- Das Syntheseziel ist zu komplex. Es kann nicht erwartet werden, daß ein ähnlich komplexes Molekül als Ausgangsmaterial zur Verfügung steht.
- Die Suchen lassen sich nicht in die gewünschte Richtung steuern. Das kann bei den Transformationssuchen der Fall sein, wenn z. B. ein Heteroatom die Zielstruktur in zwei vergleichbar große Fragmente separiert. Die Transformationssuchen werden dann immer nur für eines dieser Fragmente Ausgangsmaterialien liefern.
- Die Suchen führen zu Treffern, die nur noch bedingt auf synthetisch verwendbare Ähnlichkeit basieren. Die Transformationssuche mit dem Kriterium „Reduziertes C-Skelett“ ist dafür ein Beispiel. Unter Umständen werden damit nur gesättigte Alkane gefunden, zwar mit dem korrekten Skelett, aber ohne jegliche Funktionalität, die für erforderliche Umsetztungen wichtig wäre.

Die Aufgabe besteht dann darin, einfachere, weniger komplexe Verbindungen aus der Zielverbindung abzuleiten, für die dann erneut nach Ausgangsmaterialien gesucht werden kann. Aus solchen abgeleiteten Verbindungen (den Vorstufen) soll die Zielverbindung durch eine Reaktion darstellbar sein.

In diesen Fällen kann auf den klassischen Ansatz von Corey [1] zurückgegriffen werden, d. h. einzelne Bindungen im Syntheseziel müssen ausgewählt, gebrochen und die entstehenden Fragmente durch geeignete Abgangsgruppen oder Bindungsumordnungen in reale Reagenzien überführt werden. Es wurde bereits im Abschnitt 2.2.2 erwähnt, daß derartige Bindungen in einer Zielverbindung als *strategisch* bezeichnet werden. Ein Lösungsansatz läßt sich demnach in Teilaufgaben untergliedern:

1. Auswahl der strategischen Bindungen:
   Wo soll die Zielverbindung fragmentiert werden?
2. Ableitung der Reagenzien, die nach Bruch der strategischen Bindungen entstehen:
   Wie sollen die offenen Valenzen der fragmentarischen Moleküle gesättigt werden? Solcherart Molekülfragmente werden als Synthone verstanden.
3. Relative Wertung:
   Welche Bindung aus dem Satz aller gefundenen strategischen Bindungen ist besonders sinnvoll?

Wissens- oder transformbasierte Programme zur Syntheseplanung wie LHASA lösen diese Teilaufgaben in einem Schritt: Ein Transform, das auf eine Zielverbindung angewendet wird, definiert über eine Substruktur den *Ort* des retrosynthetischen Bindungsbruchs. Die resultierenden Reagenzien sind ebenfalls durch Substrukturen auf der Eduktseite im Transform definiert. Das Potential solcher Systeme wächst daher mit leistungsfähigeren Transforms.

Im WODCA System bestehen zwei prinzipielle Möglichkeiten zur Suche nach strategischen Bindungen: die manuelle Definition und die Benutzung von Funktionen. Auf die manuelle Erzeugung von Vorstufen wurde bereits im Abschnitt 2.2.2. hingewiesen. In diesem Kapitel werden ausschließlich neu entwickelte Funktionen abgehandelt. Zunächst sollen jedoch die prinzipiellen Möglichkeiten derartiger Funktionen diskutiert werden.

4.2 Möglichkeiten zur retrosynthetischen Analyse in WODCA

WODCA geht bei der schrittweisen Retrosynthese wie folgt vor:

- Das Syntheseziel wird analysiert, strategische Bindungen werden von programmierbaren Funktionen ausgewählt. Andere Funktionen ermitteln passende Abgangsgruppen für den Zeitpunkt, an dem die strategische Bindung gebrochen wird. Die strategischen Bindungen werden bewertet. Ein Bindungsbruch findet *nicht* statt.
- Das Resultat der WODCA-Analyse wird dem Benutzer grafisch präsentiert.* Er entscheidet, ob und welche strategische Bindung(en) gebrochen wird (werden).
- Verlangt der Benutzer den Bruch von als strategisch markierten Bindungen, werden dann die schon bei der Analyse ermittelten Abgangsgruppen an die Fragmente addiert.

Die Hauptarbeit leisten demnach die automatischen Funktionen zum Analysezeitpunkt. Dem Benutzer obliegt es anschließend, den Vorschlägen des Programms zu folgen.

Funktionen zur retrosynthetischen Analyse werden im WODCA System Version 3 wie die Transformationsregeln des Kapitels 3 als *tcl*-Skript formuliert. Dabei steht der gleiche Sprachumfang wie bei den Ähnlichkeitskriterien zur Verfügung (siehe Abschnitt 3.2.3). Es sei hier nur daran erinnert, daß dadurch eine umfassende Zahl sowohl topologischer Informationen (z. B. Atom ist Ringatom, Bindung ist aromatisch, Ring ist sechsgliedrig etc.) als auch physikochemischer Daten (z. B. Atompartialladungen) verfügbar sind. Darüber hinaus werden einige nur für die schrittweise Retrosynthese relevante zusätzliche Bindungsattribute bereitgestellt (siehe Tab. 4 - 1). Diese können gesetzt oder gelesen werden. Sie sind erforderlich und ausreichend, um flexibel strategische Bindungen und die resultierenden Reagenzien codieren zu können. Auf Einschränkungen wird weiter unten noch eingegangen.

Eine Retrosynthesefunktion des WODCA Systems geht dann in etwa so vor†: In einer Schleife über alle Bindungen werden zunächst diejenigen Bindungen ausgeschlossen, die für die gerade anzuwendende Strategie irrelevant sind (z. B. werden etwa alle aromatischen Bindungen ignoriert). Für Bindungen, die nach formalen Kriterien nicht herausfallen, können dann weitere Prüfungen stattfinden. Beispielsweise kann die Differenz in den Atompartialladungen der Bindungsatome geprüft werden, etwa weil eine solche Differenz in einem bestimmten Wertebereich eine Position signalisiert, die von strategischem Interesse sein kann. Ist auf diese Weise eine Bindung ausgewählt worden, werden Schnittmodus‡, eine Initialbewertung und erste Abgangsgruppen definiert. Unter Umständen wird auch die Zugehörigkeit zu einer Gruppe von strategischen Bindungen spezifiziert. Oft wird es so sein, daß zur Bestimmung der passenden Abgangsgruppen eine eigene Unterfunktion aufgerufen wird. Dabei ist es möglich, daß eine Initialbewertung der strategischen Bindung nachträglich korrigiert wird (Bonusse oder Malusse für die betreffende Bindung) oder wieder ganz aus dem Satz der strategischen Bindungen entfernt wird. Andere Unterfunktionen können nach Ermittlung des Satzes der strategischen Bindungen diese nochmals an Hand globaler molekularer Kriterien überprüfen (sind mehrere strategische Bindungen topologisch gleich, liegt diese im Zentrum des Molekülskeletts oder in äußeren Bereichen usw.) und Veränderungen vornehmen.

* Das gilt nur, wenn WODCA im grafischen Modus betrieben wird. Es ist jedoch auch möglich WODCA ohne grafische Benutzerschnittstelle zu betreiben.

† Dieses Szenario ist kein festes Dogma. Dem Programmierer einer solchen Funktion ist alles erlaubt, was ihm algorithmisch geeignet erscheint und den Syntax-Regeln der Sprache *tcl* entspricht.

‡ Der Schnittmodus legt fest, wie sich der Bindungsgrad verändern soll. Siehe auch Tab. 4 - 1.

Tab. 4 - 1 Zusätzliche Informationen, die von *tcl* zur Definition retrosynthetischer Schritte benutzt werden.

Information	*tcl*-Schlüsselwort	Zugriffs-art[a]
Eine Liste von Ordnungszahlen. Diese codiert die Art der Atome, die beim Bruch dieser strategischen Bindung an die freien Valenzen addiert werden. Einige Spezialcodes sind implementiert: 0 - Addiere keinerlei Atom. -1 - Entferne ein H-Atom am benachbarten C-Atom (wenn möglich). Bindungen, die nicht als strategisch markiert sind, werden durch eine leere Liste gekennzeichnet.	`B_SBADDATOMS`	L/S
Schnittmodus für eine strategische Bindung. Folgende Modi sind implementiert: 1 - Bindungsgrad wird um Eins gesenkt. 2 - Bindungsgrad wird auf Eins gesenkt. 3 - Bindung wird vollständig gebrochen. 4 - Bindungsgrad wird um Eins erhöht. 5 - Bindungsgrad wird um Zwei erhöht.	`B_SBCUTMODE`	L/S
Gruppenkennzeichnung. Durch gleiche Kennzeichnung werden mehrere strategische Bindungen zusammengefaßt. Eine Gruppe von strategischen Bindungen wird immer zusammen gebrochen.	`B_SBGROUP`	L/S
Bewertung einer strategischen Bindung	`B_SBRATING`	L/S

a. L - Lesezugriff, S - Schreibzugriff

Es wäre denkbar, daß die Retrosynthesefunktionen des WODCA Systems ähnlich wie transformbasierte Programme vorgeht, d. h. eine Transform-Substruktur im Syntheseziel sucht. Für die Substruktursuche auf dem Molekülgraphen der Zielverbindung steht ein für *tcl* implementiertes `match`-Kommando zur Verfügung. Ausgehend von den gefundenen Überlagerungen könnten dann, wie voranstehend beschrieben, Bewertungen vorgenommen werden. Dieser Weg ist derzeit jedoch in keiner Funktion realisiert.

4.3 Implementierte Retrosynthese-Strategien

4.3.1 Etablierte Prinzipien

Es gibt verschiedene Gruppen von Kriterien, die zur Ermittlung strategischer Bindungen* herangezogen werden können. Dazu zählen zum einen formale topologische Kriterien. Das SYNGEN Programm wählt strategische Bindungen in Synthesezielen nach rigorosen topologischen Kriterien aus [97]: Es sind grundsätzlich CC-Bindungen; es ist immer ein Paar von Bindungen; diese Bindungen müssen so liegen, daß die Verbindung in zwei möglichst gleich große Fragmente gespalten wird.† Diese Kriterien sind rein topologischer Natur. Auch das Programm Lilith [98] bedient sich streng formaler Kriterien. Zunächst wird das Komplexitätszentrum des Syntheseziels berechnet. Dann in dessen unmittelbarer Umgebung Bindungspaare als strategische Bindungen ausgewählt [99]. In ökonomischem Sinn wird so eine definitiv konvergente Strategie verfolgt. Die chemische Natur der Bindungen und ihrer Umgebung werden dabei aber völlig ignoriert. Daher sind Bindungsbrüche, die zu tatsächlich für eine Synthese einsetzbaren Reagenzien führen, nur zufällig zu erzielen. Auch in LHASA findet ein topologisches Kriterium Anwendung. Hier wurde ein Satz von Regeln entwickelt [100], der in komplexen Ringsystemen diejenigen Bindungen auswählt, deren retrosynthetischer Bruch ein Syntheseziel am deutlichsten vereinfacht. Sind so die *Orte* von Retro-Reaktionen festgelegt, werden anschließend nur noch diejenigen Transforms vorgeschlagen, die sich für diese Bindungen (und ihre molekulare Umgebung) eignen. Auf diese Regeln wird im Abschnitt 4.3.4 noch genauer eingegangen.

Andere Kriterien sind spezifisch für Klassen von *Verbindungen*. Beispielsweise gehören dazu viele Reaktionen aus dem Bereich der Aromatenchemie, besonders diejenigen zum Aufbau der verschiedensten Heteroaromaten. Wird ein bestimmtes Ringmuster‡ erkannt, kann das die Verwendung einer typischen Synthese signalisieren. Allerdings sind solche Reaktionen - so wertvoll und leistungsstark sie für einzelne Substanzklassen sein mögen - meist nicht generell anwendbar.

Auch einzelne, breit anwendbare *Reaktionen* können das zentrale Kriterium zur Auswahl strategischer Bindungen sein. Hierbei wird versucht, eine solche Schlüsselreaktion im Syntheseplan zu verankern. Ist die Reaktion nicht direkt verwendbar, werden u. U. eine ganze Reihe von Refunktionalisierungen vorgenommen, bis das Retron** der angestrebten Reaktion erzeugt wurde. Solche Verfahren werden als langreichweitige Strategien (*long-range strategies*) bezeichnet. In LHASA sind mehrere solche Ansätze verwirklicht, die auf einer Schlüsselreaktion basieren: Diels-Alder [101] und Chinon-Diels-Alder, Robinson-Anellierung [102], Halolactonisierung [103], Birch-Reduktion, Polyen-Cyclisierung [104], stereoselektive Olefinsynthese [105].

Schließlich gibt es noch diejenigen Kriterien, die sich an bestimmten Substitutionsmustern orientieren. Sie werden typischerweise zunächst nach der Zahl notwendiger Gruppen unterteilt (*One-Group-*, *Two-Group-Disconnection*). Eine weitere Aufgliederung bezieht sich

---

* Mit strategischer Bindung ist hier tatsächlich *nur* der Ort, d. h. die Bindung selbst, im Syntheseziel gemeint.

† Das dient dazu, auf eine konvergente Synthese hinzuarbeiten.

‡ Ein Triazin-Ring beispielweise ist natürlich auch ein topologisches Merkmal. Jedoch ist die Topologie hier von sehr spezifischem Charakter und von allgemeineren Merkmalen zu unterscheiden.

** Begriffsdefinition siehe Abschnitt 2.1 Seite 4.

dann auf die Relation von funktionellen Gruppen zueinander (*1,2-* oder *1,3-CC-Disconnection*). Diese Kriterien sind heute Lehrstoff und in vielen Textbüchern erläutert [4], [5].

Für das WODCA System Version 3 wurden drei Strategien implementiert, die in unterschiedlicher Weise verschiedene der zuvor genannten Kriterien in sich vereinen. Hinzu kommen als Kriterien im WODCA System physikochemische Eigenschaften, die ähnlich wie Relationen von funktionellen Gruppen herangezogen werden können, um strategische Bindungen auszuwählen. Diese Lösungen werden in den folgenden Abschnitten vorgestellt.

4.3.2 Strategien basierend auf CX-Bindungen

In diesem Abschnitt werden die Funktionen erläutert, die von WODCA zur Auswahl strategischer Kohlenstoff-Heteroatom-Bindungen (CX-Bindungen) herangezogen werden. Diese basieren auf Standard-Textbuch-Chemie, wie sie z. B. in [106] zu finden ist.

Generell geht diese Strategie davon aus, das *jede* CX-Bindung strategischer Natur ist. Diese Annahme ist naheliegend und durch die Vielzahl an Reaktionen, die zur Knüpfung von CX-Bindungen existieren, ausreichend begründet. Wichtig ist jedoch für die Lösung im WODCA System, daß CX-Bindungen zu freien Heterosubstituenten (wie z. B. Hydroxy- und Aminogruppen oder Halogensubstituenten) *nicht* mit dieser Strategie erfaßt werden. Denn das Ziel der Suche nach strategischen Bindungen besteht in einer *Vereinfachung des Molekülskeletts*. Ein Austausch freier Heterosubstituenten führt jedoch nur zu Umfunktionalisierungen (*FGI*), also nicht zu einer Vereinfachung. Um trotzdem variabel bei der Wahl von Verbindungen mit einem gegebenen Substitutionsmuster zu sein, können im WODCA System die Ähnlichkeitssuchen aus Kapitel 3 (vgl. besonders Abschnitt 3.3.4) herangezogen werden.

Die interessierenden CX-Bindungen können einfach herausgefiltert werden. Der Ort der strategischen Bindung ist damit festgelegt. Als nächstes müssen die entsprechenden Reagenzien abgeleitet werden. Hier stellt sich das Problem, daß für einen bestimmten Typ einer CX-Bindung, eine Reihe von Reaktionen in Syntheserichtung in Frage kommen und dementsprechend verschiedene Reagenzien. Ein ganzes Arsenal an Reaktionen steht beispielweise zur Synthese von Ethern zur Verfügung (siehe Abb. 4 - 1). Transformbasierte Programme enthalten möglichst *alle* diese Reaktionen und Verzweigen daher an entsprechender Stelle einen Syntheseplan beträchtlich in die Breite. Die resultierenden Vorstufen wiederum ähneln sich sehr stark (vgl. $ROSO_2OR$, RX, ROH in Abb. 4 - 1). Ein Gewinn für die Syntheseplanung entsteht dadurch nicht. Um diese Aufsplitterung des Syntheseplans zu vermeiden, prognostiziert WODCA immer nur *ein* Prototyp-Reagenz bzw. -Reagenzien-Paar. Die Verbindung zu den verschiedenen Synthese-Äquivalenten stellen erneut die Ähnlichkeitssuchen her.

Als letztes muß eine Retrosynthesefunktion die gefundenen strategischen Bindungen bewerten. In diesem Modell werden Grundbewertungen in Abhängigkeit vom Typ der strategischen CX-Bindung vergeben (z. B. esterartig: 100, aminanalog: 80-60; das sind hohe Anfangswerte). Mitunter werden auch physikochemische Parameter zur Berechnung einer Grundbewertung herangezogen (siehe z. B. Spezialstrategie für Sulfide in α-Position zu Carbonylgruppen). Diese Grundwerte sind in Übersicht 4 - 1 aufgeführt. Die Abstufungen erfolgen so, daß Bindungsbrüche an besonders reaktiven Zentren bevorzugt angezeigt werden. Darüber hinaus werden sämtliche gefundene strategische Bindungen am Ende der Funktion nochmals nach generellen Kriterien bewertet. Als erstes werden dabei einige Heuristiken eingesetzt, die drastischere Abstufungen vornehmen:

- Ist eines der beiden Atome der strategischen Bindung ein chirales Zentrum, erfolgt eine Abwertung um 50. Das soll dafür sorgen, daß strategische Bindungen zu chiralen Atomen nur wenn nötig gebrochen werden.

- ❏ Ist das Heteroatom der Bindung ein Ringatom, das andere Atom aber nicht in einem Ring, erhöht sich die Bewertung um 10. Das dient dazu, bevorzugt exocyclische Bindungen zu brechen.
- ❏ Zwei von einem sp^2-Kohlenstoff ausgehende strategische Bindungen, die noch die gleiche Bewertung haben, werden über die σ-Elektronegativität der beiden Bindungen gegeneinander abgestuft. Das Verhältnis der Bindungspolaritäten wird um den Faktor 10 gedämpft (nur geringe Abstufung ist erwünscht) und je zur Hälfte auf die Bewertungen aufgeschlagen bzw. abgezogen. Dieses Kriterium resultiert aus dem Wunsch, etwa bei Carbamaten bevorzugt die Bindung zum Stickstoff zu spalten, da der Rest eine typische Aminosäure-Schutzgruppe darstellt.

Desweiteren wird als weiches abstufendes Kriterium die Entfernung der strategischen Bindung vom Molekülzentrum herangezogen. Diese Entfernung wird in Bindungen gezählt. Auf diese Weise werden zentrale Bindungen gegen Bindungen in äußeren Molekülteilen abgehoben, wenn auch nur gering.

Strategische Bindungen, die gemeinsam gebrochen werden (z. B. Olefin-Strategie bei Ethern oder Aminen), erhalten beide den Wert der am höchsten bewerteten Bindung.

Schließlich werden alle Bewertungen so skaliert, daß die beste Bewertung bei 100 liegt.

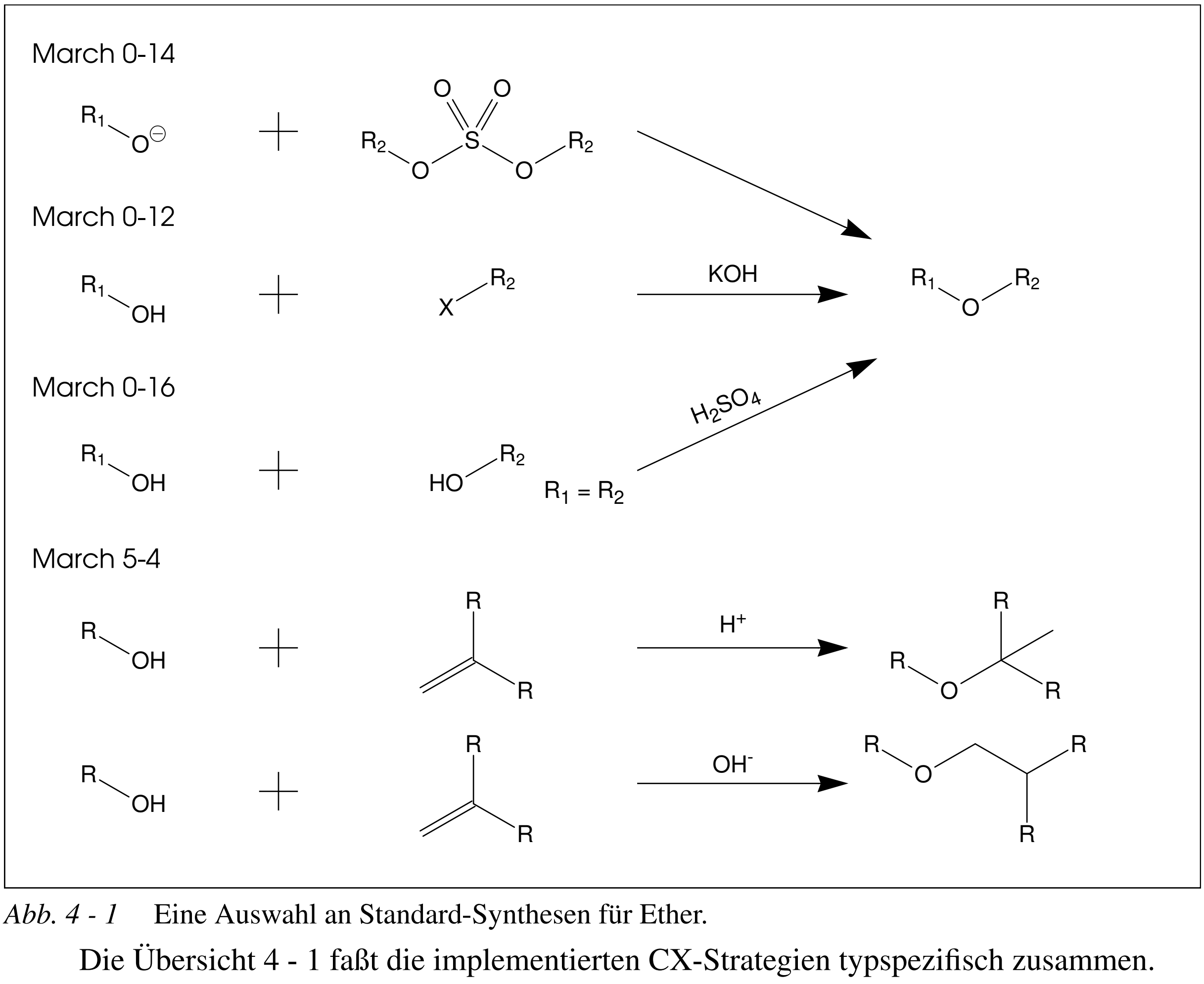
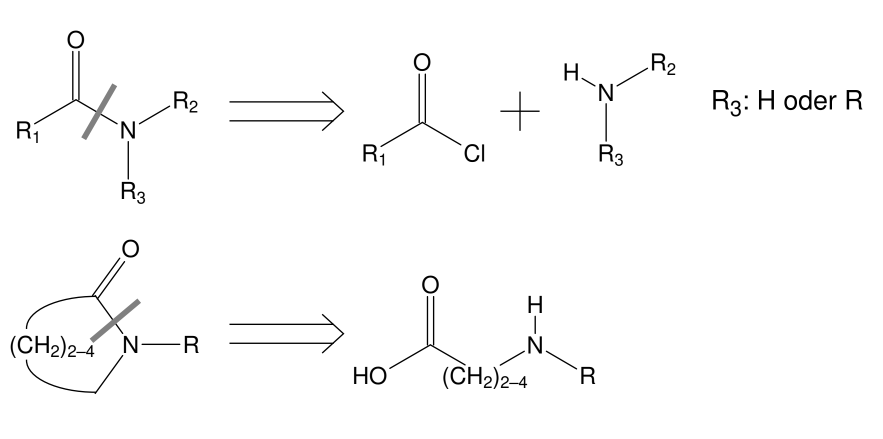


Abb. 4 - 1 Eine Auswahl an Standard-Synthesen für Ether.

Die Übersicht 4 - 1 faßt die implementierten CX-Strategien typspezifisch zusammen.

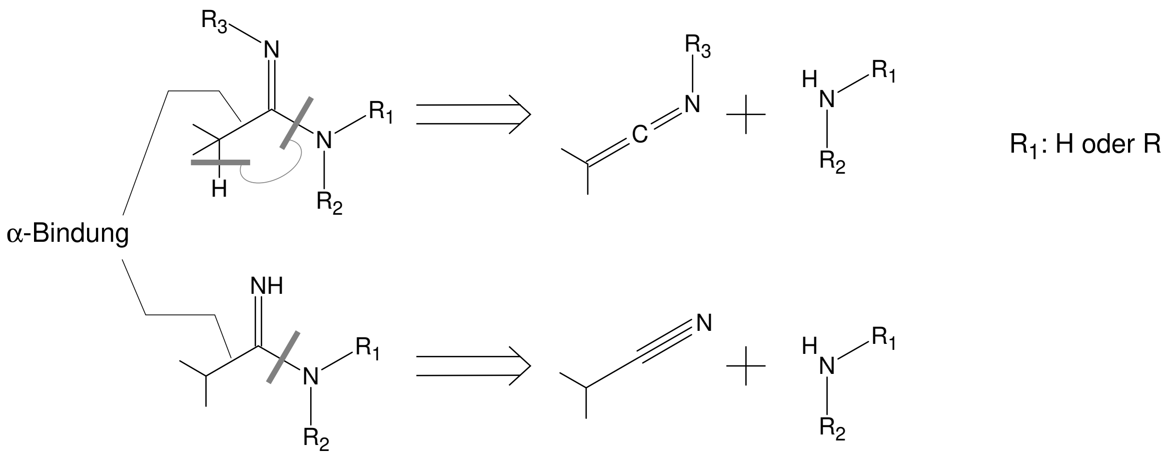
Übersicht 4 - 1 Methoden zur Auswahl strategischer CX-Bindungen im WODCA System. Fußnoten finden sich am Ende der Übersicht.

Amide (auch unsymmetrische Imide, Thiolester) (1)



Unterschieden wird zwischen Lactamen (nur 5- bis 7-gliedrige Lactamringe) und anderen Amiden (einschließlich Formamide). Thiollactame werden analog behandelt. Die Grundbewertung ist in beiden Fällen 100.

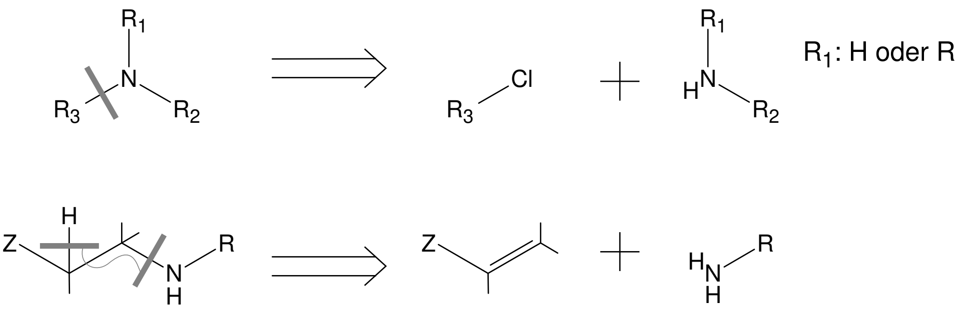
Amidine (2)



Es wird versucht, eine Strategie zu einem Nitril (unsubstituierter Imin-Stickstoff) bzw. Ketimin (substituierter Imin-Stickstoff) und einem Amin zu verfolgen. Bei der Ketimin-Strategie muß dann, wenn die Iminbindung in ein Ringsystem eingebunden ist, die Ringgröße bei acht oder größer liegen, da sonst eine zu stark gespannte Vorstufe resultieren würde. Die strategischen Bindungen werden in Abhängigkeit physikochemischer Größen bewertet, wobei letztere die in Syntheserichtung erforderliche Elektrophilie des carbonylischen C-Atoms signalisieren sollen. Herangezogen werden dabei nicht Werte für die strategische Bindung selbst, sondern solche für die benachbarte Bindung ausgehend vom carbonylischen C-Atom (α-Bindung, siehe unten).[a] Die Bewertung kann sehr hoch sein (starke unterstützende induktive oder mesomere Effekte), aber auch gering ausfallen. Mit niedrig bewerteten strategischen Bindungen signalisiert WODCA Probleme in Syntheserichtung.

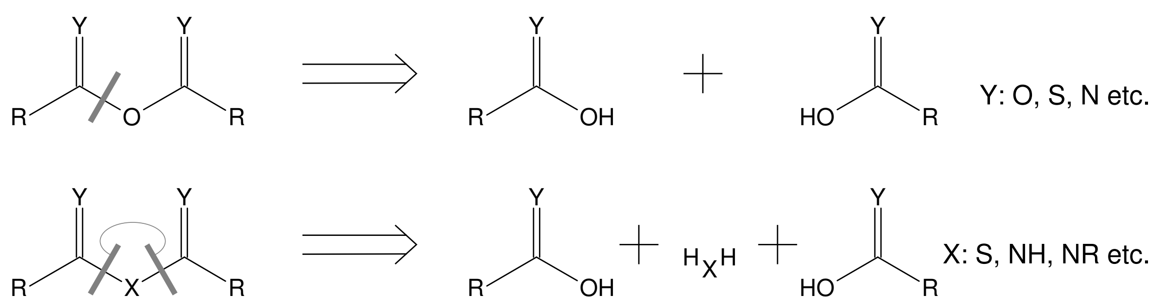
Übersicht 4 - 1 (Forts.) Methoden zur Auswahl strategischer CX-Bindungen im WODCA System. Fußnoten finden sich am Ende der Übersicht.

Amine, Phosphine etc. (3)



Je nach Substitutionsgrad des Amin-Stickstoffs und je nachdem ob es sich bei der betreffenden CX-Bindung um eine Ringbindung oder nicht handelt, wird unterschiedlich vorgegangen.
Acyclische, tertiäre oder quartäre Amine werden dealkyliert, Grundbewertung ist 80.
Acyclische, sekundäre Amine werden in ein primäres Amin und ein Olefin gespalten (Olefin-Strategie). Dafür ist allerdings eine CH-Bindung in α-Position zum Kohlenstoff der untersuchten strategischen CX-Bindung erforderlich, die durch elektronenziehende Gruppen aktiviert ist. Solche Gruppen werden durch einen physikochemischen Parameter der betroffenen CH-Bindung signalisiert.[b] Grundbewertung ist dann 80. Ist eine solche Situation nicht gegeben, wird dealkyliert mit einer niedrigeren Grundbewertung von 60 wegen problematischer Selektivität (Mehrfachalkylierung in Syntheserichtung).
Cyclische Aminbindungen in 7-gliedrigen oder größeren Ringen werden ebenfalls dealkyliert. Die Grundbewertung liegt jedoch nur bei 60. Für Aminbindungen in 3-, 5- und 6-gliedrige Ringen ist die Strategie gleich, die Grundbewertung aber höher (80). Bei 4-gliedrigen Ringen (Azetidine) ist die Grundbewertung niedriger (50), da die analogen Reaktionen hier oft nur geringe Ausbeuten erzielen.

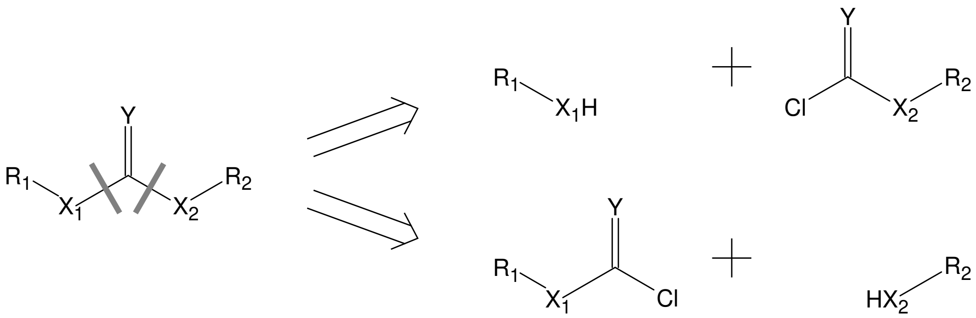
Anhydride und Analoga (Imide, Imidine) (4)



Diese Spezialstrategie kommt nur bei symmetrischen Anhydriden zur Anwendung. Dann werden Anhydride zu zwei Molekülen der freien Säure gespalten. Ist das verbindende Heteroatom kein Sauerstoff, werden beide CX-Bindungen zu diesem Atom zu einer Gruppe von strategischen Bindungen zusammengefaßt. Unsymmetrische Anhydride werden wie Ester behandelt. Die Grundbewertung ist 100.

Übersicht 4 - 1 (Forts.) Methoden zur Auswahl strategischer CX-Bindungen im WODCA System. Fußnoten finden sich am Ende der Übersicht.

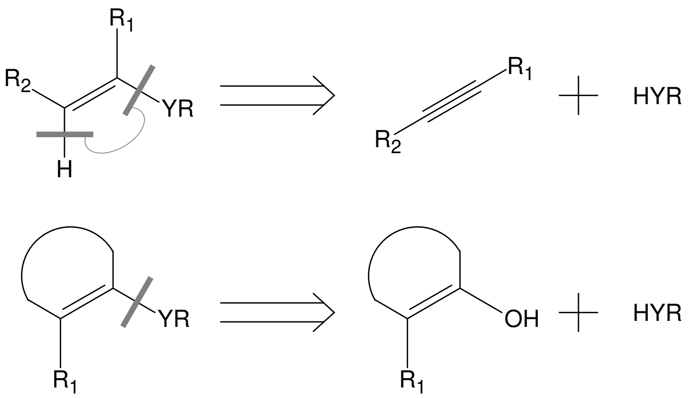
Carbamate, Harnstoffe und Analoga (5)



$X_{1,2}$: O, S, N etc.
Y: O, S, N etc.

Beide CX-Einfachbindungen zum carbonylischen Kohlenstoff werden wie Ester behandelt. Eine Unterscheidung kommt durch die Auswertung der σ-Elektronegativitätsdifferenz der jeweiligen Bindung zustande. Die Bindung mit der höheren σ-Elektronegativitätsdifferenz erhält einen Bonus von maximal fünf, die andere Bindung einen Malus von maximal zehn. Grundbewertung ist wie bei Estern 100.

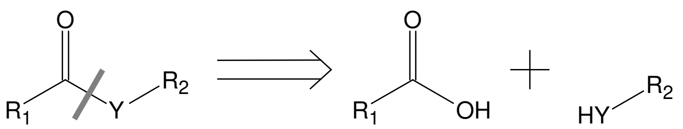
Enol- und Enolthioether (6)



$R_{1,2}$: H oder R
Y: O oder S

Bevorzugt wird die Alkin-Strategie verfolgt. Dazu darf die enolische CC-Doppelbindung nicht in Ringen mit bis zu sieben Gliedern liegen. Bei kleinen und mittleren Ringen oder wenn keine entsprechende CH-Bindung existiert, wird eine Strategie zu Ketonen/Aldehyden angewendet. Durch den Bindungsbruch erhält man die Carbonylverbindung in der Enolform. Dieses Enol kann so in den Ähnlichkeitssuchen verwendet werden, da immer eine Normalisierung (Ketonisierung) stattfindet. Grundbewertung ist in beiden Fällen 80.[c]

Ester (7)

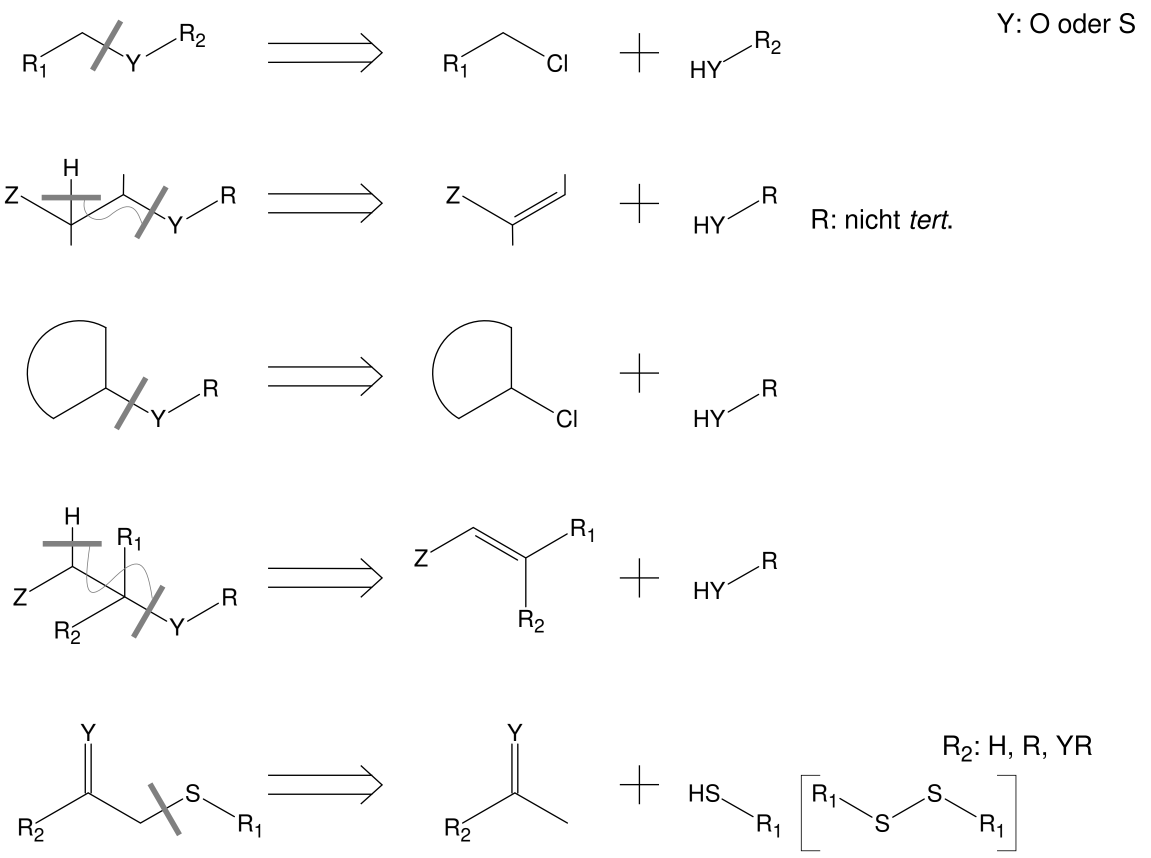


Y: O oder S

Die strategische Bindung spaltet die Verbindung in die Carbonsäure und den Alkohol bzw. das Mercaptan. Die Grundbewertung ist 100.

Übersicht 4 - 1 (Forts.) Methoden zur Auswahl strategischer CX-Bindungen im WODCA System. Fußnoten finden sich am Ende der Übersicht.

Ether und Sulfide (8)



Primäre Ether werden entsprechend einer Williamsonschen Ethersynthese gespalten. Grundbewertung ist 80.

Sekundäre aliphatische Ether werden bevorzugt in einen Alkohol und ein Olefin gespalten. Diese Olefin-Strategie wird angewendet, da sekundäre aliphatische Ether bei Williamson-Reaktionen nur in niedrigeren Ausbeuten entstehen. Grundbewertung ist 80. Würde ein tertiärer Alkohol als Vorstufe resultieren, werden auch diese Ether nach Williamson gespalten. Die Grundbewertung ist dann aber niedriger (65). Sekundäre cyclische Ether werden immer nach Williamson gespalten, da bei der Olefin-Strategie Selektivitätsprobleme entstehen können. Grundbewertung ist dann aber nur 70.

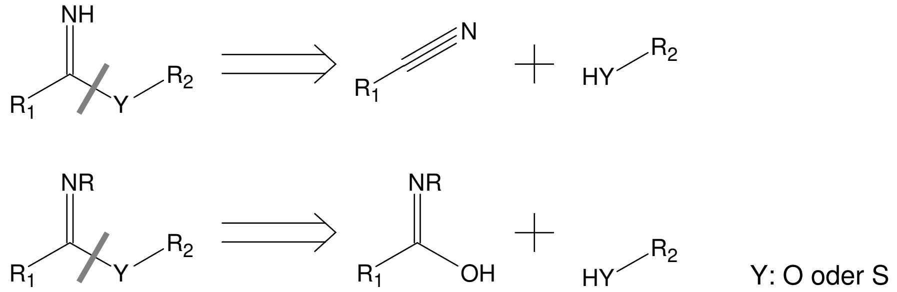
Tertiäre Ether/Sulfide werden immer nach der Olefin-Strategie behandelt. Grundbewertung ist 80. Nur für Ether gilt eine Abwertung um 20, wenn ein tertiärer Alkohol als Vorstufe gebildet werden müßte.

Die Auswahl des am leichtesten abspaltbaren H-Atoms bei der Olefin-Strategie geschieht mittels einer physikochemischen Größe.[b]

Für Sulfide in α-Position zu Carbonylgruppen ist eine Spezialstrategie implementiert. Hier werden als Vorstufen eine CH-acide Komponente und ein Mercaptan vorgeschlagen. Für die eigentliche Reaktion ist das entsprechende Disulfid erforderlich, das jedoch durch Oxidation aus dem Mercaptan entsteht. Grundbewertung ist 70, die jedoch in Abhängigkeit der elektronenziehenden Eigenschaften der Carbonylgruppe stark steigen kann. Diese Abhängigkeit wird anhand der Delokalisierungsstabilisierung (D^-) einer negativen Ladung bei angenommenem heterolytischem Bindungsbruch der CS-Bindung gemessen (siehe Abb. 4 - 2).

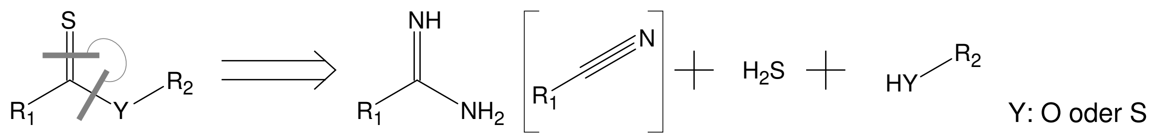
Übersicht 4 - 1 (Forts.) Methoden zur Auswahl strategischer CX-Bindungen im WODCA System. Fußnoten finden sich am Ende der Übersicht.

Imidoester	(9)



Bei freiem Imido-Stickstoff spaltet die strategische Bindung die Verbindung in ein Nitril und den Alkohol. Bei substituiertem Imido-Stickstoff entstehen als Vorstufen die Imidosäure und der Alkohol. Grundbewertung ist 100.

Thionverbindungen (Thionester, -amide, Dithioester)	(10)



Bei diesen Verbindungstypen werden sowohl die Esterbindung als auch die carbonylische CX-Doppelbindung als strategisch markiert. Aus programmtechnischen Gründen resultiert daraus dann ein freies Amidin (statt eines für die Reaktion in Syntheserichtung erforderlichen Nitrils), Schwefelwasserstoff und ein Alkohol. Das Amidin verhält sich jedoch in den Ähnlichkeitssuchen wie ein Nitril-Äquivalent. Die Grundbewertung ist 100.

a. Zur Abschätzung von Mesomerie-Effekten dient die Delokalisierungsstabilisierung D^-. Kann keine Mesomerie-Stabilisierung auftreten, werden induktive Effekte mit Hilfe der Differenz der σ-Atompartialladungen beurteilt.
b. Dieser Parameter ist die Delokalisierungsstabilisierung einer negativen Ladung bei heterolytischem Bindungsbruch D^-. Der Hintergrund zu diesem Wert wird an anderer Stelle ausführlich erläutert (siehe Abschnitt 4.3.3).
c. Bei beiden Strategien sind Fälle möglich, in denen die in synthetischer Richtung notwendige Regioselektivität zweifelhaft ist. Das herauszufinden und den entsprechenden Vorschlag abzuwerten ist eine Aufgabe für Nachbewertungsfunktionen.

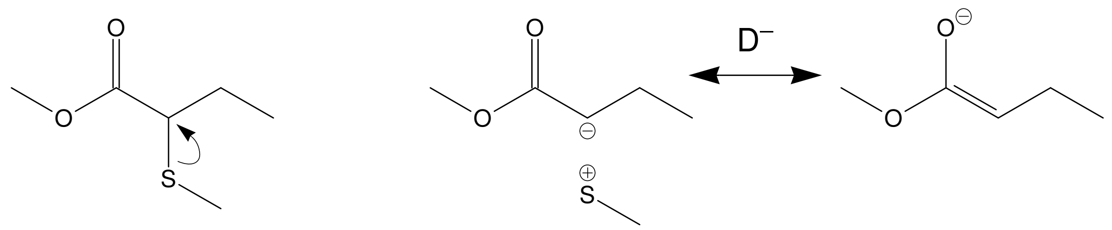
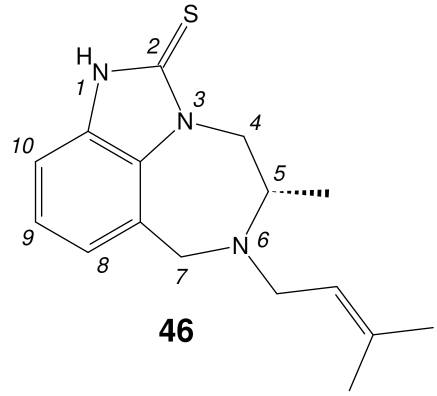


Abb. 4 - 2 Prinzip der Berechnung der Delokalisierungsstabilisierung von (negativen) Ladungen am Beispiel eines α-Mercaptocarbonsäureesters. Links die Originalverbindung. Für eine Bindung (hier die α-CS-Bindung) wird ein heterolytischer Bindungsbruch *angenommen*. Die entstehenden Ladungen (hier interessiert nur die negative Ladung) können durch Delokalisierung und/oder Hyperkonjugation stabilisiert werden. Diese Effekte berücksichtigt ein empirischer Parameter (D^-).

Man beachte, daß die Methoden aus Übersicht 4 - 1 nicht als zehn oder mehr Einzelregeln implementiert sind, sondern als eine einzige Option. Der Benutzer wählt „Carbon Hetero Atom Bond Breaks“ und erhält alle nach den Methoden möglichen Vorschläge an strategischen CX-Bindungen. Die gleichzeitig ermittelten Bewertungen helfen ihm dann bei der Entscheidung, welchen Bindungsbruch er weiterverfolgen möchte.

Ein Anwendungsbeispiel soll die Wirkungsweise der CX-Strategien verdeutlichen. Das in Abbildung 4 - 3 dargestellte 4,5,6,7-Tetrahydroimidazo[4,5,1-*jk*][1,4]benzodiazepin-2(1*H*)-thion (TIBO R82150, **46**) ist eine Prototyp-Verbindung, die in der jüngeren Vergangenheit Interesse wegen ihrer potentiellen inhibierenden Wirkung der Reversen Transkriptase des Human Immunodeficiency Virus 1 (HIV-1) gefunden hat [107].



46

Abb. 4 - 3 Beispielverbindung TIBO R82105.

Die oben beschriebene WODCA-Funktion ermittelt eine Reihe von strategischen CX-Bindungen (siehe Abb. 4 - 4). Die Bewertungen fallen sehr differenziert aus. Die drei Bindungen am N-Atom in Position 6 sind deutlich unterschiedlich. Die schwächste Bewertung erhält die Bindung zum chiralen C-Atom. Die exocyclische Bindung wird gegenüber den Ringbindungen aufgewertet. Die CN-Bindungen zum Phenylring sind nicht als strategisch markiert. Reaktionen die zu diesen Bindungen führen, würden einen nucleophilen Angriff am Aromaten bedeuten. WODCA beläßt jedoch diese Bindungen und versucht dadurch den substituierten Aromaten aus einem Ausgangsmaterial zu übernehmen.

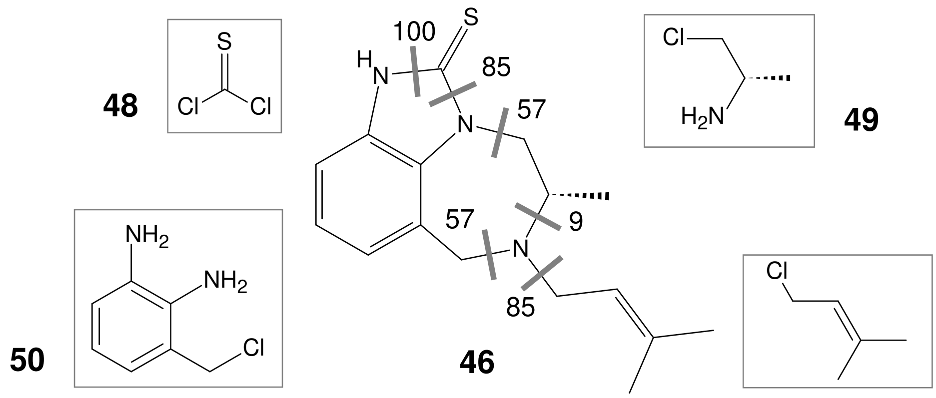


Abb. 4 - 4 Von WODCA ermittelte strategische Bindungen in TIBO (**46**) und die durch ihren Bruch entstehenden Vorstufen.

Man beachte auch, daß WODCA selbst keine strategischen Bindungen zu Gruppen zusammengefaßt hat, die dadurch immer gleichzeitig zu brechen wären. Die in Abbildung 4 - 4 dargestellten Vorstufen erhält man jedoch, wenn mehrere Bindungen *gemein-*

sam gebrochen werden. Hier kann der Chemiker eingreifen und dem Programm Ebenen im Syntheseplan ersparen, indem er vorgeschlagene strategische Bindungen zu Paaren oder Gruppen zusammenfaßt.

Die von WODCA prognostizierten Vorstufen sind klare *building blocks* für die Zielverbindung. Diese können nun mit Hilfe der Ähnlichkeitssuchen des Kapitels 3 zu kommerziellen Ausgangsmaterialien in Bezug gesetzt werden. Die folgenden Abbildungen stellen diese Zusammenhänge her.

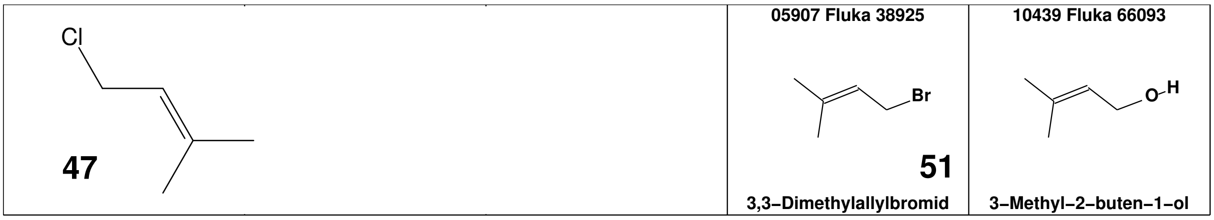


Abb. 4 - 5 Ausgangsmaterialen für die Prenylchlorid-Vorstufe (links). Die Ähnlichkeitssuche „Element- und Austauch von ZH_n-Gruppen“ ermittelt zwei Treffer im Fluka-Katalog.

Mit einem sehr selektiven Ähnlichkeitskriterium werden für die Prenyl-Vorstufe (**47**) der korrespondierende Alkohol und das Bromid gefunden (siehe Abb. 4 - 5). Das Bromid (**51**) ist für die beabsichtigte Alkylierung sehr gut geeignet.

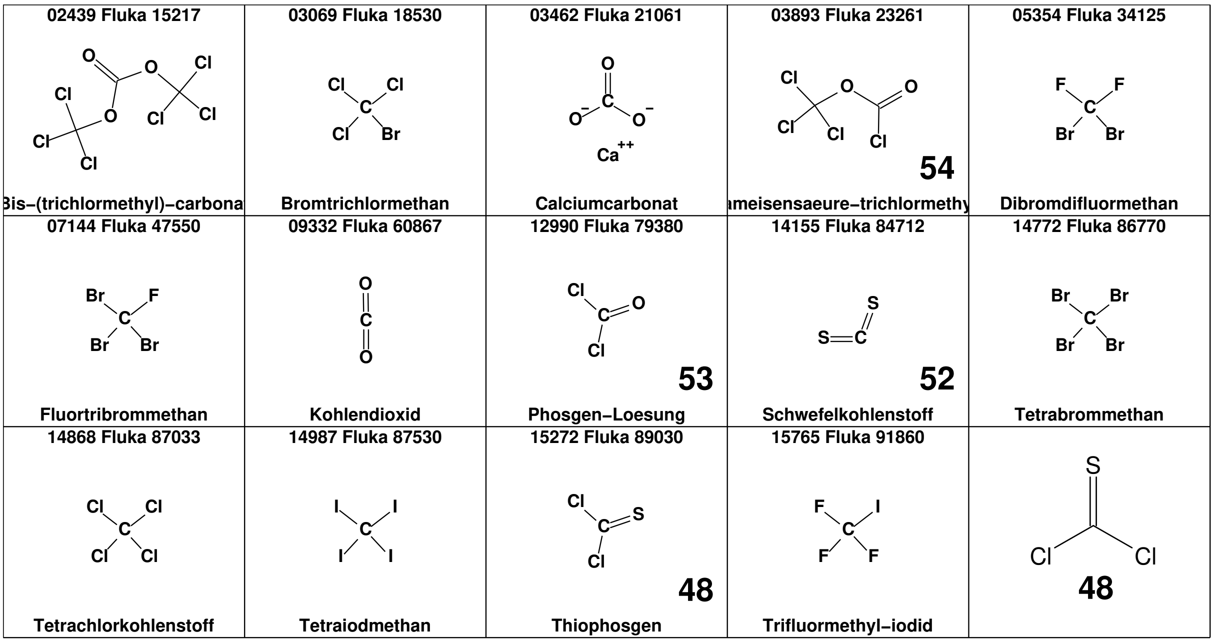


Abb. 4 - 6 Suche nach Ausgangsmaterialien für die Thiophosgen-Vorstufe (links unten). Das Kriterium „Hydrolyse“ erzielt für diese Verbindung 56 Treffer im Fluka-Katalog. Dargestellt ist eine Auswahl von 14 Verbindungen. Eine Alternative zu Thiophosgen (**48**) stellt Schwefelkohlenstoff (**52**) dar. Wäre die sauerstoff-analoge Carbonylverbindung (Phosgen, **53**) das Ziel, ist auch Chlorameisensäure-trichlormethylester (**54**) sehr gut geeignet.

Für die Thiocarbonyleinheit (**48**) finden sich mit der Hydrolyse-Transformation 56 Katalogverbindungen. Viele Verbindungen sind darin mehrfach enthalten: 26 Carbonate, siebenmal Tetrachlorkohlenstoff und viermal Schwefelkohlenstoff verzeichnet der Katalog. Schwefelkohlenstoff (**52**) ist ein typisches Reagenz für einen Thiocarbonyleinheit. Wäre die zu (**48**) sauerstoff-analoge Carbonylverbindung erforderlich, ist auch Chlorameisensäure-trichlormethylester (**54**) als Alternative geeignet. Befreit man die Liste von Dubletten, bleiben weniger als 20 Vorschläge (siehe Abb. 4 - 6) übrig. Das ist eine Menge, die vom Chemiker visuell inspiziert werden kann.

Bei der chiralen C_3-Einheit (**49**) werden mit der Transformation „Hydrolyse und Oxidation" jedoch schon 105 verwandte Verbindungen erkannt. Diese Liste enthält alle im Katalog vorhandenen C_3-Einheiten, die 1,2-difunktionalisiert sind und sich durch Hydrolyse- und Oxidationsschritte bei Brenztraubensäure (**55**) treffen. Bei Listen solcher Größenordnung ist eine Hilfestellung durch das Programm erforderlich, die dem Chemiker diejenigen Verbindungen ermittelt, welche zur Anfrageverbindung am ähnlichsten sind.

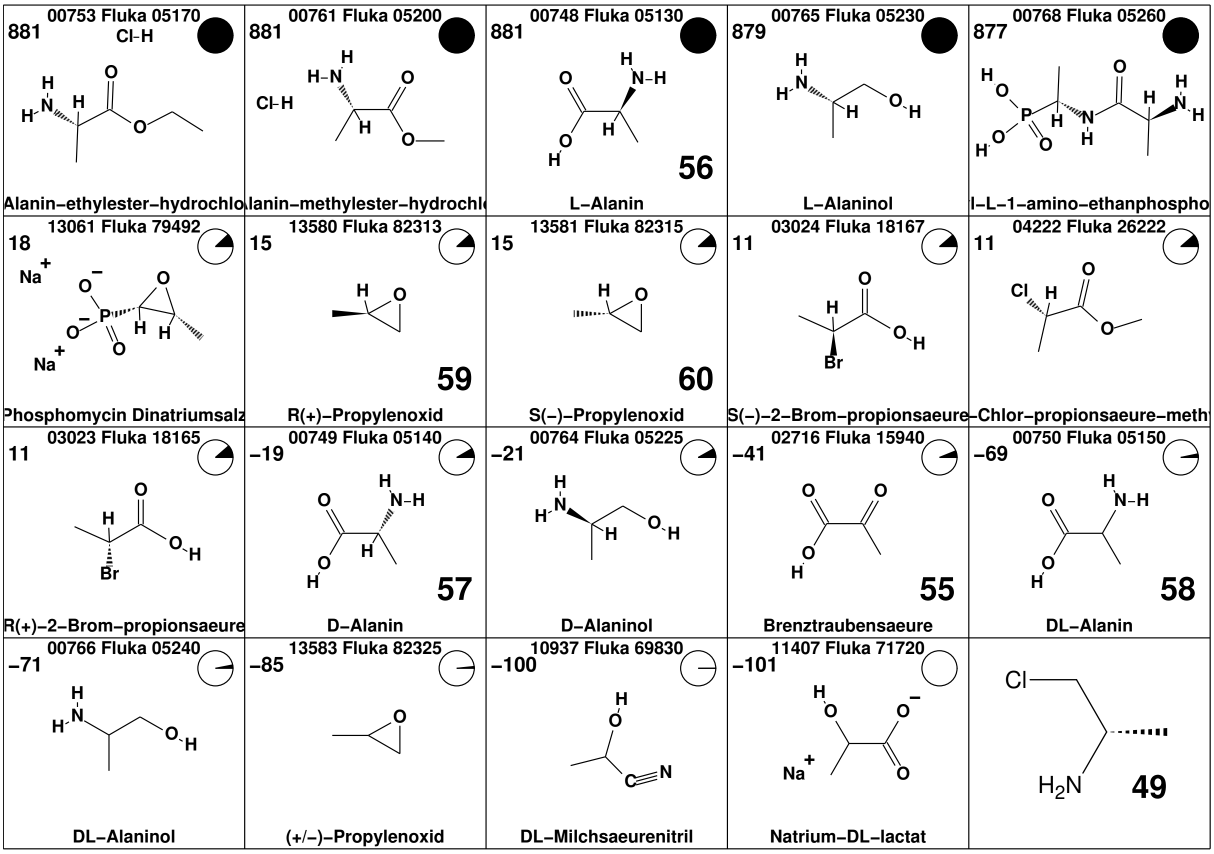


Abb. 4 - 7 Suche nach Ausgangsmaterialien für die chirale C_3-Einheit im Fluka-Katalog: Die Transformation „Hydrolyse und Oxidation" ermittelt 105 Treffer. Dargestellt ist eine Auswahl von 19 Verbindungen. Die Tortendiagramme (jeweils rechts oben) geben eine Bewertung an, daneben ist auch der numerische Wert verzeichnet (siehe Text).

Hierzu eignet sich eine Bewertungsfunktion, wie sie im Abschnitt 2.2.3 beschrieben wurde.* Wichtigstes Kriterium, das zwischen Katalogverbindung und Vorstufe übereinstimmen muß, ist die Chiralität des C-Atoms unter Beachtung der Natur des Heterosubstituenten an diesem Atom. Dieses differenziert die Liste der 105 Verbindungen deutlich. Lediglich fünf† Verbindungen stimmen in diesem Punkt mit der C_3-Vorstufe überein (siehe erste Reihe in Abb. 4 - 7). Die Tortendiagramme in Abbildung 4 - 7 spiegeln die Bewertung wieder. Je voller das Kreissegment, desto höher ist die ermittelte Bewertung. *L*-Alanin (**56**) läßt sich so leicht als das Ausgangsmaterial der Wahl für diese Vorstufe erkennen. Man beachte auch die deutlich andere Bewertung für das enantiomere *D*-Alanin (**57**, -19) und die racemische Verbindung

* Diese Bewertungsfunktion wurde von Ihlenfeldt [59] entwickelt. Sie wurde hier unverändert verwendet.

† Es sind sechs Verbindungen. *L*-Alanin ist jedoch doppelt im Katalog enthalten und daher nur einmal in Abbildung 4 - 7 dargestellt.

DL-Alanin (**58**, -69). Einleuchtend ist die starke Abwertung des falschen Stereoisomeren gegenüber der korrekten Konfiguration (881 gegenüber -19). Die noch niedrigere Bewertung für das Racemat (nur noch -69 gegenüber -19) folgt aus der Überlegung, daß bei einem Racemat in jedem Fall ein Aufarbeitungsschritt notwendig wird, während das falsche Stereoisomer in einem Reaktionsschritt möglicherweise invertiert werden kann. Bei dem Enantiomerenpaar *R(+)*- und *S(-)*-Propylenoxid (**59** bzw. **60**) führt dagegen die Konfiguration des C-Atoms nicht zu einer differenzierten Bewertung. Das liegt daran, daß das Heteroatom im Propylenoxid gegen ein N-Atom am C_2 der Anfrageverbindung (**49**) ausgetauscht werden muß. Solche Substitutionen können aber sowohl unter Retention als auch mit Inversion ablaufen. Eine Unterscheidung der Enantiomeren findet daher nicht statt.

Schließlich muß ein Ausgangsmaterial für den Benzyl-Baustein (**50**) gefunden werden. Eine substitutionsmuster-basierte Ähnlichkeitssuche ermittelt 14 Kandidaten im Fluka-Katalog. Eine hohe Oxidationsstufe der Seitenkette ist eher von Vorteil für die beabsichtigte Kondensationsreaktion. Jedoch gibt es nicht gleichzeitig zwei Stickstoff-Substituenten am Aromatenkern. Der Chlor-Substituent in 2-Chlor-3-nitro-benzoesäure (**61**) kann jedoch bei der vorgesehenen Cyclisierung verdrängt werden, die Nitrogruppe später reduziert werden.

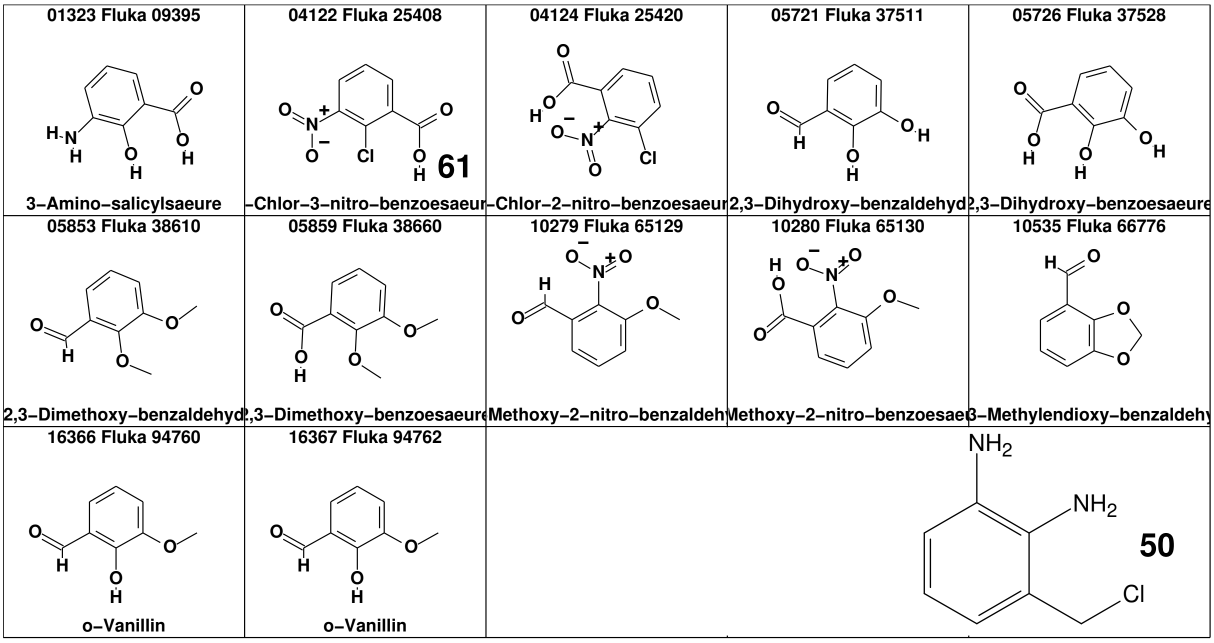


Abb. 4 - 8 14 Verbindungen des Fluka-Katalogs, die mit dem Kriterium „Substitutionsmuster (-AR+A1-CCMB-BO-MU)“ für die Benzyl-Vorstufe (unten rechts) gefunden werden. Das aussichtsreichste Ausgangsmaterial ist 2-Chlor-3-nitro-benzoesäure (**61**).

Die WODCA-Analyse wird in Abbildung 4 - 9 zusammengefaßt. Sie steht im Einklang mit einer publizierten Synthese für TIBO R82105 (**46**) [108]. Hier sei zusätzlich angemerkt, daß für zwei Bausteine der Zielverbindung bereits durch Ähnlichkeitssuchen mit dem Syntheseziel selbst (d. h. ohne vorherige Zerlegung in Vorstufen durch Bruch strategischer Bindungen), Ausgangsmaterialien ermittelt werden. Dies zeigt Abbildung 4 - 10. Für den Prenyl-Baustein (**47**) werden Ausgangsmaterialien erkannt, die auch mit der entsprechenden Vorstufe erhalten werden (vgl. Abb. 4 - 5). Hier muß aber ein (strenges) Substitutionsmuster-Kriterium verwendet werden. Das sehr restriktive Kriterium aus Abbildung 4 - 5 („Element- und Austauch von ZH_n-Gruppen“) ermittelt mit der Zielverbindung keinerlei Treffer, da diese von dieser Transformation nicht fragmentiert wird. Für den Benzyl-Baustein (**50**) kann eine Alternative zu den Vorstufen aus Abbildung 4 - 8 ermittelt werden: 2-Methyl-6-nitro-anilin (**62**). Natürlich sind in diesem Fall zusätzliche Umsetzungen notwendig, bis das Syntheseziel entsteht. Bei Wahlmöglichkeit wird daher 2-Chlor-3-nitro-benzoesäure (**61**) den Vorzug erhalten.

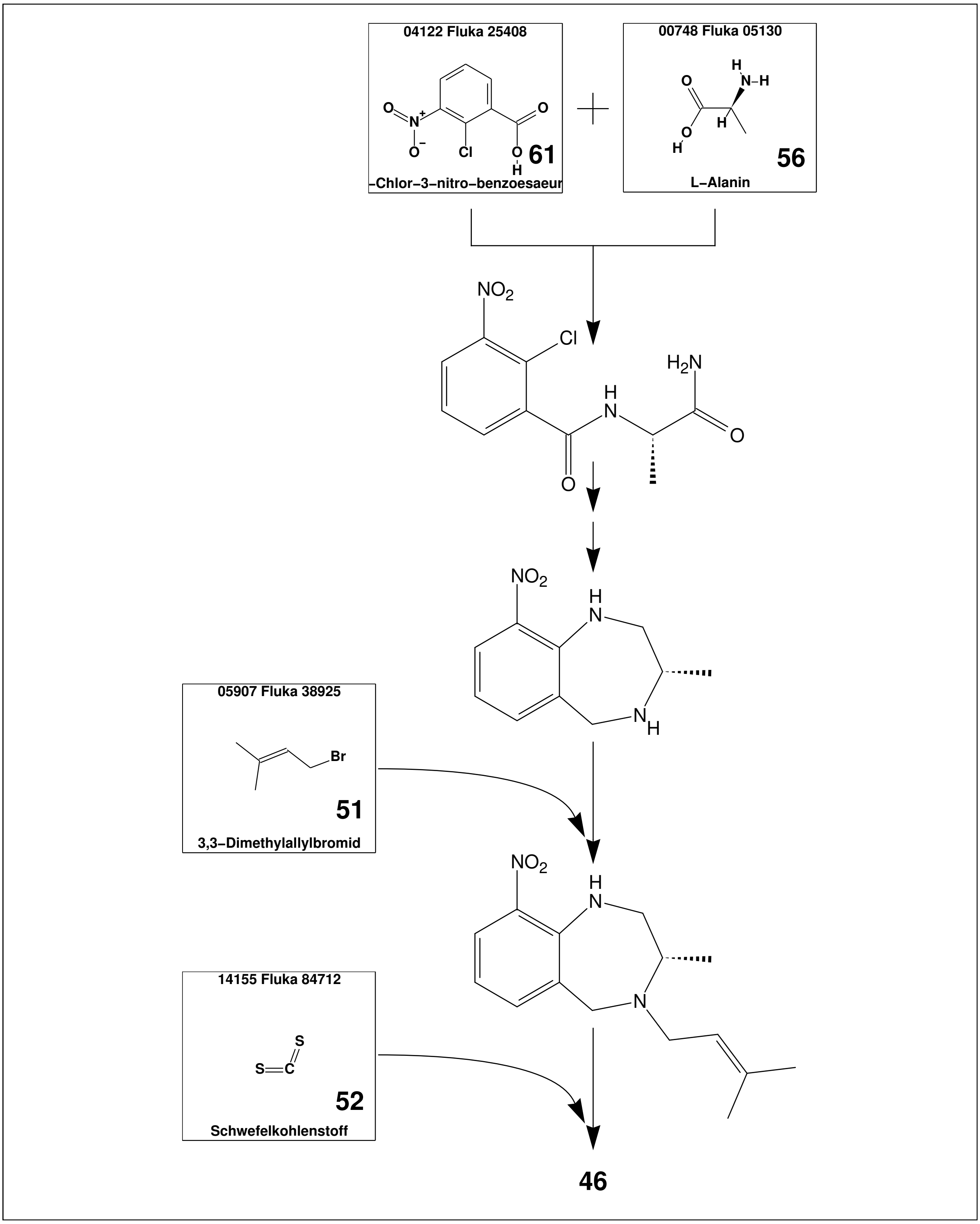


Abb. 4 - 9 Syntheseplan für TIBO R82105 (**46**). Die umrahmten Ausgangsmaterialien wurden von WODCA ermittelt. Die Ausgangsmaterialien wurden entsprechend der Reihenfolge der Umsetzungen im publizierten Syntheseweg [108] dargestellt.

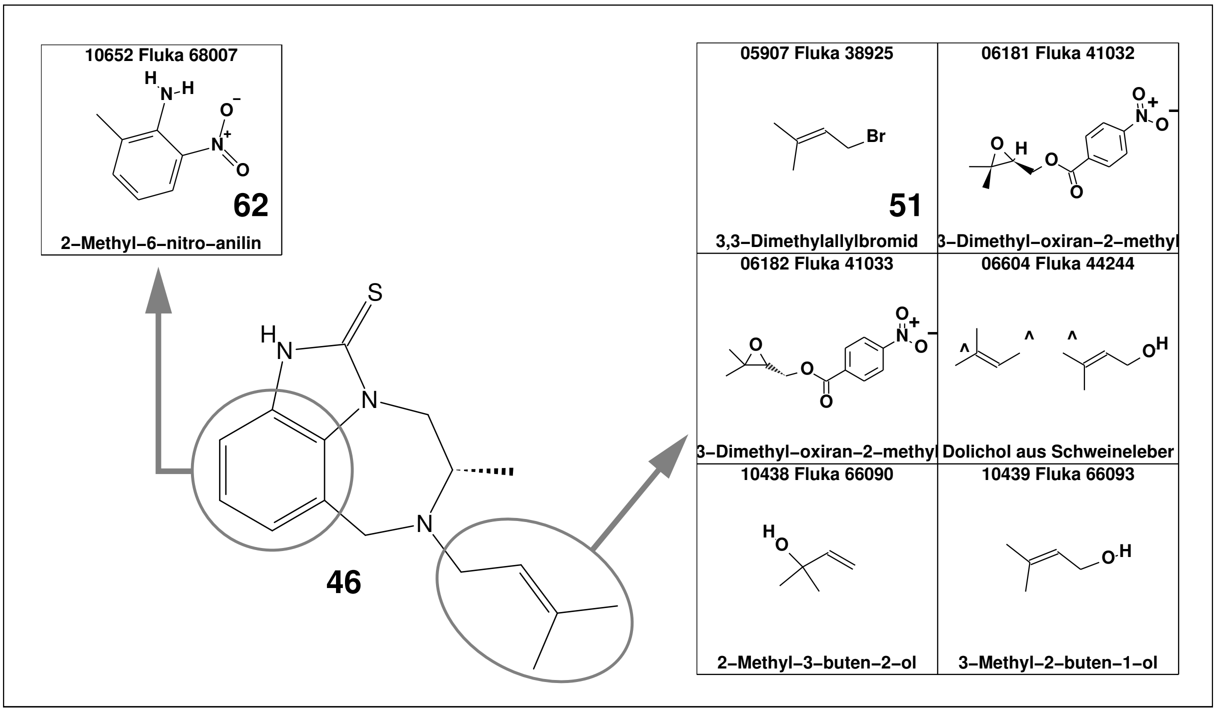


Abb. 4 - 10 Zum Vergleich: Ähnlichkeitssuchen direkt mit der Zielverbindung. Verschiedene Ähnlichkeitskriterien fokussieren auf unterschiedliche Bereiche des Syntheseziels. Zwei Bausteine werden so bereits identifiziert. Links oben: Ein Treffer mit dem Kriterium „Aromatensystem einschließlich α-Atome“. Rechts: Sechs Treffer mit dem Kriterium „Substitutionsmuster (-AR+A1+CCMB+BO+MU)“.

4.3.3 Strategie für CC-Bindungen

An dieser Stelle sollen Algorithmen beschrieben werden, die WODCA verwendet, um strategische CC-Bindungen zu ermitteln. Offensichtlich ist, das zur Planung organischer Synthesen gerade solche Strategien gefragt sind, um komplexere Verbindungen in einfachere Vorstufen zu zerlegen. Das Arsenal des organischen Chemikers zur Knüpfung von CC-Bindungen ist immens. Hier sollen daher einige sinnvolle Einschränkungen vorgenommen werden:

- Beschränkung auf nicht-aromatische CC-Bindungen. Ebenso Ausschluß von Bindungen zu aromatischen C-Atomen. In Frage kommen damit noch alle Reaktionen, die aliphatische CC-Bindungen, aber auch CC-Bindungen in nicht-aromatischen Ringen aufbauen.
- Auswahl besonders leistungsfähiger, breit anwendbarer Reaktionen. Exotische Spezialreaktionen sind zwar prinzipiell für die computer-gestützte Syntheseplanung wünschenswert, müssen aber zukünftigen Weiterentwicklungen vorbehalten bleiben.
- Erkennung der möglichen Reaktion auf der Basis von physikochemischen Merkmalen. Es kann erwartet werden, daß solche Charakteristika topologischen Eigenschaften - wozu auch eine Definition eines Reaktionszentrums über eine Substruktur (*Retron*) gehört - überlegen sind.

Der letzte Punkt bedarf eines Kommentars: Eine Beschreibung eines Reaktionszentrums durch eine Substruktur ist mit letzterer auf eine genau bestimmte Menge an molekularen Merkmalen begrenzt. Das ist erwünscht, denn die geforderten strukturellen Merkmale sind ja gerade die Voraussetzung für eine bestimmte Reaktion. Andererseits gibt es viele Reaktionen, die Gruppen mit bestimmten Eigenschaften in Nachbarschaft des Reaktionszentrums erfordern, etwa elektronenziehende Gruppen. Solche Gruppen können aber vielgestaltig sein, so daß die notwendige Substruktur, die *alle* diese Gruppen erfassen soll, sehr kompliziert wird. Demgegenüber sollten geeignete meßbare (d. h. hier: berechenbare) physikochemische Parameter unabhängig von den tatsächlich vorhandenen Nachbargruppen signalisieren, ob etwa elektronenziehende Eigenschaften vorhanden sind oder nicht. Die Substruktur als Schlüssel zur Prognose einer Reaktion sollte daher durch einen geeigneten Satz an physikochemischen Parametern ersetzbar sein.

Um die Verwendbarkeit eines solchen Ansatzes zu untersuchen, wurde als Reaktionstyp die Michael-Addition ausgewählt. Diese von dem Amerikaner Arthur Michael 1887 erstmals beschriebene Reaktion [109], bei der eine durch elektronenziehende Gruppen aktivierte CH-acide Verbindung (Methylen-Komponente) eine ebenfalls aktivierte (elektronenarme) Olefinkomponente nucleophil angreift (siehe Schema in Abb. 4 - 11), ist eine breit anwendbare, gut untersuchte Aufbaureaktion von CC-Bindungen [110], [111], [112], [113].

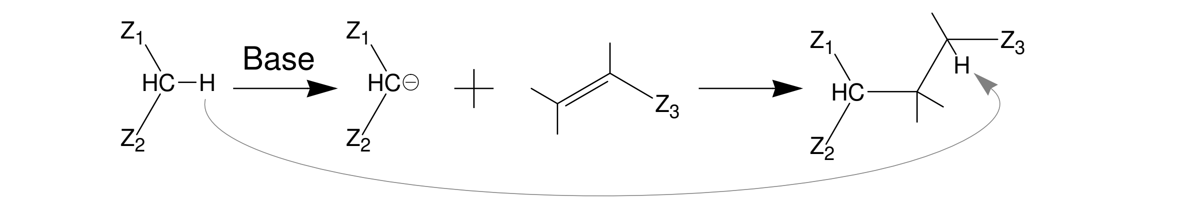


Abb. 4 - 11 Schematischer Verlauf einer Michael-Addition (Syntheserichtung).

Für die Syntheseplanung unter den oben genanntem Bedingungen besteht dann die Aufgabe darin, Michael-Typ-Bindungen im Synthese*ziel* zu erkennen. Neben der Identifikation solcher Bindungen müssen außerdem Methoden zur Absättigung der durch den retrosynthetischen Bindungsbruch entstehenden offenen Valenzen definiert werden. In diesem Fall muß ein

H-Atom auf der Seite der Methylen-Vorstufe addiert werden. Im anderen Fragment muß in α-Position zur Bruchstelle ein H-Atom entfernt werden, damit die olefinische Vorstufe gebildet werden kann (siehe Abb. 4 - 12). Und schließlich ist der Bindungsbruch mit einer Bewertung zu versehen, um verschiedene Vorschläge an strategischen Bindungen gegeneinander abwägen zu können.

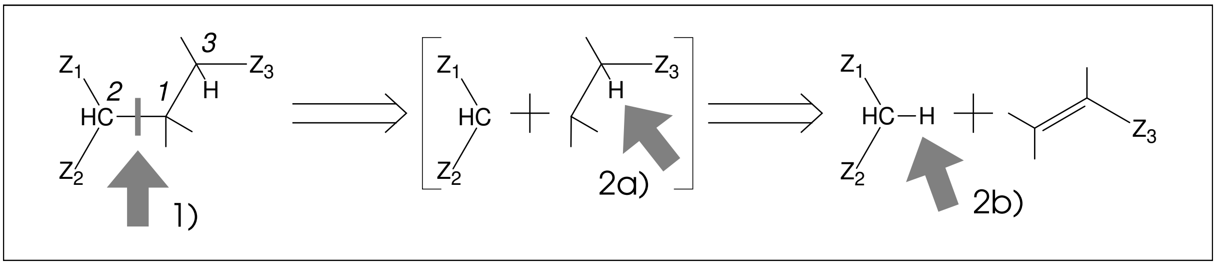


Abb. 4 - 12 Schritte der retrosynthetischen Analyse. 1.) Identifikation der strategischen Bindung; 2.) Absättigung offener Valenzen nach dem Bindungsbruch. Schließlich muß auch eine Bewertung vergeben werden. Die kursiven Zahlen an den Atomen links bezeichnen die C-Atome des Reaktionszentrums.

Zur Erstellung eines Modells für die retrosynthetische Analyse von Michael-Additionen sind dann Beispielreaktionen erforderlich. Um einen solchen, möglichst umfangreichen und variantenreichen Fundus zu erhalten, kann auf Reaktionsdatenbanken zurückgegriffen werden. Das im folgenden beschriebene Modell wurde anhand von 84 Michael-Reaktionen erstellt, die aus der ChemInform Reaktionsdatenbank [114], [115] stammen. Diese Reaktionen waren Teil eines 120 Reaktionen umfassenden Datensatzes, der für eine Studie zur automatischen Klassifizierung von Reaktionen [116] vom FIZ Chemie Berlin zur Verfügung gestellt wurde. Sämtliche 120 Reaktionen haben als gemeinsames Merkmal ein übereinstimmendes Reaktions*zentrum*[*]. Jedoch ist das allein nicht ausreichend für eine Michael-Addition. Auch Friedel-Crafts-Alkylierungen mit Olefinen oder radikalische Additionen an Olefine haben das gleiche Reaktionszentrum. Die 84 Michael-Additionen wurden durch manuelle Inspektion als solche klassifiziert.[†]

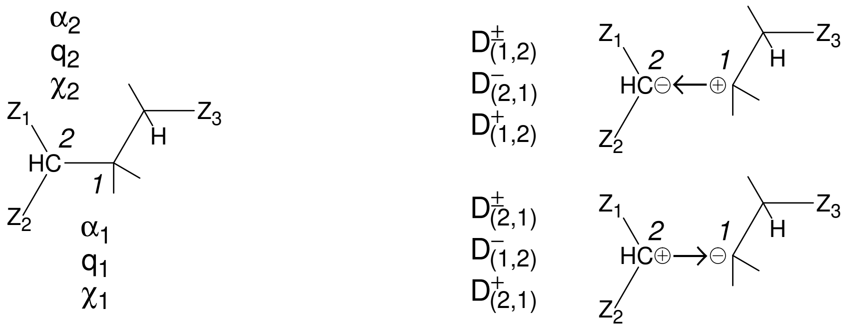
Es wurden dann zunächst Untersuchungen angestellt, welche WODCA zur Verfügung stehende physikochemische Größen (siehe Abschnitt 3.2.3) die durch die Michael-Addition entstandene Bindung im Reaktionsprodukt am besten beschreiben. Die folgende Abbildung veranschaulicht den Hintergrund der verschiedenen Werte. Die atomspezifischen Werte lassen sich als Differenzen (Ladungen, Elektronegativitäten) bzw. Mittelwerte (Polarisierbarkeiten) auch auf die Bindung projizieren. Für die Abschätzung des Mesomerieeffektes wird ein heterolytischer Bindungsbruch *angenommen*. Für das dann positiv geladene Atom wird abgeschätzt, wie diese Ladung durch benachbarte Gruppen mittels Mesomerie stabilisiert werden kann. Gleiches findet auch für die negative Ladung statt. Das Berechnungsmodell entspricht daher den Vorstellungen des Chemikers über die *Synthone*[‡] für eine Reaktion, bei der heterolytische Bindungsbrüche erfolgen. Diese Werte können natürlich für alle Bindungen/Atome eines Moleküls berechnet werden. Eine auf ihrer Grundlage identifizierbare strategische Bin-

[*] Unter dem *Reaktionszentrum* wird der minimale Satz an Atomen und Bindungen verstanden, an denen im Verlauf der Reaktion Veränderungen eintreten.

[†] Wenn im folgenden Bezug auf einige dieser Reaktionen genommen wird, werden die eindeutigen, von Molecular Design Ltd. für jeden Datenbankeintrag vergebenen Bezeichnungen verwendet. Diese haben die allgemeine Form RXCI91nnnnnn. Das identifiziert die ChemInform (CI) Reaktionsdatenbank (RX) des Datenbankjahrgangs 1991 (91). Die sechsstellige Ziffer danach bezeichnet den Eintrag eindeutig.

[‡] Gemeint sind damit die eigentlich reagierenden Spezies.

dung muß sich daher durch deren Werte klar von weniger interessanten Bindungen abgrenzen lassen.



α_i .... Polarisierbarkeit des Atoms i
q_i .... Partialladung für das Atom i
χ_i .... Elektronegativität des Atoms i
$D^{\pm}_{(i,j)}$ .... Gesamtdelokalisierungsstabilisierung von Ladungen bei heterolytischem Bindungsbruch. Der Wert ergibt sich als die Summe $D^{+}_{(i,j)} + D^{-}_{(j,i)}$.
$D^{+}_{(i,j)}$ .... Delokalisierungsstabilisierung einer positiven Ladung am Atom i bei heterolytischem Bindungsbruch
$D^{-}_{(i,j)}$ .... Delokalisierungsstabilisierung einer negativen Ladung am Atom i bei heterolytischem Bindungsbruch

Abb. 4 - 13 Physikochemische Parameter des WODCA Systems am Beispiel der bei einer Michael-Addition entstehenden CC-Bindung.

Wie aus Abbildung 4 - 13 ersichtlich ist, beschreiben die Parameter $D^{\pm}_{(1,2)}$, $D^{+}_{(1,2)}$ und $D^{-}_{(2,1)}$ die Synthonpolarität* eines Michael-Angriffs. Es kann daher erwartet werden, daß diese Größen auffällige Werte für die gesuchte strategische Bindung annehmen. Das bestätigt die Darstellung in Abbildung 4 - 14. Die $D^{\pm}_{(1,2)}$-Werte variieren in einem nach oben offenen Bereich; als untere Grenze kann 18.0 identifiziert werden. Dann werden aus typischen Michael-Reagenzien (1,3-Dicarbonylverbindung und α,β-ungesättigte Carbonylverbindung) resultierende Produkte sicher erkannt. Interessant sind die darunter liegenden Reaktionen (siehe Abb. 4 - 15). Das sind zum einen drei Reaktionen mit Chloroform als CH-acider Komponente (RXCI91143164/5/6) und eine weitere Reaktion (RXCI91124495), bei der das Michael-Produkt in der Enolform codiert ist (**63**). Wird statt der Enolform die Ketoform (**64**) analysiert, resultiert ein $D^{\pm}_{(1,2)}$-Wert von über 19, d. h. ein Wert, der oberhalb des gewählten Limits liegt. Im ersten Fall wird die Gesamtstabilisierung ausschließlich durch mesomere Effekte begründet, die auf der Seite der positiven (Synthon-)Ladung wirken. Delokalisierung der negativen Ladung durch die drei Chloratome ist nicht möglich. Der entsprechende Delokalisierungsparameter ($D^{-}_{(2,1)}$) ist daher in solchen Fällen Null. Gleichwohl stabilisieren die Halogensubstituenten eine negative Ladung jedoch durch induktive Effekte. Darauf wird gleich noch einmal zurückgekommen. In Abbildung 4 - 14 sind einige Beispiele für Michael-Produkte für den entsprechenden Wertebereich der Gesamtdelokalisierung angegeben. Man erkennt, daß in allen Fällen elektronenziehende Gruppen am Atom 2 des Reaktionszentrums präsent sind. Vor allem steigern aber Gruppen in der Umgebung von Atom 1 des Reaktionszen-

* Bei der Michael-Addition greift ein Carbanion in Position 2 eine positivierte Doppelbindung in Position 1 an (Ziffern bezeichnen die Atome des Reaktionszentrums, siehe Abb. 4 - 12).

trums, die positive Ladungen stabilisieren können, den Wert für die Gesamtdelokalisierung deutlich.

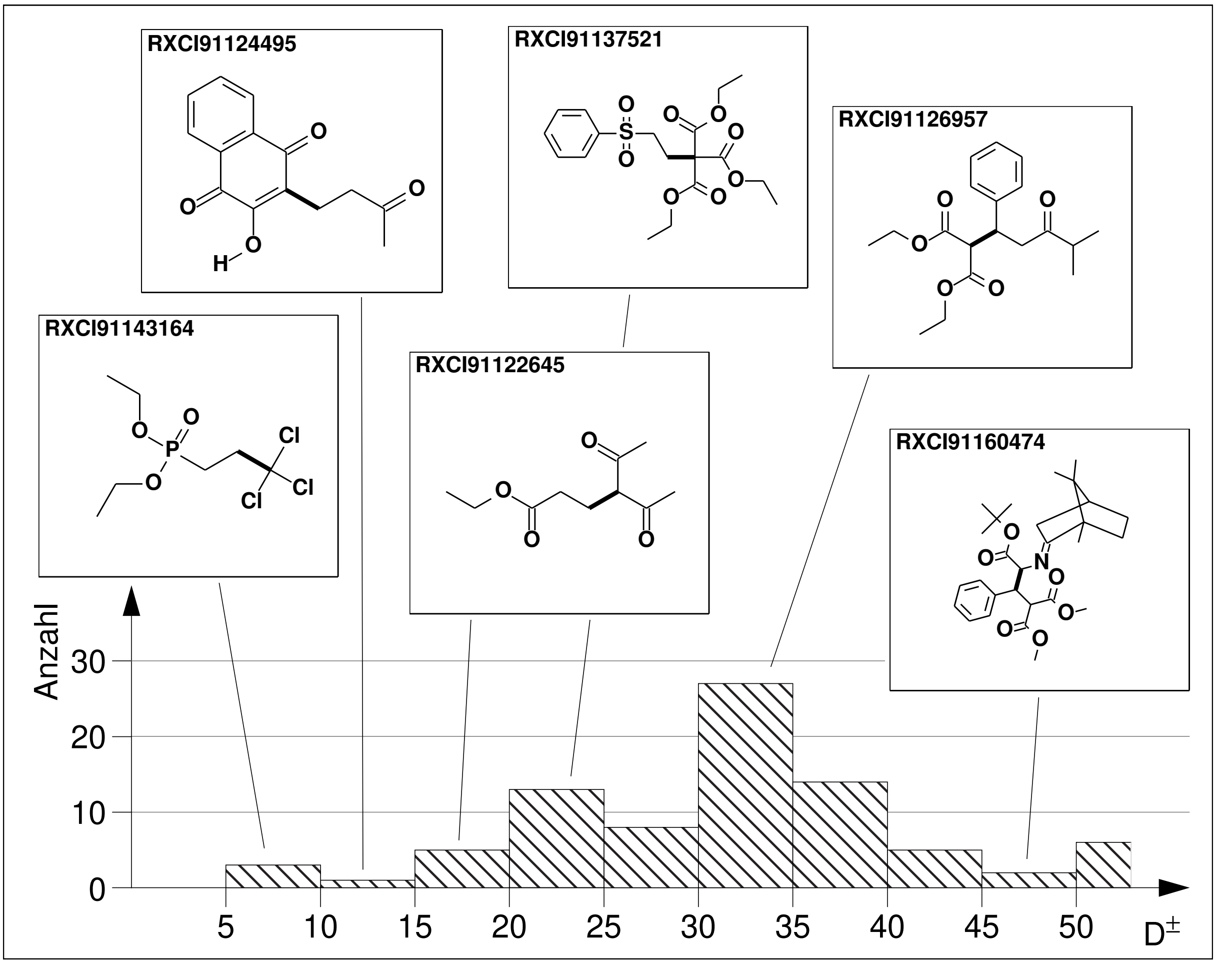


Abb. 4 - 14 Histogramm der D$^{\pm}$-Werte der gewünschten Synthonpolarität einer Michael-Addition für die strategischen Bindungen des Datensatzes. Oben sind Produkte aus einigen Wertebereichen mit Kennzeichnung der strategischen Bindung durch Fettdruck dargestellt.

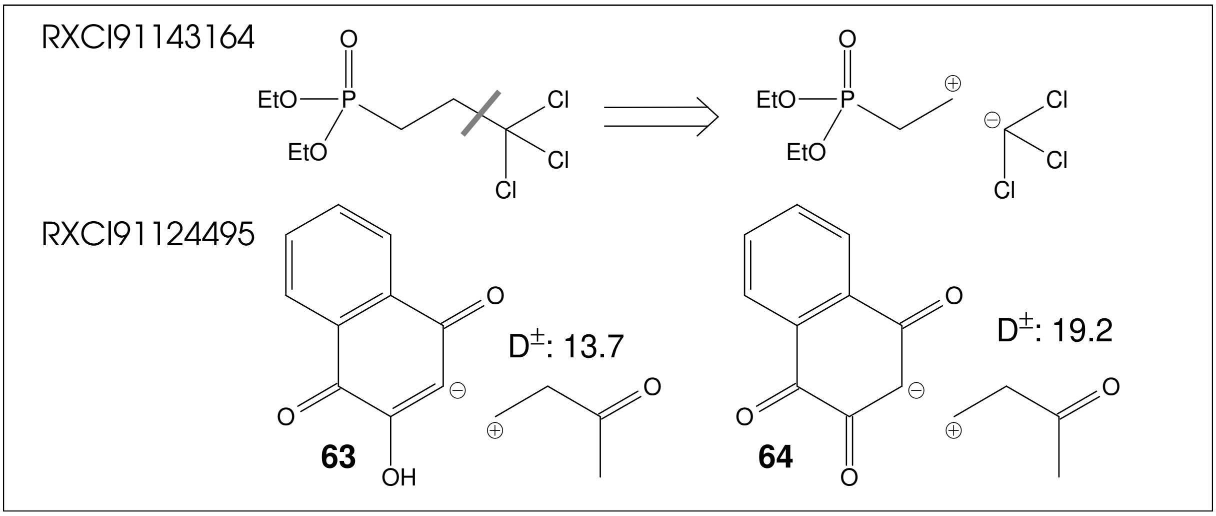


Abb. 4 - 15 Sonderfälle im Reaktionsdatensatz. Oben: Michael-Produkte, die durch Umsetzung von Chloroform als CH-acide Komponente entstehen. Unten: In der Enolform codiertes Produkt mit niedrigem D$^{\pm}$-Wert für die strategische Bindung. Daneben der analoge Wert für die Ketoform.

Offensichtlich ist der $D^{\pm}$-Parameter gut geeignet, eine Michael-Typ-Bindung in einem Syntheseziel als strategische Bindung zu erkennen. Damit dieses Modell zuverlässig arbeitet, muß jedoch auch geprüft werden, ob nicht etwa die Stabilisierung der *umgekehrten* Synthonpolarität von diesem Parameter höher bewertet wird. Tatsächlich ist das in einigen Reaktionen des Datensatzes der Fall (siehe Abb. 4 - 16).

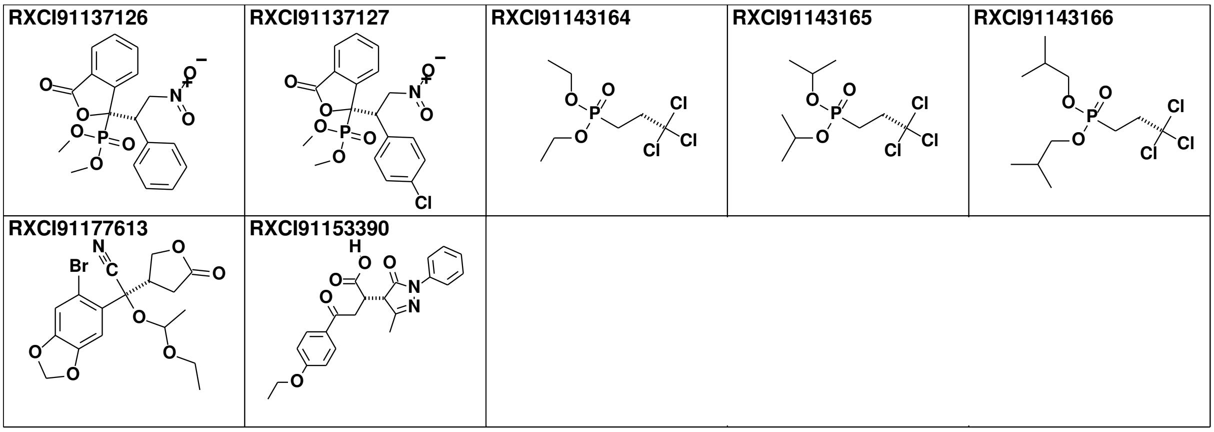


Abb. 4 - 16 Michael-Produkte von Reaktionen, bei denen der $D^{\pm}$-Wert der falschen Synthonpolarität höher ist als jener der korrekten Polarität. Die betroffenen Bindungen sind markiert.

Im Fall der Reaktionen RXCI91137126 und RXCI91137127 wird die Delokalisierungsstabilisierung einer negativen Ladung durch den Phenylring überschätzt. Der D^{-}-Wert liegt deutlich über dem D^{+}-Wert. Bei Reaktion RXCI91177613 ist es umgekehrt: Hier wird der stabilisierende Einfluß der Umgebung des eigentlich elektrophil anzugreifenden Zentrums so eingeschätzt, daß dieses als nucleophil gilt. Die Reaktionen RXCI91143164/5/6, sind jene mit Chloroform als CH-acidem Reagenz. Die Reaktion RXCI91153390 schließlich ist ein Grenzfall: Der $D^{\pm}$-Wert der falschen Synthonpolarität liegt nur geringfügig über dem der anderen Richtung (42.6 gegenüber 40.4). Offensichtlich wird die Situation am Michael-Reaktionszentrum durch mesomere Einflüsse *allein* nicht ausreichend beschrieben. Diese Reaktionen dienen daher als reduzierter Datensatz zur Ableitung eines Modells unter Einbeziehung induktiver Effekte.

Der Lösungsansatz basiert auf der Annahme, daß mesomerer und induktiver Effekt additiv zu einem Gesamteffekt führen*, der eine typische Michael-Situation beschreibt. Mesomere Effekte können ausgezeichnet durch den $D^{\pm}$-Parameter beschrieben werden (siehe oben). Zur Beschreibung induktiver Effekte eignen sich mehrere Parameter, die letztendlich alle in einem gewissen Maße korrelieren (d. h. nicht linear unabhängig voneinander sind). Hier kommt die Differenz der σ-Atompartialladungen (Δq_{σ}) als Deskriptor für die über eine Bindung wirkenden induktiven Effekte zur Anwendung. Um beide Deskriptoren ($D^{\pm}$ und Δq_{σ}) für einen additiven Effekt nutzen zu können, muß wenigstens einer von beiden so skaliert werden, das ein korrekter Effekt errechnet wird.

Um Maßzahlen für diese kombinierten Effekte zu erhalten, wurden die Reaktionen des reduzierten Datensatzes (sieben Reaktionen, siehe Abb. 4 - 16) verwendet: Diese konnten *nicht* ausschließlich durch mesomere Effekte als Michael-Additionen erkannt werden, die Einbeziehung eines Deskriptors für induktive Effekte ist daher hier *zwingend*. Da bei diesen Reaktionen die entgegengesetzte, falsche Richtung des Bindungsbruchs unter alleiniger Berücksichtigung mesomerer Effekte bevorzugt werden würde, muß der induktive Effekt also

* Die Additivität ist natürlich eine Vereinfachung. Sie erweist sich jedoch als nicht kritisch zur empirischen Beschreibung des vorliegenden Problems.

mindestens einen Bonus für den korrekten Bindungsbruch in Höhe der Differenz der mesomeren Effekte für beide Richtungen ergeben.

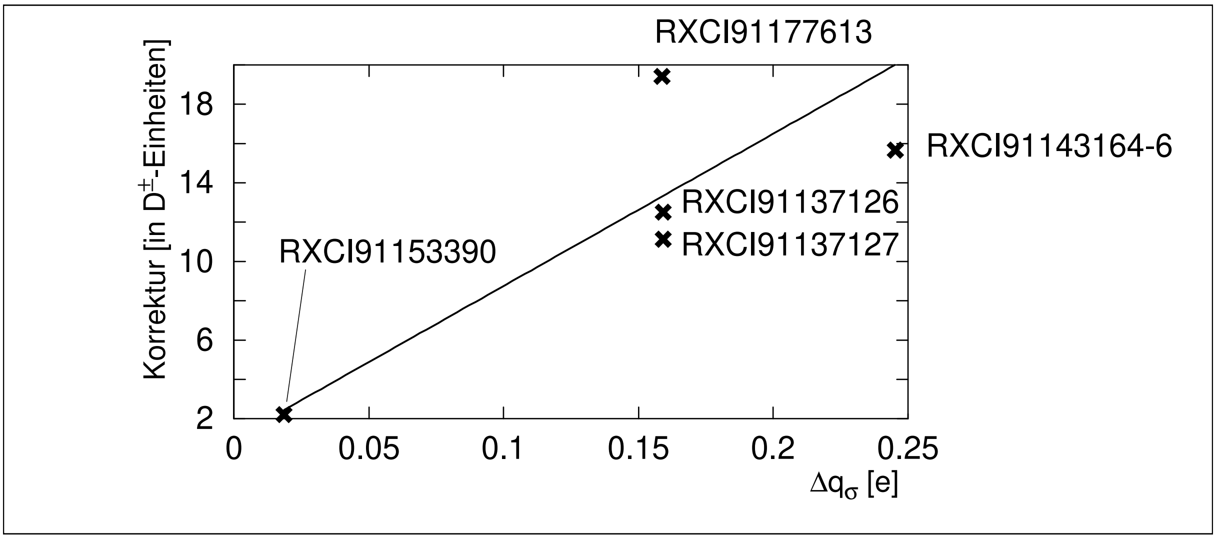


Abb. 4 - 17 Erforderliche Korrekturen aufgetragen gegen die σ-Atompartialladungsdifferenz als Maß der induktiven Effekte. Die Gerade beschreibt eine einfache lineare Funktion, die zur Berechnung der Korrekturwerte aus den σ-Ladungsdifferenzen verwendet wird.

Wie Abbildung 4 - 17 zeigt, läßt sich offensichtlich ein Zusammenhang zwischen der σ-Ladungsdifferenz und gewünschten Korrekturwerten konstruieren, der hier vereinfachender Weise als linear angenommen wird. Der Punkt für Reaktion RXCI91177613 wird jedoch ausgenommen, da er extrem außerhalb liegt.

Aus diesen Betrachtungen läßt sich folgender Zusammenhang ableiten:

$$\text{GesamtEffekt}(b) = f(b) + g(b) \tag{4-1}$$

mit

$$f(b) = D^{\pm}(b) \tag{4-2}$$

$$g(b) = \begin{cases} 77.5\,\Delta q_{\sigma}(b) + 1.0 & \text{für } \Delta q_{\sigma}(b) > 0 \text{ und} \\ 0 & \text{sonst} \end{cases} \tag{4-3}$$

b............Bindungsindex
f(b).........Deskriptorfunktion für mesomere Effekte bezüglich Bindung b
g(b)........Deskriptorfunktion für induktive Effekte bezüglich Bindung b

Mit diesem Ansatz resultieren für die Reaktionen des reduzierten Datensatzes (mit Ausnahme der Reaktion RXCI91177613) nun Werte für den Gesamteffekt an Stabilisierung von weit über 18.0 (dem ursprünglichen Grenzwert). Ein Test dieses Modells gegen den gesamten Datensatz der 84 Michael-Reaktionen bestätigt den Ansatz. In allen Fällen (mit Ausnahme der Reaktion RXCI91177613) wird nach Gleichung (4-1) die gewünschte strategische Bindung mit der korrekten Synthonpolarität erkannt.

Damit läßt sich der Algorithmus zur Identifizierung einer strategischen CC-Bindung in einem Zielmolekül definieren:

- ❏ Untersuche nur CC-Bindungen, die nicht aromatisch sind und nicht in α-Position zu aromatischen Ringen liegen.
- ❏ Prüfe für alle verbliebenen CC-Bindungen ihre Bewertung gemäß Gleichung (4-1).

❏ Verwerfe diese Richtung der CC-Bindung, wenn ihre Bewertung kleiner ist als die der Gegenrichtung. Man ermittelt also diejenige Vorzugsrichtung, bei der quantitativ mehr Stabilisierung (durch mesomere und induktive Effekte) erwartet wird.
❏ Markiere die CC-Bindung als strategisch, wenn der ermittelte Wert 18.0 überschreitet.

Dadurch wird der Ort einer strategischen Bindung festgelegt. Die folgende Abbildung veranschaulicht dieses Vorgehen.

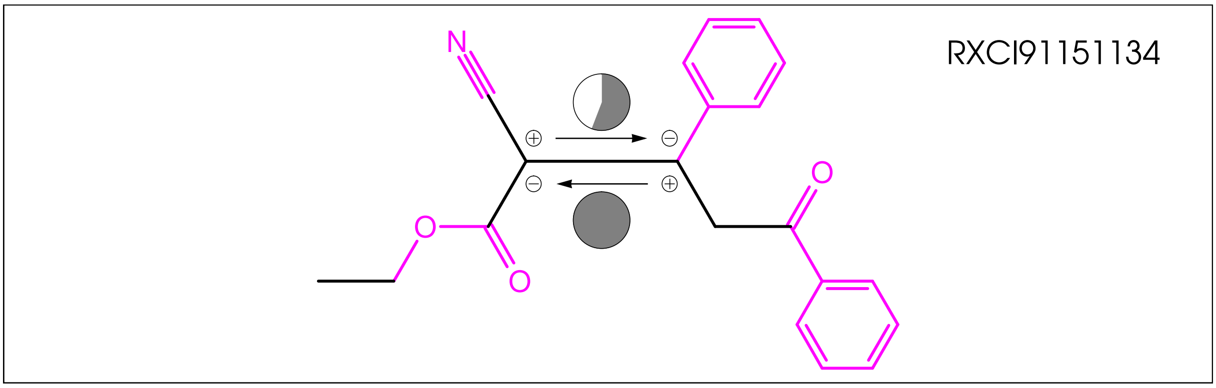


Abb. 4 - 18 Algorithmus zur Auswahl einer strategischen CC-Bindung und der Richtung des Bindungsbruchs am Beispiel der zentralen CC-Bindung einer Testreaktion. Diejenige Richtung, für die sich eine bessere Gesamtstabilisierung ergibt, wird ausgewählt. Das ist im Beispiel die unten dargestellte Richtung. Bindungen, die aus formalen Gründen nicht weiter untersucht werden sind grau dargestellt.

Nachfolgend muß nun der Algorithmus in der Lage sein, die durch den Bindungsbruch entstehenden offenen Valenzen so abzusättigen, daß geeignete Reagenzien entstehen. Für die Seite des negativ geladenen Synthons ist diese Aufgabe leicht durch Addition eines H-Atoms zu lösen. In Syntheserichtung wäre dieses H-Atom als Proton zu entfernen, es muß also in der Vorstufe eine gewisse Acidität besitzen. Acidität des addierten H-Atoms ist dann zu erwarten, wenn elektronenziehende Gruppen der Additionsstelle benachbart sind. Die elektronenziehende Wirkung kann wiederum durch mesomere Effekte oder starke induktive Effekte begründet sein. Beides wird mit Hilfe der oben beschriebenen physikochemischen Deskriptoren des Reaktionszentrums erkannt.

Komplizierter ist die Situation für das positive Synthon. Ein Chloratom wäre ein Repräsentant für eine Abgangsgruppe, die ein positiv geladenes Synthon erzeugt. Das entsprechende Edukt einer Michael-Addition ist jedoch eine elektronenarme Olefinkomponente. Um diese zu erzeugen, muß (aus retrosynthetischer Sicht) ein H-Atom in α-Position zur positiven Synthonladung entfernt werden. Damit ergibt sich im Programm durch Rekombination eine Doppelbindung in der Vorstufe. Damit solche Doppelbindungen nur dann gebildet werden, wenn es für die Syntheserichtung erforderlich ist (d. h. hier: wenn der strategischen Bindung tatsächlich eine Michael-analoge Umsetzung zugrunde liegt), muß erneut nach einem physikochemischen Parameter gesucht werden, der dieses Vorgehen signalisiert.*

Ein solcher Parameter läßt sich mit den Erfahrungen zur Auswahl der strategischen Bindung unschwer finden: Da der Doppelbindung im Edukt elektronenziehende Gruppen benachbart sein müssen und diese Gruppen auch im Produkt unverändert vorhanden sind, sollte der Delokalisierungsparameter D^- über eine CH-Bindung in α-Position zur positiven Synthonla-

* Man könnte alternativ auch die Nachbarschaft des Reaktionszentrums nach Gruppen mit solchen Eigenschaften absuchen. Jedoch ist eine solche Substruktursuche langsamer und auch unvollständiger als es die Berechnung eines geeigneten physikochemischen Deskriptors ist.

dung durch entsprechende Werte eine solche Situation diagnostizieren können (siehe Abb. 4 - 19). Tatsächlich finden sich ausgezeichnete D^--Werte für die CH-Bindungen in α-Position. Als ausreichender Grenzwert wurde 11.0 ermittelt. Damit wird sicher das Vorhandensein elektronenziehender Gruppen erkannt. Im Datensatz sind eine Reihe solcher Gruppen vertreten: Carbonsäureester, Ketone, Nitrogruppen, Sulfinyl- und Sulfonylgruppen, Phosphite - manche davon konjugiert zu Aromaten. Der Parameter beschreibt also eine große chemische Bandbreite. Dieses Maß kommt auch bei der Olefin-Strategie für Amine zur Anwendung (siehe Abschnitt 4.3.2 Seite 103).

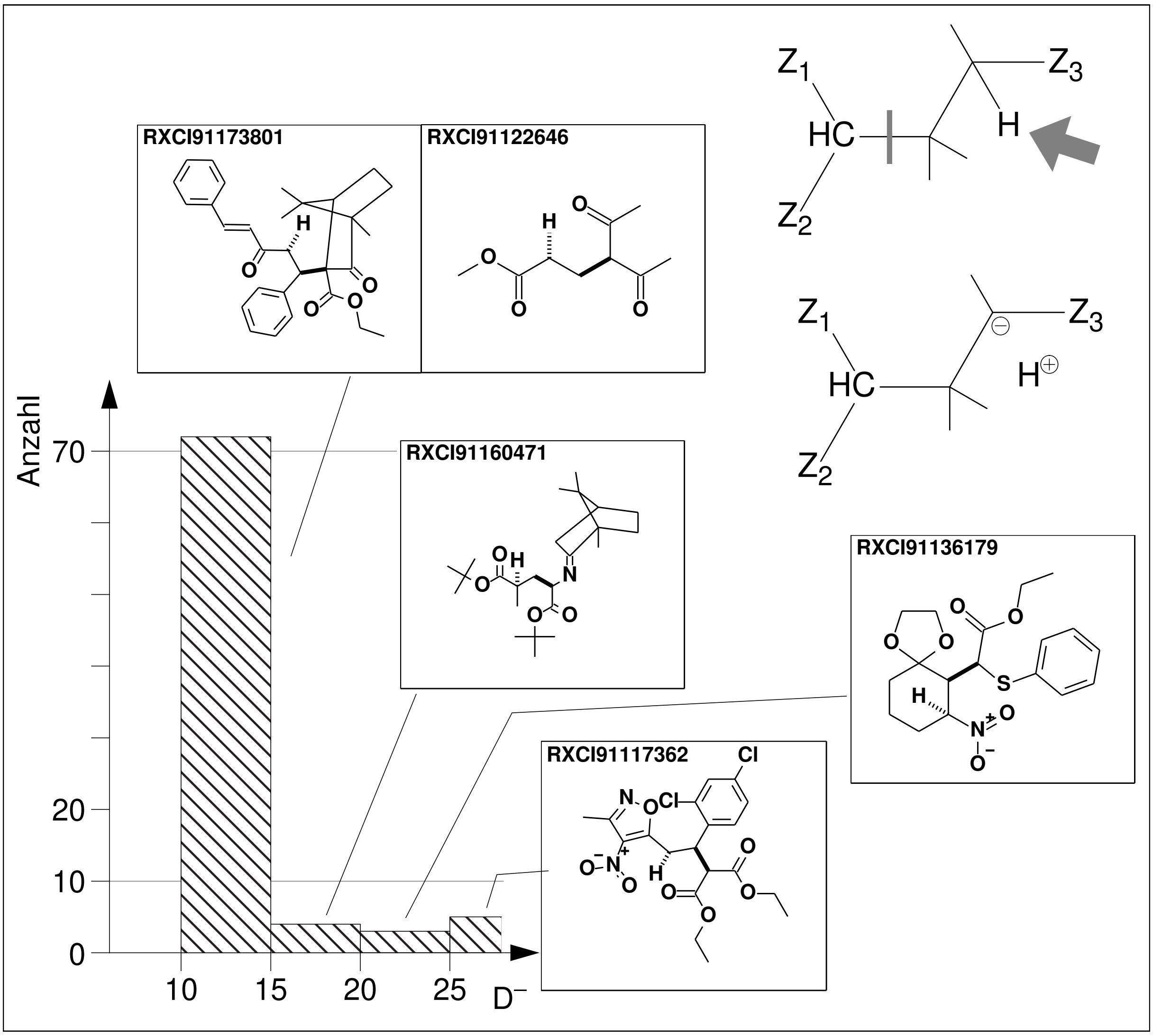


Abb. 4 - 19 Oben rechts: Topologische Position der CH-Bindung, deren D^--Wert die Anwesenheit von elektronenziehenden Gruppen in der Nachbarschaft signalisiert. Darunter das bewertete Ladungsmuster. Unten: Verteilung dieser Werte im Datensatz der 84 Michael-Additionen mit Beispielreaktionen.

Damit der Algorithmus zur Ableitung von Vorstufen nicht auf Michael-typische Situationen beschränkt bleibt, werden in Fällen, wo zwar induktive und mesomere Effekte auf eine strategische Bindung hindeuten, die Umgebung der Bindung aber nicht Michael-konform ist*, einfache Alternativen angewendet.

* Das ist beispielsweise dann der Fall, wenn es kein H-Atom am Atom 3 des Reaktionszentrums gibt.

Auf der Seite der negativen Ladung des Synthons wird dann, wenn keinerlei Stabilisierung durch mesomere Effekte vorhanden ist (D^--Parameter ist Null) und induktive Effekte allein zu gering ausfallen (meßbar durch die Ladungsdifferenz Δq_σ) ein Brom-Atom statt Wasserstoff addiert. Ein H-Atom wäre in solchen Fällen nicht acid genug für einen nucleophilen Angriff. Atome mit geringerer Elektronegativität als Wasserstoff (Metallorganyle) stellen jedoch die Bildung eines Carbanionen-Synthons sicher. Diese Metallorganyle werden aber aus den korrespondierenden Halogenverbindungen hergestellt. Da WODCA die Syntheseplanung hin zu Ausgangsmaterialien forcieren soll, wird durch die direkte Generierung der Brom-Verbindung ein Syntheseschritt übersprungen. Es obliegt dem Chemiker zu erkennen, daß die Brom-Vorstufe in ihr Metallorganyl überführt werden muß, damit die beabsichtigte Bindungsknüpfung vonstatten gehen kann.*

Fehlen elektronenziehende Gruppen in der Nachbarschaft des positiv geladenen Synthons oder fehlt ein abspaltbares H-Atom in α-Position, wird ein Chloratom als Äquivalent für eine negative Abgangsgruppe addiert. Die Ähnlichkeitssuchen des voranstehenden Kapitels können mit solchen Vorstufen leicht Beziehungen zu anderen analogen Reagenzien herstellen.

Als letzte Teilaufgabe verbleibt für eine Retrosynthesefunktion in WODCA die Vergabe einer Bewertung für verschiedene strategische Bindungen (siehe Abschnitt 4.1). Das ist hier unschwer möglich. Es wird dazu der nach Gleichung (4-1) modellierte Gesamteffekt an mesomerer und induktiver Stabilisierung eines heterolytischen Bindungsbruchs herangezogen. Wie bei der Retrosynthesefunktion für CX-Bindungen (vgl. Abschnitt 4.3.2 Seite 100) werden die Bewertungen am Ende skaliert, so daß die beste Bewertung bei 100 liegt. Vorher werden aus dem Satz an strategischen Bindungen topologisch völlig identische Dubletten entfernt. Die Zahl der Vorschläge soll damit auf ein notwendiges Maß reduziert werden.

Nachfolgend soll nun an einigen Beispielen die Anwendung der oben beschriebenen Funktion demonstriert werden. Dabei ist es besonders interessant zu prüfen, ob der Algorithmus auch bei Verbindungen verwendet werden kann, die Gruppen enthalten, die nicht im Modelldatensatz enthalten waren.

Eine ungewöhnliche elektronenziehenden Gruppe zur Aktivierung der olefinischen Komponente einer Michael-Addition stellen Bimane† (1,5-Diazabicyclo[3.3.0]octa-3,6-dien-2,8-dione) dar. Abbildung 4 - 20 stellt ein durch Michael-Addition synthetisierbares Biman-Derivat (**65**) dar. WODCA identifiziert mit obiger Retrosynthesefunktion mehrere strategische Bindungen.‡ Das mit jeweils 68 bewertete strategische Bindungspaar entspricht einer doppelten Michael-Addition (siehe resultierende Vorstufen in Abb. 4 - 20). Das beweist, daß die Retrosynthesefunktion auch andere als die elektronenziehenden Gruppen des Modelldatensatzes behandeln kann. Interessant sind auch die anderen strategischen CC-Bindungen. Wendet man zusätzlich die Retrosynthesefunktion aus Abschnitt 4.3.2 an (Suche nach strategischen CX-Bindungen), werden weitere Bindungen als strategisch identifiziert. Wird ein geeigneter Satz an Kombinationen von strategischen Bindungen ausgewählt und retrosynthetisch gebrochen, resultieren unmittelbar Reagenzien bzw. Reagenzienäquivalente, die tatsächlich für die Synthese von Bimanen eingesetzt werden [118] (siehe Abb. 4 - 22): Durch Claisen-Kondensation von zwei Molekülen Propansäureester (**66**) entsteht der β-Ketoester (**67**). Dieser wird mit Hydrazin zum 3-Ethyl-4-methyl-5-pyrazolinon (**68**) kondensiert und über zwei weitere

* Der Anwender erhält insofern eine Hilfestellung, da Brom per Konvention als Markeratom für umzupolende Reaktivität verwendet wird. Repräsentiert das Halogen eine negative Abgangsgruppe, wird vom Programm immer Chlor verwendet.

† Der Name leitet sich von *bi* - zwei und *manus* - Hand ab [117].

‡ Topologisch äquivalente Bindungen sind hier *nicht* entfernt worden.

Schritte zum 4,6-Diethyl-3,7-dimethyl-biman (**69**) umgesetzt. Bromierung und anschließende Eliminierung resultieren im erforderlichen Michael-Olefin (**70**), das mit dem entsprechenden Malonsäureester (**71**) umgesetzt wird. Das Syntheseziel entsteht daraus durch Decarboxylierung.

Ähnlichkeitssuchen sind mit den für diese Synthese notwendigen Reagenzien nicht notwendig. Propansäureester (**66**), Hydrazin und Malonsäureester (**71**) können als identische Verbindungen im Fluka-Katalog gefunden werden.

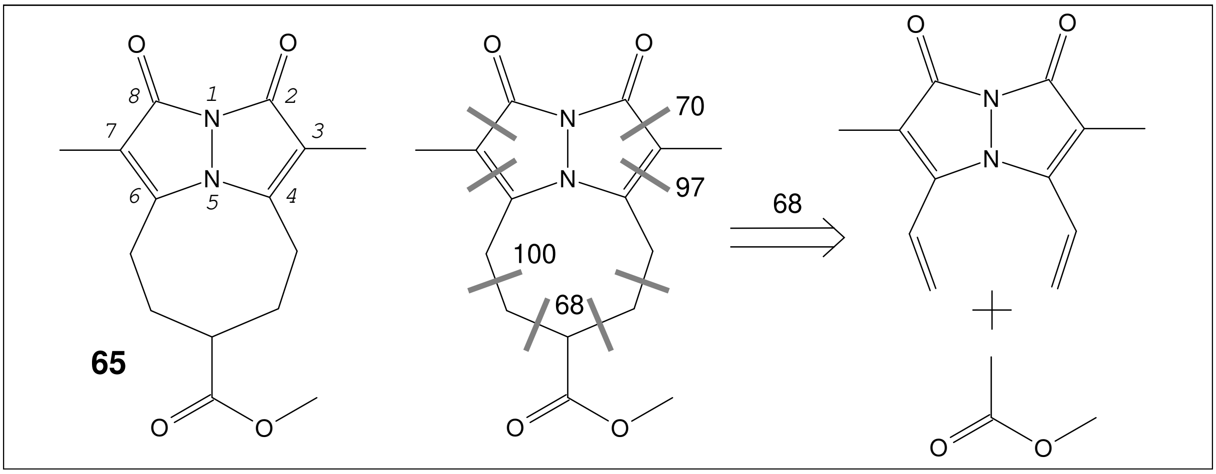


Abb. 4 - 20 Syntheseziel Biman-Derivat (links) mit Atomindizierung im Heterocyclus. In der Mitte sind die strategischen CC-Bindungen markiert, die von WODCA identifiziert werden. Die Zahlen geben die zugeordnete Bewertung an. Rechts unten sind die aus dem Bruch eines Bindungspaares (bewertet mit je 68) resultierenden Vorstufen dargestellt. Dieser retrosynthetische Schnitt entspricht einer Michael-Addition in Syntheserichtung.

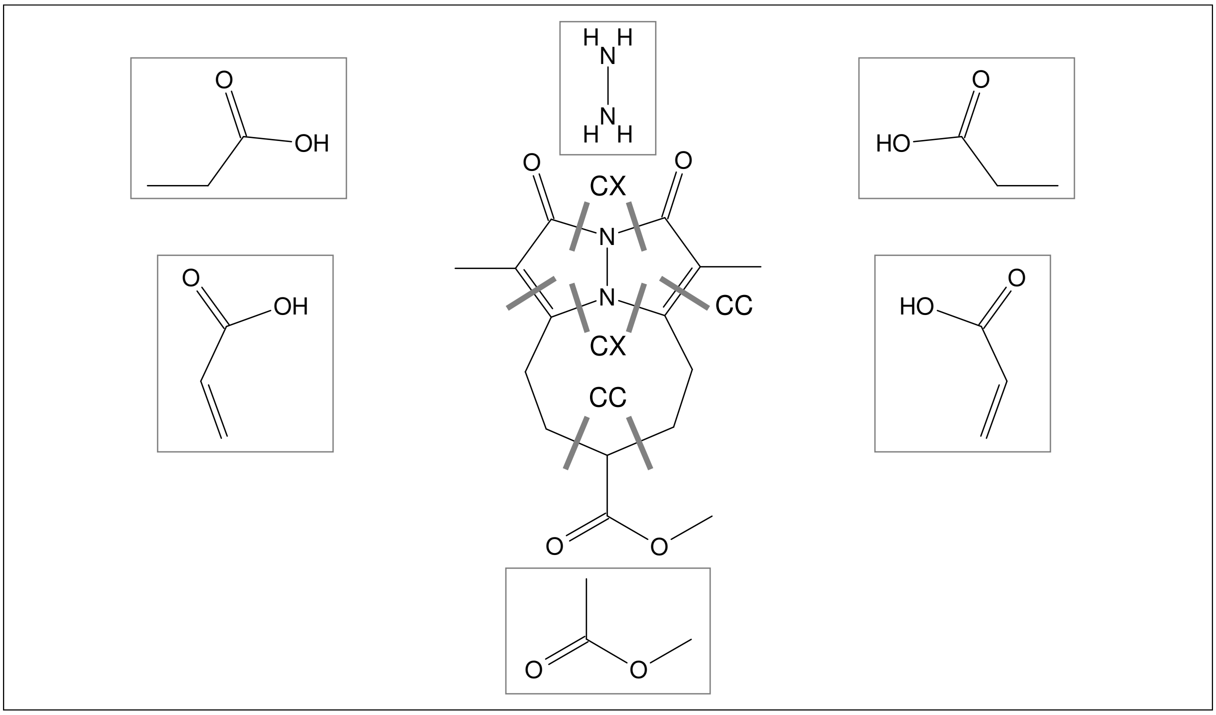


Abb. 4 - 21 Retrosynthetischer Bruch mehrerer (manuell kombinierter) strategischer Bindungen und daraus resultierende Synthesebausteine. Mit CC gekennzeichnete strategische Bindungen wurden von der in diesem Abschnitt beschriebenen Retrosynthesefunktion ermittelt, die mit CX gekennzeichneten durch die Algorithmen des Abschnitts 4.3.2.

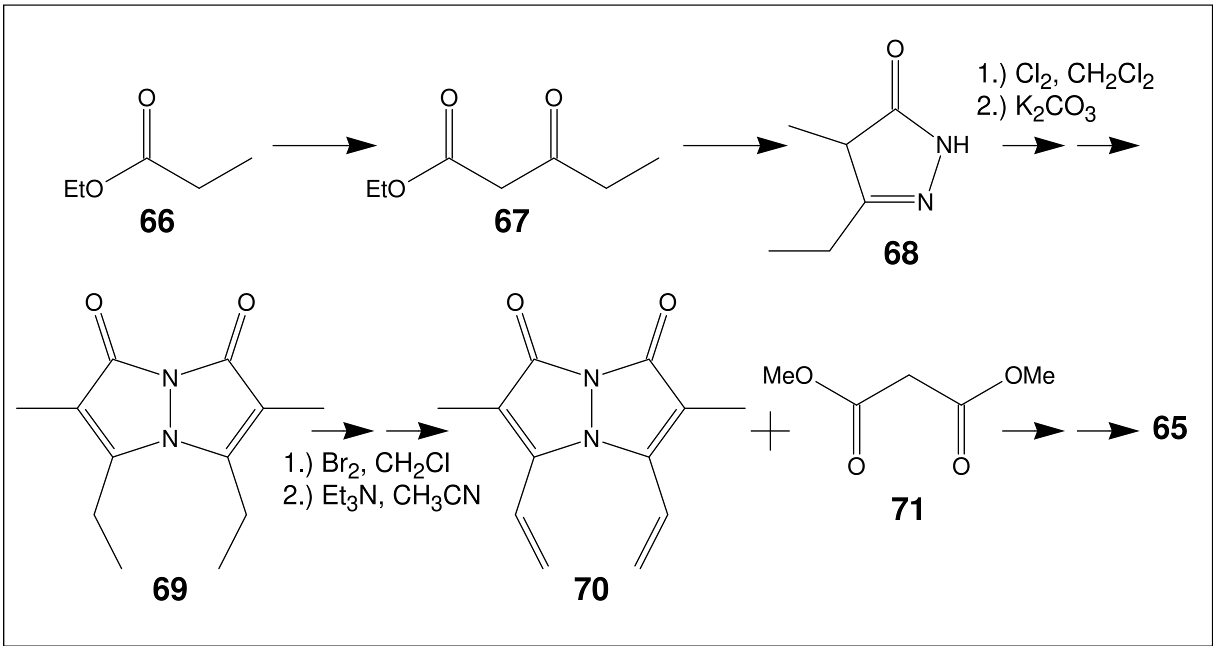


Abb. 4 - 22 Publizierter Syntheseweg für das Biman-Derivat [118].

Ein zweites Beispiel untersucht ein δ-Sultam-Derivat (**72**, siehe Abb. 4 - 23). Für diese Verbindungsklasse wurde vermutet, daß sie als Thromboxan-A_2-Rezeptorantagonisten einsetzbar wären [119]. Auch hier war eine Michael-Addition ein Schlüsselschritt bei der Darstellung dieser Verbindungen. Die hier getestete Verbindung ist eine direkte Vorstufe eines δ-Sultams. Zunächst werden strategische CX-Bindungen gesucht. WODCA identifiziert die Esterbindungen und die phenolische CO-Bindung als strategisch. Letztere und eine der beiden Esterbindungen werden retrosynthetisch gebrochen.

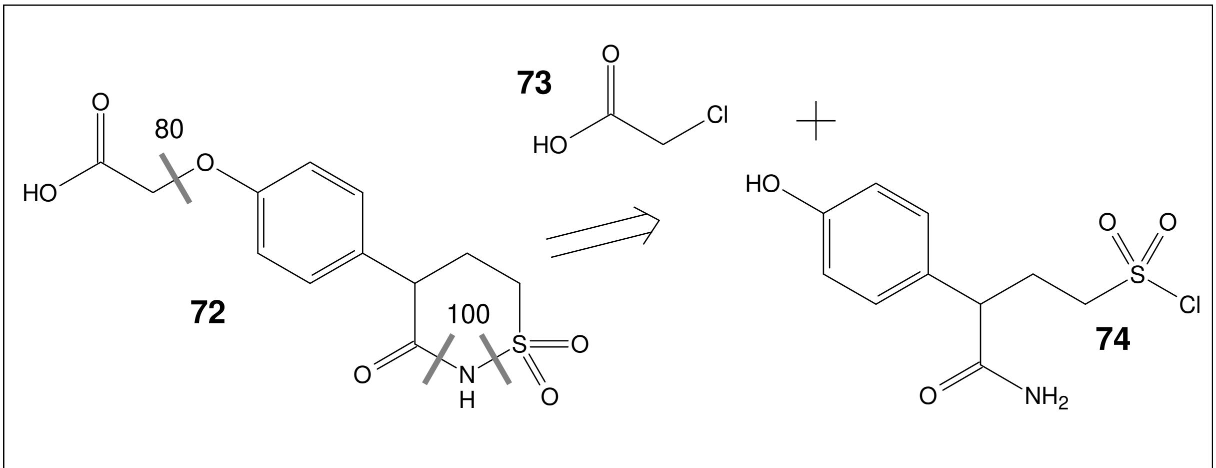


Abb. 4 - 23 Anwendungsbeispiel δ-Sultam-Derivat. Von den gefundenen drei strategischen CX-Bindungen werden zwei gebrochen und Vorstufen generiert.

Für den aliphatischen C_2-Baustein (**73**) können sofort mit einer selektiven Ähnlichkeitssuche Ausgangsmaterialien ermittelt werden. Hier kommt das Kriterium „Element- und ZH_n-Austausch“ zum Einsatz. Die geeignetsten Ausgangsmaterialien Chlor- bzw. Bromessigsäure sind in der von Duplikaten befreiten Liste von 20 Verbindungen leicht auszumachen (**73** und **75**, siehe Abb. 4 - 24).

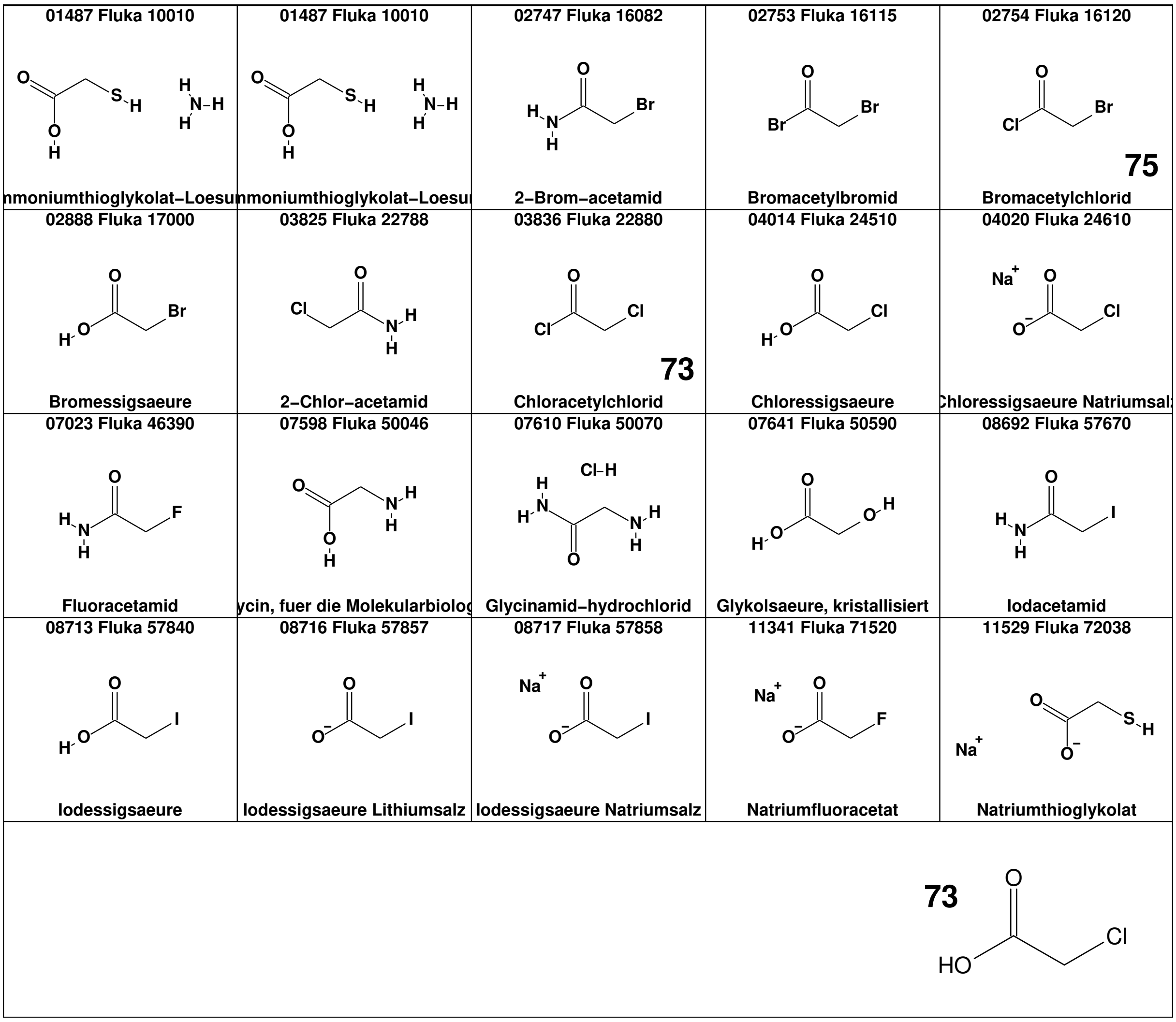


Abb. 4 - 24 Ähnlichkeitssuche mit Chloressigsäure im Fluka-Katalog mit dem Kriterium „Element und ZH_n-Austausch“. Duplikate sind entfernt. Es resultieren 20 Treffer.

Die Suche nach strategischen CC-Bindungen in der verbliebenen Vorstufe (**74**) ergibt zwei Vorschläge (siehe Abb. 4 - 25). Beide entsprechen einer Michael-Addition wie die Vorstufen zeigen, die beim Bindungsbruch generiert werden. Die höher bewertete strategische Bindung korrespondiert mit einer publizierten Synthese [120]. Für die dabei resultierenden Vorstufen Vinylsulfonylchlorid (**76**) und 4-Hydroxy-phenylessigsäureamid (**77**) können dann mit Hilfe der Transformationssuchen leicht Ausgangsmaterialien ausgewählt werden. Für Vinylsulfonylchlorid ermittelt das Ähnlichkeitskriterium „C-Skelett einschließlich α-Atome“ zwei Treffer im Fluka-Katalog. Eine klassische Substruktursuche findet ebenfalls zwei Treffer (beide Listen siehe Abb. 4 - 26). Die zweite Vorstufe bereitet ebenfalls keine Probleme: Eine Ähnlichkeitssuche basierend auf dem Substitutionsmuster (*-AR+A1-CCMB+BO+MU*) ermittelt 15 Treffer ohne Duplikate. Wird der phenolische Substituent nicht generalisiert (*-AR-__A1__-CCMB+BO+MU*), reduziert sich diese Liste auf nur acht Verbindungen (in Abb. 4 - 27 markiert mit ●). Als Alternativen können auch die Hydrolyse-Kriterien verwendet werden. Wird die Hydrolyse der phenolischen Gruppe gestattet, resultieren neun Treffer (markiert mit ⊘), andernfalls nur drei (markiert mit ◍). Diese letzte Liste läßt den Anwender unschwer das Ausgangsmaterial der Wahl erkennen: 4-Hydroxy-phenylessigsäuremethylester (**76**).

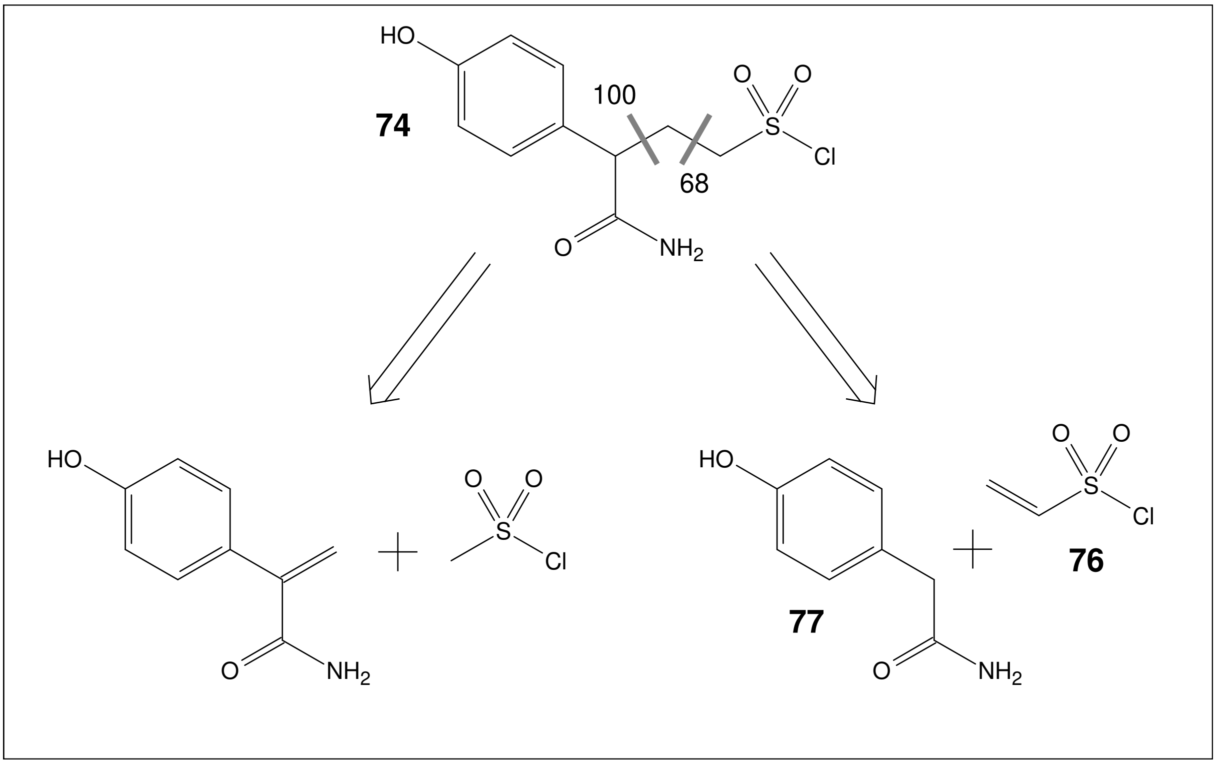


Abb. 4 - 25 Strategische CC-Bindungen in der zweiten Vorstufe. Rechts und links unten die resultierenden Reagenzien nach Bruch dieser Bindungen. Die gleichzeitig zu brechenden CH-Bindungen sind in der Zielverbindung aus Gründen der Übersichtlichkeit nicht dargestellt.

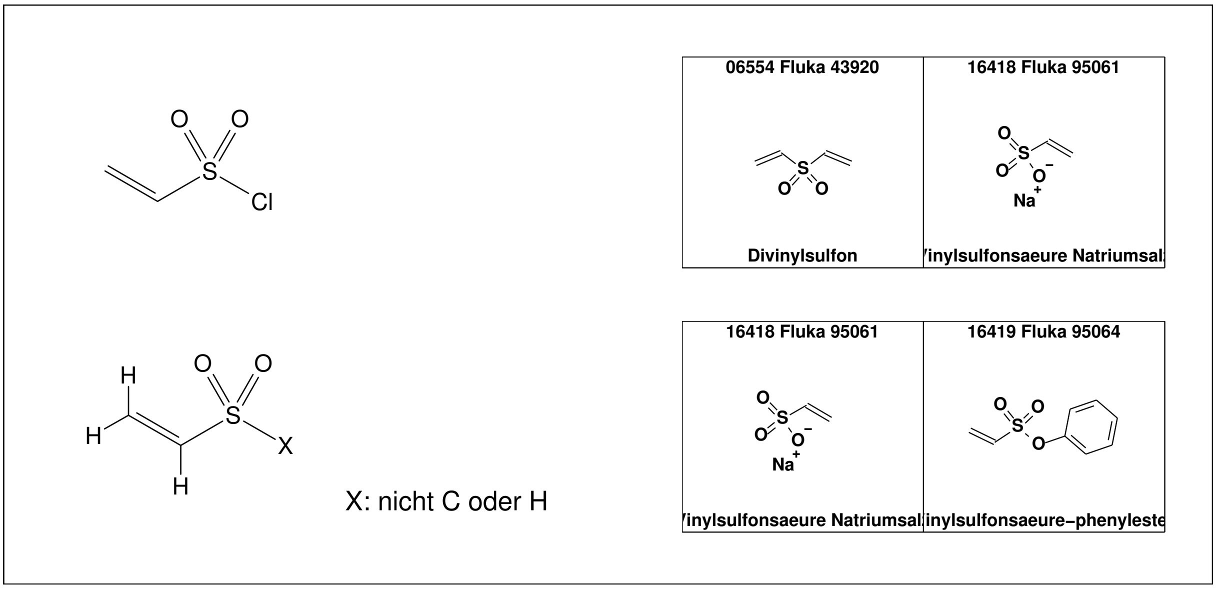


Abb. 4 - 26 Suche nach Ausgangsmaterialien für Vinylsulfonylchlorid. Oben das Resultat einer Ähnlichkeitssuche mit dem Kriterium „C-Skelett einschließlich α-Atome“, unten das Ergebnis einer Substruktursuche. Die benutzte Substruktur ist unten links dargestellt.

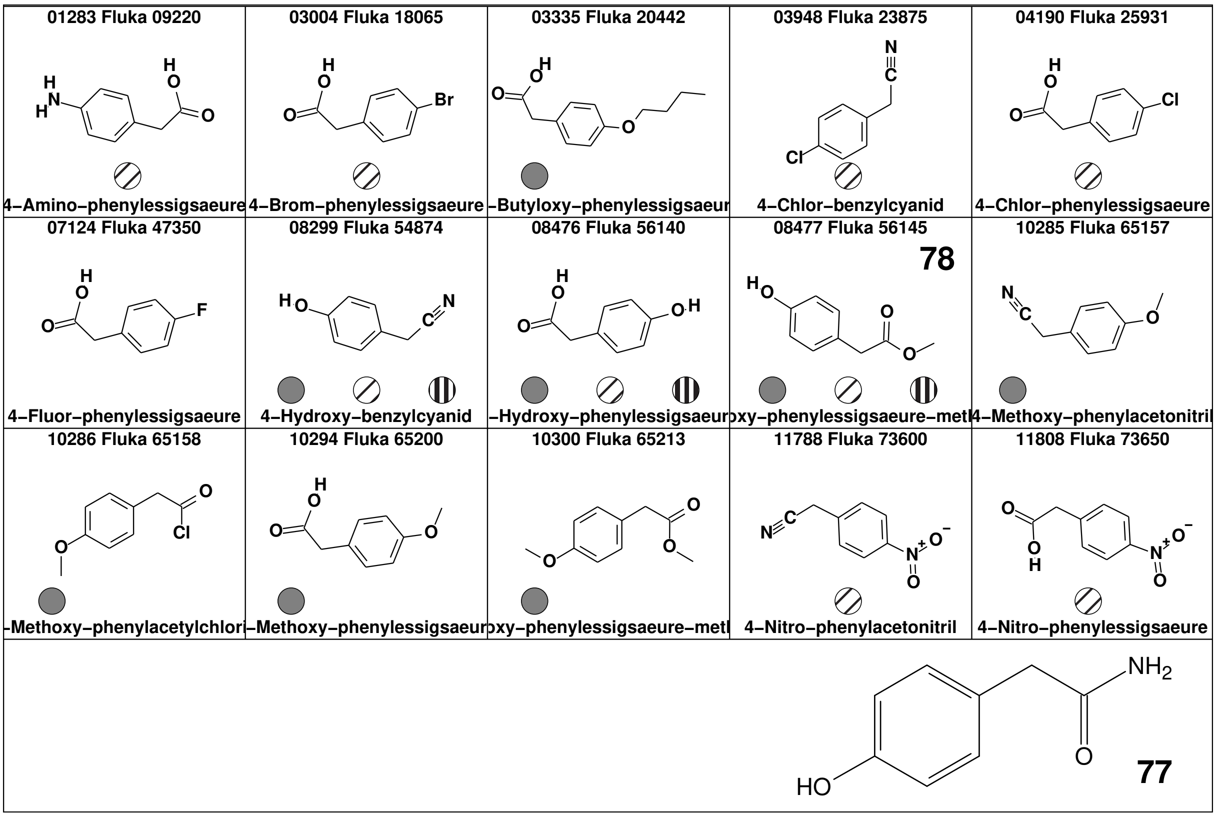


Abb. 4 - 27 Transformationssuchen mit 4-Hydroxy-phenylessigsäureamid (unten) im Fluka-Katalog. Die abgebildeten 15 Treffer resultieren mit dem Kriterium „Substitutionsmuster (-AR+A1-CCMB+BO+MU)“. Die Markierungen sind im Text erläutert.

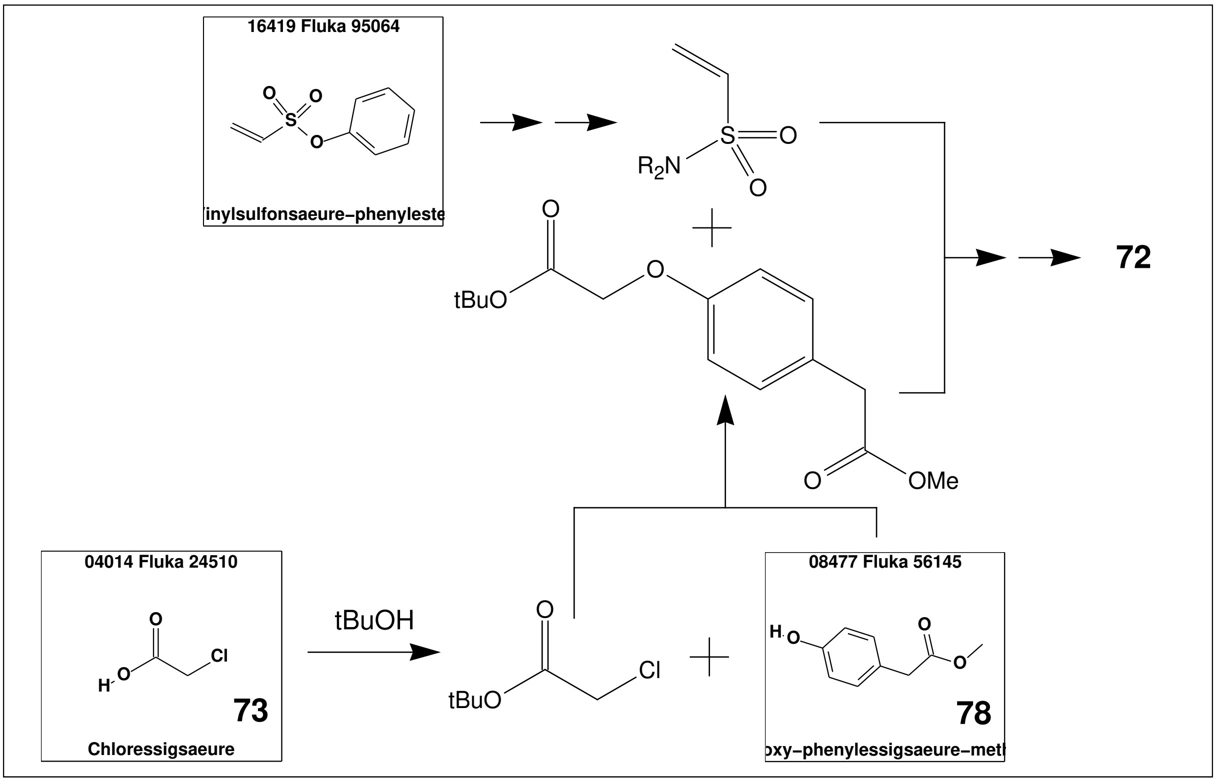


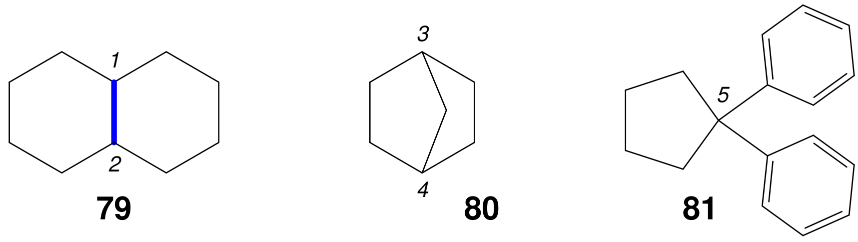
Abb. 4 - 28 Der Syntheseplan für das δ-Sultam-Derivat (**72**). Die umrahmten Verbindungen sind die von WODCA identifizierten Ausgangsmaterialien.

4.3.4 Eine formale Strategie für komplexe Ringsysteme

Dieser letzte Abschnitt soll eine für WODCA implementierte Methode beschreiben, welche die Auswahl der strategischen Bindungen ausschließlich nach formalen topologischen Kriterien betreibt. Diese Methode stützt sich auf die von Corey [100], [12] aufgestellten Regeln zur Suche nach strategischen Bindungen in komplexen Ringsystemen. Diese wurden aufgestellt, um den zahlreichen, in der Transform-Bibliothek von LHASA codierten Retro-Reaktionen eine lenkende Auswahlstrategie zur Verfügung zu stellen. Gleichwohl können diese Regeln auch ohne Computer-Unterstützung zur Planung von Synthesen angewendet werden.

Einige Begriffe, die zur Formulierung der Regeln gebraucht werden, sollen zunächst erläutert werden:

- Der SSSR (*smallest set of smallest rings*) ist diejenige Menge an möglichst kleinen Ringen, die zur Beschreibung *aller* Ringe einer Verbindung notwendig und ausreichend ist. Bei Corey heißen diese Ringe primäre Ringe.
- Ringe, die durch Kombination zweier Ringe aus dem SSSR entstehen, dabei aber nicht größer als 7-gliedrig sind, werden hier als sekundäre Ringe verstanden. Nach Corey sind sekundäre Ringe *alle* Ringe, die aus Ringpaaren des SSSR ableitbar sind.[*]
- Der für Coreys Regeln wichtige Satz an synthetisch signifikanten Ringen leitet sich dann ab aus der Summe aller primären und sekundären Ringe, die mehr als 3-gliedrig und kleiner als 8-gliedrig sind.
- Bindungen, deren retrosynthetischer Bruch zur Bildung eines neuen primären Ringes mit acht oder mehr Atomen führen würde, werden Kern-Bindungen genannt. (siehe Abb. 4 - 29)
- Brückenkopfatome sind für WODCA jene Ringatome, die zu zwei oder mehr Ringen gehören und zwei oder mehr Ringbindungen zu Nachbaratomen haben, die nicht zum gleichen Ring gehören. Dabei werden Spirozentren *nicht* als Brückenköpfe betrachtet. Von Corey wird zusätzlich verlangt, daß die zwei Brückenkopfatome verbindende Überschneidung zweier Ringe mehr als eine Bindung lang ist. Abbildung 4 - 29 veranschaulicht die Unterschiede der Definitionen.
- Der maximal verbrückte Ring ist derjenige Ring[†], der die größte Anzahl an Brückenkopfatomen aufweist.



79 **80** **81**

Abb. 4 - 29 Brückenkopfatome in verschiedenen Verbindungen. Die Atome 1 und 2 im Decalin (**79**) sind Brückenkopfatome im WODCA System, nach der Corey-Definition jedoch nicht. Die markierte Bindung ist eine Kern-Bindung. Im Norbornan (**80**) sind 3 und 4 nach beiden Definitionen Brückenköpfe. Das Atom 5 in (**81**) ist kein Brückenkopfatom.

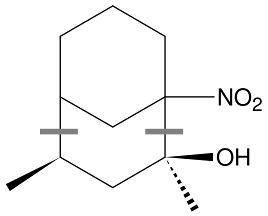
[*] In [12] werden diese sekundären Ringe auch als *periphere* Ringe bezeichnet. Die Einschränkung auf Ringe bis zu sieben Gliedern macht WODCA, weil für den Satz an synthetisch signifikanten Ringen die größeren Ringe ohnehin ausgeschlossen werden.

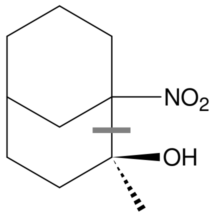
[†] Das können auch mehrere Ringe sein, welche die gleiche Anzahl Brückenkopfatome aufweisen.

Die Regeln lauten wie folgt:

1. Strategische Bindungen müssen in einem primären Ring mit vier bis sieben Ringgliedern liegen.
2a. Die strategische Bindung muß in direkter Verbindung (*exo*) zu einem weiteren Ring stehen.
2b. Die strategische Bindung darf jedoch nicht *exo* zu einem 3-gliedrigen Ring sein.
3. Die strategischen Bindungen müssen im maximal verbrückten Ring liegen.
4. Kern-Bindungen sind niemals strategisch.
5. Aromatische Bindungen sind ebenfalls nicht strategisch.
6a. Liegt die strategische Bindung auf einem Verbindungspfad zwischen zwei Brückenkopfatomen und enthält dieser Pfad ein oder mehrere chirale Zentren, wird die betreffende Bindung nicht als strategisch betrachtet (siehe markierte Bindungen im Beispiel (**82**) in Abb. 4 - 30).
6b. Existiert jedoch nur *ein* chirales Zentrum auf diesem Pfad und ist dieses Zentrum Bindungsatom der strategischen Bindung, wird diese Bindung nicht ausgeschlossen (siehe markierte Bindung im Beispiel (**83**) in Abb. 4 - 30).

82





83

Abb. 4 - 30 Beispielfälle zur Erläuterung der Regeln 6a und 6b. Im linken Beispiel werden die markierten Bindungen als strategische Bindungen nach 6a ausgeschlossen. Im rechten Fall dagegen, erlaubt 6b die hervorgehobene Bindung als strategische Bindung.

Darüber hinaus werden auch alle CX-Bindungen in die Liste der strategischen Bindungen aufgenommen, für welche die Regeln 2b, 4, 5 sowie 6a und 6b erfüllen.

Diese Regeln sind für das LHASA-Programm im Quellcode formuliert, d. h. sie sind gewissermaßen fest „verdrahtet“ im Gegensatz etwa zu den Transforms, die mittels *CHMTRN* codiert sind. Für WODCA werden auch diese Regeln, wie die anderen Methoden zur Suche nach strategischen Bindungen (siehe vorangehende Abschnitte) oder die Transformationssuchen (siehe Kapitel 3) in *tcl* codiert. Eventuell erforderliche Veränderungen werden damit *jedem* Anwender möglich.

Um die Formulierung der obigen Regeln in *tcl* zu bewerkstelligen, wurde das `ring` Kommando als zusätzliche Erweiterung implementiert (siehe Abschnitt 3.2.3). Die Größenbegrenzung im Begriff der synthetisch signifikanten Ringe basiert auf der Erfahrung, daß Großringe (ab acht Atome) schwieriger zu synthetisieren sind. Auch der Begriff der Kern-Bindung wurde aus diesem Grund eingeführt. Diese Begrenzung ist jedoch nicht unumstritten. Deslongchamps [121] fordert beispielsweise gerade die Entwicklung von Synthesemethoden für Ringe mittlerer und großer Gliederzahl. Für diese Arbeit wurde Coreys Begrenzung beibehalten. Durch die externe Codierung der Regeln ist es jedoch problemlos und einfach für jeden Anwender möglich, derartige Beschränkungen zu verändern.

Die WODCA-Implementation weicht in einzelnen Punkten von den Regeln Coreys ab.

❑ Die Definitionen des maximal verbrückten Rings unterscheiden sich, weil die Definitionen von Brückeköpfen verschieden sind (siehe oben).

- ❑ Die Regeln 6a und 6b werden nicht angewendet. Seit Aufstellung der Regeln durch Corey sind viele Reaktionen so weiterentwickelt worden, daß sie stereokontrolliert ablaufen können. Es erscheint daher nicht länger erforderlich, diese Einschränkung zu berücksichtigen. In [12] werden diese Regeln ebenfalls dahingehend modifiziert, daß die Entfernung der Stereozentren dem retrosynthetischen Bruch der Ringbindung *vorangehen* soll.

In Abbildung 4 - 31 wurde die WODCA-Implementation der Corey-Regeln auf Beispielverbindungen der Originalpublikation [100] angewendet. In diesen Fällen sind die Ergebnisse identisch. Einzige Ausnahme ist das Bicyclo[3.2.1]octan-Derivat (**84**), für das WODCA eine zusätzliche strategische Bindung lokalisiert, die bei Corey aufgrund der Regel 6a eliminiert wird.

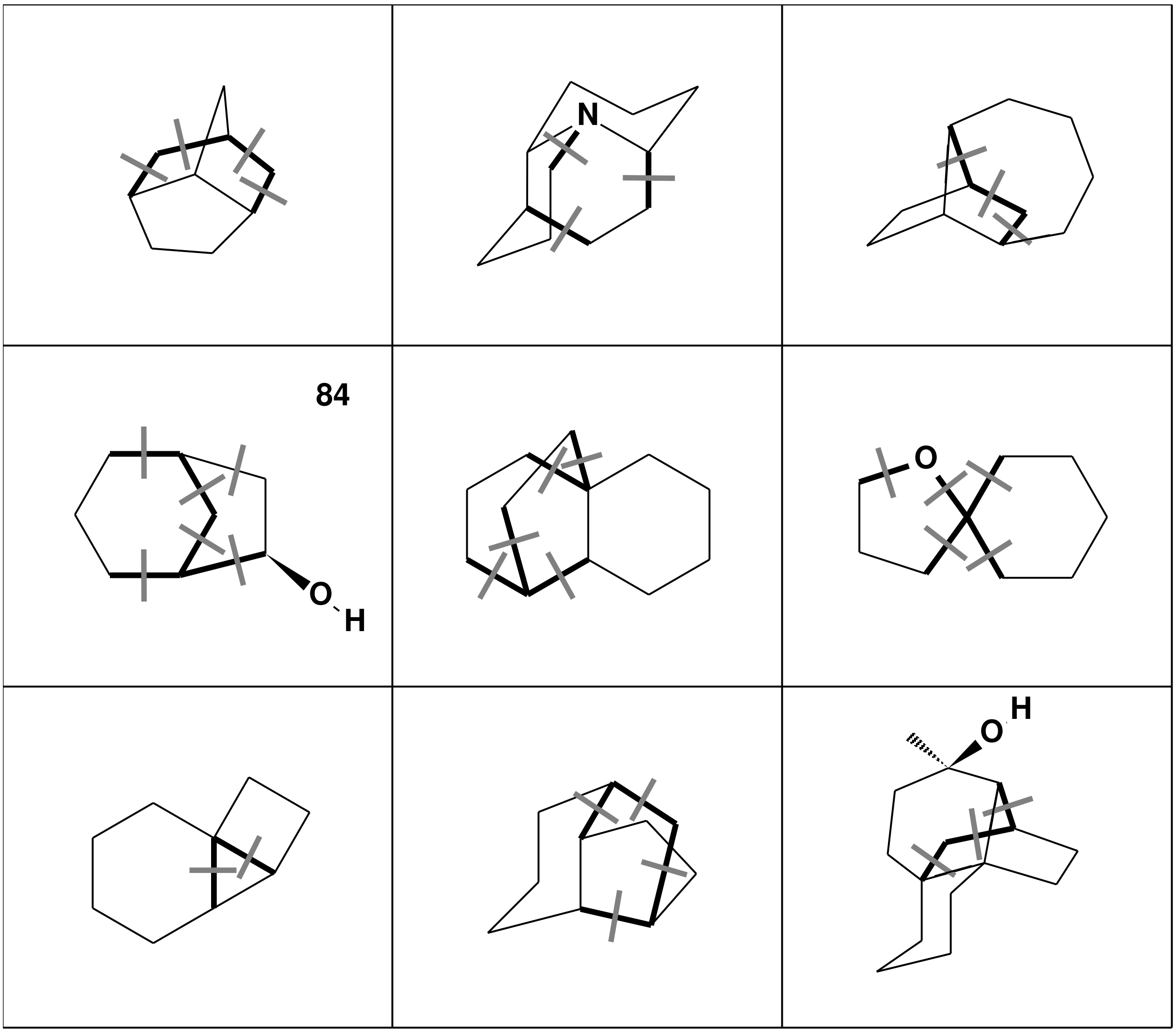


Abb. 4 - 31 Anwendung der WODCA-Implementation der Regeln zur Auswahl strategischer Bindungen in komplexen Ringsystemen. Die fett markierten Bindungen sind die in der Originalpublikation als strategisch erkannten Bindungen. Der Querstrich markiert die von WODCA ermittelten strategischen Bindungen.

Die beschriebenen Regeln selektieren nur den *Ort* eines retrosynthetischen Bindungsbruchs. Sie liefern keinerlei Information über die zugrundeliegende Reaktion oder notwendige Abgangsgruppen. In LHASA werden mit den vorgegebenen strategischen Bindungen nun solche Retro-Reaktionen in der Transform-Bibliothek gesucht, welche an der geforderten Position anwendbar sind. WODCA kennt keine Transform-Bibliothek. Stattdessen werden nun Funktionen verwendet, welche die Generierung von Vorstufen in den voranstehenden Abschnitten bewerkstelligt haben: Handelt es sich um CX-Bindungen, werden die Methoden aus Abschnitt 4.3.2 verwendet, bei CC-Bindungen diejenigen aus Abschnitt 4.3.3. Es ist dabei durchaus möglich, daß für eine gegebene strategische Bindung keine Vorstufen ableitbar sind (die physikochemischen Parameter signalisieren keine ausreichende Aktivierung bzw. die atomare Umgebung der strategischen Bindung entspricht keiner erwarteten Situation). Die strategische Bindung wird in solchen Fällen nicht verworfen, jedoch mit einer sehr niedrigen Bewertung (10) versehen. Der Anwender kann dann selbst geeignete Vorstufen ableiten oder aber diese strategische Bindung von weiteren Untersuchungen ausschließen.

Auch hier soll die Anwendung dieser Algorithmen an einem Beispiel demonstriert werden. Dazu dient das Bicyclo[2.2.2]octan-Derivat (**85**) aus Abbildung 4 - 32. Man beachte die chiralen Zentren. Die Verbindung hat zwei primäre Ringe mit sechs Gliedern sowie einen sekundären Ring mit ebenfalls sechs Atomen. In jedem der Ringe gibt es zwei Brückenkopfatome (Atome 1 und 4 in Abb. 4 - 32).

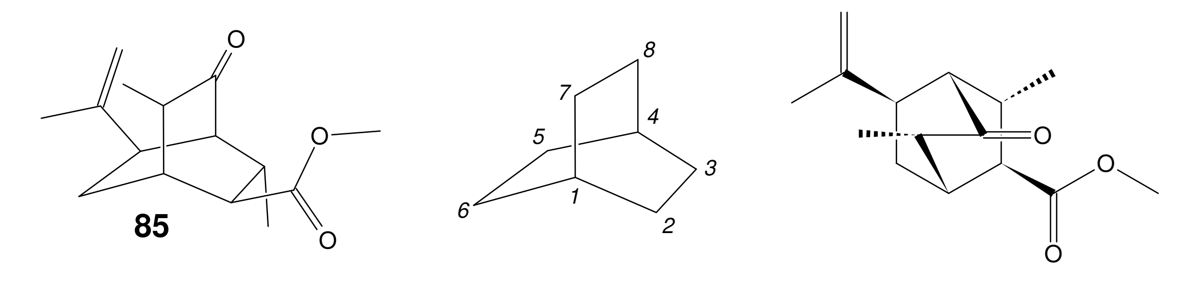


Abb. 4 - 32 Anwendungsbeispiel Bicyclo[2.2.2]octan-Derivat (**85**). Links eine perspektivische Darstellung der Zielverbindung, in der Mitte die verwendete Numerierung für das Ringsystem. Rechts eine alternative Darstellung der Zielverbindung mit entsprechenden Keilbindungen. Um die Stereochemie korrekt zu spezifizieren, *muß* die Verbindung mit gekeilten Bindungen eingegeben werden. Ein Computerprogramm kann allein aus der Konnektivitätsinformation, wie sie in der linken Darstellung enthalten ist, keine eindeutige Zuordnung der chiralen Zentren vornehmen.

Sämtliche zu diesen Brückenkopfatomen *exo* stehende Bindungen werden von WODCA als strategisch angesehen. Bei Anwendung der Regeln 6a und 6b wären es nur zwei. Diese Bindungen sind in Abbildung 4 - 33 fett markiert. Für vier der strategischen Bindungen (5-4. 1-2, 3-4 und 1-7) vergibt WODCA solide Bewertungen* (100, 94, 91 und 88). Für zwei Bindungen fehlen Vorstellungen zu erforderlichen Abgangsgruppen (4-6 und 1-6, Bewertung 10). Folgt man den Vorschlägen von WODCA und bricht die strategischen Bindungen 1-2 und 3-4, resultieren die Vorstufen Carvon (**86**) und Crotonsäuremethylester (**87**).

* Das sind nach Gleichung (4-1) ermittelte und innerhalb des Satzes der strategischen CC-Bindungen skalierte Werte analog dem Vorgehen in Abschnitt 4.3.3.

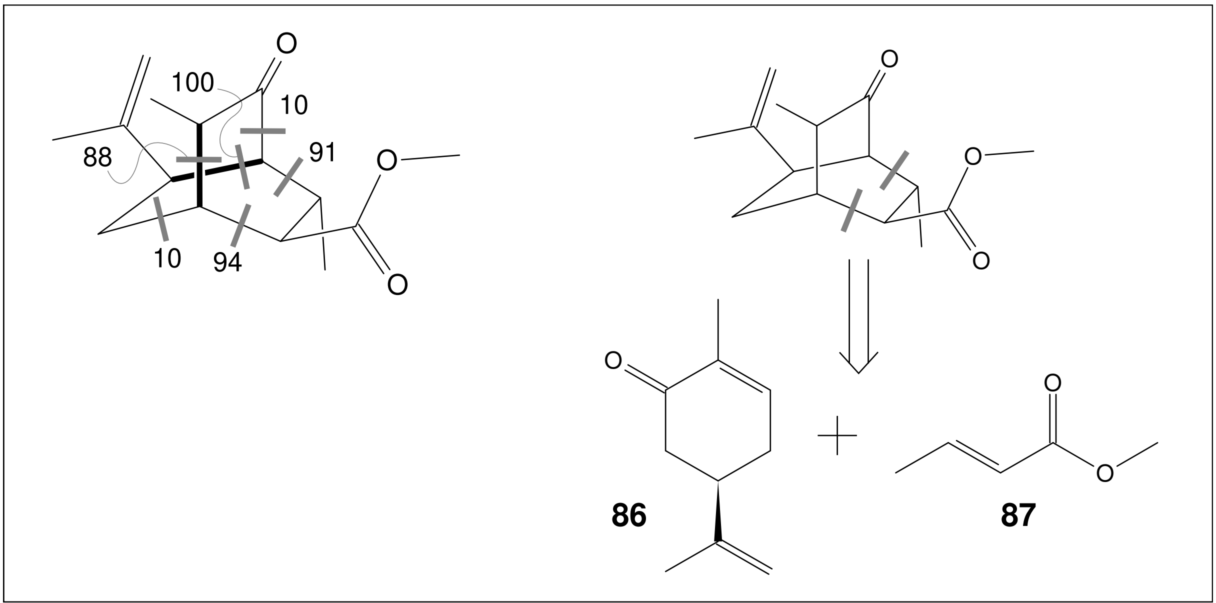


Abb. 4 - 33 Links alle strategischen Bindungen in Verbindung (**85**). Die fett gezeichneten Bindungen resultieren nach dem Corey-Algorithmus. Die markierten Bindungen erkennt WODCA als strategisch. Die Zahlen neben den Markierungen sind die zugeordneten Bewertungen. Rechts die Vorstufen, die beim Bruch zweier strategischer Bindungen resultieren.

Diese Vorstufen entsprechen einer doppelten Michael-Addition. Bricht man zunächst nur eine dieser beiden Bindungen (3-8) und wendet auf die resultierende Vorstufe die Retrosynthesefunktion des Abschnitts 4.3.3 an, wird das noch deutlicher: WODCA charakterisiert wiederum die Bindung (6-7) als strategisch, nun mit etwas anderen Bewertungen (siehe Abb. 4 - 34). Daneben werden weitere strategische Bindungen gefunden. Der Bruch der am höchsten bewerteten Bindung führt dann zu den bekannten Vorstufen (**86**) und (**87**).

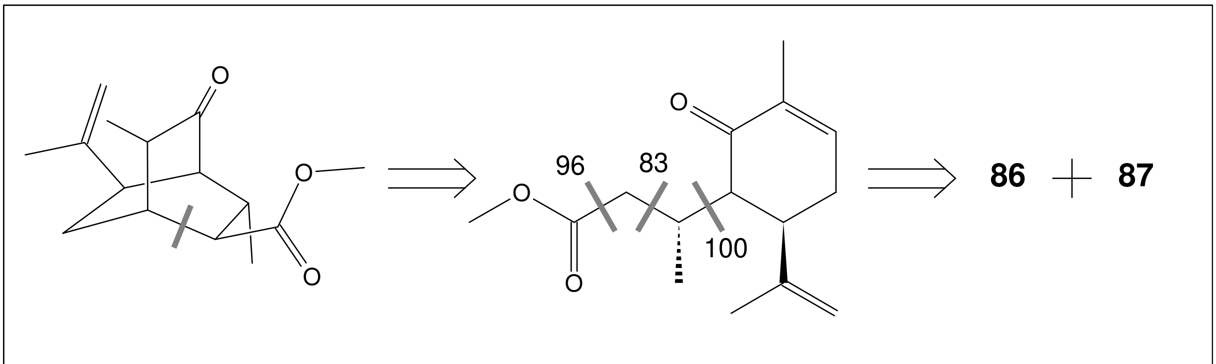


Abb. 4 - 34 Vorstufengenerierung durch Bruch nur einer strategischen Bindung und erneute Anwendung einer Analysefunktion. Die letztendlich resultierenden Vorstufen sind gleich. Dafür wird der doppelte Michael-Angriff so deutlicher.

Dieser von WODCA vorgeschlagene Syntheseweg entspricht einer publizierten Route [122]. Die Autoren erzielten bei ihrer Umsetzung von (-)-Carvon mit Crotonsäuremethylester eine Ausbeute von über 70 %. Die absolute Konfiguration der Zielverbindung (**85**) konnte mittels 2D-^{1}H-NMR-Spektroskopie bestätigt werden.

Die Vorstufen (**86**) und (**87**) werden direkt im Fluka-Katalog gefunden. Interessant ist die Fragestellung, ob es zum (-)-Carvon (**86**) alternative Ausgangsmaterialien gäbe, wenn dieses selbst nicht kommerziell verfügbar wäre. Dazu wird mit dem Ähnlichkeitskriterium

„C-Skelett einschließlich α-Atome und Reduktion“ im Fluka-Katalog gesucht. Zehn Treffer erhält man auf diese Weise. Diese werden mittels der in Abschnitt 2.2.3 vorgestellten Funktion auf ihre relative Übereinstimmung mit der Anfrageverbindung (-)-Carvon (**86**) bewertet und abfallend geordnet. Diese Reihenfolge gibt Abbildung 4 - 35 wieder. Das klar am besten bewertete Ausgangsmaterial ist natürlich (-)-Carvon selbst. Aber auch das (-)-Carvyl-acetat (**88**) stimmt mit seiner Konfiguration und Funktionalitätsverteilung exzellent mit der Anfrage überein und erhält daher eine nur wenig niedrigere Bewertung. Wäre (-)-Carvon nicht im Katalog enthalten, stellte diese Verbindung eine bestens geeignete Alternative dar. Alle anderen Ausgangsmaterialien fallen dagegen deutlich ab.

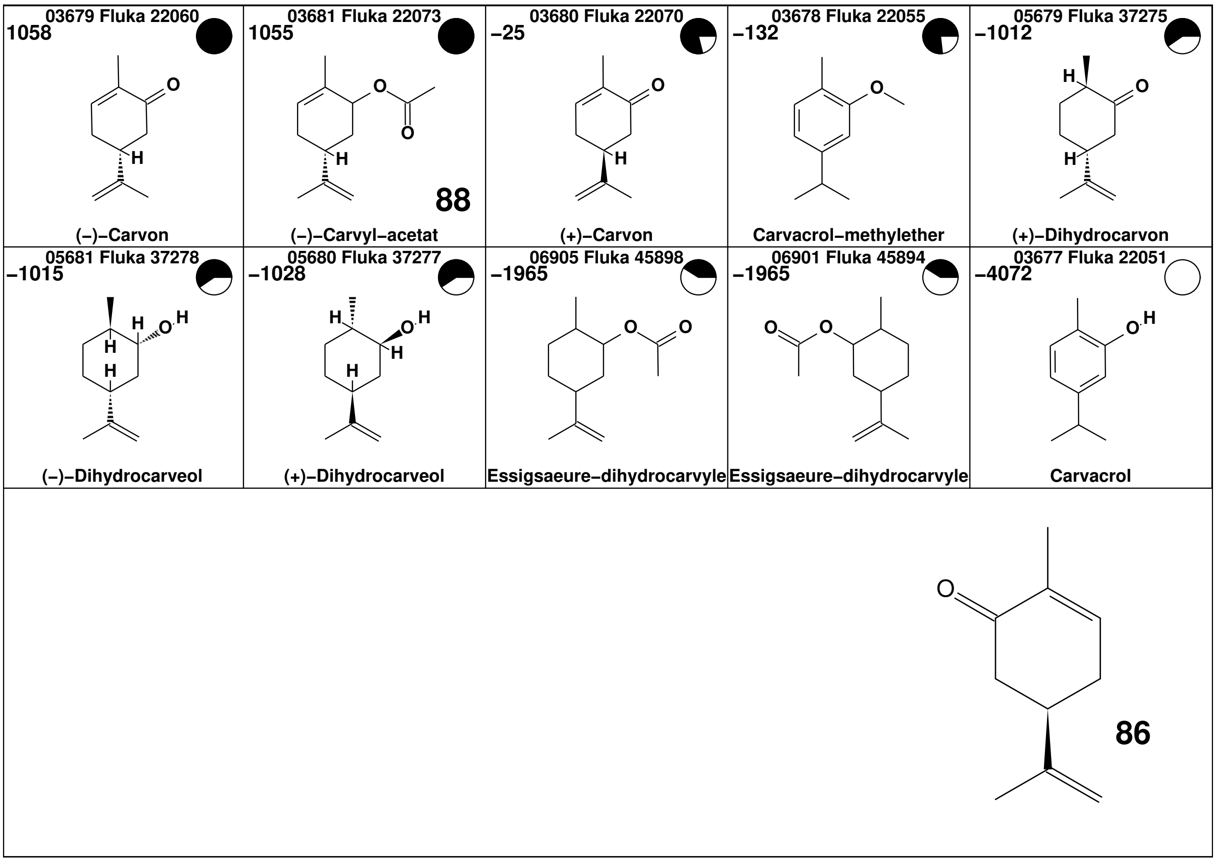


Abb. 4 - 35 Nach ihrer Bewertung geordnete Ausgangsmaterialien für (-)-Carvon (**86**, unten rechts). Die Bewertung ist als Tortendiagramm (rechts oben) wie auch numerisch (links oben) angegeben. Die Liste der zehn Verbindungen wurde mit der Ähnlichkeitssuche „C-Skelett einschließlich α-Atome und Reduktion“ ermittelt.

In einem zweiten Beispiel soll nun die Retrosynthesefunktion dieses Abschnitts auf das TIBO-Beispiel (**46**) aus Abschnitt 4.3.2 angewendet werden. Das Resultat zeigt Abbildung 4 - 36. Zu den strategischen Bindungen, die allein durch die CX-Strategien gefunden werden, kommen zwei weitere strategische Bindungen hinzu. Eine davon ist eine CC-Bindung, die durch die CX-Strategien nicht gefunden werden kann. Die Bindung des N_1-Atoms zum Phenylring wurde nur hier ermittelt, weil diese Bindung *exo* zu einem Ring steht. Die CX-Strategie allein vermeidet jedoch den Bruch von zu Aromaten α-ständigen Bindungen. Für die Syntheseplanung kann man daraus die Schlußfolgerung ziehen, daß Bindungen, die von *mehreren* Methoden als strategisch erkannt werden, zu besonders aussichtsreichen Reaktionswegen führen.

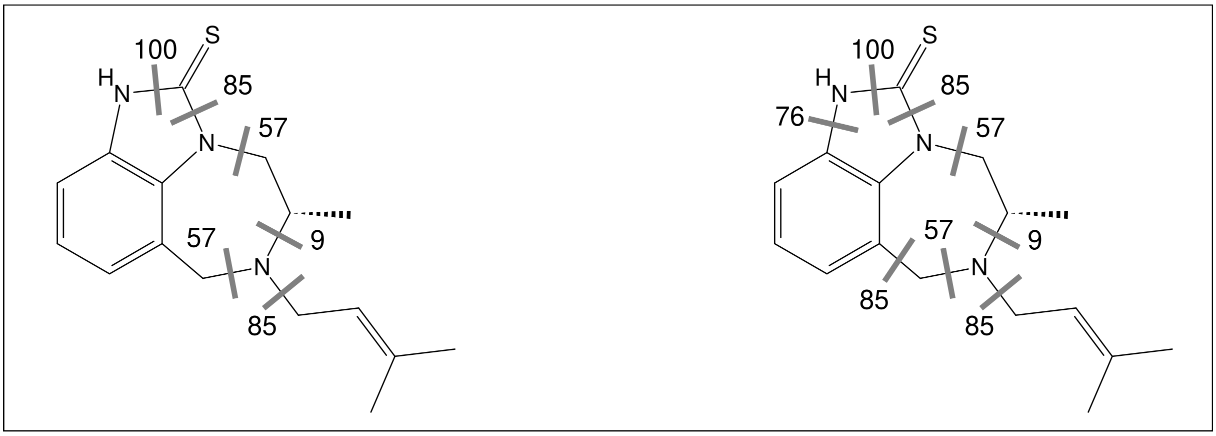


Abb. 4 - 36 Strategische Bindungen in der Zielverbindung TIBO (**46**). Links die mit den Algorithmen des Abschnitts 4.3.2 ermittelten Vorschläge. Rechts die mittels der formalen Strategie dieses Abschnitts gefundenen Empfehlungen.

4.4 Kommunikation mit dem Benutzer

An dieser Stelle soll ein kurzer Überblick über die Interaktion des Benutzers mit dem Programm gegeben werden. Eine WODCA Sitzung läuft nach folgendem typischen Schema ab:

- ❏ Eingabe des Syntheseziels. Das kann mittels eines graphischen Editors erfolgen. Die Zielverbindung kann aber auch aus einer Datei eingelesen werden. Als Moleküleditor kommt der CACTVS Editor [123] zum Einsatz (siehe Abb. 4 - 38 rechts oben).
- ❏ Der Benutzer entscheidet dann, welche Analysemethoden von WODCA eingesetzt werden sollen. Er kann zunächst versuchen, mittels der in Abschnitt 3.3 beschriebenen Ähnlichkeitssuchen, direkte Relationen zwischen seinem Syntheseziel und Ausgangsmaterialien herzustellen. In welchem der vorhandenen Kataloge er suchen möchte liegt ebenfalls in seiner Hand. Wie der Zugriff auf die verschiedenen Ähnlichkeitskriterien erfolgt und welche Hilfen dabei zur Verfügung stehen, wurde bereits in Abschnitt 3.3.11 erläutert.
- ❏ Werden Listen von Ausgangsmaterialien gefunden, können diese an ein Anzeigeprogramm exportiert werden. Dieses Anzeigeprogramm ist der CACTVS Browser [123] (siehe Abb. 4 - 38 links). Mit diesem Tool kann der Benutzer in den ermittelten Listen blättern, einzelne Verbindungen auswählen, um sie zu löschen, in Dateien zu speichern oder auszudrucken etc. Es ist auch möglich Verbindungen an den Editor oder zurück an das WODCA Kernsystem zu transferieren.
- ❏ Führen die Ähnlichkeitssuchen nicht zu befriedigenden Ergebnissen, werden die Methoden zur Suche nach strategischen Bindungen angewendet. Diese Algorithmen werden einem Anwender über ein Dialogfenster zur Verfügung gestellt (siehe Abb. 4 - 37). Darin werden alle Funktionalitäten, welche die Arbeit mit strategischen Bindungen betreffen, für den Chemiker bereitgehalten. Links wird die gerade untersuchte Zielverbindung dargestellt. In diesem Fall sind bereits strategische Bindungen ermittelt worden (mittels der Option „Carbon Hetero Atom Bond Breaks“), die durch Balken quer zu den betreffenden Bindungen markiert sind. Alternativ können statt der Balken auch die Bewertungen oder der Rang der betreffenden strategischen Bindung dargestellt werden. Die eigentliche Analyse löst der Benutzer durch Auswahl einer Funktion aus der rechts angebotenen Liste und anschließendes Betätigen der *Search*-Knopfes aus. Durch Spezifikation einzelner oder mehrerer strategischer Bindungen im Eingabefeld *Cut strategic bond(s)* wählt der Chemiker diejenigen strategischen Bindungen, die tatsächlich gebrochen werden sollen. Die dabei entstehenden Vorstufen werden in dem linken Fenster dargestellt. Wurden Bindungsbrüche vorgenommen, können diese mittels *Undo* rückgängig gemacht werden. Einzelne oder alle strategische Bindungen können schließlich auch gelöscht werden (Eingabefeld *Delete strategic bond(s)*).
- ❏ Rechercheergebnisse, erzeugte Vorstufen oder modifizierte Zwischenstufen können jederzeit in Dateien gespeichert und bei Bedarf zu einem späteren Zeitpunkt erneut geladen werden.
- ❏ Dieses Szenario läßt sich dann iterativ solange wiederholen, bis für alle Teilprobleme Lösungen ermittelt wurden.

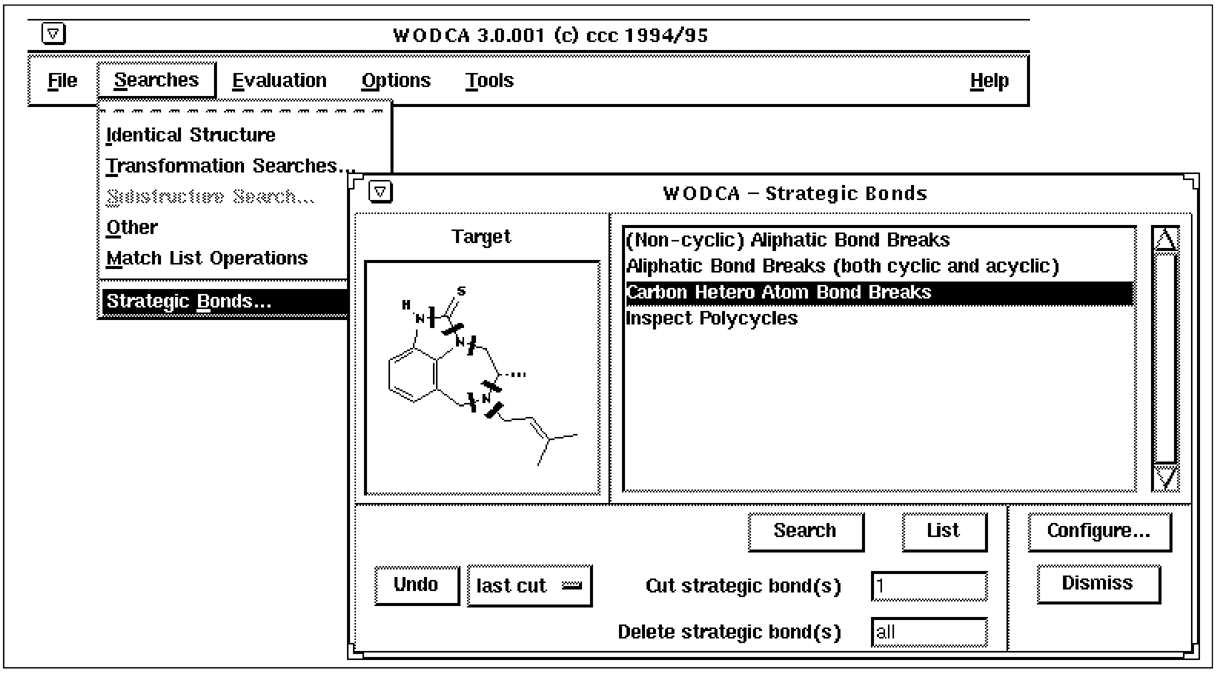


Abb. 4 - 37 Zugriff auf die Methoden zur Arbeit mit strategischen Bindungen im WODCA System Version 3. Das Dialogfenster ähnelt dem aus Abbildung 3 - 38 auf Seite 90 für die Transformationssuchen. Die Komponenten sind im Text erläutert.

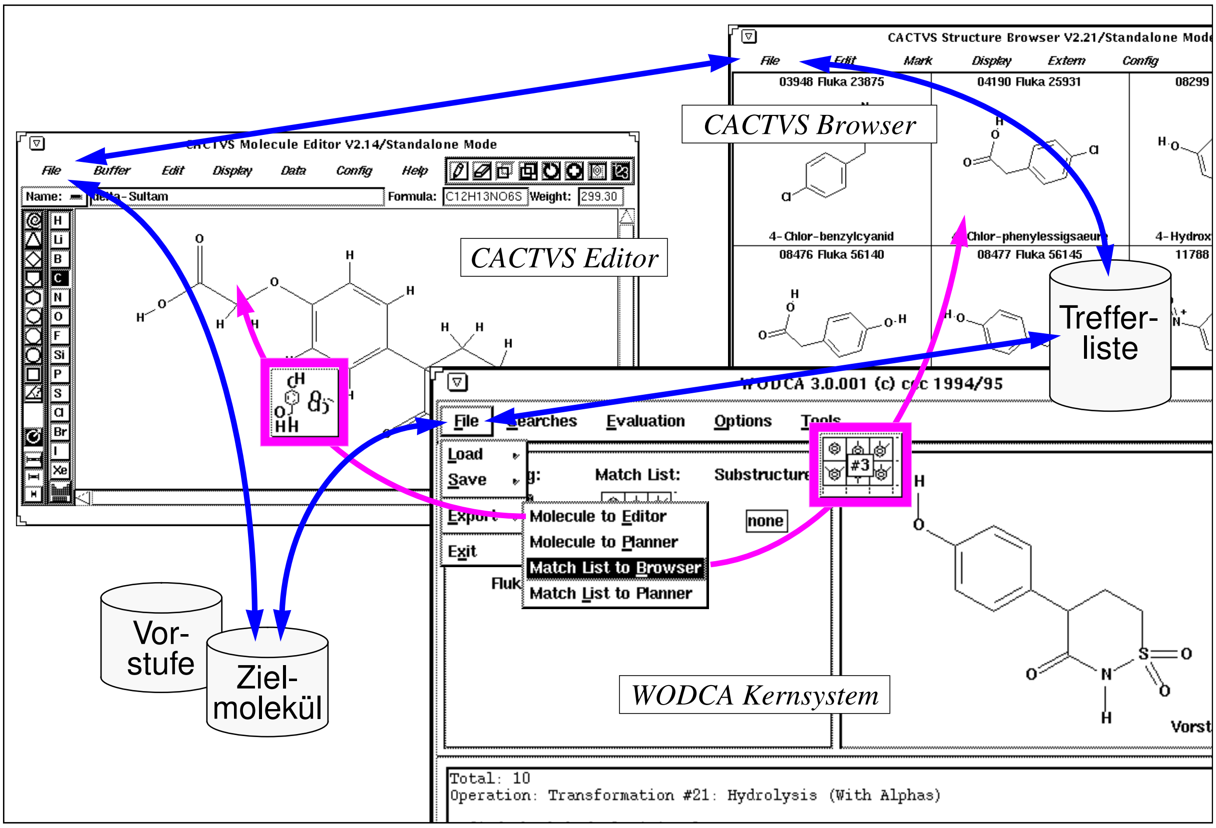


Abb. 4 - 38 Möglichkeiten des Datentransfers zwischen den Programmkomponenten. Klassischer Datenaustausch über Dateien ist ebenso möglich wie Drag&Drop oder direktes Senden per Knopfdruck. Man beachte, daß das WODCA Kernsystem und die CACTVS Tools separate Programme sind.

Der Transfer an Daten zwischen den verschiedenen Programmen wird in Abbildung 4 - 38 veranschaulicht. Die Möglichkeiten beginnen beim klassischen Schreiben und Lesen von Dateien. Zusätzlich werden aber auch modernere Methoden des Datenaustauschs unterstützt, die jeder Computeranwender kennt: Drag&Drop von Molekülen oder ganzen Listen zwischen verschiedenen Programmen oder direktes Senden, ohne daß der Benutzer eine Zwischendatei erzeugen muß.

Die Schnittstelle zwischen Programm und dem Anwender ist für jedwede bereitgestellte Funktionalität sehr einfach und gerade dadurch intuitiv gestaltet worden. Auch dem gelegentlichen Anwender des Programms wird damit die Benutzung deutlich erleichtert.

Das WODCA System Version 3 ist für die UNIX-Betriebssysteme SunOS 4.1.3 und Solaris 2.4 auf Sun Workstations verwirklicht. Andere UNIX-Umgebungen (z. B. Silicon Graphics unter IRIX, Hewlett-Packard unter HP-UX oder IBM unter AIX) können mit minimalem Aufwand unterstützt werden. Auch Portierungen auf Personal Computer sind möglich geworden. Um letztere Architektur vollständig zu unterstützen, ist jedoch noch ein höherer Programmieraufwand notwendig. Durch die bestehende bzw. potentielle Verfügbarkeit für eine Vielzahl an Rechner-Umgebungen, läßt sich das WODCA System auch zusammen mit anderen für die Syntheseplanung nützlichen Werkzeugen einsetzen. Beispielsweise ist die ISIS® Programm-Familie von Molecular Design Ltd. (ISIS® Base, ISIS® Draw) - ein System, das u. a. chemische Struktur- und Reaktionsdatenbanken zur Verfügung stellt - auf Silicon Graphics Workstations präsent. Dadurch können z. B. auf einfache Weise von WODCA vorgeschlagene Vorstufen als Anfrage für eine Suche in Reaktionsdatenbanken an ISIS® Base übergeben werden. Der Chemiker erhält damit sofort zusätzliche Hinweise über eine Reaktion wie z. B. Reaktionsbedingungen.

Zusammenfassung

Zwei leistungsfähige Konzepte zur Planung organischer Synthesen - die Ähnlichkeitssuche in Form der Transformationssuche und die Suche nach strategischen Bindungen mit nachfolgender Generierung von Synthesevorstufen - wurde mit einem Maximum an Flexibilität bezüglich der Erweiterbarkeit dieser Konzepte in einem Computerprogramm realisiert. Dieses Programm kommuniziert interaktiv mit dem Benutzerfür und steht für weit verbreitete Rechnerumgebungen zur Verfügung.

Die Flexibilität erlaubt es jedem Anwender, eigene Vorstellungen und Definitionen für strukturelle Ähnlichkeit zu definieren, Funktionen zur Erkennung strategischer Bindungen zu modifizieren oder neu zu entwickeln oder eigene Bewertungsalgorithmen zu verwirklichen. Das wird durch die Bereitstellung dieser Fähigkeiten in Form externer Dateien erreicht, in denen wiederum die betreffenden Methoden mittels einer leistungsfähigen Sprache (*tcl*) implementiert werden. Der administrative Aufwand zur Erweiterung oder Modifikation der Programmfähigkeiten wurde dadurch minimiert.

Durch die strenge Berücksichtigung von Standards bei der Durchführung der Programmierung wurde die Fixierung der Vorläufersystems auf einen einzigen Rechnertyp überwunden. Das realisierte Programmsystem steht prinzipiell für jedwede UNIX-Architektur zur Verfügung, die zusätzlich das X Window System als graphische Oberfläche bereitstellt. Darüber hinaus ist auch eine Portierung in eine Microsoft-Windows-basierte PC-Architektur mit überschaubarem Aufwand möglich geworden.

Ein Satz an Ähnlichkeitskriterien, mit denen Verwandtschaften zwischen Synthesezielen und verfügbaren Ausgangsmaterialien erkannt werden können, ist entwickelt worden. Die Definitionen dieser Kriterien erfolgt in externen Dateien mittels *tcl*, wodurch ggf. Erweiterungen oder Modifikationen einfach zu bewerkstelligen sind. Die Leistungsfähigkeit und breite Anwendbarkeit dieser Kriterien im Bereich der Syntheseplanung konnte anhand vielfältiger Beispiele gezeigt werden.

Drei auf ganz unterschiedlichen Prinzipien beruhende Algorithmen zur Suche nach strategischen Bindungen wurden verwirklicht. Der erste Algorithmus basiert auf gesicherter Textbuch-Chemie aus dem Bereich der Knüpfung von Kohlenstoff-Heteroatom-Bindungen und vereinigt eine Vielzahl an Einzelreaktionen in einem allgemeinen Ansatz. Ein zweites Modell basiert auf berechenbaren physikochemischen Parametern zur Abschätzung von mesomeren und induktiven Effekten, um auf deren Grundlage praktikable retrosynthetische Spaltungen von Kohlenstoff-Kohlenstoff-Bindungen vorzuschlagen. In einem weiteren Ansatz schließlich wurden generelle strukturelle Merkmale in Ringsystemen in Abwandlung eines ursprünglichen Verfahrens herangezogen, um in komplexen Ringsystemen effektive Vorstufen abzuleiten. Auch diese Funktionen werden dem Programmsystem über externe *tcl*-Skripts bereitgestellt.

An Hand einer Reihe von Beispielen konnte die Anwendbarkeit und Potenz dieser Methoden belegt werden.

Damit steht ein Programmsystem zur Syntheseplanung zur Verfügung, das in der Summe der Charakteristika Benutzerfreundlichkeit, Verfügbarkeit für verschiedene Rechnerarchitekturen, Flexibilität bezüglich seiner Weiterentwicklung, Angebot an Planungskonzepten und deren Leistungsfähigkeit jedem Vergleich standhält. Es bleibt zu hoffen, daß in Zukunft dieses System möglichst vielen Synthesechemikern zur Verfügung steht und jene unvoreingenommen das Programm als Werkzeug begreifen, das ihre tägliche Arbeit erleichtern kann.

Literaturverzeichnis

[1] Corey, E. J.
„General Methods for the Construction of Complex Molecules“
Pure Appl. Chem. 14 (1967) 19-37

[2] Corey, E. J.; Wipke, W. T.
„Computer-Assisted Design of Complex Organic Syntheses“
Science 166 (1969) 178-192

[3] Corey, E. J.
„Computer-assisted Analysis of Complex Synthetic Problems“
Quart. Rev. Chem. Soc. 25 (1971) 455-482

[4] Warren, S.
„Organic Synthesis: The Disconnection Approach“
John Wiley & Sons, Chichester (1982)

[5] Fuhrhop, J.; Penzlin, G.
„Organic Synthesis - Concepts, Methods, Starting Materials“
VCH, Weinheim (1983)

[6] Corey, E. J.; Long, A. K.; Rubenstein, S. D.
„Computer-Assisted Analysis of Organic Synthesis“
Science 228 (1985) 408-418

[7] Vléduts, G. E.
„Concerning One System of Classification of Organic Reactions“
Inform. Stor. Retr. 1 (1963) 117-146

[8] Corey, E. J.; Wipke, W. T.; Cramer III, R. D.; Howe, W. J.
„Computer-Assisted Synthetic Analysis. Facile Man-Machine Communication of Chemical Structure by Interactive Computer Graphics“
J. Am. Chem. Soc. 94 (1972) 421-430

[9] Corey, E. J.; Wipke, W. T.; Cramer III, R. D.; Howe, W. J.
„Techniques for Perception by a Computer of Synthetically Significant Structural Features in Complex Molecules“
J. Am. Chem. Soc. 94 (1972) 431-439

[10] Corey, E. J.; Wipke, W. T.; Cramer III, R. D.; Howe, W. J.
„Computer-Assisted Synthetic Analysis for Complex Molecules. Methods and Procedures for Machine Generation of Synthetic Intermediates“
J. Am. Chem. Soc. 94 (1972) 440-459

[11] Corey, E. J.; Petersson, G. A.
„An Algorithm for Machine Perception of Synthetically Significant Rings in Complex Cyclic Organic Structures“
J. Am. Chem. Soc. 94 (1972) 460-465

[12] Corey, E. J.; Cheng, X. M.
„The Logic of Chemical Synthesis“
John Wiley & Sons, New York (1989)

[13] Wipke, W. T.; Dyott, T. M.
„Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry“
J. Am. Chem. Soc. 96 (1974) 4825-4834

[14] Wipke, W. T.; Braun, H.; Smith, G.; Choplin, F.; Sieber, W.
In: „Computer-Assisted Organic Synthesis“ Wipke, W. T.; Howe, W. J. (Hrsg.)
American Chemical Society (1977), ACS Symposium Series 61, 97

[15] Gelernter, H.; Sridharan, N. S.; Hart, A. J.; Yen, S. C.; Fowler, F. W.; Shue, H. J.
„The Discovery of Organic Synthetic Routes by Computer“
Top. Curr. Chem. 41 (1973) 113-150

[16] Bersohn, M.
„Automatic problem solving applied to synthetic chemistry“
Bull. Chem. Soc. Jap. 45 (1972) 1897-1903

[17] Takahashi, M.; Dogane, I.; Yoshida, M.; Yamachika, H.; Takabatake, T.; Bersohn, M.
„The performance of a noninteractive synthesis program“
J. Chem. Inf. Comput. Sci. 30 (1990) 436-441

[18] Nakayama, T.
„Computer-Assisted Knowledge Acquisition System for Synthesis Planning“
J. Chem. Inf. Comput. Sci. 31 (1991) 495-503

[19] Moll, R.
„Context Description in Synthesis Planning“
J. Chem. Inf. Comput. Sci. 34 (1994) 117-119

[20] Gelernter, H. K.; Sanders, A. F.; Larsen, D. L.; Agarwal, K. K.; Boivie, R. H.; Spritzer, G. A.; Searleman, J. E.
„Empirical Explorations of SYNCHEM“
Science 197 (1977) 1041-1049

[21] Braun, H.
„Syntheseplanung. Strategische Werkzeuge für den Chemiker“
Chem. Ind. (Düsseldorf) 40 (1988) 43-53

[22] Dugundji, J.; Ugi, I.
„An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs“
Top. Curr. Chem. 39 (1973) 19-64

[23] Blair, J.; Gasteiger, J.; Gillespie, C.; Gillespie, P. D.; Ugi, I.
„Representation of the Constitutional and Stereochemical Features of Chemical Systems in the Computer Assisted Design of Syntheses"
Tetrahedron 30 (1974) 1845-1859

[24] Bauer, J.; Ugi, I.
„Chemical reactions and Structures without Precedent Generated by Computer Program"
J. Chem. Res. (1982) S 298; M 3101-3196, 3201-3260

[25] Bauer, J.
„IGOR2: a PC-Program for Generating New Reactions and Molecular Structures"
Tetrahedron Comput. Methodol. 2 (1989) 269-280

[26] Fontain, E.; Reitsam, K.
„The Generation of Reaction Networks with RAIN. 1. The Reaction Generator"
J. Chem. Inf. Comput. Sci. 31 (1991) 96-101

[27] Gasteiger, J.; Jochum, C.; Marsili, M.; Thoma, J.
„The synthesis planning program EROS"
MATCH 6 (1979) 177-199

[28] Herges, R.
„Reaction Planning: Prediction of New Organic Reactions"
J. Chem. Inf. Comput. Sci. 30 (1990) 377-383

[29] Hendrickson, J. B.; Grier, D. L.; Toczko, A. G.
„A Logic-Based Program for Synthesis Design"
J. Am. Chem. Soc. 107 (1985) 5228-5238

[30] Hendrickson, J. B.
„A Systematic Characterization of Structures and Reactions for Use in Organic Synthesis"
J. Am. Chem. Soc. 93 (1971) 6847-6854

[31] Hendrickson, J. B.
„Systematic Synthesis Design. IV. Numerical Codification of Construction Reactions"
J. Am. Chem. Soc. 97 (1975) 5784-5800

[32] Hendrickson, J. B.
„Descriptions of reactions: their logic and applications"
Recl. Trav. Chim. Pays-Bas 111 (1992) 323-334

[33] Hendrickson, J. B.; Sander, T.
„COGNOS: A Beilstein-Type System for Organizing Organic Reactions"
J. Chem. Inf. Comput. Sci. 35 (1995) 251-260

[34] Funatsu, K.; Sasaki, S. I.
„Computer-Assisted Organic Synthesis Design and Reaction Prediction System, "AIPHOS""
Tetrahedron Comput. Methodol. 1 (1988) 27-37

[35] Wipke, W. T.; Rodgers, D.
„Artificial Intelligence in Organic Synthesis. SST: Starting Material Selection Strategies. An Application of Superstructure Search"
J. Chem. Inf. Comput. Sci. 24 (1984) 71-81

[36] Hanessian, S.; Franco, J.; Larouche, B.
„The Psychobiological Basis of Heuristic Synthesis Planning - Man, Machine and the Chiron Approach“
Pure Appl. Chem. 62 (1990) 1887-1910

[37] Salatin, T. D.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 1. Overview“
J. Org. Chem. 45 (1980) 2043-2051

[38] Röse, P.; Gasteiger, J.
„Automated derivation of reaction rules for the EROS 6.0 system for reaction prediction“
Anal. Chim. Acta 235 (1990) 163-168

[39] Salatin, T. D.; McLaughlin, D. R.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 3. Ylide Chemistry and the Organometallic Chemistry of Lithium, Magnesium, and Lithium Cuprates“
J. Org. Chem. 46 (1981) 5284-5294

[40] Peishoff, C. E.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 4. Organosilicon Chemistry“
J. Org. Chem. 48 (1983) 1970-1979

[41] Peishoff, C. E.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 9. Reactions of Unsaturated Electrophiles Including Nucleophilic Aromatic Substitution“
J. Org. Chem. 50 (1985) 1056-1068

[42] Metivier, P.; Gushurst, A. J.; Jorgensen, W. L.
„A General Treatment of Nucleophilic Chemistry (Computer-Assisted Mechanistic Evaluation of Organic Reactions. 13.)“
J. Org. Chem. 52 (1987) 3724-3738

[43] Gushurst, A. J.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 14. Reactions of Sulfur and Phosphorus Ylides, Iminophosphoranes, and P=X-Activated Anions“
J. Org. Chem. 53 (1988) 3397-3408

[44] Laird, E. R.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 17. Free Radical Chain Reactions“
J. Org. Chem. 55 (1990) 9-27

[45] Bures, M. G.; Ross-Kozel, B. L.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 11. Electrophilic Aromatic Substitution“
J. Org. Chem. 50 (1985) 4490-4498

[46] Bures, M. G.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 15. Heterocycle Synthesis“
J. Org. Chem. 53 (1988) 2504-2520

[47] Burnier, J. S.; Jorgensen, W. L.
„Computer-Assisted Mechanistic Evaluation of Organic Reactions. 7. Six-Electron Cycloadditions“
J. Org. Chem. 48 (1983) 3923-3941

[48] Burnier, J. S.; Jorgensen, W. L.
„General Treatment of Periselectivity“
J. Org. Chem. 49 (1984) 3001-3020

[49] Paderes, G. D.; Jorgensen, W. L.
„Computer-Assisted Evaluation of Oxidation Reactions“
J. Org. Chem. 54 (1989) 2058-2085

[50] Blake, J. E.; Dana, R. C.
„CASREACT: more than a million reactions“
J. Chem. Inf. Comput. Sci. 30 (1990) 394-399

[51] Bersohn, M.; Mackay, K.
„Steps toward the automatic compilation of synthetic organic reactions“
J. Chem. Inf. Comput. Sci. 19 (1979) 137-141

[52] Gelernter, H. K.; Rose, J. R.; Chen, C.
„Building and Refining a Knowledge Base for Synthetic Organic Chemistry via the Methodology of Inductive and Deductive Machine Learning“
J. Chem. Inf. Comput. Sci. 30 (1990) 492-504

[53] Rose, J. R.; Gasteiger, J.
„HORACE: An automatic system for the hierarchical classification of chemical reactions“
J. Chem. Inf. Comput. Sci. 34 (1994) 74-90

[54] Jauffret, P.; Tonnelier, C.; Hanser, T.; Kaufmann, G.
„Machine Learning of Generic Reactions: 1. Scope of the Project; the GRAMS Program“
Tetrahedron Comput. Methodol. 3 (1990) 323-333

[55] Ihlenfeldt, W.-D.; Gasteiger, J.
„Computergestützte Planung organisch-chemischer Synthesen: die zweite Programmgeneration“
Angew. Chem. 107 (1995) 2807-2829

[56] Gasteiger, J.; Ihlenfeldt, W.-D.; Röse, P.
„A collection of computer methods for the synthesis design and reaction prediction“
Recl. Trav. Chim. Pays-Bas 111 (1992) 270-290

[57] Ihlenfeldt, W.-D.; Gasteiger, J.
„Hash Codes for the Identification and Classification of Molecular Structure Elements“
J. Comp. Chem. 15 (1994) 793-813

[58] Ihlenfeldt, W.-D.
„Computergestützte Syntheseplanung durch Erkennung synthetisch nutzbarer Ähnlichkeit von Molekülen“
Dissertation, TU München (1991) S. 105

[59] Wie [58], S. 78

[60] Ihlenfeldt, W.-D.
„An Efficient Approach toward a Flexible and General Knowledge Definition and Program Control Language System for a Synthesis-Planning Program"
J. Chem. Inf. Comput. Sci. 34 (1994) 872-880

[61] Weise, A.
„Ableitung organisch-chemischer Reaktionen mit dem Simulationsprogramm AHMOS"
Z. Chem. 15 (1975) 333-340

[62] Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S.
„CAMEO: a program for the logical prediction of the products of organic reactions"
Pure Appl. Chem. 62 (1990) 1921-1932

[63] Hamm, P.; Jauffret, P.; Kaufmann, G.
„Generalization and representation of synthesis pathways in the COSYMA system"
Recl. Trav. Chim. Pays-Bas 111 (1992) 317-322

[64] Zefirow, N. S.; Gordeeva, E. V.; Tratch, S. S.
„Problems of molecular design and computer. 11. The FLAMINGOES program system for the nonempirical solution of structural problems of organic chemistry. The BASIC program oriented for the microcomputer"
J. Chem. Inf. Comput. Sci. 28 (1988) 188-193

[65] Long, A. K.; Kappos, J. C.; Rubenstein, S. D.; Walker, G. E.
„Computer-Assisted Synthetic Analysis: A Generalized Procedure for Subgoal Transform Selection Based on a Two-dimensional Pattern Language"
J. Chem. Inf. Comput. Sci. 34 (1994) 922-933

[66] Sello, G.
„Lilith: From Childhood to Adolescence"
J. Chem. Inf. Comput. Sci. 34 (1994) 120-129

[67] Moreau, G.
„MASSO, un programme d'aide a la synthèse organique, utilisant des demi-réactions"
Nouv. J. Chim. 2 (1978) 187-193

[68] Azario, P.; Barone, R.; Chanon, M.
„Microcomputer and Organic Synthesis. 3. The MARSEIL/SOS Expert System, a New Graphic Approach. An Electronic Lab Note for Organic Synthesis"
J. Org. Chem. 53 (1988) 720-724

[69] Der Autor hatte während eines Praktikums 1989 Gelegenheit, das Programm RDSS in der Computer-Chemie-Abteilung des damaligen Bitterfelder Chemie-Kombinats kennenzulernen.

[70] Persönliche Mitteilung von W. T. Wipke während eines Workshops zu Computergestützter Syntheseplanung, der vom 14.05. bis 16.05.1995 in Goslar-Hahnenklee stattfand.

[71] Persönliche Mitteilung von J. R. Rose, einem ehemaligen Mitarbeiter von Gelernter, der als Postdoktorand im Arbeitskreis Gasteiger tätig war.

[72] Blower Jr., P. E.; Whitlock Jr., H. W.
„An Application of Artificial Intelligence to Organic Synthesis“
J. Am. Chem. Soc. 98 (1976) 1499-1510

[73] Barone, R.; Arbelot, M.; Chanon, M.
„SCORE: a program to explore the SCOpe of a given intramolecular REaction. How to find systematically the relevancy of a transform to various classes of structures. Application to the DeMayo reaction“
Tetrahedron Comput. Methodol. 1 (1988) 3-14

[74] Sussman, G. J.; Steele, J.
„Scheme: an interpreter for extended lambda calculus“
MIT (1975)

[75] Abelson, H.; Sussman, G. J.; Sussman, J.
„Structure and Interpretation of Computer Programs“
MIT Press, Cambridge (1985)

[76] Clinger, W.; Rees, J.
„Revised^4 Report on the Algorithmic Language Scheme“
(1991)

[77] Wall, L.; Schwartz, R. L.
„Programming PERL“
O'Reilly & Associates (1991)

[78] Wall, L.; Schwartz, R. L.
„Programmieren in perl“
Hanser (1993)

[79] Ousterhout, J. K.
„Tcl and the Tk Toolkit“
Addison-Wesley (1994)

[80] Welch, B. B.
„Practical Programming in Tcl and Tk“
Prentice Hall (1995)

[81] Cahn, R. S.; Sir Ingold, C.; Prelog, V.
„Spezifikation der molekularen Chiralität“
Angew. Chem. 78 (1966) 413-447

[82] Prelog, V.; Helmchen, G.
„Grundlagen des CIP-Systems und Vorschläge für eine Revision“
Angew. Chem. 94 (1982) 614-631

[83] Gasteiger, J.; Hutchings, M. G.
„Quantification of Effective Polarizability. Applications to Studies of X-Ray Photoelectron Spectroscopy and Alkylamine Protonation“
J. Chem. Soc., Perkin Trans. 2 (1984) 559-564

[84] Gasteiger, J.; Marsili, M.
„Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges“
Tetrahedron 36 (1980) 3219-3288

[85] Gasteiger, J.; Saller, H.; Löw, P.
„Elucidating Chemical Reactivity by Pattern Recognition Methods“
Anal. Chim. Acta 191 (1986) 111-123

[86] Fröhlich, A.
„Modelle zur Beschreibung elektronischer Effekte in π-Systemen und ihr Einsatz in der Modellierung der elektrophilen aromatischen Substitution“
Dissertation, TU München (1993)

[87] Hutchings, M. G.; Gasteiger, J.
„Residual Electronegativity - Am Empirical Quantification of Polar Influences and Its Application to the Proton Affinity of Amines“
Tertahedron Lett. 24 (1983) 2541-2544

[88] Gasteiger, J.; Saller, H.
„Berechnung der Ladungsverteilung in konjugierten Systemen durch eine Quantifizierung des Mesomeriekonzeptes“
Angew. Chem. 97 (1985) 699-701

[89] Ihlenfeldt, W. D.
„CACTVS System Documentation“
Computer-Chemie-Centrum, Universität Erlangen

[90] Ihlenfeldt, W.-D.; Fick, R.
„Reference Manual for Chemistry-Aware Tcl“
Computer-Chemie-Centrum, Universität Erlangen

[91] „Keyword Reference Manual for Gasteiger Clear Text Files“
Version 1.12 (1995) Computer-Chemie-Centrum, Universität Erlangen

[92] Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J.
„Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited“
J. Chem. Inf. Comput. Sci. 32 (1992) 244-255

[93] Bebak, H.; Buse, C.; Donner, W. T.; Hoever, P.; Jacob, H.; Klaus, H.; Pesch, J.; Roemelt, J.; Schilling, P.
„The standard molecular data format (SMD format) as an integration tool in computer chemistry“
J. Chem. Inf. Comput. Sci. 29 (1989) 1-5

[94] Barnard, J. M.
„Draft Specification for Revised Version of the Standard Molecular Data (SMD) Format“
J. Chem. Inf. Comput. Sci. 30 (1990) 81-96

[95] Bernstein, F. C.; et al,
„The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures“
J. Mol. Biol. 112 (1977) 535-542

[96] Basis der automatischen Erzeugung der Substitutionsmuster list die Identifizierung der verschiedenen Verbindungen durch ihre Hashcodes. Siehe dazu [Ihlenfeldt_1994_793][57].

[97] Hendrickson, J. B.; Braun-Keller, E.
„Systematic Synthesis Design. 8. Generation of Reaction Sequences“
J. Comp. Chem. 1 (1980) 323-333

[98] Baumer, L.; Sala, G.; Sello, G.
„The LILITH approach to organic synthesis planning“
Anal. Chim. Acta 235 (1990) 209-214

[99] Baumer, L.; Sala, G.; Sello, G.
„Organic Synthesis Planning: A New Algorithm For Strategic Bond Perception“
Tetrahedron 44 (1988) 1195-1206

[100] Corey, E. J.; Howe, W. J.; Orf, H. W.; Pensak, D. A.; Petersson, G.
„General Methods of Synthetic Analysis. Strategic Bond Disconnections for Bridged Polycyclic Structures“
J. Am. Chem. Soc. 97 (1975) 6116-6124

[101] Corey, E. J.; Howe, W. J.; Pensak, D. A.
„Computer-Assisted Synthetic Analysis. Methods for Machine Generation of Synthetic Intermediates Involving Multistep Look-Ahead“
J. Am. Chem. Soc. 96 (1974) 7724-7737

[102] Corey, E. J.; Johnson, A. P.; Long, A. K.
„Computer-Assisted Synthetic Analysis. Techniques for Efficient Long-Range Retrosynthetic Searches applied to the Robinson Annulation Process“
J. Org. Chem. 45 (1980) 2051-2057

[103] Corey, E. J.; Long, A. K.; Mulzer, J.; Orf, H. W.; Johnson, A. P.; Hewett, A. P. W.
„Computer-assisted synthetic analysis. Long-range search procedures for antithetic simplification of complex targets by application of the halolactonization transform“
J. Chem. Inf. Comput. Sci. 20 (1980) 221-230

[104] Boiten, J. W.; Noordik, J. H.; Groen, M. B.
„Cationic polyene cyclizations. A computer assisted synthesis approach“
J. Chem. Inf. Comput. Sci. 33 (1993) 727-735

[105] Corey, E. J.; Long, A. K.
„Computer-assisted synthetic analysis. Performance of long-range strategies for stereoselective olefin synthesis“
J. Org. Chem. 43 (1978) 2208-2216

[106] March, J.
„Advanced Organic Chemistry - Reactions, Mechanisms, and Structure“
John Wiley & Sons, New York (1992)

[107] Pauwels, R.; Andries, K.; Desmyter, J.; Schols, D.; Kukla, M. J.; Breslin, H. J.; Raeymaeckers, A.; Gelder, J.; Woestenborghs, R.; Heykants, J.; Schellekens, K.; Janssen, M. A. C.; Clercq, E.; Janssen, P. A. J.
„Potent and selective inhibition of HIV-1 replication in vitro by a novel series of TIBO derivatives“
Nature 343 (1990) 470-474

[108] Parker, K. A.; Coburn, C. A.
„A Four-Step Synthesis of TIBO R82150“
J. Org. Chem. 56 (1991) 4600-4601

[109] Michael, A.
J. Prakt. Chem. 36 (1887) 113-114

[110] Bergman, E. D.; Ginsburg, D.; Pappo, R.
Org. React. 10 (1959) 179-560

[111] Oare, D. A.; Heathcock, C. H.
„Stereochemistry of the Base-Promoted Michael Addition Reaction“
Top. Stereochem. 19 (1989) 227-407

[112] Oare, D. A.; Heathcock, C. H.
„Acyclic Stereocontrol in Michael Addition Reactions of Enamines and Enol Ethers“
Top. Stereochem. 20 (1991) 87-170

[113] „Additions to and Substitutions at C-C π-Bonds“
In: „Comprehensive Organic Synthesis. Selectivity, Strategy and Efficiency in Modern Organic Chemistry“, Vol. 4, Semmelhack, M. F. (Hrsg.)
Pergamon Press, Oxford (1991)

[114] Parlow, A.; Weiske, C.; Gasteiger, J.
„ChemInform - An Integrated Information System on Chemical Reactions“
J. Chem. Inf. Comput. Sci. 30 (1990) 400-402

[115] Gasteiger, J.; Weiske, C.
„ChemInform - Vom gedruckten Referatedienst zur Reaktionendatenbank“
Nach. Chem. Tech. Lab. 40 (1992) 1114-1120

[116] Chen, L.; Gasteiger, J.; Rose, J. R.
„Automatic extraction of chemical knowledge from organic reaction data: Addition of carbon-hydrogen bonds to carbon-carbon double bonds“
J. Org. Chem. 60 (1995) 8002-8014

[117] Kosower, E. M.; Pazhenchevsky, B.; Hershkowitz, E.
„1,5-Diazabicyclo[3.3.0]octadienediones (9,10-dioxabimanes). Strongly fluorescent *syn* isomers“
J. Am. Chem. Soc. 100 (1978) 6516-6518

[118] Kosower, E. M.; Zbaida, D.; Baud'huin, M.; Marciano, D.; Goldberg, I.
„Bimanes. 23. The Synthesis and Properties of Vinyl-9,10-dioxabimanes“
J. Am. Chem. Soc. 112 (1990) 7305-7314

[119] Narisada, M.; Ohtani, M.; Watanabe, F.; Uchida, K.; Arita, H.; Doteuchi, M.; Hanasaki, K.; Kakushi, H.; Otani, K.; Hara, S.
„Synthesis and in vitro activity of various derivatives of a novel thromboxane receptor antagonist, *(+-)-(5Z)*-7-[3-endo[(phenylsulfonyl)amino]bicyclo[2.2.1]hept-2-exo-yl]heptenoic acid“
J. Med. Chem. 31 (1988) 1847-1854

[120] Morris, J.; Wishka, D. G.
„Vinyl Sulfonyl Esters and Amides in the Synthesis of Substituted δ-Sultams and δ-Sultones“
J. Org. Chem. 56 (1991) 3549-3556

[121] Deslongchamps, P.
„The Concept Of Strategy In Organic Synthesis“
Aldrichimica Acta 17 (1984) 59-71

[122] Zhao, R. B.; Zhao, Y. F.; Song, G. Q.; Wu, Y. L.
„Double Michael reaction of carvone and its derivatives“
Tetrahedron Lett. 31 (1990) 3559-3562

[123] Die Programme das CACTVS Systems sind frei über das World Wide Web (unter `http://schiele.organik.uni-erlangen.de/cactvs/`) erhältlich.
Siehe auch:
Ihlenfeldt, W.D.; Takahashi, Y.; Abe, H.; Gasteiger, J.
„Rapid and Portable Development of Visualization Tools for Computational Chemistry Applications“
J. Mol. Graphics

Anhang

Dieser Anhang enthält Tabellen und Abbildungen mit umfangreichem Zahlenmaterial das den Rahmen einzelner Abschnitte sprengen würde und daher an dieser Stelle zusammengefaßt wird.

- Abschnitt A.1 stellt Rechenzeiten für verschiedene Algorithmen des WODCA Systems zusammen.
- In Abschnitt A.2 sind Tabellen mit Ergebnissen von substitutions-basierten Ähnlichkeitssuchen zu finden.

A.1 Rechenzeiten

In diesem Abschnitt sind Rechenzeiten für verschiedene Algorithmen des WODCA Systems zusammengefaßt.

Tabelle A - 1 (Seite A-3ff) enthält Rechenzeiten von Transformationen von Verbindungen mittels *tcl*-Skripts. Wo möglich, sind Vergleichszahlen für die analogen, compilierten *Fortran*-Funktionen angegeben.

Tabelle A - 2 auf Seite A-9f stellt Rechenzeiten der Analysefunktionen zur Suche nach strategischen Bindungen (*tcl*- und *Forth*-basiert) gegenüber.

Tab. A - 1 Rechenzeiten für verschiedene Tranformationsregeln, Kataloge und Maschinen

Transformation	Katalog	Rechner Sun Sparc	Programm-version[a]	Prozeß-zeit[b] [s]	System-zeit[c] [s]	CPU-Zeit[d] [s]	Zeit[e] [h:min:s]	Zeit je Eintrag [s]	f
Identität	Janssen	10/40	SunOS	2.445,7	3,0	2.448,7	0:41:10	0,3	•
	Fluka	10/512	SunOS	6.198,0	5,6	6.203,6	1:44:05	0,4	•
	Trivialnamen	10/512	SunOS	10.103,6	9,3	10.112,9	2:53:06	0,5	•
Elementaustausch und Austausch von ZH_n-Gruppen	Janssen	10/40	SunOS	2.536,0	3,5	2.539,5	0:42:46	0,3	•
	Fluka	10/40	SunOS	10.091,7	9,6	10.101,3	2:50:16	0,6	•
	Trivialnamen	10/40	SunOS	15.948,8	7,0	15.955,8	4:27:20	0,8	•
Größtes Ringsystem	Janssen	10/512	Solaris	8.664,2	2,4	8.666,6	2:24:38	1,0	
	Fluka	10/512	Solaris	17.166,8	6,0	17.172,8	4:47:27	1,0	
	Trivialnamen	10/40	SunOS	30.884,7	9,7	30.894,4	8:35:40	1,5	•
Substitutionspositionen auf dem reduzierten Ringsystem	Janssen	10/512	SunOS	6.116,4	3,9	6.120,3	1:45:00	0,7	•
	Fluka	10/512	Solaris	18.866,3	5,7	18.872,0	5:17:44	1,1	
	Trivialnamen	10/512	SunOS	34.192,8	11,4	34.204,2	9:32:42	1,7	•
Substitutionspositionen auf dem Ringsystem	Janssen	10/40	SunOS	5.661,9	3,4	5.665,3	1:36:02	0,8	•
	Fluka	10/51	SunOS	14.150,7	7,4	14.158,1	4:00:43	0,8	•
Ringe und C-Skelett	Janssen	10/512	Solaris	9.370,2	2,6	9.372,8	2:36:23	1,1	
	Fluka	10/512	Solaris	22.301,7	5,9	22.307,6	6:12:44	1,3	
	Trivialnamen	10/512	SunOS	36.884,8	10,6	36.895,4	10:38:18	1,8	
C-Skelett	Janssen	10/512	Solaris	8.143,2	3,3	8.146,5	2:16:30	1,0	
	Fluka	10/512	Solaris	18.184,5	6,1	18.190,6	5:07:47	1,1	
	Trivialnamen	10/512	SunOS	31.068,0	13,7	31.081,7	9:18:46	1,5	

Tab. A - 1 (Forts.) Rechenzeiten für verschiedene Tranformationsregeln, Kataloge und Maschinen

Transformation	Katalog	Rechner Sun Sparc	Programm-version[a]	Prozeß-zeit[b] [s]	System-zeit[c] [s]	CPU-Zeit[d] [s]	Zeit[e] [h:min:s]	Zeit je Eintrag [s]	f
Reduziertes C-Skelett	Janssen	10/512	Solaris	9.721,1	3,1	9.724,2	2:43:35	1,2	
	Fluka	10/512	Solaris	20.903,7	6,1	20.909,8	5:52:25	1,2	
	Trivialnamen	10/512	SunOS	37.821,6	12,1	37.833,7	10:31:49	1,8	
Aromatensystem einschließlich α-Atome	Janssen	10/512	Solaris	7.651,0	2,9	7.653,9	2:07:38	0,9	
	Fluka	10/51	SunOS	11.878,3	7,5	11.885,8	3:24:23	0,7	•
	Trivialnamen	10/512	SunOS	20.762,4	11,2	20.773,6	5:49:35	1,0	•
C-Skelett einschließlich α-Atome	Janssen	10/512	Solaris	12.860,7	3,0	12.863,7	3:38:21	1,5	
	Fluka	10/512	Solaris	28.742,4	6,3	28.748,7	14:43:46	1,7	
	Trivialnamen	10/51	SunOS	45.648,4	11,0	45.659,4	12:55:49	2,2	
C-Skelett und Aromaten einschließlich α-Atome	Janssen	10/512	Solaris	12.228,7	3,0	12.231,7	3:26:37	1,5	
	Fluka	10/512	Solaris	27.470,7	6,7	27.477,4	7:43:11	1,6	
	Trivialnamen	10/51	SunOS	45.237,7	10,0	45.247,7	12:45:56	2,2	
Substitutionsmuster (-AR+A1-CCMB-BO-MU)	Janssen	10/51	Solaris	10.058,0	2,8	10.060,8	2:51:06	1,2	
	Fluka	10/512	Solaris	22.149,6	7,0	22.156,6	6:14:29	1,3	
	Trivialnamen	10/512	SunOS	32.884,0	10,4	32.894,4	9:09:47	1,6	
Substitutionsmuster (-AR+A1+CCMB-BO-MU)	Janssen	10/51	Solaris	10.290,6	2,5	10.290,6	2:54:39	1,2	
	Fluka	10/40	Solaris	19.475,6	7,7	19.483,3	5:27:44	1,2	
		10/512	Solaris	22.336,1	6,6	22.342,7	6:16:11	1,3	
	Trivialnamen	10/512	SunOS	43.879,2	11,6	43.890,8	12:15:29	2,2	
Substitutionsmuster (-AR+A1-CCMB-BO+MU)	Janssen	10/51	Solaris	10.135,2	2,3	10.137,5	2:52:06	1,2	
	Fluka	10/512	Solaris	22.918,3	8,2	22.926,5	6:32:20	1,4	
	Trivialnamen	10/512	SunOS	32.659,5	10,1	32.669,6	9:06:35	1,6	

Tab. A - 1 (Forts.) Rechenzeiten für verschiedene Tranformationsregeln, Kataloge und Maschinen

Transformation	Katalog	Rechner Sun Sparc	Programm-version[a]	Prozeß-zeit[b] [s]	System-zeit[c] [s]	CPU-Zeit[d] [s]	Zeit[e] [h:min:s]	Zeit je Eintrag [s]	f
Substitutionsmuster (-AR+A1+CCMB-BO+MU)	Janssen	10/51	Solaris	12.481,7	2,4	12.484,1	3:31:38	1,5	
	Fluka	10/512	Solaris	22.317,5	7,4	22.324,9	6:38:50	1,3	
	Trivialnamen	10/40	SunOS	42.416,8	11,3	42.428,1	11:53:04	2,1	
Substitutionsmuster (-AR+A1-CCMB+BO+MU)	Janssen	10/51	Solaris	10.273,0	2,7	10.275,7	2:56:02	1,2	
	Fluka	10/512	Solaris	22.052,5	6,6	22.059,1	6:10:52	1,3	
	Trivialnamen	10/40	SunOS	42.358,5	10,6	42.369,1	13:20:15	2,1	
Substitutionsmuster (-AR+A1+CCMB+BO+MU)	Janssen	10/51	Solaris	10.287,2	2,9	10.290,1	2:54:36	1,2	
	Fluka	10/512	Solaris	23.076,6	7,6	23.084,2	6:27:41	1,4	
	Trivialnamen	10/40	SunOS	41.850,6	7,9	41.858,5	11:41:03	2,0	
Substitutionsmuster (-AR-A1-CCMB-BO-MU)	Janssen	10/512	Solaris	10.849,1	3,2	10.852,3	3:04:48	1,3	
	Fluka	10/512	Solaris	23.249,9	4,9	23.254,8	6:32:14	1,4	
	Trivialnamen	10/40	SunOS	44.350,5	7,9	44.358,4	12:22:40	2,2	•
Fortran-Rountine	Trivialnamen	10/512	SunOS	5.385,7	173,9	5.559,6	1:40:25	0,3	•
Substitutionsmuster (-AR-A1+CCMB-BO-MU)	Janssen	10/512	Solaris	11.180,3	2,9	11.183,2	3:07:49	1,3	
	Fluka	10/512	Solaris	23.020,4	4,4	23.024,8	6:28:35	1,4	
	Trivialnamen	10/40	SunOS	54.314,9	8,1	54.323,0	15:09:50	2,7	•
Fortran-Routine	Trivialnamen	10/512	SunOS	5.336,9	194,7	5.531,6	1:39:35	0,3	•
Substitutionsmuster (-AR-A1-CCMB-BO+MU)	Janssen	10/512	Solaris	10.823,7	3,3	10.827,0	3:01:15	1,3	
	Fluka	10/512	Solaris	24.264,7	6,7	24.271,4	6:53:00	1,4	
	Trivialnamen	10/40	SunOS	44.534,2	8,9	44.543,1	12:40:32	2,2	•
Fortran-Routine	Trivialnamen	10/512	SunOS	5.356,7	174,5	5.531,2	1:39:53	0,3	•

Tab. A - 1 (Forts.) Rechenzeiten für verschiedene Tranformationsregeln, Kataloge und Maschinen

Transformation	Katalog	Rechner Sun Sparc	Programm-version[a]	Prozeß-zeit[b] [s]	System-zeit[c] [s]	CPU-Zeit[d] [s]	Zeit[e] [h:min:s]	Zeit je Eintrag [s]	f
Substitutionsmuster (-AR-A1+CCMB-BO+MU)	Janssen	10/512	Solaris	10.908,5	2,8	10.911,3	3:02:22	1,3	
	Fluka	10/512	Solaris	23.090,5	5,7	23.096,2	6:41:02	1,4	
	Trivialnamen	10/40	SunOS	44.326,7	8,4	44.335,1	12:22:39	2,2	
Substitutionsmuster (-AR-A1-CCMB+BO+MU)	Janssen	10/512	Solaris	11.159,2	2,9	11.162,1	3:06:11	1,3	
	Fluka	10/512	Solaris	22.639,9	4,4	22.644,3	6:20:12	1,4	
	Trivialnamen	10/40	SunOS	44.154,7	8,9	44.163,6	12:20:47	2,2	
Substitutionsmuster (-AR-A1+CCMB+BO+MU)	Janssen	10/512	Solaris	10.918,2	3,0	10.921,2	3:02:13	1,3	
	Fluka	10/512	Solaris	22.877,2	4,4	22.881,6	6:23:44	1,4	
	Trivialnamen	10/40	SunOS	43.887,9	8,3	43.896,2	12:14:32	2,2	
Maximale Oxidation	Janssen	10/512	SunOS	6.889,0	3,1	6.892,1	1:55:18	0,8	•
	Fluka	10/512	SunOS	32.168,6	7,2	32.175,8	8:59:44	1,9	•
	Trivialnamen	10/40	SunOS	54.102,2	9,1	54.111,3	15:05:07	2,6	•
C-Skelett einschließlich α-Atome und maximale Oxidation	Janssen	10/40	SunOS	15.686,9	3,7	15.690,6	4:22:31	1,9	•
	Fluka	10/40	SunOS	60.072,2	9,4	60.081,6	16:46:06	3,6	•
	Trivialnamen	10/40	SunOS	100.227,2	15,2	100.242,4	28:46:35	4,9	•
Maximale Reduktion	Janssen	10/40	SunOS	9.061,8	3,6	9.065,4	2:32:05	1,1	•
	Fluka	10/40	SunOS	25.063,3	8,3	25.071,6	7:00:17	1,5	•
	Trivialnamen	10/512	SunOS	37.901,9	11,1	37.913,0	10:32:27	1,8	•
C-Skelett einschließlich α-Atome und maximale Reduktion	Janssen	10/40	SunOS	14.387,2	3,6	14.390,8	4:00:47	1,7	•
	Fluka	10/40	SunOS	44.767,5	12,2	44.779,7	12:42:09	2,7	•
	Trivialnamen	10/40	SunOS	79.735,6	14,0	79.749,6	22:16:09	3,9	•

Tab. A - 1 (Forts.) Rechenzeiten für verschiedene Tranformationsregeln, Kataloge und Maschinen

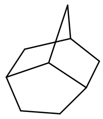
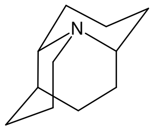
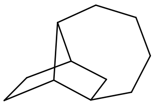
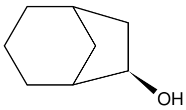
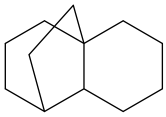
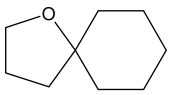
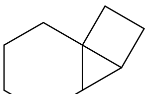
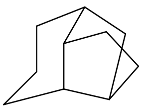
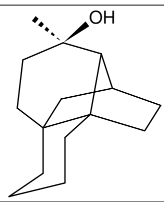
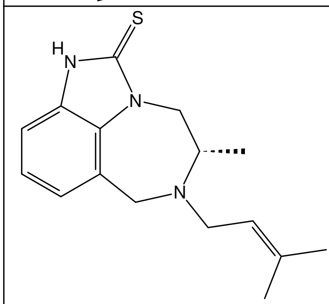
Transformation	Katalog	Rechner Sun Sparc	Programm-version[a]	Prozeß-zeit[b] [s]	System-zeit[c] [s]	CPU-Zeit[d] [s]	Zeit[e] [h:min:s]	Zeit je Eintrag [s]	f
Maximale Reduktion (ohne Aromaten)	Janssen	10/40	SunOS	7.010,9	3,5	7.014,4	1:57:41	0,8	•
	Fluka	10/40	SunOS	21.988,2	9,5	21.997,7	6:14:55	1,3	•
	Trivialnamen	10/40	SunOS	40.289,1	12,3	40.301,4	18:47:57	2,0	•
Maximale Reduktion (ohne C-Aromaten)	Janssen	10/40	SunOS	7.115,5	3,4	7.118,9	1:59:22	0,8	•
	Fluka	10/40	SunOS	22.242,6	7,9	22.250,5	6:13:18	1,3	•
	Trivialnamen	10/512	Solaris	31.508,1	9,8	31.517,9	8:45:54	1,5	
Hydrolyse ohne Aromatensubstituenten	Janssen	10/512	SunOS	5.600,7	2,3	5.603,0	1:33:55	0,7	•
	Fluka	10/40	SunOS	20.644,1	7,6	20.651,7	5:46:27	1,2	•
	Trivialnamen	10/51	SunOS	30.332,1	8,5	30.340,6	8:38:36	1,5	•
Fortran-Routine	Trivialnamen	10/512	SunOS	5.405,0	220,7	5.625,7	2:00:25	0,3	•
Hydrolyse einschließlich Aromatensubstituenten	Janssen	10/40	SunOS	7.403,2	3,3	7.406,5	2:04:01	0,9	•
	Fluka	10/40	SunOS	21.858,4	7,3	21.865,7	6:06:44	1,3	•
	Trivialnamen	10/40	SunOS	36.791,7	7,9	36.799,6	10:16:35	1,8	•
Fortran-Routine	Trivialnamen	10/512	SunOS	5.576,1	211,3	5.787,4	1:55:33	0,3	•
Hydrolyse ohne Aromatensubstituenten und Oxidation	Janssen	10/40	SunOS	13.803,0	3,5	13.806,5	3:51:08	1,6	•
	Fluka	10/40	SunOS	49.525,1	9,4	49.534,5	13:50:23	3,0	•
	Trivialnamen	10/40	SunOS	85.920,9	17,0	85.937,9	24:17:49	4,2	•
Fortran-Routine	Trivialnamen	10/512	SunOS	6.063,6	227,7	6.291,3	2:14:18	0,3	•
Hydrolyse einschließlich Aromatensubstituenten und Oxidation	Janssen	10/512	SunOS	12.292,3	5,0	12.297,3	5:06:05	1,4	•
	Trivialnamen	10/512	SunOS	69.912,7	10,9	69.923,6	19:29:48	3,4	•
Fortran-Routine	Trivialnamen	10/512	SunOS	5.705,3	203,7	5.909,0	1:54:21	0,3	•

Tab. A - 1 (Forts.) Rechenzeiten für verschiedene Tranformationsregeln, Kataloge und Maschinen

Transformation	Katalog	Rechner Sun Sparc	Programm-version[a]	Prozeß-zeit[b] [s]	System-zeit[c] [s]	CPU-Zeit[d] [s]	Zeit[e] [h:min:s]	Zeit je Eintrag [s]	[f]
Eliminierung	Janssen	10/512	SunOS	5.244,1	7,9	5.252,0	1:33:31	0,6	
	Fluka	10/512	Solaris	12.702,8	28,7	12.731,5	3:32:59	0,7	
Eliminierung und Substitutionsmuster	Janssen	10/512	Solaris	14.594,7	3,4	14.598,1	4:03:31	1,7	
	Fluka	10/512	Solaris	32.233,8	27,5	32.261,3	9:00:35	1,9	
	Trivialnamen	10/512	SunOS	60.610,9	24,2	60.616,1	16:53:20	3,0	
C-Skelett einschließlich Aromaten mit α-Atomen und Eliminierung	Janssen	10/512	SunOS	14.895,7	5,2	14.900,9	5:16:51	1,8	●
	Fluka	10/512	Solaris	26.841,1	13,7	26.854,8	7:33:56	1,6	
	Trivialnamen	10/51	SunOS	70.311,6	24,7	70.336,3	19:55:10	3,4	●
Fortran-Routine	Trivialnamen	10/512	SunOS	9.266,8	203,6	9.470,4	2:44:32	0,5	●
Ozonolyse	Janssen	10/51	SunOS	4.113,8	3,3	4.117,1	1:10:23	0,5	●
	Fluka	10/40	SunOS	14.715,4	7,3	14.722,7	4:06:28	0,9	●
	Trivialnamen	10/40	SunOS	27.647,2	7,4	27.654,6	7:42:56	1,4	●
Fortran-Routine	Trivialnamen	10/51	SunOS	6.137,6	162,6	6.300,2	2:06:08	0,3	●
Ozonolyse und Substitutionsmuster	Janssen	10/51	Solaris	10.356,8	3,2	10.360,0	3:02:29	1,2	
	Fluka	10/512	Solaris	22.384,0	7,0	22.391,0	6:16:44	1,3	
	Trivialnamen	10/40	SunOS	45.541,2	8,8	45.550,0	12:39:01	2,2	

a. Die verwendeten Rechner benutzen als Betriebssystem Solaris 2.4 (kurz: Solaris). Es ist jedoch möglich auch für SunOS 4.1.3 (kurz: SunOS) compilierte Programme unter diesem Betriebssytem laufen zu lassen. Das bedeutet jedoch eine langsamere Ausführung des Programms.
b. CPU-Zeit des eigentlichen Prozesses.
c. CPU-Zeit, die der Prozesses für Systemfunktionen benötigte.
d. Summe aus Prozeß- und Systemzeit.
e. Verstrichene Zeit für den Prozeß, das ist die reale Zeit die dieser Prozeß zusammen mit anderen Prozessen brauchte.
f. Mit ● sind Zeiten markiert, die in Abbildung 3 - 2 auf Seite 38 verwendet wurden.

Tab. A - 2 Rechenzeiten[a] von Retrosynthesefunktionen auf der Basis der alten (*Forth*) bzw. neuen (*tcl*) Implementation auf einer Sun Sparc 10/20 unter SunOS. Anmerkungen siehe folgende Seite.

Verbindung	Methode	Implementation	
		Forth [b]	*tcl* [c]
	Ring-Strategie	2	0,8
	Ring-Strategie	2	0,9
	Ring-Strategie	1	0,8
	Ring-Strategie	1	1,0
	Ring-Strategie	1	1,0
	Ring-Strategie	1	1,0
	Ring-Strategie	1	0,9
	Ring-Strategie	1	0,9
	Ring-Strategie	2	1,0
	CX-Strategie	3	1,2

a. Gemessen wurde die real verstrichene Zeit (*elapsed time*) zur Abarbeitung der entsprechenden Funktionen, jedoch nicht die textuelle oder graphische Ausgabe der Resultate.
b. Zeitmessung mittels der *Fortran77* `time`-Funktion. Die Zeit wird als Differenz zwischen dem Anfangszeitpunkt der *Forth*-Funktion und ihrem Ende ermittelt. Diese `time`-Funktion arbeitet auf Sekunden genau.
c. Zeitmessung mittels der *tcl* `time`-Funktion. Anfangs- und Endzeitpunkt werden hier in Mikrosekunden gemessen. Die Angaben sind auf Sekunden gerundet.

A.2 Transformationssuchen

Dieser Abschnitt stellt Suchergebnisse von Transformationssuchen mit verschiedenen Ähnlichkeitskriterien in diversen Katalogen zusammen.

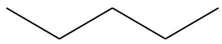
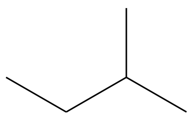
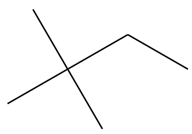
Tabelle A - 3 (Seite A-12) enthält Suchergebnisse für skelett-basierte Transformationen mit acyclischen C_5-Skeletten.

In Tabelle A - 4 ab Seite A-13 werden Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten, acyclischen C_5-Skeletten zusammengefaßt.

Die analogen Ergebnisse für C_6-Skelette umfassen die Tabellen A - 5, Seite A-21 (skelett-basierte Transformationen) und A - 6 ab Seite A-22 (substitutionsmuster-basierte Transformationen).

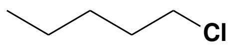
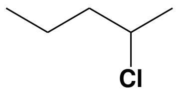
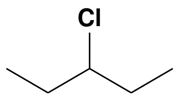
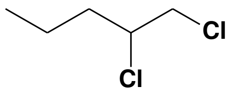
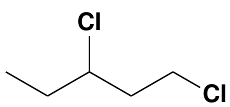
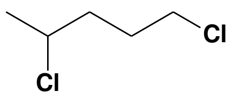
In Tabelle Tab. A - 7 auf Seite A-49f schließlich, sind Ergebnisse für substitutionsmuster-basierte Transformationen mit Cyclohexanen zusammengestellt.

Tab. A - 3 Suchergebnisse für skelett-basierte Transformationen mit acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

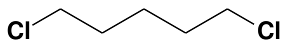
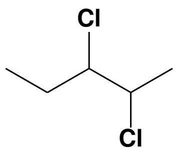
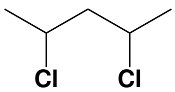
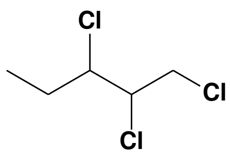
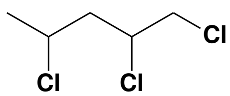
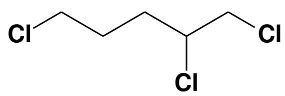
ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		a	b	a	b	a	b
C5-00000		284	451	526	674	363	427
C5-0000-0		65	92	105	144	68	108
C5-000-0-0		32	30	42	37	25	16

a. Transformationssuche „C-Skelett“
b. Transformationssuche „Reduziertes C-Skelett“

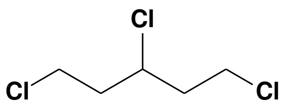
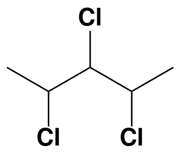
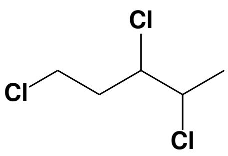
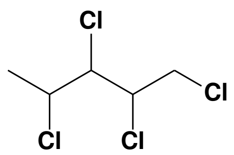
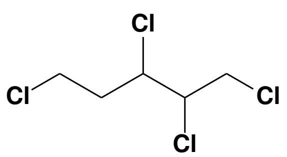
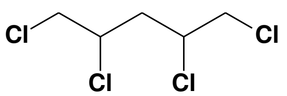
Tab. A - 4 Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-00001		21	17	60	45	4	4
C5-00010		7	4	11	7	—	—
C5-00100		6	3	9	5	—	—
C5-00011		7	6	10	12	4	5
C5-00101		1	1	5	5	—	—
C5-01001		16	14	11	8	4	2

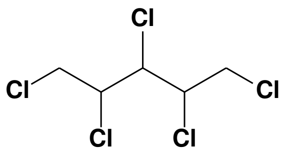
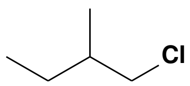
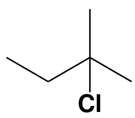
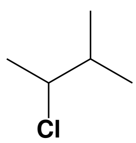
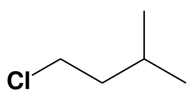
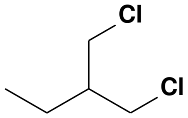
Tab. A - 4 (Forts.) Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-10001		66	58	66	57	48	32
C5-00110		1	2	1	4	—	1
C5-01010		3	2	24	5	2	2
C5-00111		—	4	1	6	—	—
C5-01011		1	2	2	4	3	3
C5-10011		58	59	147	148	81	79

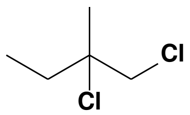
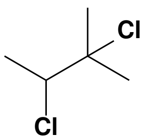
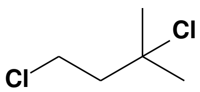
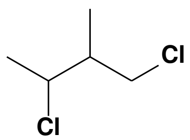
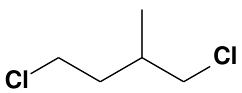
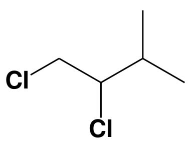
Tab. A - 4 (Forts.) Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-10101		29	27	26	24	11	9
C5-01110		—	3	1	22	—	—
C5-01101		—	3	—	4	13	13
C5-01111		—	3	—	4	2	7
C5-10111		12	18	22	25	33	32
C5-11011		3	5	10	13	17	21

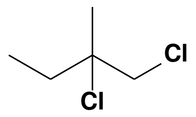
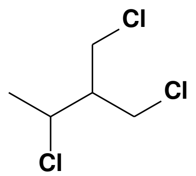
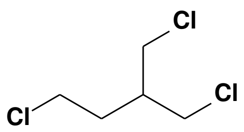
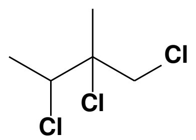
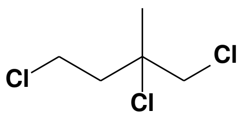
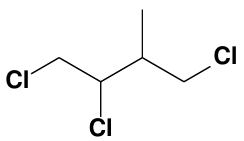
Tab. A - 4 (Forts.) Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-11111		40	45	64	84	104	121
C5-0001-0		7	6	16	11	5	—
C5-0010-0		7	4	10	8	3	1
C5-0100-0		3	3	2	2	1	1
C5-1000-0		16	11	21	14	16	6
C5-0001-1		3	2	3	3	1	—

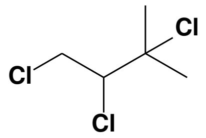
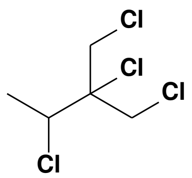
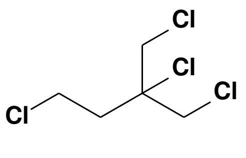
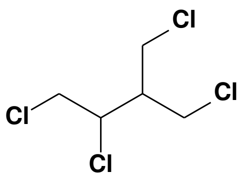
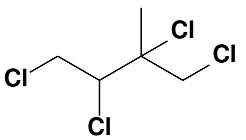
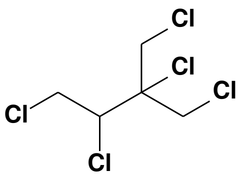
Tab. A - 4 (Forts.) Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-0011-0		1	1	1	4	7	6
C5-0110-0		1	2	1	3	—	1
C5-1010-0		—	—	3	3	2	1
C5-0101-0		3	3	—	—	4	4
C5-1001-0		14	7	19	6	7	1
C5-1100-0		10	10	28	30	11	9

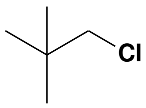
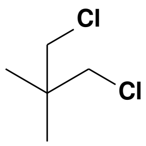
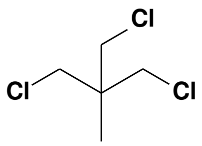
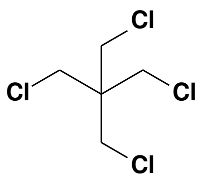
Tab. A - 4 (Forts.) Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-0011-1		1	1	1	1	—	—
C5-0101-1		2	1	1	1	4	2
C5-1001-1		2	2	1	—	5	5
C5-0111-0		1	2	—	3	—	5
C5-1011-0		—	2	4	5	1	1
C5-1101-0		2	1	1	3	3	1

Tab. A - 4 (Forts.) Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

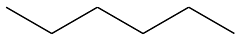
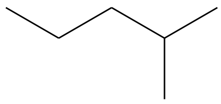
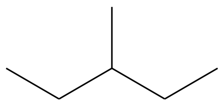
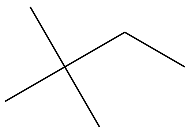
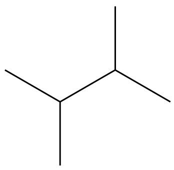
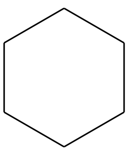
ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-1110-0		8	15	13	21	15	27
C5-0111-1		—	2	—	—	—	3
C5-1011-1		—	3	—	4	2	4
C5-1101-1		1	—	7	3	29	2
C5-1111-0		—	5	—	11	1	12
C5-1111-1		—	1	—	5	2	30

Tab. A - 4 (Forts.) Suchergebnisse für substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_5-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C5-001-0-0		13	13	23	23	8	8
C5-001-1-0		6	6	10	10	5	5
C5-001-1-1		4	4	4	4	2	2
C5-101-1-1		9	9	5	5	6	6

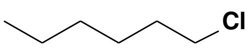
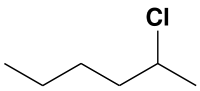
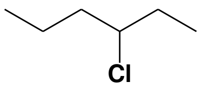
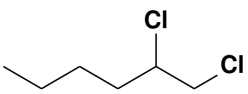
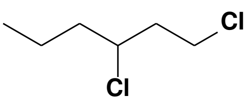
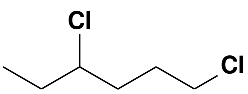
a. Variiert wurde nur der Ort der Substitution. Keine Beachtung finden dagegen die Zahl der Substituenten an einem Kohlenstoffatom oder die Bindungsordnung der Substituentenbindung.
b. Transformationssuche „Substitutionsmuster (-AR+A1-CCMB-BO-MU)“. Die Chloratome stehen für beliebige Substituenten, gleich welcher Zahl, chemischer Natur oder Bindungsordnung. Mehrfachbindungen zwischen Kohlenstoffatomen werden nicht als Substitutionsorte angesehen.
c. Transformationssuche „Substitutionsmuster (-AR+A1+CCMB-BO-MU)“. Wie oben. Mehrfachbindungen werden hier jedoch als Substitutionsorte angesehen.

Tab. A - 5 Suchergebnisse für skelett-basierte Transformationen mit acyclischen C_6-Skeletten und Cyclohexan in den Katalogen Janssen, Fluka und Trivialnamen

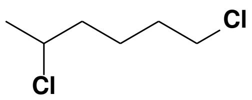
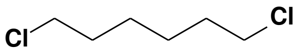
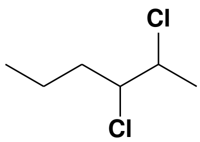
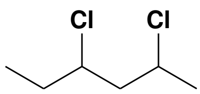
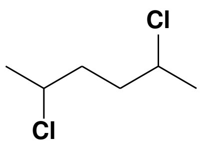
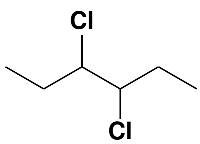
ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		a	b	a	b	a	b
C6-000000		176	247	454	492	776	758
C6-00000-0		47	89	75	125	50	90
C6-0000-00		31	63	86	118	63	82
C6-0000-0-0		10	12	48	53	36	37
C6-0-0000-0		8	18	11	19	4	15
Cyclo-C6-000000		70	936	113	1.170	115	809

a. Transformationssuche „C-Skelett“
b. Transformationssuche „Reduziertes C-Skelett“

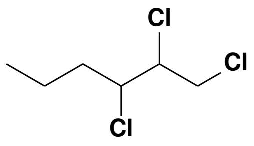
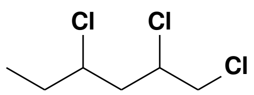
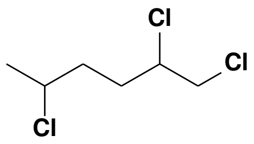
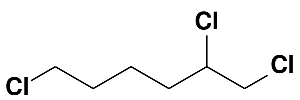
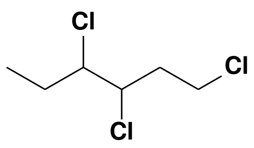
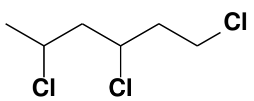
Tab. A - 6 Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-000001		29	18	64	53	16	7
C6-000010		3	2	5	4	—	—
C6-000100		4	2	3	2	—	—
C6-000011		3	4	8	12	5	5
C6-000101		1	1	2	2	—	—
C6-001001		—	—	—	—	3	2

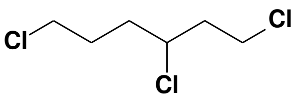
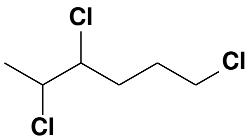
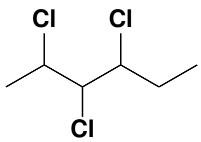
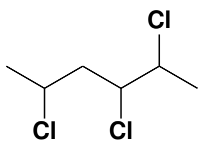
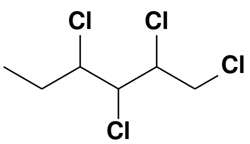
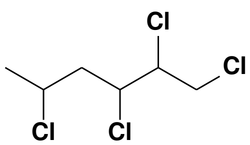
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-010001		5	5	1	1	10	8
C6-100001		34	31	49	46	34	28
C6-000110		—	2	1	4	—	—
C6-001010		—	—	—	—	—	—
C6-010010		5	4	6	5	2	2
C6-001100		—	1	1	4	—	—

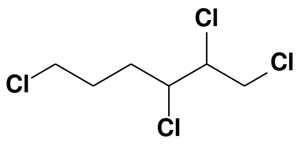
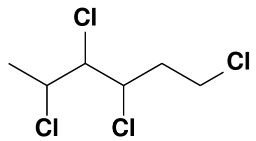
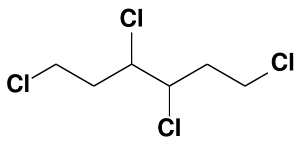
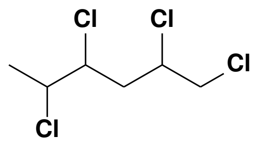
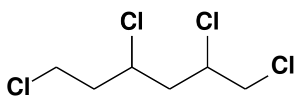
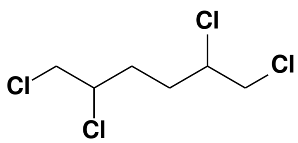
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-000111		—	6	1	4	—	1
C6-001011		—	—	—	—	—	—
C6-010011		—	1	—	1	—	—
C6-100011		21	20	42	44	30	29
C6-001101		—	2	—	2	—	2
C6-010101		2	—	1	—	1	—

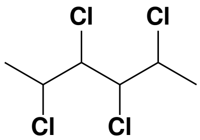
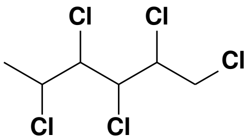
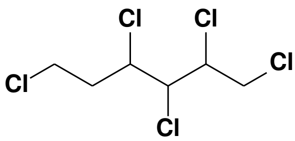
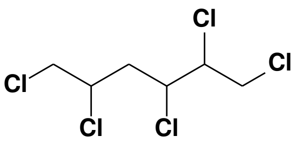
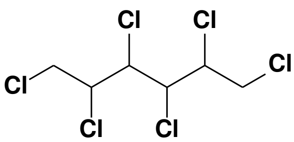
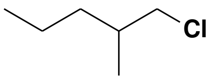
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-100101		3	3	3	3	4	4
C6-011001		—	—	—	—	6	5
C6-001110		—	—	—	1	—	—
C6-010110		—	—	—	—	—	—
C6-001111		—	1	—	—	—	—
C6-010111		—	—	1	1	5	5

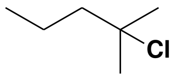
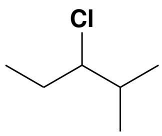
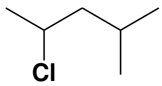
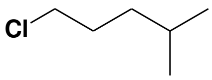
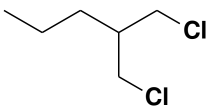
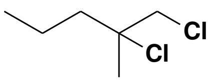
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-100111		—	—	1	—	3	2
C6-011101		2	2	4	3	28	28
C6-101101		—	2	—	1	2	2
C6-011011		—	1	1	2	30	30
C6-101011		—	—	—	—	5	2
C6-110011		1	3	5	7	28	23

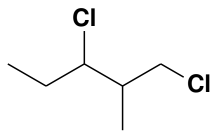
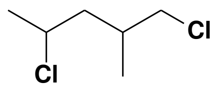
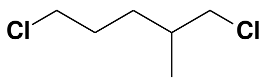
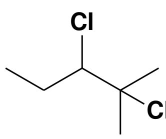
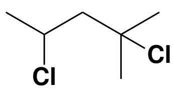
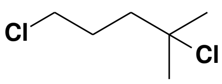
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-011110		—	4	—	3	—	—
C6-011111		3	9	7	11	64	74
C6-101111		3	2	7	4	4	4
C6-110111		1	1	1	—	8	14
C6-111111		76	80	251	259	490	503
C6-00001-0		1	1	2	2	—	—

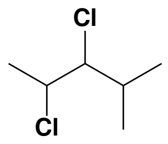
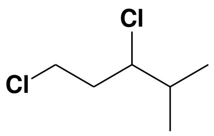
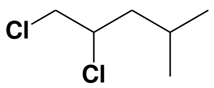
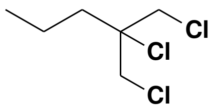
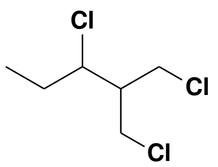
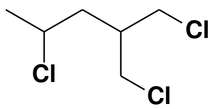
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-00010-0		1	1	1	1	—	—
C6-00100-0		1	1	3	3	—	—
C6-01000-0		4	3	4	3	4	1
C6-10000-0		1	1	3	3	1	—
C6-00001-1		1	—	3	1	1	1
C6-00011-0		—	1	—	1	—	—

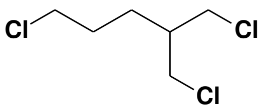
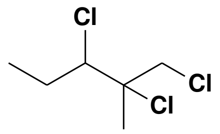
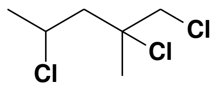
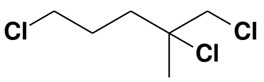
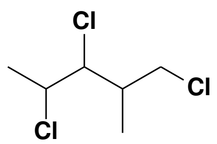
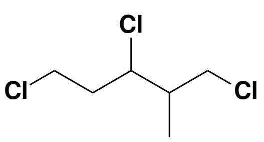
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-00101-0		—	—	—	—	1	1
C6-01001-0		1	1	1	1	—	—
C6-10001-0		3	3	6	6	2	1
C6-00110-0		—	1	—	1	—	—
C6-01010-0		6	6	7	7	4	4
C6-10010-0		1	1	1	1	—	—

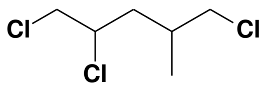
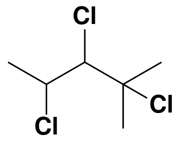
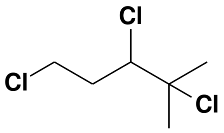
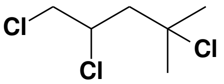
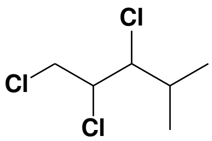
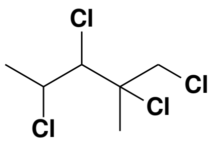
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-01100-0		—	1	—	—	—	—
C6-10100-0		1	1	1	1	—	—
C6-11000-0		13	14	33	35	15	15
C6-00011-1		—	—	—	—	—	—
C6-00101-1		—	—	—	—	—	—
C6-01001-1		—	—	—	—	—	—

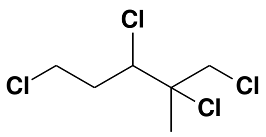
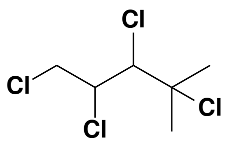
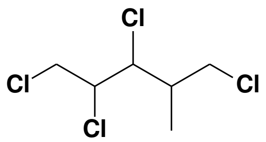
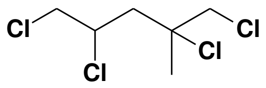
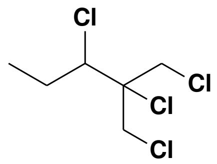
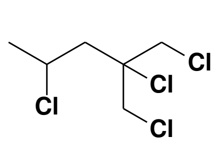
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-10001-1		13	10	6	2	9	3
C6-00111-0		—	—	—	—	—	—
C6-01011-0		—	—	—	—	—	1
C6-10011-0		2	2	2	2	1	1
C6-01101-0		—	—	—	—	—	—
C6-10101-0		—	—	—	—	—	—

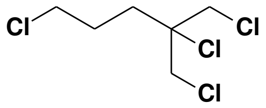
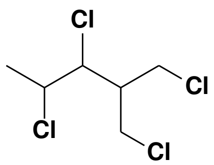
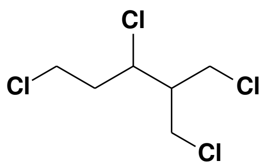
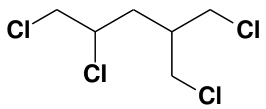
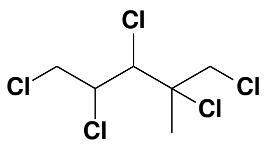
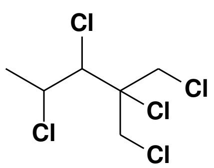
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-11001-0		—	—	—	—	3	2
C6-01110-0		—	1	—	1	—	2
C6-10110-0		—	—	—	—	—	1
C6-11010-0		—	—	1	1	—	—
C6-11100-0		—	—	1	1	4	4
C6-01111-0		—	1	—	2	—	—

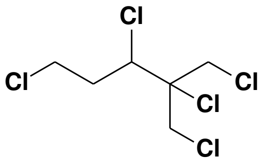
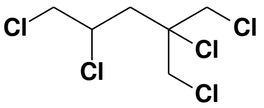
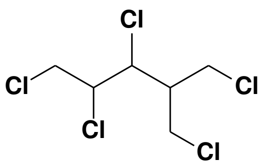
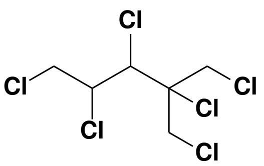
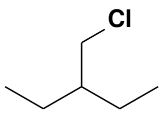
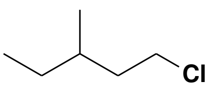
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-10111-0		—	—	—	—	—	—
C6-11110-0		—	—	—	1	—	—
C6-11101-0		—	—	—	—	—	—
C6-11011-0		—	—	—	—	1	1
C6-00111-1		—	—	—	—	—	—
C6-01011-1		—	—	—	—	1	1

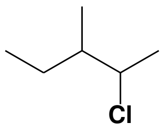
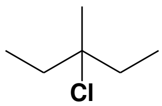
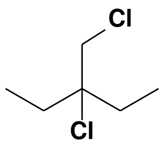
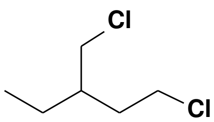
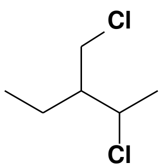
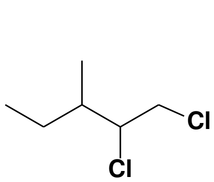
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-10011-1		—	—	—	—	—	1
C6-01101-1		—	—	—	—	—	—
C6-10101-1		2	2	3	3	2	—
C6-11001-1		—	1	2	4	—	—
C6-11111-0		—	—	—	—	8	7
C6-01111-1		—	—	—	—	—	—

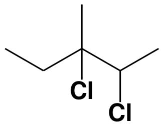
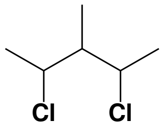
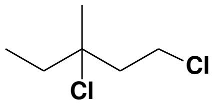
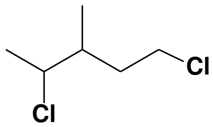
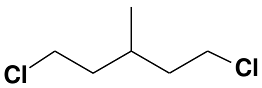
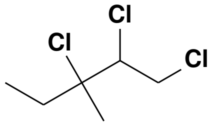
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-10111-1		—	2	—	1	—	4
C6-11011-1		—	—	—	—	1	1
C6-11101-1		—	—	—	—	1	2
C6-11111-1		—	1	—	3	3	6
C6-0001-00		3	3	4	4	1	—
C6-0000-01		1	1	5	3	—	—

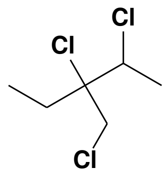
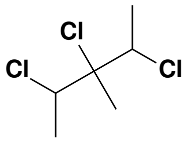
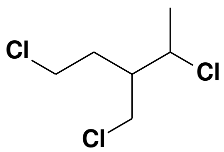
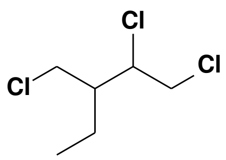
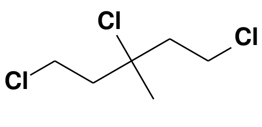
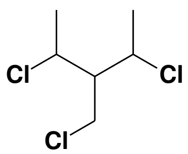
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0000-10		2	2	—	—	—	—
C6-0010-00		2	1	2	1	1	—
C6-0011-00		1	1	—	1	2	2
C6-0001-01		1	—	—	—	—	—
C6-0001-10		—	—	1	1	—	—
C6-0000-11		4	5	18	19	12	12

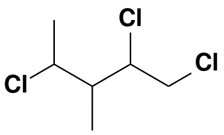
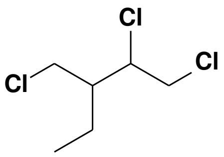
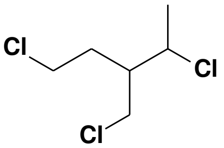
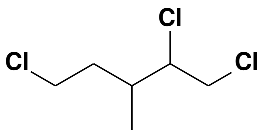
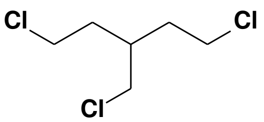
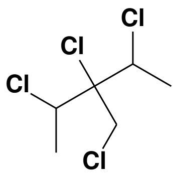
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0010-10		—	—	—	2	—	—
C6-0100-10		—	—	—	—	—	—
C6-0001-01		—	—	—	—	—	—
C6-0100-01		—	—	—	—	2	—
C6-1000-01		5	3	6	6	2	1
C6-0010-11		—	1	—	1	—	1

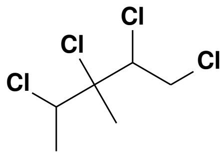
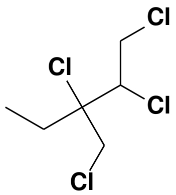
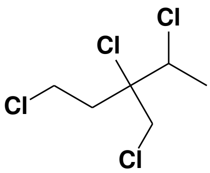
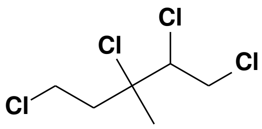
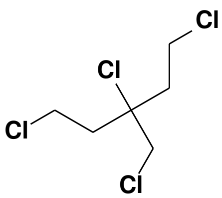
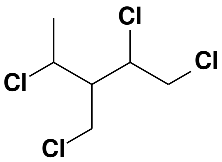
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0011-10		—	—	—	—	—	1
C6-0110-10		—	—	—	—	—	—
C6-1001-10		—	—	—	—	—	—
C6-1011-00		—	—	—	—	—	—
C6-0101-01		1	1	4	4	8	8
C6-0101-10		—	—	1	1	—	—

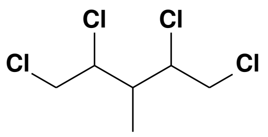
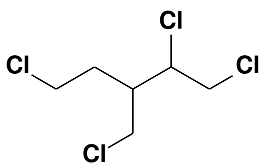
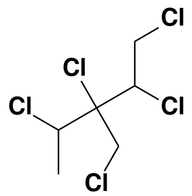
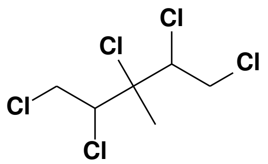
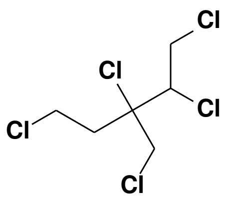
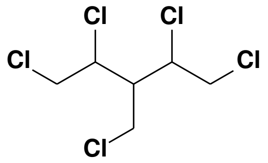
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-1100-10		—	—	—	—	2	2
C6-1011-00		—	—	—	—	2	1
C6-1001-10		2	2	1	1	—	—
C6-1100-01		—	—	—	—	2	2
C6-1001-01		7	5	8	5	11	6
C6-0111-10		—	—	—	—	—	—

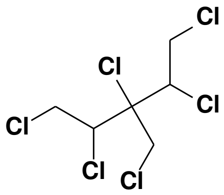
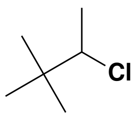
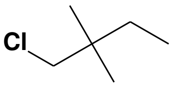
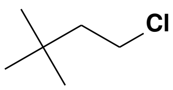
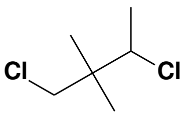
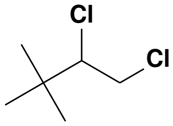
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-1110-10		—	—	—	—	—	2
C6-1111-00		—	1	—	—	—	—
C6-0111-01		—	—	—	—	—	—
C6-1110-01		—	2	—	—	—	1
C6-1011-01		4	4	32	32	10	10
C6-1101-10		—	—	—	—	—	—

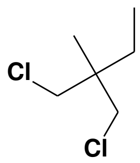
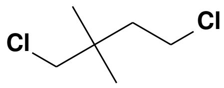
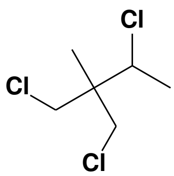
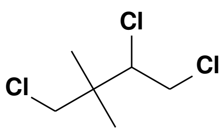
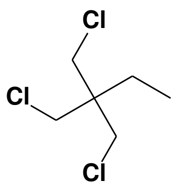
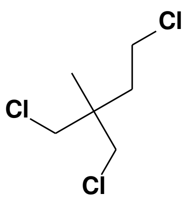
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-1100-11		—	—	—	1	—	—
C6-1101-01		3	3	3	3	9	9
C6-1111-10		—	—	—	—	3	3
C6-1110-11		—	—	—	2	—	—
C6-1111-01		—	2	2	5	4	9
C6-1101-11		—	—	—	—	—	—

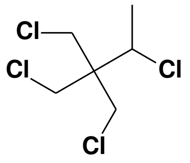
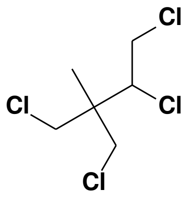
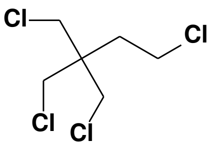
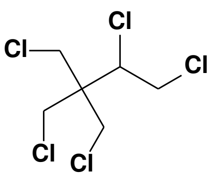
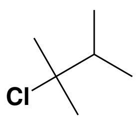
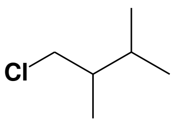
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-1111-11		—	—	—	—	—	1
C6-0100-0-0		2	2	3	2	2	2
C6-0001-0-0		—	—	—	—	—	—
C6-1000-0-0		4	4	6	5	—	—
C6-0101-0-0		—	—	—	—	—	—
C6-1100-0-0		—	1	15	20	4	5

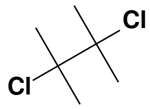
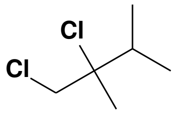
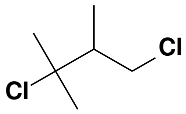
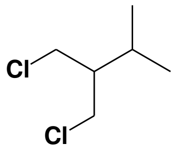
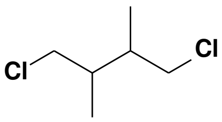
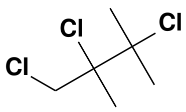
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0001-1-0		1	1	—	—	—	—
C6-1001-0-0		1	1	1	1	—	—
C6-0101-1-0		—	—	—	—	—	—
C6-1101-0-0		—	—	6	6	16	16
C6-0001-1-1		1	1	4	4	2	2
C6-1001-1-0		—	—	—	—	—	—

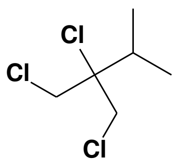
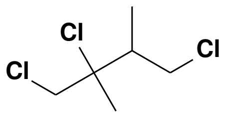
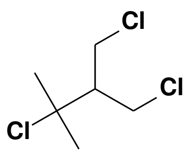
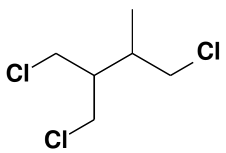
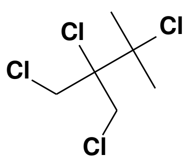
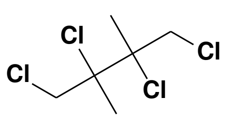
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0101-1-1		—	—	—	—	—	—
C6-1101-1-0		—	—	—	—	—	—
C6-1001-1-1		—	—	—	—	—	—
C6-1101-1-1		—	—	—	—	—	—
C6-0-0010-0		1	1	4	4	—	—
C6-0-0001-0		—	—	—	—	—	—

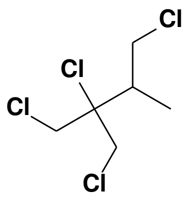
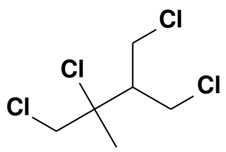
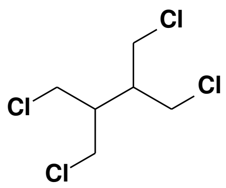
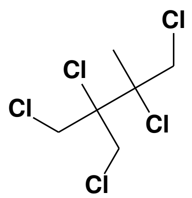
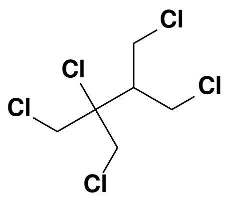
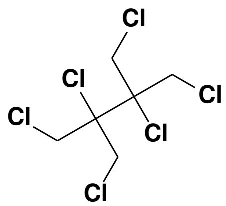
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0-0110-0		3	5	1	3	1	1
C6-0-0011-0		—	1	—	1	2	1
C6-0-0101-0		—	—	—	—	—	—
C6-0-0001-1		—	—	1	—	—	—
C6-0-1001-0		3	2	2	1	2	—
C6-0-0111-0		—	—	—	—	—	—

Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0-0011-1		—	—	—	—	—	—
C6-0-1011-0		—	—	—	—	—	—
C6-0-0101-1		—	—	—	—	—	—
C6-0-1001-1		—	—	—	—	—	—
C6-0-0111-1		—	—	—	1	—	—
C6-0-1111-0		—	2	—	2	—	2

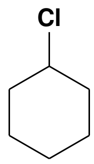
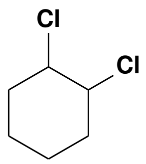
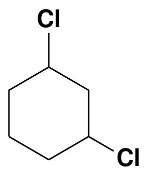
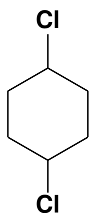
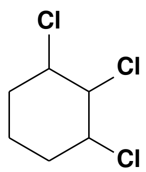
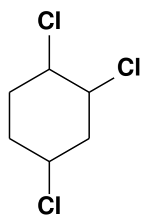
Tab. A - 6 (Forts.) Suchergebnisse für Substitutionsmuster-basierte Transformationen mit verschieden substituierten[a], acyclischen C_6-Skeletten in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		b	c	b	c	b	c
C6-0-1011-1		—	—	—	—	—	—
C6-0-1101-1		—	—	—	—	—	—
C6-1-1001-1		2	1	3	1	2	2
C6-0-1111-1		—	—	—	—	—	—
C6-1-1011-1		—	—	—	—	—	—
C6-1-1111-1		—	1	2	4	—	1

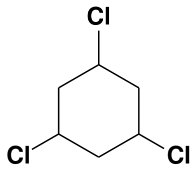
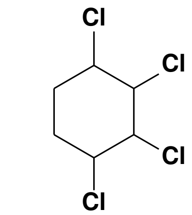
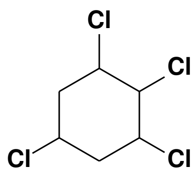
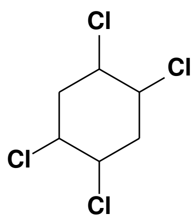
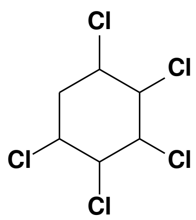
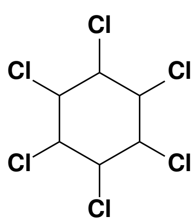
a. Variiert wurde nur der Ort der Substitution. Keine Beachtung finden dagegen die Zahl der Substituenten an einem Kohlenstoffatom oder die Bindungsordnung der Substituentenbindung.

b. Transformationssuche „C-Skelett“
c. Transformationssuche „Reduziertes C-Skelett“

Tab. A - 7 Suchergebnisse für Transformationssuchen mit substituierten Cyclohexanen in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		a	b	a	b	a	b
Cyclo-C6-000001		42	34	63	59	20	20
Cyclo-C6-000011		16	20	21	24	7	4
Cyclo-C6-000101		3	2	3	2	2	2
Cyclo-C6-001001		6	4	9	5	14	2
Cyclo-C6-000111		2	6	—	3	1	—
Cyclo-C6-001011		6	—	11	—	6	—

Tab. A - 7 (Forts.) Suchergebnisse für Transformationssuchen mit substituierten Cyclohexanen in den Katalogen Janssen, Fluka und Trivialnamen

ID	Verbindung	Suchergebnis (Treffer)					
		Janssen		Fluka		Trivialnamen	
		a	b	a	b	a	b
Cyclo-C6-010101		—	—	2	2	—	—
Cyclo-C6-001111		1	2	1	1	8	1
Cyclo-C6-010111		2	—	3	—	—	—
Cyclo-C6-011011		1	3	—	1	7	1
Cyclo-C6-011111		—	—	—	—	16	17
Cyclo-C6-111111		12	17	16	26	72	106

a. Transformationssuche „C-Skelett“
b. Transformationssuche „Reduziertes C-Skelett“

Publikationsliste

Gasteiger, J.; Ihlenfeldt, W. D.; Fick, R.; Rose, J. R.
„Similarity Concepts for the Planning of Organic Reactions and Syntheses“
J. Chem. Inf. Comput. Sci. 32 (1992) 700-712

Fick, R.; Gasteiger, J.; Ihlenfeldt, W. D.
„Synthesis Planning in the 90's: The WODCA System“
In: E.C.C.C.1 Computational Chemistry F.E.C.S. Conference,
Nancy, France May 1994; Bernardi, F.; Rivail, J.-L. (Hrsg.)
AIP Press, Woodbury (1995) 526-531

Fick, R.; Ihlenfeldt, W.-D.; Gasteiger, J.
„Computer-assisted design of syntheses for heterocyclic compounds“
Heterocycles 40 (1995) 993-1007

Lebenslauf

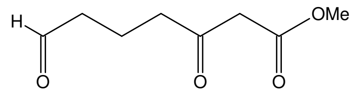
Vor- und Zuname:	Ralf Fick
Geburtsdatum und -ort:	06.04.1964 in Eilenburg
Eltern:	Heinz Fick und Hildegard Fick, geb. Wollmann
Staatsangehörigkeit:	deutsch
Familienstand:	ledig
Schulbildung:	1970-1978 Allgemeinbildende Oberschule Laußig 1978-1982 Erweiterte Oberschule (Gymnasium) Eilenburg
Wehrdienst:	November 1982 bis Oktober 1985
Studium:	Chemiestudium vom November 1985 bis August 1990 an der Technischen Hochschule Merseburg Diplomarbeit 1990 bei Prof. Dr. H. Kehlen
Berufliche Tätigkeit:	Wissenschaftlicher Assistent an der TH Merseburg vom September 1990 bis Oktober 1991
Promotion:	Begonnen im November 1991 bei Prof. Dr. J. Gasteiger an der TU München, im Februar 1994 Wechsel mit dem Doktorvater an die Universität Erlangen-Nürnberg

Das ist ein Referenzdokument, das Bestandteil des Thesis-Buches ist. Es dient dazu, fortlaufend numerierte Moleküle zu referenzieren. Jedes irgendwo referenzierte Molekül ist hier durch einen Absatz vom Typ MolNummer vertreten. Zusätzlich zur automatisch generierten Nummer ist eine Bezeichnung der Verbindung angegeben, u. U. auch eine Grafik.

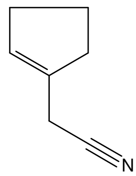
<u>Dieses Dokument wird niemals gedruckt!</u>

Wenn zusätzlich Moleküle referenziert werden sollen oder Abbildungen verschoben werden, müssen natürlich *hier* ebenfalls entsprechende Änderungen vorgenommen werden.

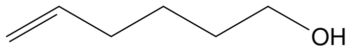
1 3-Oxo-heptandicarbonsäure (1,3,7-funktionalisierte C_7-Einheit)



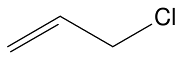
2 Cyclopenten-1-essigsäurenitril (Cyclopenten-Derivat)



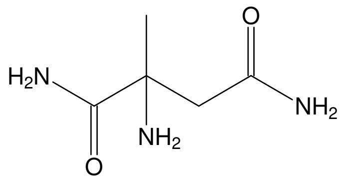
3 6-Hydroxy-hex-1-en



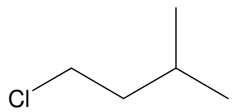
4 Allylchlorid



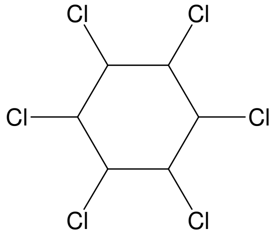
5 2-Amino-2-methyl-bernsteinsäurediamid



6 3-Methyl-3-buten-1-ol (Treffer)
7 Oxetan-Derivat (Treffer)
8 Isopentylchlorid

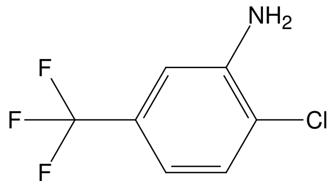


9 Isopentylbromid (Treffer)
10 Isopentylamin (Treffer)
11 Isopentylalkohol (Treffer)
12 1,2,3,4,5,6-Hexachlor-cyclohexan

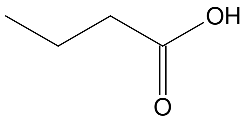


13 *meso*-Inosit (Treffer)

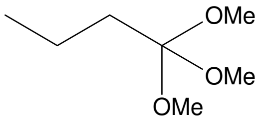
14 *meso*-Inosit-Derivat (Treffer)
15 Rhodizonsäure (entfernter Treffer)
16 Tetrafluoro-*p*-Benzochinon (Treffer)
17 3-Amino-4-chlor-benzotrifluorid



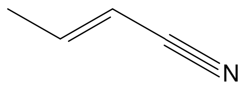
18 Buttersäure



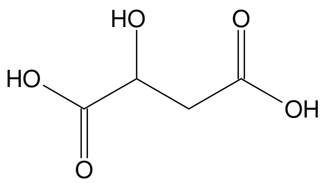
19 Ortho-Buttersäuremethylester



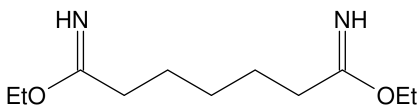
20 Crotonsäurenitril



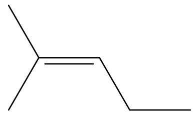
21 Cholesterylbutyrat (Treffer)
22 N(6),2‘-O-Dibutyryl-adenosin-3‘,5‘-cyclophosphat Natriumsalz Monohydrat (Treffer)
23 Äpfelsäure



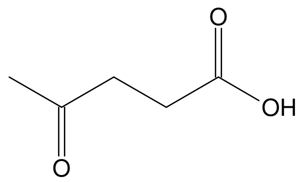
24 4-Trichlor-acetessigsäure-ethylester (Treffer)
25 Pimelindiimidsäurediethylester



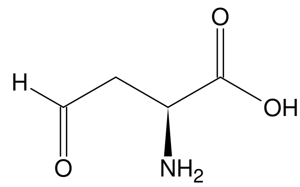
26 Pimelinsäureester (Treffer)
27 Pimelindiimidsäuredimethylester (Treffer)
28 Pimelinsäuredinitril (Treffer)
29 2-Methyl-2-penten



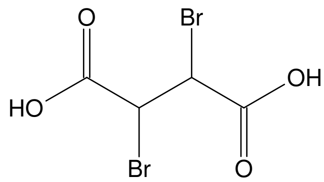
30 Lävulinsäure



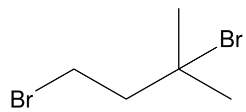
31 α-Angelicalacton (Treffer)
32 Methylen-Verbindung (γ-Methylen-γ-butyrolacton) (Treffer)
33 Asparaginsäure-semialdehyd



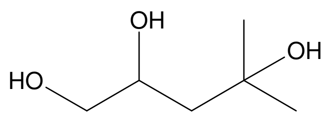
34 *L*-Homserin (Treffer)
35 2-Amino-4-pentensäure (Treffer)
36 2,3-Dibrombernsteinsäure



37 (Pyrimidin-Derivat) 5-Amino-2,4-dihydroxypyrimidin-6-carbonsäure (5-2,6-4) (ausgeschlossener Treffer)
38 (Hydrazin-Derivat1) (ausgeschlossener Treffer)
39 (Hydrazin-Dervat2) 4,5-Dichlor-1-phenylhydrazin-6-on (ausgeschlossener Treffer)
40 2,4-Dibrom-2-methyl-butan

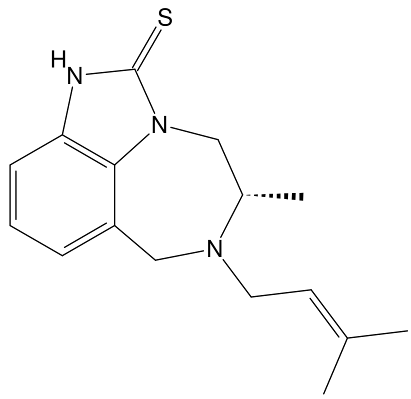


41 3-Hydroxy-3-methyl-but-1-en (Treffer)
42 (Triol) 1,2,4-Hydroxy-4-methyl-pentan

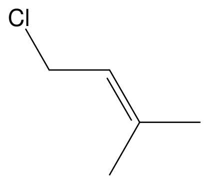


43 2-Methyl-2,4-pentandiol (Treffer)
44 3,4-Dihydro-2,2-dimethyl-γ-pyron-6-carbonsäure (Treffer)
45 Butylester (3,4-Dihydro-2,2-dimethyl-*g*-pyron-6-carbonsäurebutylester) (Treffer)

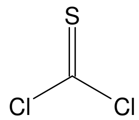
46 TIBO (4,5,6,7-Tetrahydroimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-thion TIBO R82150)



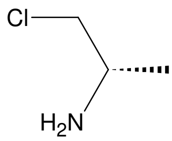
47 Prenyl-Vorstufe



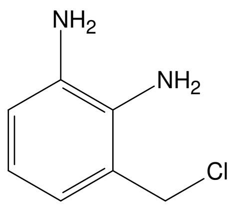
48 Thiophosgen-Vorstufe



49 chirale C_3-Einheit (Alanyl-Vorstufe)

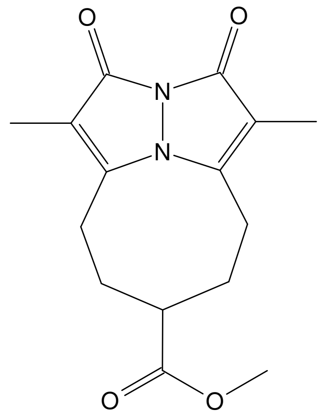


50 Benzyl-Baustein

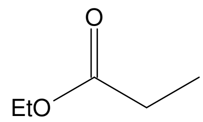


51 Prenyl-Bromid (3,3-Dimethylallylbromid, 1-Brom-3-methyl-pent-2-en) (Treffer)
52 Schwefelkohlenstoff (Treffer)
53 Phosgen (Treffer)
54 Chlorameisensäure-trichlormethylester (Treffer)
55 Brenztraubensäure (Treffer)
56 *L*-Alanin (Treffer)
57 *D*-Alanin (Treffer)
58 *DL*-Alanin (Treffer)
59 R(+)-Propylenoxid (Treffer)
60 S(-)-Propylenoxid (Treffer)
61 2-Chlor-3-nitro-benzoesäure (Treffer)
62 2-Methyl-6-nitro-anilin (Treffer)

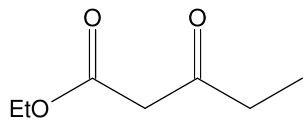
63 Enolform
64 Ketoform
65 Biman-Derivat



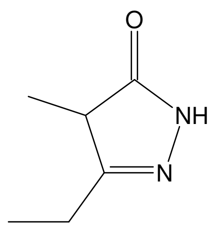
66 Propansäureester



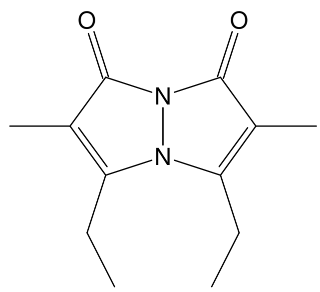
67 β-Ketoester



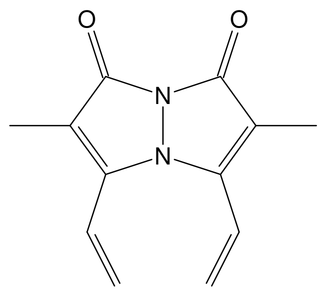
68 3-Ethyl-4-methyl-5-pyrazolinon



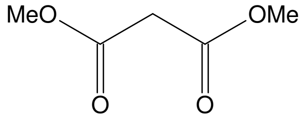
69 4,6-Diethyl-3,7-dimethyl-biman



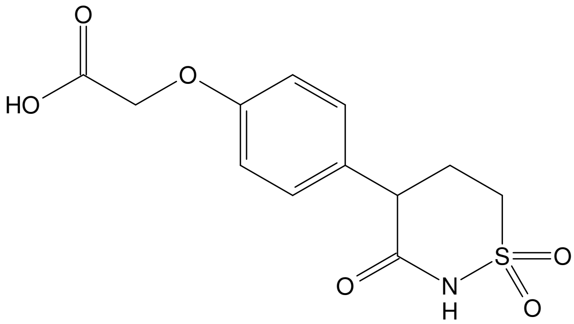
70 Michael-Olefin Biman



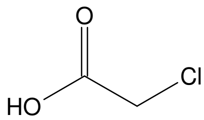
71 Malonsäureester



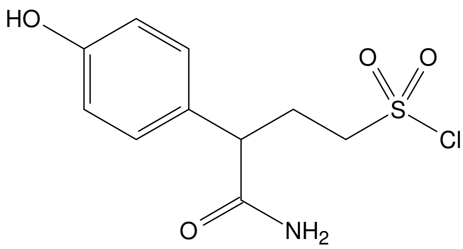
72 δ-Sultam-Derivat



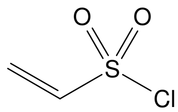
73 C_2-Baustein (Chloressigsäure)



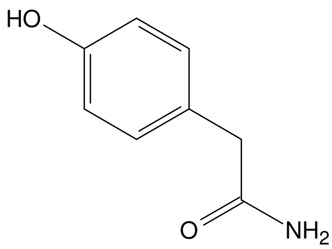
74 2. Vorstufe (Michael-Produkt)



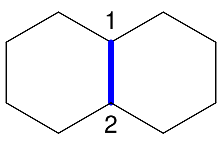
75 Bromessigsäure (Treffer)
76 Vinylsulfonylchlorid



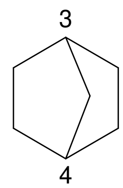
77 4-Hydroxy-phenylessigsäureamid



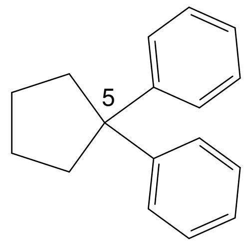
78 4-Hydroxy-phenylessigsäure-methylester (Treffer)
79 Dekalin



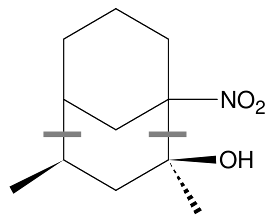
80 Norbonan



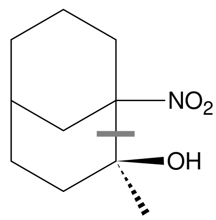
81 Cyclopentan-Derivat



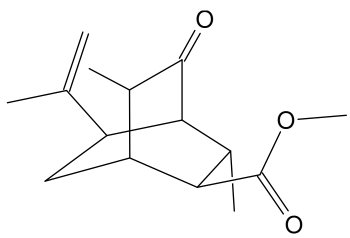
82 Negativ-Beispiel für Regel 6a



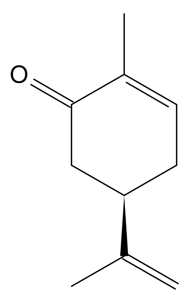
83 Positiv-Beipiel für Regel 6b



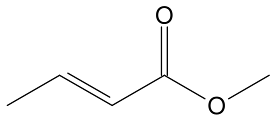
84 Bicyclo[3.2.1]octan-Derivat (Corey-Beispiel)
85 Anwendungsbsp. Bicyclo[2.2.2]octan-Derivat



86 (-)-Carvon



87 Crotonsäuremethylester (Methyl-methacrylat)



88 (-)-Carvyl-acetat (Treffer)