



„Entwicklung und Implementierung einer Prozedur zur
automatischen Parametrisierung des MOMO-Kraftfeldes“

Dissertation zur Erlangung des Doktorgrades der
Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie
der Goethe-Universität in Frankfurt am Main

von
Guido Wagner

aus Frankfurt am Main

Frankfurt am Main 2009
(D30)

vom Fachbereich Biochemie, Chemie und Pharmazie der
Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Dieter Steinhilber

Gutachter : 1. Prof. Dr. Ernst Egert

2. Prof. Dr. Gisbert Schneider

Datum der Disputation : 15.12.2009

Die vorliegende Arbeit wurde im Zeitraum vom Mai 2004 bis Dezember 2008 am Fachbereich Biochemie, Chemie und Pharmazie der Goethe-Universität in Frankfurt am Main unter Anleitung von Herrn Prof. Dr. Ernst Egert angefertigt.

Mein ganz besonderer Dank gilt Herrn Prof. Dr. Ernst Egert für die Überlassung des interessanten Themas, die sehr guten Arbeitsbedingungen, die zahlreichen Anregungen und Diskussionen sowie sein stetes Interesse am Fortgang der Arbeit.

Mein weiterer herzlicher Dank gilt insbesondere:

- Allen Mitgliedern des Arbeitskreises für das außergewöhnlich gute Arbeitsklima
- Meinen Eltern für die Unterstützung während der Promotion
- Den Beteiligten der Open Source Softwareprojekte, die als Grundlage diese Arbeit erst ermöglicht haben

Inhaltsverzeichnis

1 Einleitung.....	1
2 Aufbau von Kraftfeldern.....	5
3 Aufbau des MOMO-Kraftfeldes.....	7
3.1 Bindende Wechselwirkungen.....	7
3.1.1 Das Bindungslängendeformationspotenzial.....	7
3.1.2 Das Bindungswinkeldeformationspotenzial.....	8
3.1.3 Das Torsionswinkeldeformationspotenzial.....	12
3.1.4 Out-of-plane-Winkel.....	16
3.2 Nichtbindende Wechselwirkungen.....	18
3.2.1 Wasserstoffbrückenbindungen im MOMO-Kraftfeld.....	18
3.2.2 Das Van-der-Waals-Potenzial.....	23
3.2.3 Das Coulomb-Potenzial.....	25
3.3 Ladungsberechnung.....	26
3.3.1 SCF-Rechnung.....	26
3.3.2 Ladungsberechnung im σ -Gerüst.....	27
3.3.2.1 Die PEOE-Methode von Gasteiger und Marsili.....	27
3.3.2.2 ABR27.....	29
3.4 Die Standardbildungsenthalpie.....	33
4 Die Parametrisierung von Kraftfeldern.....	36
4.1 Auswahl der Referenzdaten.....	37
4.2 Abschätzen vorläufiger Parameter.....	40
4.3 Optimierung der Parameter.....	41
5 Optimierungsverfahren.....	42
5.1 Übersicht.....	42
5.2 Evolutionäre Algorithmen.....	45
5.3 Die genetischen Operatoren.....	47
5.4 Evolutionsstrategische Algorithmen.....	55
5.4.1 Der (1+1) evolutionsstrategische Algorithmus.....	55
5.4.2 (m, n) evolutionsstrategische Algorithmen.....	56
5.4.3 (m+n) evolutionsstrategische Algorithmen.....	56
5.5 Genetische Algorithmen.....	56
5.6 Verfeinerungstechniken.....	57
5.7 Probleme bei der automatischen Parametrisierung.....	58
5.8 Mehrkriterielle Optimierung.....	59
6 Das MOMO-Programmsystem.....	62
6.1 Die MOMO-Versionen.....	63
6.2 Anforderungen an Bibliotheken zum Molecular Modelling.....	64
6.3 Das Molecular Modelling Toolkit.....	65

6.4 Die Implementierung von MOMO für MMTK.....	68
6.4.1 Vorüberlegungen.....	68
6.4.2 Aufbereitung des molekularen Systems für das MOMO-Kraftfeld.....	69
6.5 Verwendete Hilfsprogramme.....	79
7 Angewandter Teil.....	85
7.1 Parametrisierung des MOMO-Wasserstoffbrückenpotenzials.....	85
7.1.1 Referenzdaten für Wasserstoffbrücken.....	86
7.1.2 Energieberechnung von Wasserstoffbrückenbindungen.....	86
7.1.3 Testrechnungen mit verschiedenen Ab-initio-Methoden.....	87
7.1.4 Die untersuchten Wasserstoffbrücken-Typen.....	89
7.1.4.1 C–H...O= - Wasserstoffbrücken.....	90
7.1.4.2 C–H...O- -Wasserstoffbrücken.....	97
7.1.4.3 C–H...N= -Wasserstoffbrücken.....	103
7.1.4.4 O–H...N(Aromat)-Wasserstoffbrücken.....	110
7.1.5 Parametrisierungsprozedur für Wasserstoffbrücken.....	117
7.1.6 Automatische Generierung der Referenzdaten.....	117
7.1.7 Die Parametrisierung des Wasserstoffbrückenpotenzials.....	120
7.2 Parametrisierung von atomaren Punktladungsmodellen.....	130
7.2.1 Atomare Punktladungsmodelle.....	130
7.2.2 Anforderungen an Ladungsmodelle als Referenz.....	132
7.2.3 Zusammenstellung der Referenzdaten.....	133
7.2.4 Der Referenzdatensatz.....	137
7.2.5 Die Codierung (XML-Tags und Vergleichsanweisungen).....	149
7.2.6 Analyse des Testdatensatzes mit den verfügbaren Ladungsmodellen.....	150
7.2.7 Vergleich der Ladungsmodelle MK und NBO.....	151
7.2.8 Vergleich der Ladungsmodelle GM und ABR27.....	157
7.2.9 Die Parametrisierung der Punktladungsmodelle.....	164
7.2.10 Parametrisierung des Gasteiger-Marsili-Modells.....	164
7.2.11 Parametrisierung des ABR27-Modells.....	185
7.3 Parametrisierung der bindenden Kraftfeldterme.....	190
7.3.1 Speicherung der Referenzdaten.....	190
7.3.2 Der Referenzdatensatz.....	191
7.3.3 Das Parametrisierungsprogramm.....	193
7.3.4 Evaluierung des Kraftfeldes und Ergebnisse der Parametrisierung.....	196
8 Zusammenfassung und Ausblick.....	200

Abkürzungsverzeichnis

ABR27	modifiziertes Abraham-Modell von S. Monz
AO	Atomorbital
B3LYP	(Becke, three-parameter, Lee-Yang-Parr) exchange-correlation functional
BSSE	basis set superposition error
CI	configuration interaction
CML	chemical markup language
Coul.	Coulomb
DFT	Dichtefunktional-Theorie
<i>e</i>	Elementarladung
EN	Elektronegativität
ESP	elektrostatisches Potenzial
FEOE	full equalization of orbital electronegativity
GM	Modell von Gasteiger und Marsili
HB	Wasserstoffbrücke
HF	Hartree Fock
HTML	hypertext markup language
IR	infrarot
LCAO-MO	linear Combination of Atomic Orbitals to Molecular Orbitals
MAD	mittlere absolute Abweichung
MD	molecular dynamics
MK	Merz-Kollman(-Verfahren)
MMTK	das Molecular Modelling Toolkit
MOMO	Kraftfeldprogramm
MP2	Møller-Plesset 2. Ordnung
NBO	natural bond orbital(-Methoden)
NFL	no free lunch
OOP	out-of-plane
PEOE	partial equalization of orbital electronegativity
PIMM	Kraftfeldprogramm
R-Wert	Korrelationskoeffizient
RMSD	root mean square
SCF	Methode des selbstkonsistenten Feldes
SEE	Standardabweichung des Schätzwertes

STO	Slater-type orbitals
SVD	Singulärwertzerlegung
VdW	van-der-Waals
ZDO	zero differential overlap
XML	extensible markup language

1 Einleitung

Modelle spielen allgemein eine zentrale Rolle für unser Verständnis der Chemie, zu den einfachsten davon gehören mechanische Molekülbaukästen. Der Computereinsatz eröffnet neue, wichtige Möglichkeiten zur Modellierung von Molekülen [Leach2001]: Die Computermodelle basieren auf mathematischen Gleichungen, die die Positionen der Atomkerne und die elektronischen Eigenschaften der Moleküle vorhersagen. Dabei können unerwartete Wechselwirkungen, Elektronendichteverteilungen, Orbitalüberlappungen oder sterische Einflüsse direkt sichtbar gemacht werden. Man unterscheidet zwischen klassischen mechanischen und quantenmechanischen Verfahren. Die mechanischen Modelle fassen Moleküle vereinfacht als Gebilde aus Kugeln (Atomkerne) und Federn (Bindungen oder andere Wechselwirkungen) auf. Informationen über die „Steifigkeit“ der Federn und die Größe der Kugeln werden genutzt, um deren beste Anordnung (die mit der geringsten Energie) zu finden.

Entscheidend ist hier die Erkenntnis, in welchen Bereichen das Modell zuverlässige Daten liefert, denn ansonsten läuft man Gefahr, an der Realität vorbei zu simulieren.

Über die Anwendbarkeit eines Rechenmodells bestimmen sowohl die hineingesteckten Gleichungen als auch deren Parameter. Gerade bei Kraftfeldern für das Molecular Modelling gibt es zwischen den Beiträgen der verschiedenen Gleichungen Wechselwirkungen, die durch geeignete Parameter ausbalanciert werden müssen. Teile eines Kraftfeldes lassen sich daher nur eingeschränkt isoliert voneinander ändern. Nicht immer beschreibt solch ein Modell die Wirklichkeit genau genug, um mit nur einem Satz Parameter universell anwendbar zu sein. Dies hat zur Folge, dass die Qualität der Modellierung in einem Anwendungsgebiet ggf. nur auf Kosten eines anderen verbessert werden kann. Lässt sich ein Modell an sich nicht optimieren, weil dies z. B. nicht hinnehmbare Nachteile bei der Rechengeschwindigkeit nach sich zöge, so kann man sich durch genaue Kenntnis dieser Grenzen und mehrere spezielle Parametersätze helfen.

Diese Eigenschaften der Modelle gestalten die Entwicklung von Kraftfeldern schwierig. Methoden, die bei deren Parametrisierung und Evaluierung unterstützen, sind deshalb sehr wichtig. In der Literatur ist eine Vielzahl unterschiedlichster Optimierungsverfahren bekannt [NumRecipes]. Werden diese geschickt kombiniert und angewendet, können mit Computerunterstützung Parameter nach bestimmten Qualitätskriterien verbessert und nicht miteinander verträgliche Anwendungsgebiete identifiziert werden. Für die Kraftfeldentwicklung bringt dies enorme Vorteile.

Die Auswirkungen einer Änderung am Kraftfeld können so umfassend sichtbar gemacht werden, und die Optimierungsverfahren vereinfachen die aufwändige Ausbalancierung der Parametersätze.

Eine solche Automatisierung zwingt den Benutzer zudem, sorgfältig Vorgaben und Qualitätskriterien zu definieren, da Optimierungsstrategien oftmals überraschende Schlupflöcher in der ihnen vorgegebenen Spezifikation ausnutzen, und fördert so ein tieferes Verständnis.

Das Programm MOMO [Beck1989] wird seit 1986 in unserem Arbeitskreis entwickelt. Es fußt auf dem PIMM85-Kraftfeld und dient dazu, eine Umgebung für spezielle Erweiterungen bereitzustellen. Dazu gehören z. B. eine Potenzialfunktion für Wasserstoffbrücken, Modelle zur Berechnung atomarer Punktladungen, die Berechnung supramolekularer Komplexe oder die Behandlung elektrostatischer Wechselwirkungen durch Multipolmomente. Immer wieder war es notwendig, das Programm an neue Rechnergenerationen, Betriebssysteme und Programmierwerkzeuge anzupassen. Die letzte große Anpassung war die von mir durchgeführte Portierung auf Linux während meiner Diplomarbeit. Dabei blieb der bisherige Aufbau inklusive der grafischen Benutzeroberfläche erhalten. Das Programm bietet so zwar die Möglichkeit, Arbeitsabläufe durch Batchdateien zu automatisieren, allerdings sind diese Fähigkeiten für die vorliegende Arbeit bei Weitem nicht ausreichend. Darüber hinaus macht sich störend bemerkbar, dass sich Bindungen nicht explizit abspeichern lassen und die Dateien keine Zusatzinformationen aufnehmen können.

Einen Ausweg bildet die Implementierung des Kraftfeldkerns als Bibliothek für das Molecular Modelling Toolkit [Hinsen2000] von Konrad Hinsen. Dieses Toolkit besteht aus einer Sammlung von Programmen in der Skriptsprache Python [van_Rossum1995], die grundlegende Funktionen zum Umgang mit chemischen Daten in Kraftfeldprogrammen bereitstellen, und einigen aus Geschwindigkeitsgründen in der Programmiersprache C geschriebenen Bibliotheken, die u. a. das AMBER-Kraftfeld implementieren sowie Funktionen zur Energieminimierung und für die molekulare Dynamik bereitstellen. Modellierungsprobleme werden durch kurze Python-Skripte bearbeitet, die durch die umfangreichen Bibliotheken recht einfach ausfallen können. Durch die Verwendung einer „echten“ Programmiersprache ist es aber leicht möglich, auch sehr komplexe, spezielle Lösungen zu realisieren. Da Python im Bereich der Chemie- und Bioinformatik recht beliebt ist, existiert eine Vielzahl nützlicher Bibliotheken zum Zeichnen von Diagrammen, für mathematische Standardverfahren, zum Umgang mit speziellen Dateiformaten, zur Visualisierung von Molekülen und vielem mehr, die die Arbeit zusätzlich erleichtern [Langtangen2004]. Von

diesen Möglichkeiten wird in der vorliegenden Arbeit ausgiebig Gebrauch gemacht, nebenbei wurden dadurch eine Reihe von Fehlern in der Kraftfeldimplementierung aufgedeckt, die sich nur unter speziellen Bedingungen auswirken und daher bisher unentdeckt geblieben sind.

Um Daten über Moleküle sinnvoll für Parametrisierungsprogramme nutzen zu können, ist es nicht mehr ausreichend, nur eine Liste von Atomkoordinaten zu speichern. Hier werden Zusatzinformationen wie eine Liste der Bindungen und verschiedene Referenzen wie experimentelle Daten oder Ergebnisse aus quantenmechanischen Rechnungen benötigt. Ein weit verbreitetes universelles Dateiformat ist XML [Bray], eine erweiterbare Auszeichnungssprache für hierarchisch strukturierte Daten in Form von Textdateien – für die Problemstellung also genau das Richtige. Hier können alle Daten sowohl für den Menschen als auch für den Rechner lesbar abgespeichert werden, und Programmbibliotheken für XML ermöglichen eine einfache und effiziente Implementierung. Eine XML-basierte Sprache ist die CML (Chemical Markup Language) [Murray-Rust1999], für die Bedürfnisse von MOMO sind allerdings so viele Erweiterungen und Änderungen nötig, dass das Dateiformat den Namen CML nicht mehr verdient.

Für den eigentlichen Parameterisierungsvorgang werden etablierte Optimierungsverfahren herangezogen, die sich speziell an der Problemstellung orientieren. Besonders vielseitig erscheint hier die Klasse der evolutionären Algorithmen. Möchte man einen Vorteil gegenüber den Standardmethoden erzielen, so muss man spezielles Wissen über das bearbeitete Optimierungsproblem nutzen. Für den speziellen Fall kann so die Leistung verbessert werden, dies geht allerdings nur, wie es das NFL (no free lunch) Theorem [Wolpert1997] beschreibt, auf Kosten des allgemeinen Falles. Es besteht die Gefahr, dass das hineingesteckte „Expertenwissen“ Fehler enthält, allerdings ist man bei Problemen mit einem sehr großen Suchraum auf solche Verfahren angewiesen. Die Bestimmung von Kraftfeldparametern ist ein mehrkriterielles Optimierungsproblem, d. h. es müssen mehrere Kriterien gut erfüllt werden, die sich sogar widersprechen können. Hierbei ergeben sich viele Lösungen, bei denen sich ein Kriterium nicht besser erfüllen lässt, ohne woanders eine Verschlechterung in Kauf zu nehmen. Dies kann der Fall sein, wenn verschiedene Verbindungsklassen Atomtypen im Kraftfeld gemeinsam haben. Eine solche Lösung nennt man Pareto-optimal [Fonseca1995], alle diese Lösungen bilden im Suchraum die Paretofront. Ein Optimierungsalgorithmus kann den Verlauf dieser Paretofront annähern, die Abwägung der möglichen Kompromisslösungen bleibt Sache des Anwenders.

1 Einleitung

In der vorliegenden Arbeit werden so Potenzialfunktionen für Wasserstoffbrückenbindungen, empirische Modelle für atomare Punktladungen und Potenzialfunktionen für interne Koordinaten optimiert. Ein großes Problem dabei ist die eingeschränkte Datenbasis, deren Vergrößerung auf jeden Fall wünschenswert ist. Trotzdem steht nun eine funktionierende Sammlung von Tools zur Parametrisierung und Evaluierung des MOMO-Kraftfeldes zur Verfügung, die sich durch Verwendung der bereits erwähnten Programmierumgebung leicht auf neue Potenzialfunktionen oder gar ganze Kraftfelder ausdehnen lässt und so die anfangs geschilderten Probleme bei der Kraftfeldentwicklung lindert.

2 Aufbau von Kraftfeldern

Ein Kraftfeldprogramm für das Molecular Modelling berechnet die Energie eines chemischen Systems in Abhängigkeit von seiner Geometrie. Dies erfolgt meist durch eine Zerlegung in interne Koordinaten mit anschließendem Aufsummieren verschiedener klassisch physikalischer Wechselwirkungsenergien zu einer Gesamtenergie, die zur Beurteilung der relativen Stabilität verschiedener Konformationen und Konfigurationen dient. Jede dieser Wechselwirkungen wird separat als sogenannter Term behandelt. In eine solche Gleichung werden die internen Koordinaten der Wechselwirkung und deren Kraftkonstanten eingesetzt. Die partiellen Ableitungen der Terme nach den Atomkoordinaten dienen schließlich dazu, die Energieminimierung (Bewegung der Atomkerne entlang der Gradienten) und die molekulare Dynamik (Integration der Newtonschen Bewegungsgleichungen) zu ermöglichen. Die entsprechenden zweiten Ableitungen können zur Vorhersage von Schwingungsspektren des Moleküls genutzt werden. Man kann eine grobe Einteilung der Terme in bindende und nichtbindende Wechselwirkungen vornehmen.

Sogenannte Kreuzterme koppeln interne Koordinaten miteinander. Verbreitet ist z. B. die Kopplung eines Bindungswinkels mit den angrenzenden Bindungen. Die gebräuchlichsten Kreuzterme beziehen sich auf zwei interne Koordinaten, es gibt aber auch Beispiele mit mehr gekoppelten Koordinaten. Je nachdem, welche Arten von Termen in einem Kraftfeld enthalten sind, ordnet man es einer bestimmten Klasse zu [Leach2001].

- *Klasse I:* Enthält nur zueinander orthogonale quadratische Terme
- *Klasse II:* Enthält auch nicht harmonische funktionale Formen und Kreuzterme zwischen mehreren internen Koordinaten.
- *Klasse III:* Hier werden zusätzlich noch chemische Eigenschaften und Effekte berücksichtigt. Dies können z. B. Hyperkonjugation und Elektronegativität sein.

Tabelle 1: Einteilung von Kraftfeldern in Klassen

Addiert man die Energiewerte aller Terme eines Kraftfeldes, so erhält man eine „sterische Energie“, die als Energiedifferenz zu einer hypothetischen Konformation und Konfiguration, in der die Koordinaten für alle Kraftfeldterme ihre Idealwerte einnehmen, aufgefasst werden kann. So wird die Berechnung relativer Energien zwischen verschiedenen Konfigurationen und Konformationen einer Verbindung im Kraftfeld möglich.

2 Aufbau von Kraftfeldern

Einen zwischen verschiedenen Kraftfeldern, ab-initio-Verfahren und Experimenten vergleichbaren Wert stellt die Standardbildungsenthalpie dar. Um sie zu berechnen, müssen die Bildungsenthalpien des Bindungsgerüsts zur sterischen Energie des Moleküls addiert werden. Der Beitrag des Bindungsgerüsts wird als Summe konstanter Enthalpie-Beiträge aus submolekularen Bauelementen gedeutet [Quinkert1995]. Die Werte der einzelnen Beiträge lassen sich empirisch durch den Vergleich der Verbrennungs- oder Hydrierungsenthalpien und der konsequenten Anwendung des Heßschen Satzes („Die Standardenthalpie einer Reaktion ist gleich der Summe der Standardenthalpien einer Folge von Reaktionen, in die sich die betreffende Reaktion formal zerlegen lässt“ [Atkins2001]) bestimmen.

Ein dritter in Kraftfeldern üblicher Energiewert ist die „Spannungsenergie“. Mit ihr können verschiedene Moleküle in einem Kraftfeld miteinander verglichen werden. Hier wird von der sterischen Energie eines Moleküls dessen Energie in einem spannungslosen Referenzzustand abgezogen.

Die Zusammensetzung eines typischen Kraftfeldes für das Molecular Modelling soll im Folgenden anhand des MOMO-Kraftfeldes besprochen werden.

3 Aufbau des MOMO-Kraftfeldes

Das MOMO-Kraftfeld wird hier sehr detailliert besprochen, da diese Einzelheiten für dessen Parametrisierung bedeutend sind. Die genaue Funktion wurde durch Analyse des original Quellcodes gewonnen, die im Rahmen dieser Arbeit neu implementierte Version des Kraftfeldes arbeitet genauso. Die sterische Energie setzt sich im MOMO-Kraftfeld additiv aus folgenden Termen zusammen:

$$E_{ster} = \sum_r E_r + \sum_\theta E_\theta + \sum_\omega E_\omega + \sum_{vdw} E_{vdw} + \sum_{es} E_{es} + \sum_{HB} E_{HB} + \sum_{oop} E_{oop}$$

E_{ster} Sterische Energie

Bindende Wechselwirkungen:

E_r Bindungslängendeformationsenergie
 E_θ Bindungswinkeldeformationsenergie
 E_ω Torsionsenergie
 E_{oop} Energie für „out of plane“ Deformation

Nichtbindende Wechselwirkungen:

E_{vdw} Van-der-Waals Energie
 E_{es} Elektrostatische Energie
 E_{HB} Wasserstoffbrückenbindungsenergie

Formel 3.1: Energierme des MOMO-Kraftfeldes

3.1 Bindende Wechselwirkungen

3.1.1 Das Bindungslängendeformationspotenzial

Die Energie einer Bindung wird durch das Morse-Potenzial beschrieben. Dieses kann im Kraftfeld durch eine harmonische Funktion angenähert werden, da Bindungen dem Programm bei der Eingabe vorgegeben werden. Auf diese Weise werden durch das Programm keine Bindungen geknüpft oder gebrochen. MOMO unterscheidet σ - und π -Bindungen. Die Bindungslänge einer π -Bindung hängt linear von der Bindungsordnung ab, welche mit der in Kapitel 3.3.1 beschriebenen π -SCF-Methode bestimmt wird.

$$E_r = k_r (r - r_0)^2$$

k_r	Kraftkonstante der Bindungslängendeformation
r	Bindungslänge
r_0	Idealwert der Bindungslänge

Formel 3.1.1: Bindungslängendeformationspotenzial für σ -Bindungen

Besteht die Bindung zwischen entgegengesetzt geladenen Ionen, so wird r_0 mit einem Korrekturfaktor multipliziert. Für π -Bindungen gilt:

$$r_0 = a - bp_{ij}$$

a, b	Parameter
p_{ij}	π -Bindungsordnung

Formel 3.1.2: Bindungslängendeformationspotenzial für π -Bindungen

Für den Fall, dass die Atome der Bindung genau vier π -Substituenten haben, wird r_0 um einen empirischen Parameter $r_{IS,JS}^{inc}$ inkrementiert.

Die Gradienten dieses Potenzials zwischen den Atomen I und J berechnen sich wie folgt:

$$E'_r = k_r \cdot 2(r - r_0) \quad (3.1.3)$$

$$\overrightarrow{grad}_I = E'_r \frac{\overrightarrow{IJ}}{r_{ij}} \quad ; \quad \overrightarrow{grad}_J = -\overrightarrow{grad}_I \quad (3.1.4)$$

3.1.2 Das Bindungswinkeldeformationspotenzial

Die Bindungswinkeldeformation kann in der Nähe des Gleichgewichts durch ein harmonisches Potenzial charakterisiert werden. Bei großen sterischen Spannungen kann es zu erheblichen Deformationen der Bindungswinkel kommen, hier müssen anharmonische Effekte berücksichtigt werden; dies lässt sich durch Polynome höheren Grades erreichen, z. B. bei der Darstellung der Potenzialfunktion als Taylor-Reihe (3.1.5).

$$E_\theta = E_0 + k_1(\theta - \theta_0) + k_2(\theta - \theta_0)^2 + k_3(\theta - \theta_0)^3 + \dots \quad (3.1.5)$$

3 Aufbau des MOMO-Kraftfeldes

Im Kraftfeldmodell werden E_0 und k_1 generell gleich Null gesetzt. Ungerade Polynome haben den Effekt, dass die Energie bei zunehmender Abweichung vom Idealwert geringer wird. Verwendet man sie, so darf dies nur in der Nähe des Idealwertes geschehen. Beim Wechsel auf nicht invertierbare Potenziale muss dafür gesorgt werden, dass die resultierende Funktion stetig differenzierbar bleibt. Dazu wird im Übergangsbereich ein Bruchteil genutzt, der die Anteile linear zwischen 0 und 100% ändert. MOMO verwendet in der Nähe von θ_0 einen Bindungswinkeldeformationsterm mit negativem Glied 4. Grades (3.1.6), bei größeren Abweichungen wird nur der quadratische Term benutzt.

$$E_\theta = k_\theta(\theta - \theta_0)^2 - k'_\theta(\theta - \theta_0)^4 \quad (3.1.6)$$

Der Übergang zwischen den beiden Funktionen in 3.1.7 wird durch eine Switchfunktion (3.1.8) realisiert:

$$E_\theta = k_\theta(\theta - \theta_0)^2 - k'_\theta(\theta - \theta_0)^4 \cdot (3 \cdot \text{switch}^2 - 2 \cdot \text{switch}^3) \quad (3.1.7)$$

$$\text{switch} = \frac{\Delta_{\text{umsch}} - \Delta_\theta}{10}$$

Δ_{umsch}

Winkelabweichung im Bogenmaß, ab der nur das quadratische Potenzial genutzt werden soll

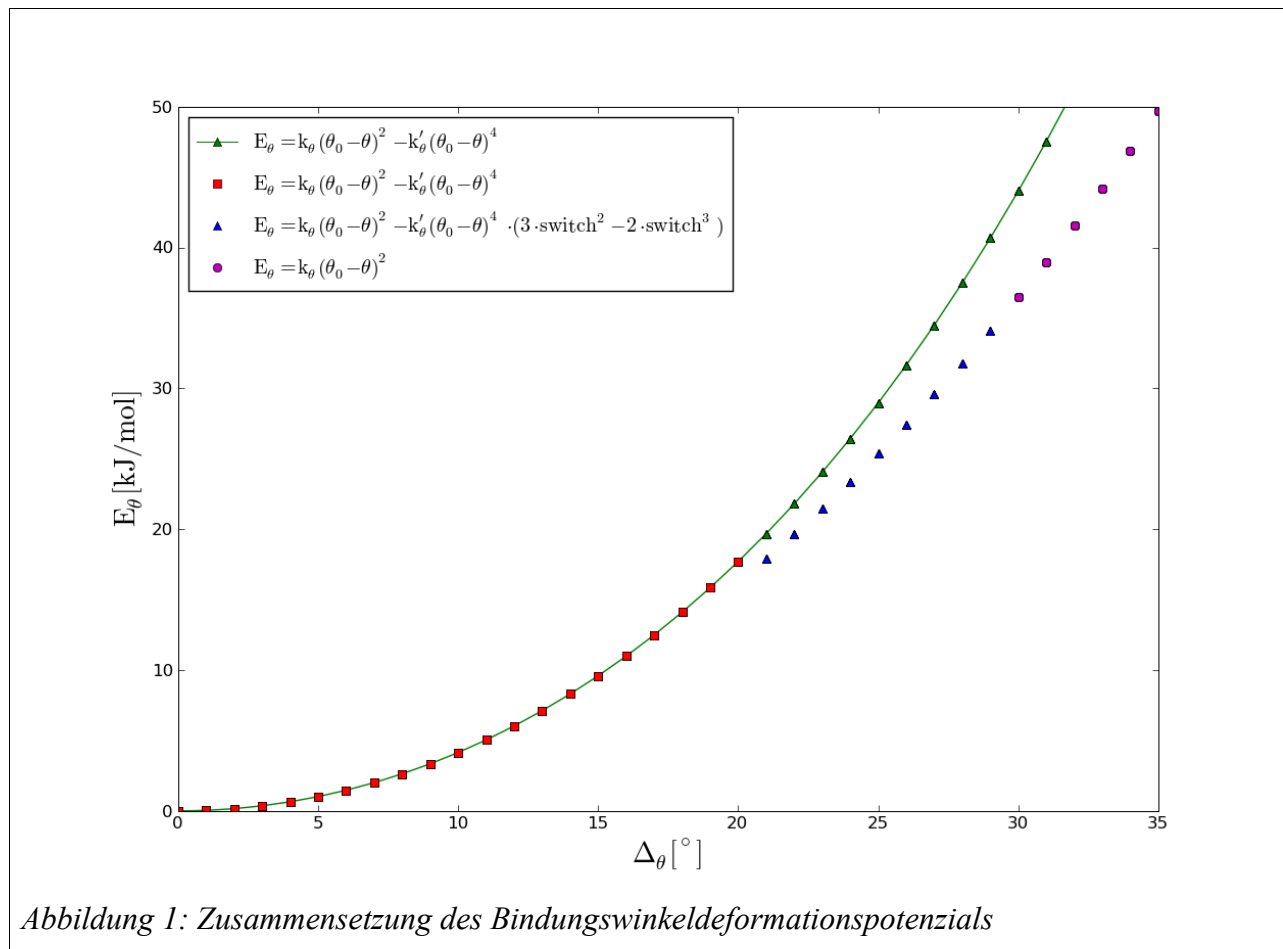
Δ_θ

Betrag der Abweichung vom Idealwinkel im Bogenmaß

Formel 3.1.8: Switchfunktion des Bindungswinkeldeformationspotenzials

Der Anteil des kubischen Gliedes wird nach Formel (3.1.8) bestimmt. Leider gibt es mit der Umschaltfunktion zurzeit noch Probleme, sodass eine stetige Differenzierbarkeit nicht gewährleistet ist. In der aktuellen Version wird deshalb nur die mit dem negativen Glied 4. Grades benutzt. Nimmt man die Parameter aus einer älteren MOMO-Version (V1.0), so ergibt sich am Beispiel des sp^2 -Kohlenstoffs der Funktionsgraph aus Abb. 1.

MOMO verwendet für einige Spezialfälle besondere Standardbindungswinkel. Es können bestimmte Winkel durch drei Atomtypen definiert werden, für die dann ein quadratisches Potenzial



mit Sonderparametern genutzt wird. Für jeden Atomtyp kann unterschieden werden, ob beide Substituenten Wasserstoffe sind oder der Winkel Teil eines kleinen Rings (3-, 4- oder 5-Ring) ist. Wenn eine dieser beiden Bedingungen erfüllt ist, kann ein alternativer Standardbindungswinkel vorgegeben werden. Hiervon wird bisher nur beim sp^3 -Kohlenstoff Gebrauch gemacht. Für dreifach koordinierte Stickstoffatome wird der Idealwinkel θ_0^N in Abhängigkeit von der π -Ladungsdichte nach Formel (3.1.9) berechnet.

$$\theta_0^N = 120^\circ - 100^\circ (P - q_N^0)$$

P π -Ladungsdichte

Formel 3.1.9: Modifizierung des Standardbindungswinkels

3 Aufbau des MOMO-Kraftfeldes

So können mit nur einem Atomtyp pyramidal koordinierter Stickstoff in Aminen und trigonalplanar koordinierter in Amiden unterschieden werden. Diese Differenzierung kann prinzipiell für jeden Atomtyp durch entsprechende Parameter vollzogen werden.

Nun muss noch die erste Ableitung von E_θ (Formel 3.1.7) bestimmt werden:

$$E'_\theta = k_\theta \cdot 2(\theta - \theta_0) + k'_\theta \cdot 3 \cdot (\theta - \theta_0)^2 \cdot \text{switch} \quad (3.1.10)$$

Damit können die Gradienten, wie in Abb. 2 illustriert, berechnet werden:

- Die drei den Bindungswinkel definierenden Atome sind durch die Punkte I, J und K symbolisiert. Als Erstes wird der senkrecht auf der von ihnen gebildeten Ebene stehende Vektor \vec{dI} berechnet:

$$\vec{dI} = \vec{IJ} \times \vec{IK}$$

- Mithilfe von \vec{dI} lassen sich nun \vec{dIij} und \vec{dIik} berechnen, die senkrecht auf der von \vec{dI} und \vec{IJ} (rot getönt) bzw. \vec{dI} und \vec{IK} (blau getönt) aufgespannten Ebene stehen:

$$\vec{dIij} = \frac{\vec{dI} \times \vec{IJ}}{|\vec{dI}| \cdot |\vec{IJ}|^2}$$

$$\vec{dIik} = \frac{\vec{dI} \times \vec{IK}}{|\vec{dI}| \cdot |\vec{IK}|^2}$$

- Jetzt können die auf die Atome wirkenden Kraftvektoren ermittelt werden:

$$\vec{DJ} = \vec{dIij} \cdot E'_\theta$$

$$\vec{DK} = \vec{dIik} \cdot E'_\theta$$

$$\vec{DI} = -(\vec{dIij} + \vec{dIik}) \cdot E'_\theta$$

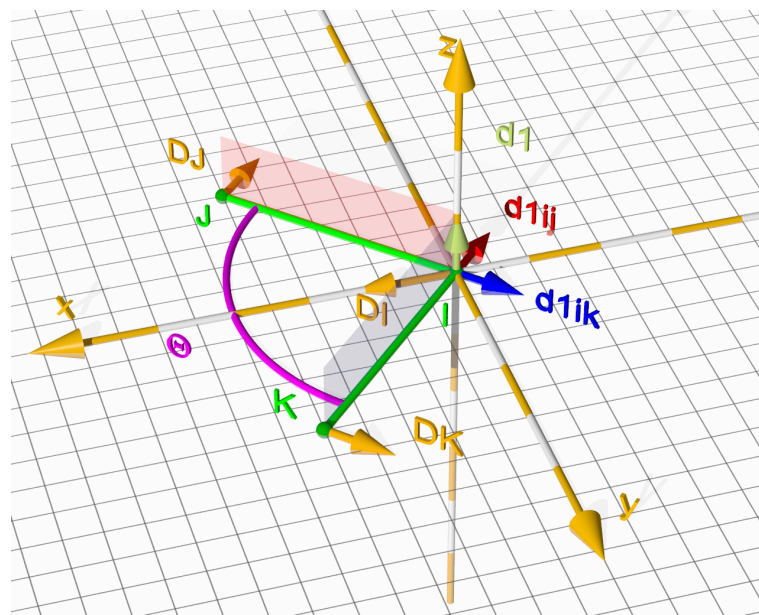


Abbildung 2: Hilfsvektoren zur Berechnung des Bindungswinkeldeformationspotenzials

3.1.3 Das Torsionswinkeldeformationspotenzial

Betrachtet man die Potenziale für Bindungslängen- und Bindungswinkeldeformation, so fällt auf, dass dort für eine signifikante Änderung der Molekülgeometrie sehr hohe Energien aufgewendet werden müssen. Die Energieunterschiede verschiedener Konformationen sind deshalb im Wesentlichen auf das Zusammenspiel von Torsionspotenzialen und nichtbindenden Wechselwirkungen zurückzuführen.

Das Torsionspotenzial in MOMO wird in einer Fourierreihe (3.1.11) entwickelt.

$$E_{\omega} = k_1(\cos \omega + 1) + k_2(\cos(2\omega) - 1) + k_3(\cos(3\omega) + 1) + \dots$$

k_x	Kraftkonstanten
ω	Torsionswinkel

Formel 3.1.11: Fourierreihe des Torsionspotenzials

In MOMO wird die Reihe (3.1.11) spätestens nach dem dritten Glied abgebrochen. Für σ -Bindungen wird nur das dritte Glied verwendet (3.1.13). Für P-O- und S-S- Einfachbindungen werden drei Glieder verwendet (3.1.12).

$$E_{\omega} = k_1(\cos \omega + 1) + k_2(\cos(2\omega) - 1) + k_3(\cos(3\omega) + 1) \quad (3.1.12)$$

$$E_{\omega} = k_{\sigma}^{1,4} k_{\sigma}^{2,3} (\cos(3\omega) + 1)$$

$k_{\omega}^{1,4}$ Kraftkonstante für die 1,4-Wechselwirkung
 $k_{\sigma}^{2,3}$ Kraftkonstante für die 2,3-Wechselwirkung

Formel 3.1.13: Torsionspotenzial für σ -Bindungen

π -Bindungen werden durch ein separates Torsionspotenzial (3.1.14) beschrieben.

Die Bindungsordnung p_{ij} wird nicht direkt aus der π -SCF-Rechnung übernommen, sondern nach Gleichung (3.1.15) und (3.1.16) aus den Kraftfeldparametern a , b und der π -SCF-Bindungsordnung P_{ij} berechnet.

$$p_{ij} = \frac{a - r_0}{b} \quad (3.1.15)$$

$$r_0 = a - b \cdot P_{ij} \quad (3.1.16)$$

$$E_{\omega} = k_{\pi}(\cos(2\omega) - 1) + \sin^2(\omega) \cdot (p_{ij} + p_{ij}^2 \cdot 0,15) \cdot k_4$$

k_{π} Kraftkonstante für 1,4-Wechselwirkungen bei π -Bindungen
 k_4 Kraftkonstante für 2,3-Wechselwirkungen bei π -Bindungen
 p_{ij} π -Bindungsordnung

Formel 3.1.14: Torsionspotenzial für π -Bindungen

Für den Fall, dass die Atome der zentralen Bindung des Torsionswinkels genau vier π -Substituenten haben, wird r_0 um einen von den beteiligten Atomtypen abhängigen empirischen Parameter r^{inc} inkrementiert:

$$r_0 = a - b \cdot P_{ij} + r_{IS,JS}^{inc} \quad (3.1.17)$$

a und b sind empirische Parameter des Bindungslängenpotenzials (vgl. Formel 3.1.1).

Nun muss noch die erste Ableitung berechnet werden. Für σ -Bindungen:

3 Aufbau des MOMO-Kraftfeldes

$$E'_\omega = -k_\sigma^{1,4} k_\sigma^{2,3} (\sin(3\omega)) \cdot 3 \quad (3.1.18)$$

Für P-O- und S-S- Einfachbindungen:

$$E'_\omega = -k_1 (\sin \omega) - k_2 (\sin(2\omega)) \cdot 2 - k_3 (\sin(3\omega)) \cdot 3 \quad (3.1.19)$$

Für π -Bindungen:

$$E'_\omega = -k_\pi \sin(2\omega) \cdot 2 + \cos(\omega) \cdot \sin(\omega) \cdot 2 \cdot (p_{ij} + p_{ij}^2 \cdot 0,15) \cdot k_4 \quad (3.1.20)$$

Damit lassen sich nun die Gradienten berechnen. Der Torsionswinkel und einige Hilfsvektoren werden wie folgt erzeugt:

Die Hilfsvektoren sind in Abb. 3 illustriert.

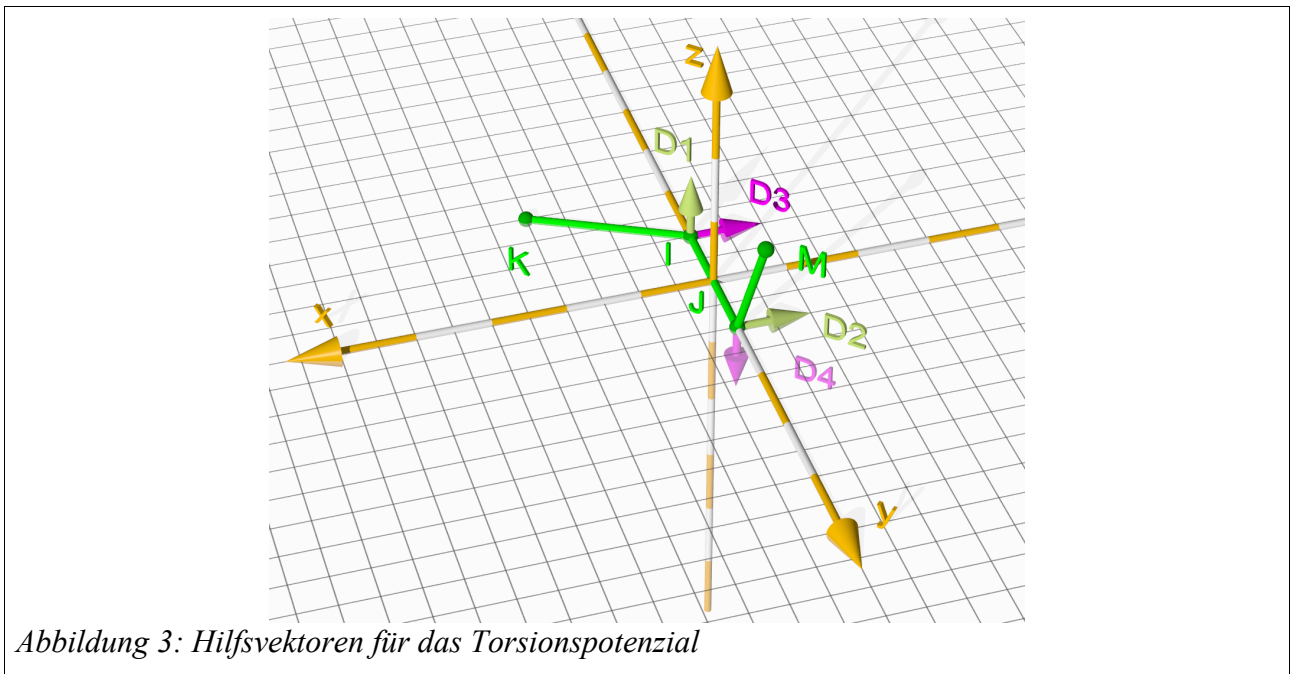


Abbildung 3: Hilfsvektoren für das Torsionspotenzial

Zuerst werden die auf den Ebenen I, J, K und I, J, M senkrecht stehenden Vektoren $\vec{D1}$ und $\vec{D2}$ berechnet:

$$\vec{D1} = \vec{KI} \times \vec{IJ} \quad \vec{D2} = \vec{IJ} \times \vec{JM} \quad (3.1.21)$$

$$\cos(\omega) = \frac{\vec{D1} \cdot \vec{D2}}{|\vec{D1}| \cdot |\vec{D2}|} \quad (3.1.22)$$

Das Vorzeichen des Winkels wird über die Händigkeit des Systems $\vec{IJ}; \vec{D1}; \vec{D2}$ bestimmt. Im linkshändigen Fall wird das Vorzeichen negativ. Dazu wird folgendes Spatprodukt bestimmt:

3 Aufbau des MOMO-Kraftfeldes

$$DET = \vec{IJ} \cdot (\vec{D1} \times \vec{D2}) \quad (3.1.23)$$

Für die folgenden Rechnungen wird der normierte Vektor $\overline{RN\vec{IJ}}$, der die Richtung der Verbindungslinie der Atome I und J angibt, bestimmt. Die Vektoren $\vec{D1}$ und $\vec{D2}$ werden durch das Quadrat ihres Betrages geteilt.

$$\overline{RN\vec{IJ}} = \frac{\vec{IJ}}{|\vec{IJ}|} ; \quad \vec{D1} = \frac{\vec{D1}}{|\vec{D1}|^2} ; \quad \vec{D2} = \frac{\vec{D2}}{|\vec{D2}|^2} \quad (3.1.24)$$

Die Ergebnisse der folgenden Kreuzprodukte $\vec{D3}$ und $\vec{D4}$ sind durch dieses Vorgehen gleich Normalvektoren.

$$\vec{D3} = \overline{RN\vec{IJ}} \times \vec{D1} \quad \vec{D4} = \overline{RN\vec{IJ}} \times \vec{D2} \quad (3.1.25)$$

Weiter werden folgende Hilfsvektoren berechnet:

$$\vec{D3\vec{IJ}} = \vec{D3} \times \vec{IJ} ; \quad \vec{D3\vec{KI}} = \vec{D3} \times \vec{KI} ; \quad \vec{D4\vec{JM}} = \vec{D4} \times \vec{JM} ; \quad \vec{D4\vec{IJ}} = \vec{D4} \times \vec{IJ} \quad (3.1.26)$$

Mit diesen Vektoren können schließlich die Gradientenvektoren (Abb. 4) generiert werden:

$$Grad_I = E'_\omega \cdot (\vec{D3\vec{IJ}} + \vec{D3\vec{KI}} + \vec{D4\vec{JM}}) \quad (3.1.27)$$

$$Grad_K = -E'_\omega \cdot \vec{D3\vec{IJ}} \quad (3.1.28)$$

$$Grad_J = E'_\omega \cdot (-\vec{D3\vec{KI}} - \vec{D4\vec{JM}} - \vec{D4\vec{IJ}}) \quad (3.1.29)$$

$$Grad_M = E'_\omega \cdot \vec{D4\vec{IJ}} \quad (3.1.30)$$

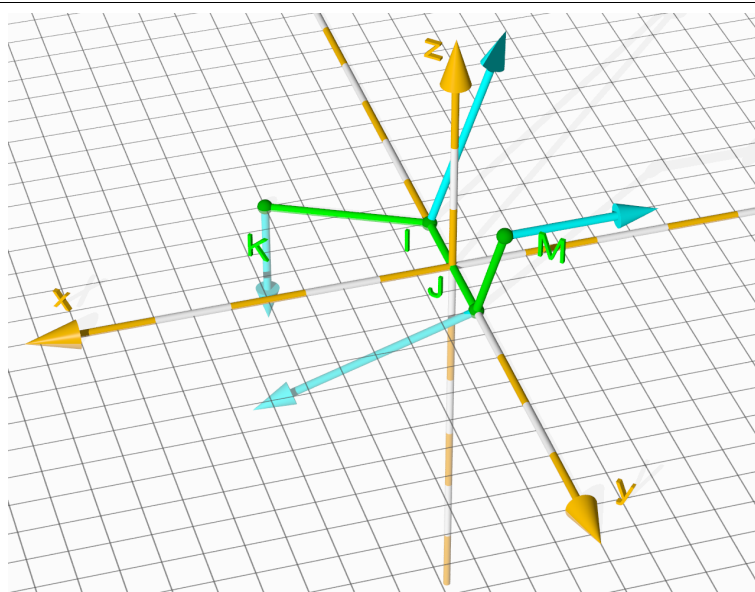


Abbildung 4: Gradientenvektoren des Torsionspotenzials

3.1.4 Out-of-plane-Winkel

An trigonalen Zentren muss zwischen Deformationen in der Ebene bei einer planaren Umgebung und einer aus der Ebene herausführenden Deformation unterschieden werden. So wird nebenbei die elektronische Struktur des Moleküls mit berücksichtigt, die durch die beiden Arten der Deformation in unterschiedlicher Weise beeinflusst wird. Zur Beschreibung der out-of-plane-Wechselwirkung verwendet MOMO ein zusätzliches Potenzial (3.1.31).

$$E_{oop} = k_{oop} \delta^2$$

k_{oop} Über das Zentralatom definierte Kraftkonstante.
 δ Winkel zwischen der von Zentralatom und zwei Substituenten aufgespannten Ebene und dem dritten Substituenten.

Formel 3.1.31: MOMO-OOP Potenzial

Mithilfe dieses Potentials wird der entsprechende Bereich im Molekül planarisiert, falls das Zentralatom zu einem π -System gehört, andernfalls ist es abgeschaltet. Die pyramidale Form kommt dann durch die Bindungswinkeldeformationspotenziale um das Zentralatom herum zustande.

Jetzt können Ableitung und Gradienten berechnet werden:

$$E'_{oop} = k_{oop} 2 \delta \quad (3.1.32)$$

Zur Bestimmung des Winkels und der Gradientenvektoren müssen einige Hilfsvektoren berechnet werden. Diese sind in Abb. 5 gezeigt.

Im ersten Schritt wird ein senkrecht auf der Ebene I, J, K stehender Vektor \vec{DI} berechnet.

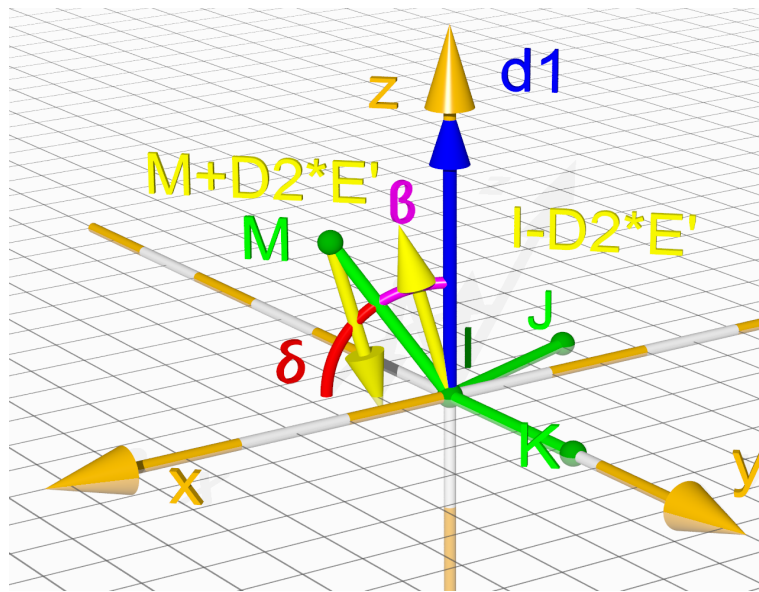


Abbildung 5: Hilfsvektoren des OOP-Potenzials (Teil 1)

$$\vec{D1} = \vec{IJ} \times \vec{IK} \quad (3.1.33)$$

Nun kann der Winkel zwischen $\vec{D1}$ und der Bindung \vec{IM} bestimmt werden.

$$\cos(\beta) = \vec{D1} \cdot \vec{IM} ; \quad \delta = 90^\circ - \beta \quad (3.1.34)$$

Nun wird der Hilfsvektor $\vec{D2}$ erzeugt, der den Gradienten für Atom M und die erste Komponente des Gradienten für Atom I bildet:

$$\vec{D2} = (\vec{D1} - SC2 \cdot \vec{IM}) \cdot SC1 \quad (3.1.35)$$

$$SC1 = \frac{E'_{oop}}{|D1| \cdot |IM| \cdot \sqrt{1 - \cos^2(\beta)}} \quad (3.1.36)$$

$$SC2 = \cos(\beta) \cdot \frac{|D1|}{|IM|} \quad (3.1.37)$$

Jetzt folgt die Berechnung der Komponenten des Gradienten für die Atome K und J:

$$SC2_{K,J} = \frac{\cos(\beta)}{|D1| \cdot |IM|} \quad (3.1.38)$$

$$\vec{D2}_K = (\vec{IM} \times \vec{IJ} - SC2_{K,J} \cdot (\vec{D1} \times \vec{IJ})) \cdot SC1 \quad (3.1.39)$$

$$\vec{D2}_J = (\vec{IK} \times \vec{IM} - SC2_{K,J} \cdot (\vec{IK} \times \vec{D1})) \cdot SC1 \quad (3.1.40)$$

Die Komponenten der Gradienten werden schließlich wie folgt aufsummiert:

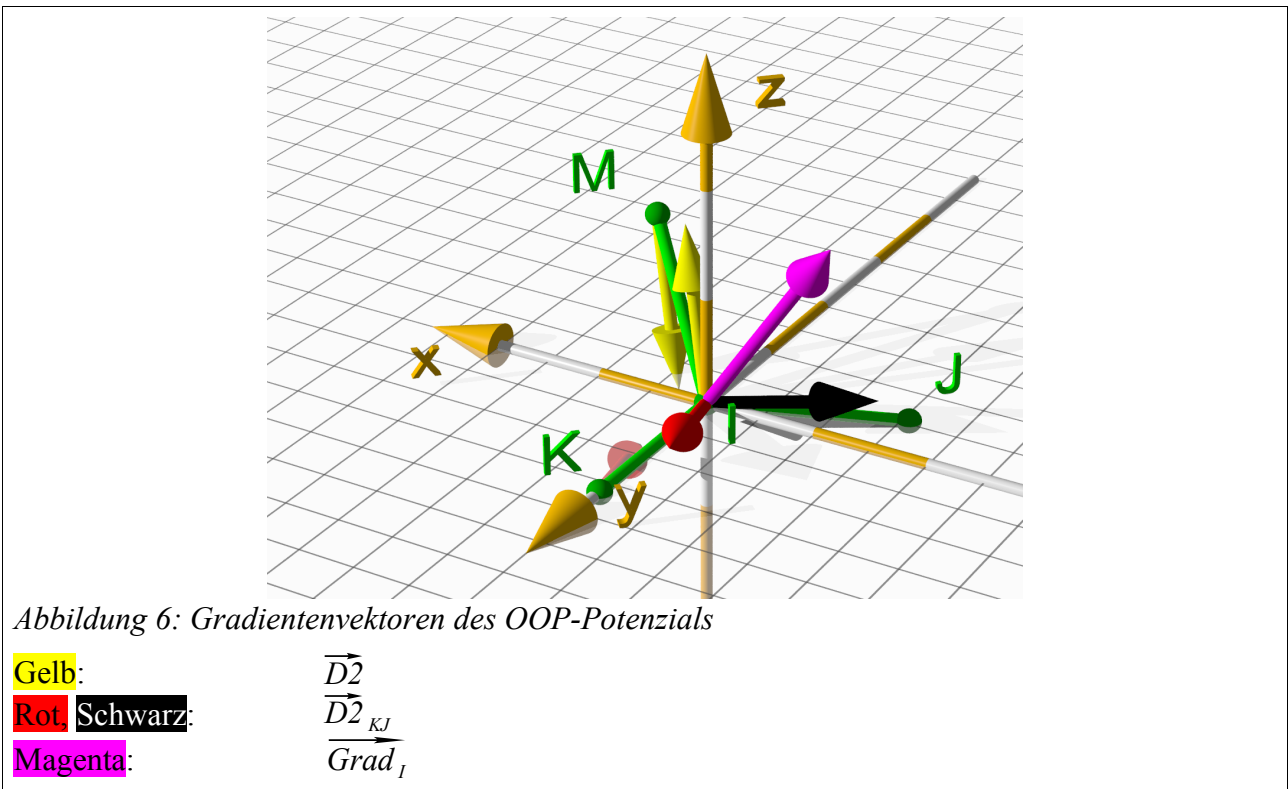
$$\text{Grad}_M = \overrightarrow{D2} \quad (3.1.41)$$

$$\text{Grad}_K = \overrightarrow{D2}_K \quad (3.1.42)$$

$$\text{Grad}_J = \overrightarrow{D2}_J \quad (3.1.43)$$

$$\text{Grad}_I = -\overrightarrow{D2} - \overrightarrow{D2}_K - \overrightarrow{D2}_J \quad (3.1.44)$$

Jede Komponente wird, wie oben ersichtlich, einmal mit positivem und einmal mit negativem Vorzeichen verwendet. Die resultierenden Gradientenvektoren sind in Abb. 6 dargestellt.



3.2 Nichtbindende Wechselwirkungen

3.2.1 Wasserstoffbrückenbindungen im MOMO-Kraftfeld

Die Wasserstoffbrückenbindung beruht hauptsächlich auf elektrostatischen Wechselwirkungen. Um die Vorzugsgeometrie dieser Bindung zu berücksichtigen, kann in MOMO ein spezielles Potenzial genutzt werden. In der Standardeinstellung werden unterhalb eines bestimmten Abstandes zwischen Wasserstoffatom und Akzeptor 0.0125 Elementarladungen vom Wasserstoff auf den Akzeptor

3 Aufbau des MOMO-Kraftfeldes

verschoben. Für den Fall, dass Donor und Akzeptor Teil eines π -Systems sind, wird der Betrag um die Hälfte reduziert.

$$q_H^\sigma(HB) = q_H^\sigma + 0.0125 C_{HB}; \quad q_{Donor}^\sigma(HB) = q_{Donor}^\sigma - 0.0125 C_{HB}$$

C_{HB} Skalierungsfaktor, der für den Fall, dass Donor und Akzeptor Teil von π -Systemen sind, 0.5 und andernfalls 1 beträgt. Die Wasserstoffbrücke wird so durch eine Anhebung der attraktiven Coulomb-Wechselwirkung berücksichtigt, welche im Gegensatz zur „echten“ Wasserstoffbrücke aber ungerichtet ist.

Formel 3.2.1: Ladungsübertragung bei Wasserstoffbrücken im MOMO-Kraftfeld

Daneben kann noch ein von Vedani und Dunitz [Vedani1985] entwickeltes Potenzial (3.2.2) herangezogen werden, bei dem die Energie neben dem Donor-Akzeptor-Abstand noch von der Ausrichtung abhängt.

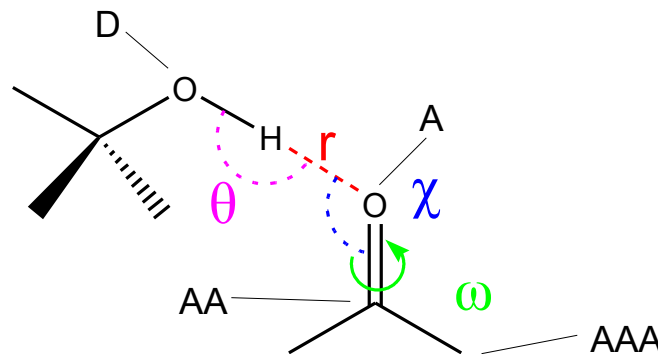


Abbildung 7: Parameter im Wasserstoffbrückenpotenzial von Vedani und Dunitz

D	=	Donor	AAA	=	zweiter Nachbar des Akzeptors oder, wenn das nicht möglich ist, Nachbar des Nachbarn des Akzeptors.
A	=	Akzeptor			
AA	=	Nachbar des Akzeptors			

Für χ_0, ω_0 lassen sich separate Parameter für zu π -Systemen gehörende Atome definieren. Ist einer der Winkel im Molekül nicht realisierbar, wird der entsprechende Term nicht berücksichtigt. Für die Terme der Potenzialfunktion werden bei der Berechnung einige Sonderfälle behandelt, deshalb werden sie im Folgenden präzisiert.

$$E_{HB} = \left(\frac{a}{r_{HA}^i} - \frac{c}{r_{HA}^j} \right) \cos^k(\theta) \cos^m(\chi - \chi_0) \cos^n(\omega - \omega_0)$$

r_{HA}	Abstand zwischen Wasserstoff und Akzeptor
a, c	Von Donor und Akzeptor abhängige Parameter
θ	Winkel D-H...A
χ	Winkel H...A-AA
ω	Torsionswinkel H...A-AA-AAA
χ_0, ω_0	Idealwerte der betreffenden Winkel

Formel 3.2.2: Wasserstoffbrückenpotenzial von Vedani und Dunitz

Der abstandsabhängige Term

Zuerst wird überprüft, ob r_{HA} kleiner als 2 \AA ist, ansonsten wird die Berechnung abgebrochen. Ist $1 \text{ \AA} < r_{HA} < 1.5 \text{ \AA}$, so wird nach folgender Gleichung berechnet:

$$E_r = \frac{a}{r^i} - \frac{c}{r^j} \quad (3.2.3)$$

$$E'_r = -i \cdot \frac{a}{r^i \cdot r} + j \cdot \frac{c}{r^j \cdot r} \quad (3.2.4)$$

Für $r < 1 \text{ \AA}$ wird für die Energie konstant gesetzt (3.2.5), die Ableitung wird aber weiterhin nach der Gleichung (3.2.4) berechnet.

$$E_r = a - c \quad (3.2.5)$$

Im Bereich $1.5 \text{ \AA} < r_{HA} < 2 \text{ \AA}$ erfolgt eine Dämpfung der Potenzialfunktion:

$$E_r = \frac{a}{r^i} - \frac{c}{r^j} \cdot 4(r-2)^2 \cdot (4r-5) \quad (3.2.6)$$

$$E'_r = \left[-i \cdot \frac{a}{r^i \cdot r} + j \cdot \frac{c}{r^j \cdot r} \right] \cdot 24(r-2) \cdot (2r-3) \quad (3.2.7)$$

Die Gradienten werden analog zu denen des Bindungslängenpotenzials berechnet:

$$\text{Grad}_A = \vec{HA} \cdot E'_r \cdot E'_\theta \cdot E'_\chi \cdot E'_\omega \quad (3.2.8)$$

$$\text{Grad}_H = -\text{Grad}_A \quad (3.2.9)$$

Winkelabhängiger Term für den Winkel D-H...A

Zunächst wird abgefragt, ob der Winkel größer als ein für den Wasserstoffbrückentyp vorgegebener Parameter θ_0 ist. Wenn dies nicht der Fall ist, erfolgt keine Berechnung. Der Term und dessen Ableitung werden nach folgenden Gleichungen bestimmt:

$$E_\theta = \cos^k(\theta) \quad (3.2.10)$$

$$E'_\theta = -k \cdot \cos^{k-1}(\theta) \cdot \sin(\theta) \quad (3.2.11)$$

Die Berechnung der Gradienten unterscheidet sich von der für das Bindungswinkelpotenzial.

$$\vec{Grad}_D = \left(\vec{HA} - \left(\frac{\vec{HD} \cdot \cos(\theta)}{|\vec{HD}| \cdot |\vec{HA}|} \right) \right) \cdot SCI \quad (3.2.12)$$

$$\vec{Grad}_A = \left(\vec{HD} - \left(\vec{HA} \cdot \cos(\theta) \cdot \frac{|\vec{HD}|}{|\vec{HA}|} \right) \right) \cdot SCI \quad (3.2.13)$$

$$Grad_H = -\vec{Grad}_D - \vec{Grad}_A \quad (3.2.14)$$

$$SCI = E_r \cdot E'_\theta \cdot E_x \cdot E_\omega \cdot \frac{-1}{(1 - \cos^2(\theta)) \cdot |\vec{HD}| |\vec{HA}|} \quad (3.2.15)$$

Die Berechnung ähnelt der beim OOP-Potenzial.

Winkelabhängiger Term für den Winkel H...A-AA

Die Berechnung erfolgt nur, wenn $\chi > \chi_{min}$ gilt. Solange $\chi > \chi_{min} + 5^\circ$, wird wie folgt berechnet:

$$E_\chi = \cos^m(\chi - \chi_0) \cdot G \quad (3.2.16)$$

$$E'_\chi = -m \cdot \cos^{m-1}(\chi - \chi_0) \cdot \sin(\chi - \chi_0) \cdot DG \quad (3.2.17)$$

Falls es sich beim Akzeptor um ein zu einem π -System gehörendes Atom handelt, kann für χ_0 ein alternativer Wert angegeben werden.

G und DG sind Dämpfungsfaktoren. Sie werden nach folgenden Gleichungen berechnet:

$$G = 3((\chi - \chi_0) \cdot 0.2)^2 - 2 \cdot ((\chi - \chi_0) \cdot 0.2)^3 \quad (3.2.18)$$

$$DG = 68.7^\circ \cdot 1.2 \cdot ((\chi - \chi_0) \cdot 0.2) - ((\chi - \chi_0) \cdot 0.2)^2 \quad (3.2.19)$$

3 Aufbau des MOMO-Kraftfeldes

Die Gradienten werden analog zum Winkel D-H...A berechnet. Lediglich der Skalierungsfaktor SC1 (3.2.20) muss angepasst werden.

$$SC1 = E_r \cdot E_\theta \cdot E'_\chi \cdot E_\omega \cdot \frac{-1}{(1 - \cos^2(\theta)) \cdot |AH| |AAA|} \quad (3.2.20)$$

Vom Torsionswinkel H...A-AA-AAA abhängiger Term

Für Torsionswinkel $\omega > 90^\circ$ wird ein Parameter ω_0 subtrahiert. Falls es sich beim Akzeptor um ein zu einem π -System gehörendes Atom handelt, kann für ω_0 ein alternativer Wert angegeben werden.

Für $\omega > 90^\circ$:

$$E_\omega = \cos^n(\omega - \omega_0) \quad (3.2.21)$$

$$E'_\omega = -n \cdot \cos^{(n-1)}(\omega - \omega_0) \cdot \sin(\omega - \omega_0) \cdot DT \quad (3.2.22)$$

Für $\omega \leq 90^\circ$:

$$E_\omega = \cos^n(\omega) \quad (3.2.23)$$

$$E'_\omega = -n \cdot \cos^{(n-1)}(\omega) \cdot \sin(\omega) \cdot DT \quad (3.2.24)$$

DT	Vorzeichen des Torsionswinkels H...A-AA-AAA
ω	Betrag des Torsionswinkels H...A-AA-AAA

Die Gradienten werden abweichend zum Torsionswinkelpotenzial wie folgt berechnet:

Zunächst werden zwei Hilfsvektoren berechnet, die senkrecht auf den den Torsionswinkel einschließenden Ebenen stehen.

$$\vec{V1} = \vec{AA} \times \vec{AH} \quad (3.2.25)$$

$$\vec{V2} = \vec{AA} \times \vec{AAA} \quad (3.2.26)$$

Der Torsionswinkel ergibt sich daraus über

$$\cos(\omega) = \vec{V1} \cdot \vec{V2} \quad (3.2.27)$$

Die weitere Berechnung geschieht nur unter der Bedingung $1 - \cos^2(\omega) > 1 \cdot 10^{-6}$.

$$SCI = \frac{-E_r \cdot E_\theta \cdot E_\chi \cdot E_\omega}{|V1| \cdot |V2| \cdot \sqrt{1 - \cos^2(\omega)}} \quad (3.2.28)$$

$$\overrightarrow{Grad}_H = SCI \cdot (\overrightarrow{V2} \times \overrightarrow{AA}) - \frac{\cos(\omega) \cdot SCI}{|V1| \cdot |V2|} \cdot (\overrightarrow{V1} \times \overrightarrow{AA}) \quad (3.2.29)$$

$$\overrightarrow{Grad}_{AAA} = SCI \cdot (\overrightarrow{V1} \times \overrightarrow{AA}) - \frac{\cos(\omega) \cdot SCI \cdot |V1|}{|V2|} (\overrightarrow{V2} \times \overrightarrow{AA}) \quad (3.2.30)$$

$$\begin{aligned} \overrightarrow{Grad}_A = & SCI \cdot [(\overrightarrow{V2} \times \overrightarrow{AAH}) + (\overrightarrow{V1} \times \overrightarrow{AAA})] \\ - \cos(\omega) \cdot SCI \cdot & \left[\frac{\overrightarrow{V1} \times \overrightarrow{AAH}}{|V1| \cdot |V2|} + (\overrightarrow{V2} \times \overrightarrow{AAA}) \cdot \frac{|V1|}{|V2|} \right] \end{aligned} \quad (3.2.31)$$

$$\overrightarrow{Grad}_{AA} = -(\overrightarrow{Grad}_A + \overrightarrow{Grad}_H + \overrightarrow{Grad}_{AAA}) \quad (3.2.32)$$

3.2.2 Das Van-der-Waals-Potenzial

In MOMO wird eine Mischform aus Lennard-Jones- und Buckingham-Potenzial eingesetzt. Die

Gleichung setzt sich aus einem repulsiven ($\frac{1}{r^{12}}$ Lennard-Jones; e^r Buckingham) und einem

attraktiven Term ($\frac{1}{r^6}$) zusammen.

$$E_{vdw} = \varepsilon \left(\frac{r_0}{r} \right)^{12} - 2 \varepsilon \left(\frac{r_0}{r} \right)^6 = \frac{A}{r^{12}} - \frac{B}{r^6} \quad \text{Lennard-Jones-Potenzial}$$

$$E_{vdw} = Ae^{-br} - \frac{C}{r^6} \quad \text{Buckingham-Potenzial}$$

$$E_{vdw} = \frac{ae^{-b(r-r_0)}}{r^n} - \frac{c}{r^6} \quad \text{MOMO-Potenzial}$$

Formel 3.2.33: Das MOMO van-der-Waals-Potenzial

Folgende Sonderfälle werden speziell behandelt:

Zunächst kann eine Korrektur des Atomabstands nach Formel 3.2.34 vorgenommen werden:

3 Aufbau des MOMO-Kraftfeldes

$$r_{kor} = AR_I \cdot AR_J \cdot r_{IJ} - BR_I - BR_J$$

r_{kor} korrigierter Abstand zwischen den Atomen I und J

AR, BR empirische Parameter, vom Atomtyp des Atoms im Index abhängig

Formel 3.2.34: Korrektur des Atomabstandes.

Zurzeit wird davon nur bei den Atomtypen Mg und Ca Gebrauch gemacht.

Eine Berechnung erfolgt nur, wenn $r_{kor} < 6.5 \text{ \AA}$ und wenn die Wechselwirkung für das betreffende Atompaar nicht abgeschaltet wurde. Eine Abschaltung erfolgt immer, wenn ein Atompaar Teil einer Bindung, eines Bindungswinkels oder eines Torsionswinkels ist.

Ein Mindestabstand ist als Parameter für jeden Atompaartyp festgelegt. Sofern der tatsächliche Abstand geringer ausfällt, wird stets mit dem Mindestabstand operiert.

Die erste Ableitung wird nach Gleichung (3.2.35) berechnet:

$$E'_{vdw} = \frac{ae^{-b(r-r_0)}}{r^n} \cdot b - \frac{\frac{ae^{-b(r-r_0)}}{r^n} \cdot n - \frac{6 \cdot c}{r^6}}{r} \quad (3.2.35)$$

Für Distanzen zwischen $6.0 \text{ \AA} < r_{kor} < 6.5 \text{ \AA}$ wird das Potenzial wie folgt gedämpft:

$$E_{vdw_{gedämpft}} = E_{vdw} \cdot [3((6.5 \text{ \AA} - r) \cdot 2)^2 - 2((6.5 \text{ \AA} - r) \cdot 2)^3] \quad (3.2.36)$$

$$E'_{vdw_{gedämpft}} = E'_{vdw} \cdot G + E_{vdw} \cdot DG \quad (3.2.37)$$

$$G = 3((6.5 \text{ \AA} - r) \cdot 2)^2 - 2((6.5 \text{ \AA} - r) \cdot 2)^3 \quad (3.2.38)$$

$$DG = 12 \cdot [((6.5 \text{ \AA} - r) \cdot 2)^2 - (6.5 \text{ \AA} - r) \cdot 2] \quad (3.2.39)$$

Für Distanzen $r < r_{min}$:

$$G = 3 \cdot \left(\frac{r}{r_{min}} \right)^2 + 2 \cdot \left(\frac{r}{r_{min}} \right)^3 \quad (3.2.40)$$

$$DG = \frac{6 \cdot \left(\left(\frac{r}{r_{min}} \right) - \left(\frac{r}{r_{min}} \right)^2 \right)}{r_{min}} \quad (3.2.41)$$

$$E_{vdw_{gedämpft}} = E'_{vdw} \cdot r \cdot G + E_{vdw} - E'_{vdw} \cdot r_{min} \quad (3.2.42)$$

$$E'_{vdw_{gedämpft}} = E'_{vdw} \cdot (G \cdot r \cdot DG) \quad (3.2.43)$$

Die Gradienten werden analog zum Bindungslängenpotenzial ermittelt.

3.2.3 Das Coulomb-Potenzial

MOMO charakterisiert die Ladungsverteilung in einem Molekül durch an den Atomkernen lokalisierte Punktladungen. Zur Berechnung der elektrostatischen Wechselwirkung zwischen diesen wird ein aus dem Coulomb-Gesetz hergeleitetes Potenzial verwendet.

$$E_{coul} = \frac{N_L}{4 \pi \epsilon_0 \epsilon} \frac{q_i q_j}{r_{ij}}$$

N_L	Avogadro-Konstante
q_i, q_j	Ladungen an den Atomen I und J
r_{ij}	Abstand der Atome I und J voneinander
ϵ	Relative Dielektrizitätskonstante. $\epsilon = d r_{ij}^n$; d und n können vom Benutzer angegeben werden
ϵ_0	Dielektrizitätskonstante des Vakuums

Formel 3.2.44: Das MOMO-Coulomb-Potenzial

Die Coulomb-Energien werden für alle Paare von Atomen berechnet, die nicht über eine Bindung miteinander verknüpft sind. Durch Verwendung einer relativen Dielektrizitätskonstante kann diese mit wachsendem Abstand des Punktladungspaares voneinander wachsen. So kann die Polarisation eines zwischen den beiden Zentren liegenden Mediums, z. B. eines Lösungsmittels, berücksichtigt werden. Die Berechnung der Coulomb-Energien kann nach einem einstellbaren Maximalabstand der Zentren übersprungen werden.

Die Ableitung lautet:

$$E'_{coul} = \frac{E_{coul} \cdot (n+1)}{r_{ij}} \quad (3.2.45)$$

Bei Konformationsanalysen kann die Startstruktur in ungünstigen Fällen so gewählt werden, dass Atome nahezu an identischen Positionen liegen. Damit das Kraftfeld in solchen Situationen keinen

Überlauf erzeugt und die Berechnung abbricht, wird das Potenzial für Abstände $0.2 \text{ \AA} > r > 0.1 \text{ \AA}$ gedämpft:

$$T = (r - 0.1) \cdot 10 \quad (3.2.46)$$

$$G = 3 \cdot T^2 - 2 \cdot T^3 \quad (3.2.47)$$

$$DG = 60 \cdot (T - T^2) \quad (3.2.48)$$

$$T1 = \frac{N_L}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon r_{ij}} \quad (3.2.49)$$

$$T2 = \frac{N_L}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon 0.1} \quad (3.2.50)$$

$$E_{coul_{gedämpft}} = T1 \cdot G + T2 \cdot (1 - G) \quad (3.2.51)$$

$$E'_{coul_{gedämpft}} = \frac{-(n+1) \cdot T1 \cdot G}{r_{ij}} + (T1 - T2) \cdot DG \quad (3.2.52)$$

Für Abstände $0.1 \text{ \AA} > r$ wird mit $r = 0.1 \text{ \AA}$ gerechnet. Die Gradienten werden analog zum Bindungslängenpotenzial erzeugt.

3.3 Ladungsberechnung

Im Programm MOMO stehen verschiedene empirische Modelle zur Verfügung, um atomare Punktladungen zu berechnen. Allen gemeinsam ist die getrennte Behandlung von π - und σ -Ladungen. Die π -Ladungen werden durch eine SCF-Rechnung [PIMM1991] bestimmt, wobei nur p_z -Orbitale verwendet werden. Für die Ladungsberechnung im σ -Gerüst stehen u. a. das Gasteiger-Marsili- [Gasteiger1980] und eine von S. Monz modifizierte Variante [Monz1997] des Abraham-Verfahrens zur Verfügung [Abraham1982][Abraham1984][Abraham1985][Abraham1988][Abraham1988#45][Abraham1988#46][Abraham1989][Abraham1991][Abraham1992].

3.3.1 SCF-Rechnung

MOMO führt für konjugierte π -Systeme SCF-Rechnungen durch. Hierbei werden Bindungsordnungen und Ladungen berechnet. Da keine genauen Eigenwerte der Zustände ermittelt werden müssen, kann mit minimalem Basissatz unter groben Näherungen gearbeitet werden. Die π -MOs werden aus jeweils einem $2p_z$ -AO der n am π -System beteiligten Atome (Basissatz) gebildet. Zur weiteren Vereinfachung wird hier eine ZDO-Näherung gemacht, wobei die

Überlappungsmatrix zur Einheitsmatrix wird. Weiterhin werden lediglich die Zweizentrenintegrale berechnet. Durch diese Maßnahmen wächst der Rechenaufwand nur noch quadratisch und nicht mehr mit n^4 , wie bei der klassischen Hartree-Fock-Methode. Zusätzlich wird die Problemgröße im Programm klein gehalten, weil lokale π -Systeme voneinander getrennt behandelt werden. Da die σ -Ladungen das π -System polarisieren und umgekehrt auch die π -Ladungen die relativen Elektronegativitäten der Atome, die zur Berechnung der Ladungen im σ -Gerüst genutzt werden, beeinflussen, muss zwischen den π -SCF-Zyklen eine σ -Ladungsberechnung erfolgen.

3.3.2 Ladungsberechnung im σ -Gerüst

3.3.2.1 Die PEOE-Methode von Gasteiger und Marsili

Der Standardalgorithmus zur Berechnung der σ -Ladungen in MOMO ist die PEOE-Methode (Partial Equalisation of Orbital Electronegativity) von Gasteiger und Marsili. Die Implementierung wurde aus dem PIMM-Kraftfeld übernommen, jedoch wurde eine Reparametrisierung durchgeführt. Für die Berechnung wird lediglich die Topologie des Moleküls berücksichtigt. Grundannahme dieses Modells ist die Dominanz der Elektronegativitätsdifferenzen benachbarter Orbitale an kovalent gebundenen Atomen über die Ladungsverschiebung. Als Konsequenz der Ausbildung chemischer Bindungen kommt es zu einem partiellen Ausgleich der Orbitalelektronegativitäten, der die treibende Kraft für einen Ladungsfluss vom elektropositiven zum elektronegativen Zentrum ist. Hierbei wird die von Jaffe und Hinze [Hinze1962] erweiterte Definition der Elektronegativität Mullikens verwendet:

$$\chi_m = \frac{1}{2}(I_m + A_m)$$

χ_m	Elektronegativität nach Jaffe und Hinze
I_m	Ionisierungspotenzial nach Jaffe und Hinze
A_m	Elektronenaffinität nach Jaffe und Hinze
m	Valenzorbital

Formel 3.3.1: Elektronegativität nach Jaffe und Hinze

Die Elektronegativität lässt sich als Potenzreihe (Formel 3.3.2) der Nettoladung eines Atoms entwickeln:

$$\chi_m = a + bq_m + cq_m^2 + \dots \quad (3.3.2)$$

Anhand dieser Gleichung und den Anregungsenergien für ionische Valenzzustände können die Koeffizienten der Reihenentwicklung für verschiedene Hybridisierungszustände gefunden werden. Normalerweise wird die Reihenentwicklung nach dem quadratischen Term abgebrochen. Stellt man nun ein Gleichungssystem auf (eine Gleichung für jede kovalente Bindung), so lässt sich die Ladungsverteilung berechnen. Dazu müssen die Gleichungen aber einige Randbedingungen einhalten:

- Die Ladungsverschiebung zwischen zwei kovalent gebundenen Atomen ist proportional zur Differenz ihrer Elektronegativitäten.
- Die Elektronegativität des elektropositiveren Atoms in seinem Valenzzustand fungiert als Skalierungsfaktor für die Ladungsverschiebung.
- Es wird ein Dämpfungsfaktor verwendet, der den vollständigen Ausgleich der Elektronegativitäten aller Atome des Moleküls verhindert.

$$\delta q_{m,n} = \left| \left(\frac{1}{2} \right)^k \frac{(\chi_m - \chi_n)}{\chi_+} \right| \quad (3.3.3)$$

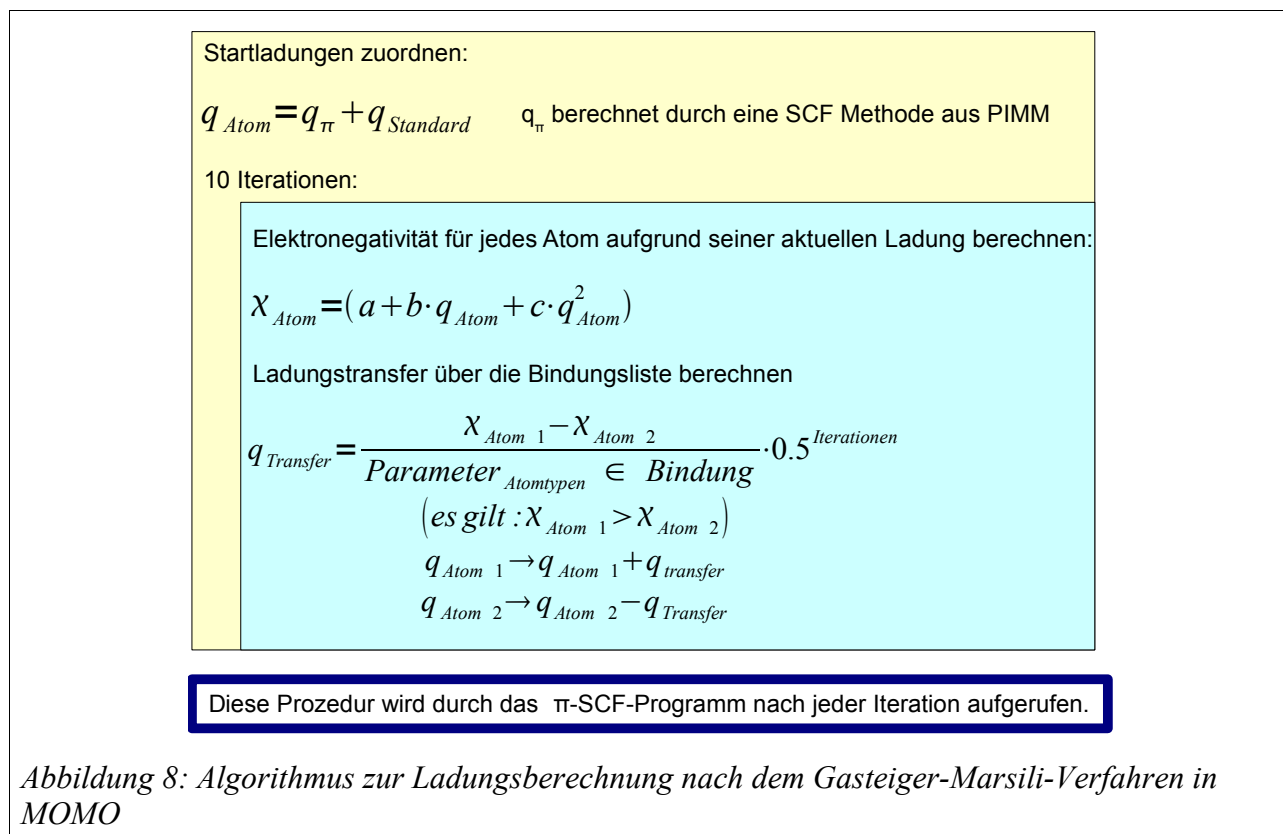
$$\delta q_m = -\delta q_{m,n} \frac{(\chi_m - \chi_n)}{(|\chi_m - \chi_n|)} \quad (3.3.4)$$

$$\delta q_n = -\delta q_m \quad (3.3.5)$$

Die Gleichungen werden iterativ gelöst. Nach dem k-ten Schritt werden die Ladungen $q_m(k-1)$ durch Addition der $\delta q_m(k)$ -Werte aktualisiert:

$$q_m^\sigma(k) = q_m^\sigma(k-1) + \sum \delta q_m(k) \quad (3.3.6)$$

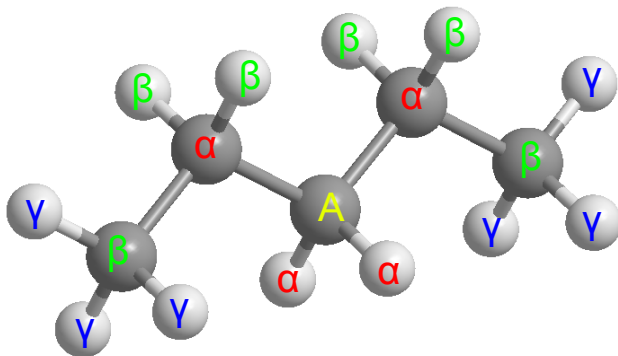
Daraus ergeben sich neue Elektronegativitäten für den nächsten Iterationsschritt. Der Dämpfungsfaktor garantiert Konvergenz nach 10 Zyklen. Die Separierung von π - und σ -Ladungen ist zulässig, da mit Elektronegativitäten von Valenzzuständen gearbeitet wird. Der Einfluss der π -Ladungsdichte auf die σ -Elektronen wird berücksichtigt, indem vor dem ersten Iterationsschritt die π -Ladungen über eine π -SCF-Rechnung bestimmt werden. π - und σ -Ladungen werden schließlich addiert. Die dem Programm PIMM nachempfundene Implementierung arbeitet, wie in Abb. 8 dargestellt.



3.3.2.2 ABR27

ABR27 ist das von S. Monz modifizierte Abraham-Modell. Dieses beruht auf dem Konzept des induktiven Effektes für gesättigte Moleküle und der in MOMO implementierten π -SCF-Rechnung für π -Systeme. Wechselwirkungen zwischen π - und σ -System werden hier nicht berücksichtigt, σ - und π -Ladungen aber schließlich aufaddiert.

Im σ -Gerüst wird die Ladung mit einem Inkrementsystem bestimmt. Hierzu wird jedes Atom in einem 4-Atom-Fragment betrachtet. Für jedes Atom werden die induktiven Effekte aufsummiert.



$$q_a^\sigma = q_a(\alpha) + q_a(\beta) + q_a(\gamma)$$

Abbildung 9: Bezeichnungen der Inkremente beim ABR27 Algorithmus

Der α -Effekt

Nach Mulliken [Mulliken1962], Smith, Magee und Eyring [Smith1951] kann die Induktion einer Ladung am Atom A in einer Bindung A-B nach Gleichung (3.3.7) berechnet werden.

$$q_a = \left(\frac{A}{2\beta} \right) \left\{ \left(T_a - \frac{1}{2} Q_{aa} \right) - \left(T_b - \frac{1}{2} Q_{ab} \right) \right\}$$

A	Konstante, ca. 0.5 für die von Mulliken untersuchten Bindungen
β	Ein negativer Parameter, der von Austausch und Überlappungsintegral der Bindung A-B abhängt
T	Ionisierungspotenzial
Q	Coulomb-Integral

Formel 3.3.7: Induktion einer Ladung nach Smith, Magee und Eyring [Smith1951]

Spätere Arbeiten verwenden die von Mulliken definierten Orbitalelektro negativitäten des Ionisierungspotenzials. Das Coulomb-Integral wird in erster Näherung vernachlässigt. Damit vereinfacht sich die Gleichung wie folgt:

$$q_a = \left(\frac{A}{2\beta} \right) (\chi_A - \chi_B) \quad (3.3.8)$$

Noch weiter vereinfacht wird, indem man das Überlappungsintegral mit einer Konstanten a beschreibt (3.3.9). Die Elektro negativität wird unabhängig von ihrer Definition mit E bezeichnet.

$$q_A = \frac{E_B - E_A}{a_{AB}}$$

E_A, E_B Elektronegativitäten der Atome A und B

a_{AB} Von Überlappungs- und Austauschintegral der Bindung A-B abhängige Konstante

Formel 3.3.9: Vereinfachte Form von Formel (3.3.7)

Für $E_A > E_B$ wird eine positive Ladung am Atom B und eine gleich große entgegengesetzte Ladung an Atom A induziert. Die Beiträge aller direkt an A gebundenen Atome werden summiert.

$$q_A(\alpha) = \sum_{\alpha} \frac{E_{\alpha} - E_A}{a_{A\alpha}} \quad (3.3.10)$$

Der β -Effekt

Der β -Effekt wird als proportional zur Differenz der Elektronegativität von Atom β und Atom α sowie der Polarisierbarkeit von Atom α angenommen. Die Ladung ist proportional zur EN-Differenz der beteiligten Zentren.

$$q_A(\beta) = \sum_{\beta} \frac{|E_{\beta} - E_{\alpha}| q_{\beta} P_{\alpha}}{b_{A\beta}}$$

$b_{A\beta}$ Von den Atomen A und β abhängiger Parameter. Mit ihm können, da von beiden Atomtypen abhängig, auch Fälle berücksichtigt werden, in denen Atome bevorzugt extrem positive oder negative Ladungen aufweisen und der Ladungstransfer zwischen ihnen nur eine untergeordnete Rolle spielt.

P_{β} Atomare Polarisierbarkeit von Atom β

Formel 3.3.11: Induktion durch den β -Effekt

Die Induktion zweier etwa gleich elektronegativer Elemente ist so geringer als die solcher mit stark unterschiedlicher EN, was chemisch sinnvoll ist. Damit die Ladung immer mit dem richtigen Vorzeichen induziert wird, multipliziert man den Betrag der EN-Differenz mit der Gesamtladung des β -Substituenten. Nebenbei wird so berücksichtigt, dass extrem negativ geladene Zentren einen größeren Beitrag leisten als solche mit geringerer Elektronendichte. Die atomare Polarisierbarkeit

des α -Substituenten dient als Skalierungsfaktor, um ein Aufschaukeln der Induktion zu verhindern. Für positive Zentren beschreibt er das dem Ladungstransfer zwischen Atom A und dem β -Substituenten entgegenwirkende Potenzial, negative Zentren erleichtern hingegen den Ladungsfluss. P_β wird über Gleichung (3.3.12) bestimmt.

$$P = P_0(1 + 1.5 \cdot (q_0 - q_\sigma - q_\pi)) \quad (3.3.12)$$

Eine Abfrage setzt den Wert auf null, falls er negativ werden sollte, auch die π -Ladung muss mit einbezogen werden. Da sich die Polarisierbarkeit mit der Ladung ändert, muss die Berechnung iterativ erfolgen.

Der γ -Effekt

Der γ -Effekt wird als direkt proportional zum β -Effekt betrachtet. Eine Konformationsabhängigkeit der Ladungsverteilung wird durch einen umgekehrt zur vierten Potenz des Abstandes zum γ -Substituenten proportionalen Terms berücksichtigt. Wegen des geringen Einflusses der 1,4-Wechselwirkungen ist der konformationsabhängige Beitrag eher gering.

$$q_A(\gamma) = \sum_{\gamma} \frac{(E_\gamma - E_A) P_\beta P_\alpha}{r_{A\gamma}^4 c_A c_\gamma} \quad (3.3.13)$$

Bei diesem Inkrementsystem kann eine Diskrepanz zwischen der Summe der atomaren Punktladungen und der Gesamtladung des Moleküls entstehen. Diese Differenz wird am Ende der Rechnung gleichmäßig auf alle Nicht-Wasserstoffatome verteilt.

Die in MOMO implementierte Realisierung des Verfahrens funktioniert wie in Abb. 10 veranschaulicht.

$$q_a(\alpha) = \sum_{\alpha} \frac{E_{\alpha} - E_a}{a_{aj}} \quad E: \text{ Elektronegativität}$$

Der β - und der γ -Effekt hängen von der Polarisierbarkeit der Nachbaratome ab, welche wiederum von deren Ladung abhängt. Deshalb ist eine iterative Berechnung erforderlich.

Setze P_A auf den Standardwert

10 Iterationen:

$$q_a(\beta) = \sum_c \frac{(E_{\beta} - E_a) q_{\beta} P_{\alpha}}{b_{a\beta}} \quad q_a(\gamma) = \sum_d \frac{(E_{\gamma} - E_a) P_{\beta} P_{\alpha}}{r_{ay}^4 c_a c_{\gamma}}$$

Entspricht die Summe der Partialladungen nicht der Ladung des Moleküls, so wird die Differenz gleichmäßig auf alle Nicht-Wasserstoffatome verteilt.

Neuberechnung der atomaren Polarisierbarkeiten

$$P_a = P_a^0 (1 + 1.5 \cdot (q_a^0 - q_a^{\sigma} - q_a^{\pi}))$$

Diese Prozedur wird durch das π -SCF-Programm nach jeder Iteration aufgerufen

Abbildung 10: Der ABR27-Algorithmus in MOMO

3.4 Die Standardbildungsenthalpie

$$\Delta H_f = E_{ster} + E_{\sigma} + E_{\pi} + \Delta H_a$$

ΔH_f	Standardbildungsenthalpie
E_{ster}	Sterische Energie. Diese beinhaltet alle Terme des MOMO Kraftfeldes, ohne das Bindungslängenpotenzial und den π -Anteil des Coulomb-Potenzials
E_{σ}	σ -Bindungsenergie
E_{π}	π -Energie
ΔH_a	Atomisierungsenergie der Elemente im Standardzustand (3.4.12)

Formel 3.4.1: Standardbildungsenthalpie in MOMO

MOMO verwendet zur Bestimmung der Standardbildungsenthalpie einen Inkrementansatz. Die berechnete Standardbildungsenthalpie bezieht sich auf die Moleküle unter Standardbedingungen in der Gasphase. Eine Konformationsabhängigkeit von ΔH_f ergibt sich aus der sterischen Energie

und der Abhängigkeit der π -Energie von den Abständen der π -Atome untereinander. Die einzelnen Energiebeiträge errechnen sich in MOMO wie folgt:

σ -Bindungsenergie:

$$E_{\sigma} = (E_{\text{Bindungen}} + E_{\text{Gruppen}}) \cdot 96,5333 \quad (3.4.2)$$

E_{σ} wird hier mit dem Faktor 96.5333 von der Einheit eV in kJ/mol umgerechnet.

$$E_{\text{Bindungen}} = \sum DE_{IS, JS} \cdot TM - TQ \quad (3.4.3)$$

$$TQ = TT \cdot TM \quad (3.4.4)$$

$$TT = (Q_{\sigma}(I) \cdot Q_{\sigma}(J) + Q_{\sigma}(I) \cdot Q_{\pi}(J) + Q_{\pi}(I) \cdot Q_{\sigma}(J)) \cdot DF_{IS, JS} \cdot 14,397 \cdot \frac{1}{r_{ij}} \quad (3.4.5)$$

$$TM = 1 - \left(1 - e^{AME_{IS, JS} \cdot (AL_{IS, JS} - r_{ij})}\right)^2 \quad (3.4.6)$$

$$E_{\text{Gruppen}} = E_{\pi\text{-Systeme}} + E_{\text{repulsive Gruppen}} \quad (3.4.7)$$

$E_{\pi\text{-Systeme}}$ Für alle Bindungen, deren beteiligte Atome π -Atome sind und die zusammen genau vier weitere Substituenten haben, die ebenfalls π -Atome sind, wird ein von den beiden an der Bindung beteiligten Atomtypen abhängiger Parameter $GIP_{IS, JS}$ addiert.

$E_{\text{repulsive Gruppen}}$ Für alle Bindungen, an denen kein Wasserstoffatom beteiligt ist, werden die als repulsive Elektronenpaare tragend markierten Substituenten gezählt und folgender Energiebeitrag addiert: $E_{\text{repulsive Gruppen}} = 0,025 \cdot N_{\text{Substituenten}}^2$

$DE_{IS, JS}$ Dissoziationsenergie einer polaren σ -Bindung [eV]

$DF_{IS, JS}$ Korrekturfaktor für Ladungseffekte in polaren σ -Bindungen

$AME_{IS, JS}$ Morse-Konstante für die Einfachbindung

$AL_{IS, JS}$ Standardbindungslänge

$GIP_{IS, JS}$ Gruppeninkremente für π -Systeme

IS, JS Atomtypen für die an der betrachteten Bindung beteiligten Atome

Tabelle 2: Parameter zu Gleichung (3.4.2) bis (3.4.7).

π -Energie

$$E_{\pi} = E_{\pi 1,1} + DT \cdot 96,5333 \quad (3.4.8)$$

3 Aufbau des MOMO-Kraftfeldes

$$E_{\pi 1,1} = \sum_{\pi \text{ Atome}} P_i \cdot RV_{IS} \cdot 0,25 - H_i \cdot P_i + HV_{IS} \cdot ZV_{IS} - (ZV_{IS} - 1) \cdot RV_{IS} - 0,5 \cdot (Q_{\sigma} - QQ_{IS}) \cdot RV_{IS} \cdot (P_i - ZV_{IS}) \quad (3.4.9)$$

$$DT = \sum_{\text{Paare } \pi \text{ Atome}} 2 \cdot P_{IJ} \cdot (H_{IJ} + Q_{\sigma}(I) \cdot Q_{\sigma}(J) \cdot RW) - 0,5 \cdot P_{IJ}^2 \cdot RW + RW \cdot (P_{IJ} - ZV_{IS}) \cdot (P_{IJ} - ZV_{JS}) \quad (3.4.10)$$

$$RW = 14,397 / \sqrt{r_{IJ}^2 + 0,25 \cdot \left(\frac{14,397 \cdot RV_{JS}}{RV_{IS} + 14,397} \right)^2} \quad (3.4.11) \text{ (Zwei-Zentren-Coulomb-Integral)}$$

RV_{IS}	Ein-Zentren-Coulomb-Integral
QQ_{IS}	Standard σ -Ladung
HV_{IS}	Ionisierungsenergie des Valenzzustandes
ZV_{IS}	Effektive Kernladung
P	Bindungsordnungsmatrix aus der π -SCF-Rechnung
H	Hückel-Matrix aus der π -SCF-Rechnung

Tabelle 3: Parameter zu Gleichung (3.4.8) bis (3.4.11) .

I und J sind die an einer Bindung beteiligten Atome IS und JS. Sie bezeichnen die jeweiligen Atomtypen im MOMO-Kraftfeld.

$$\Delta H_a = \sum_{\text{alle Atome}} DHFA_{IS} + \sum_{\text{alle geladenen Atome}} HFQINC_{IS}(Q)$$

ΔH_a	Atomisierungsenergie der Elemente im Standardzustand
$DHFA_{IS}$	Standardbildungsenthalpie für freie Atome
$HFQINC_{IS}(Q)$	Korrektur der Standardbildungsenthalpie in Abhängigkeit von der Ladung

Formel 3.4.12: Atomisierungsenergie der Elemente im Standardzustand

4 Die Parametrisierung von Kraftfeldern

Das ideale Kraftfeld ist konsistent, mit nur einer funktionalen Form und einem Parametersatz soll es alles mit akzeptabler Genauigkeit vorhersagen. Teilgebiete sind die Vorhersage von Schwingungsspektren, der Struktur und der Bildungsenthalpie. Für jedes dieser Teilprobleme sind andere Eigenschaften der Energiehyperfläche wichtig. Aus diesem Grund kann ein Kraftfeld in einem Gebiet hervorragende Ergebnisse liefern, während es in den anderen versagt [Allinger1996#53]. Ein viel versprechender Ansatz war zunächst, Kraftfelder auf die Vorhersage vom Schwingungsspektren zu trimmen, andere Parameter sollten durch die Fülle der in einem Spektrum enthaltenen Informationen implizit mit abgedeckt werden. Allerdings werden Spektren von der exakten Form der Energiehyperfläche in der Umgebung eines Energieminimums bestimmt. Die Struktur wird von der groben Form der Energiehyperfläche und der Position der Energieminima charakterisiert, und die Bildungsenthalpie hängt von der Position der Energiehyperfläche auf der zu ihr vertikalen Energieachse ab. Das Kraftfeldproblem zerfällt damit in drei lose miteinander gekoppelte Teilprobleme.

Um aussagekräftige Ergebnisse mit einem Kraftfeld treffen zu können, werden auf das Problem passende Potenzialfunktionen und die dazu erforderlichen Parameter benötigt. Potenzialfunktionen lassen sich durch physikalische Überlegungen ableiten, wobei ein Kompromiss zwischen mathematischer Komplexität, Anwendungsbereich und Parameterzahl gefunden werden muss.

Für die Parametrisierung gibt es mehrere mögliche Quellen:

- Direkte Ableitung aus physikalischen Messgrößen
- Anpassen der Parameter, um in einer Simulation experimentell bestimmbare Größen zu reproduzieren
- Anpassen der Parameter, um in einer quantenchemischen Rechnung gewonnene Ergebnisse zu reproduzieren

Die erforderlichen Parameter lassen sich grob in bindende (Valenz) und nichtbindende Terme unterteilen. Die Valenz-Terme können dazu dienen, um lokale Moleküleigenschaften vorherzusagen, wozu Struktur und die Frequenzen möglicher Molekülschwingungen gehören. Nichtbindende Terme dominieren intermolekulare Wechselwirkungen und thermodynamische Eigenschaften.

Leider lassen sich die Terme kaum isoliert voneinander optimieren, da zwischen ihnen vielfältige Wechselwirkungen existieren. Betrachtet man jedoch die Beiträge der einzelnen Terme zur Gesamtenergie im Kraftfeld, so lässt sich eine Strategie für das Vorgehen bei der Parametrisierung ableiten. Die Kopplung der Energiebeiträge ist nicht in jedem Fall gleich stark ausgeprägt. „Harte“ Freiheitsgrade eines Moleküls, wie zum Beispiel Bindungslängen und -winkel, können oft separat behandelt und sogar ohne Änderung zwischen verschiedenen Kraftfeldern übernommen werden ([Leach2001] S. 229). „Weiche“ Freiheitsgrade wie nichtbindende Wechselwirkungen und Torsionswinkelpotenziale beeinflussen sich dagegen erheblich untereinander. Man geht deshalb bei der Parametrisierung schrittweise vor.

4.1 Auswahl der Referenzdaten

Zunächst muss man sich darüber im Klaren sein, wo und wie das zu parametrisierende Kraftfeld eingesetzt werden soll. Wie eingangs erwähnt, sind für Teilaspekte des Kraftfeldproblems verschiedene Eigenschaften der Energiehyperfläche maßgebend. Vernünftige Strukturen lassen sich meist mit relativ einfachen Mitteln berechnen, zur Vorhersage von Schwingungsspektren sind eine komplexere funktionale Form und Parameter, die die Form der Energiehyperfläche genau wiedergeben, erforderlich.

Kritisch für die Qualität eines Kraftfeldes sind die zur Verfügung stehenden Referenzdaten. Kraftfelder sind empirisch und können daher nicht besser als die zugrunde liegende Referenz sein. Eine zentrale Rolle spielt die Struktur, da andere Prognosen beim Molecular Modelling darauf aufbauen. Für alle anderen vorherzusagenden Daten, wie zum Beispiel relative Energien, ist es ebenfalls erforderlich, dass sie mit in die Parametrisierung einfließen. Ziel ist es, die Referenzdaten mit dem fertigen Kraftfeld möglichst gut zu reproduzieren, weshalb diese sorgfältig validiert werden müssen. Es erscheint vorteilhaft, eine große Bandbreite von Verbindungsklassen abzudecken, allerdings sollten diese nicht so divers sein, dass eine Reproduktion mit dem beschränkten Atomtypensatz des Kraftfeldes nicht mehr möglich ist. Je größer die Datenbasis, desto weniger wirken sich zufällige Fehler aus.

Kraftfelder simulieren meist isolierte Moleküle im Vakuum, aus der Gasphase ermittelte experimentelle Daten sind deshalb als Referenz zu bevorzugen. Leider liegen oft nur in kondensierter Phase gewonnene Daten vor. Außerdem sind experimentell bestimmte Eigenschaften häufig als „Mittelwert“, zu dem verschiedene Konformationen beitragen, anzusehen. Zudem gibt es zwischen den experimentellen Methoden prinzipielle Unterschiede. Beugungsmethoden bestimmen

Abstände zwischen mittleren Atompositionen, wohingegen spektroskopische Methoden wie die Mikrowellenspektroskopie Trägheitsmomente und damit mittlere Abstände bestimmen. Eine verbreitete These ist das gegenseitige Aufheben des Einflusses von Populationen und Umgebungseinflüssen in großen Datensätzen. Auf Kosten größerer Rechenzeiten lassen sich diese Effekte auch bei der Parametrisierung berücksichtigen [Norrby1998].

Neben dem Experiment kommen auch quantenchemische Berechnungen als Datenquelle in Betracht. Sie eignen sich gut zur Parametrisierung, da sie wie die Kraftfelder isolierte Moleküle im Vakuum betrachten und nicht um entropische oder Schwingungsbeiträge korrigiert werden müssen. Daneben lassen sich jederzeit auch nichtstationäre Punkte der Energiehyperfläche und elektrostatische Eigenschaften berechnen. Als Referenz kommen aber nur aufwendigere Verfahren mit großen Basissätzen in Betracht. In der Literatur wird oft der HF/6-31G*-Basissatz verwendet. Die so erhaltenen Ergebnisse müssen allerdings noch skaliert werden, um experimentelle Ergebnisse zu reproduzieren [Maple1994][Maple1988]. Hybride DFT-Methoden (B3LYP [Stephens1994]) bringen hier eine Verbesserung, ohne die Rechenzeiten zu sehr zu erhöhen.

Strukturdaten

Einkristallstrukturdaten sind oft genutzte Quellen für Strukturdaten. Daneben spielen auch durch Mikrowellenspektroskopie aus der Gasphase gewonnene Daten eine Rolle, auch wenn diese nicht für viele Moleküle zur Verfügung stehen. Quantenchemisch berechnete Strukturen eignen sich ebenfalls unter den erwähnten Vorbehalten als Referenz. Bei der Verwendung von Kristallstrukturdaten [Norrby1998] sollten jedoch einige Einschränkungen beachtet werden:

So wird die Umgebung eines Moleküls im Kristall in der Regel dessen Struktur beeinflussen, weiche Koordinaten wie Torsionswinkel und weitreichende Wechselwirkungen sind hier stärker betroffen als die widerstandsfähigeren Bindungslängen und Bindungswinkel. Ein vollständiger Fit der Kristallstruktur erscheint daher nicht als Qualitätskriterium für das Kraftfeld geeignet, einzelne interne Koordinaten lassen sich, angemessen gewichtet, gut vergleichen. Für genaue Strukturen müssen die unterschiedlichen Paradigmen für Experiment und Kraftfeld beachtet werden. Im Experiment wirken sich z. B. bei der Messung gemittelte Molekülschwingungen aus.

Strukturen enthalten keine Energiedaten, dennoch lassen sich aus Abweichungen interner Koordinaten in verwandten Strukturen Informationen über Kraftkonstanten gewinnen.

Energien verschiedener Konformationen

Ein direkter Vergleich mit den Energien im Kraftfeld ist nur bei quantenchemischen Rechnungen möglich. Bei experimentellen Daten ist eine Korrektur notwendig, da diese thermische Beiträge enthalten und Mittelwerte mehrerer Konformationen widerspiegeln können. Die Berechnung der verschiedenen experimentell zugänglichen Energiearten mit einem Kraftfeldprogramm kann sehr zeitaufwendig sein, es ist aber möglich, aus wenigen solcher Rechnungen einen Korrekturfaktor zu bestimmen, der während der Parameterverfeinerung genutzt werden kann.

Eigenschaften der Energiehyperfläche in der Nähe der Minima

Ableitungen der Energiehyperfläche in der Nähe ihrer Minima lassen sich experimentell mithilfe der Schwingungsspektroskopie (IR und Raman) erzeugen. Ebenso liefern quantenchemische Rechnungen solche Daten direkt und vermeiden das Problem der Zuordnung des Spektrums [Halgren1996].

Ladungen

Elektrostatische Wechselwirkungen werden in Kraftfeldern meist anhand von Atomladungen, auf den Bindungen zentrierten Dipolen [Dinur1995] oder Multipolmomenten berechnet. Diese Größen sind allerdings keine Observablen, sondern müssen aus anderen, observablen Größen bestimmt werden [Dudek1995]. Dies kann z. B. auf der Grundlage des molekularen Multipolmomentes, Lösungseigenschaften, Messung der Elektronendichteverteilung mittels Röntgenbeugung oder quantenchemischen Rechnungen geschehen.

Über ab-initio-Rechnungen lässt sich die Elektronendichteverteilung um ein Molekül ermitteln. Um daraus Atomladungen und atomare Multipolmomente zu bestimmen, existieren verschiedene Rechenverfahren, die unterschiedliche Ergebnisse liefern [Sigfridsson1998]. Die Wahl eines dieser Verfahren orientiert sich am Einsatzgebiet des Kraftfeldes. Meistens sollen die elektrostatischen Eigenschaften eines Fragments außerhalb der Moleküloberfläche nachgebildet werden. Diese Eigenschaften hängen auch von Molekülgeometrie und Umgebung ab, „polarisierbare“ Kraftfelder stoßen daher zunehmend auf Interesse. Die Parametrisierung solcher Ladungsmodelle erfordert aber sehr diverse Referenzdaten oder mit ab-initio-Methoden berechnete Polarisierbarkeiten.

Van-der-Waals-Kräfte

Van-der-Waals-Kräfte machen sich vor allem in Messgrößen bemerkbar, die supramolekulare Wechselwirkungen repräsentieren. Dazu gehören z. B. Sublimationswärmen oder Gitterenergien in

Kristallen. Die Balance mit den elektrostatischen Wechselwirkungen ist hier sehr wichtig. Besondere Vorsicht ist bei der Ableitung von van-der-Waals-Parametern aus quantenchemischen Rechnungen geboten:

Die London-Kraft kann durch HF-Rechnungen und die meisten Formen der DFT nicht wiedergegeben werden. Darüber hinaus sind quantenchemische Berechnungen intermolekularer Wechselwirkungen sehr anfällig für „basis set superposition errors“ (BSSE). Dieser Fehler ist oft größer als die eigentlichen Van-der-Waals-Kräfte. Berechnungen mit sehr großen Basissätzen liefern sinnvolle Ergebnisse, kommen aber nur für kleine Moleküle in Betracht [Norrby2001].

4.2 Abschätzen vorläufiger Parameter

Mithilfe von Optimierungsmethoden lassen sich die Parameter in einem Kraftfeld schrittweise verbessern, es wird aber ein Ausgangsparametersatz benötigt. Je besser dieser gewählt ist, desto eher wird die Optimierung konvergieren. Ein schlechter Parametersatz kann sogar zu einem nicht-physikalischen lokalen Optimum hin konvergieren. Am Anfang der Optimierung kann es vorteilhaft sein, Untereinheiten des Parametersatzes separat zu optimieren. Die Startparameter sollen vor allem sinnvolle Strukturen erzeugen, damit die beobachteten Strukturen Energieminima im Kraftfeld darstellen.

Parameter für einen im Kraftfeld vorhandenen Atomtyp sind ggf. ähnlich genug zu einem neuen Atomtyp, um die Parameter zumindest als Startwert einfach zu übernehmen. Oft werden umfangreiche Ersetzungsregeln angewendet, der Erfolg hängt von der Ähnlichkeit der Atomtypen und den übernommen Parametern ab:

Für die Bindungslänge sind ähnliche kovalente Radien nötig, für Bindungswinkel gleiche Hybridisierungszustände der betreffenden Zentralatome, für Ladungsmodelle müssen beide Atomtypen eine ähnliche Elektronegativität aufweisen.

Aus den Referenzwerten an sich lassen sich bereits Parameter abschätzen. Als Standardwerte für Bindungslängen und Bindungswinkel können zum Beispiel die mittleren beobachteten Werte dienen. Aus der Verteilung der Abweichungen von den Mittelwerten ist auch eine grobe Schätzung der betreffenden Kraftkonstanten möglich. Torsionsparameter lassen sich aus den beobachteten Minima und der Rotationsbarriere ableiten. Bei jedem Parametertyp gibt es gewisse Einschränkungen, so sollten z. B. die Strukturen zur Abschätzung der Bindungswinkelparameter relativ ungespannt sein. Schwierig ist dann aber die Abschätzung der Kraftkonstante: Die

beobachtete tetraedrische Struktur von Methan wird z. B. unabhängig von der Kraftkonstante im Kraftfeld realisiert, sodass sich diese so nicht verfeinern lässt. Besonders einfach hat man es bei quantenchemisch berechneten Referenzdaten:

Bei ungespannten Systemen können die berechneten Werte für Bindungslängen und -winkel direkt als Standardwerte übernommen werden, Kraftkonstanten können berechnet werden. Atomladungen und Multipolmomente lassen sich ebenfalls nach verschiedenen Paradigmen aus der berechneten Elektronendichte ableiten.

4.3 Optimierung der Parameter

Um die Parameter automatisiert zu verfeinern, muss zunächst mathematisch definiert werden, wann Parameter gut sind. Üblich ist das Aufstellen einer Bewertungsfunktion, die für Abweichungen vom Soll Strafpunkte vergibt. Die Zahl der Strafpunkte soll möglichst klein werden. Diese Funktion kann zum Beispiel die quadratischen Abweichungen von der Referenz multipliziert mit einem Faktor, der Wichtigkeit, Maßeinheit und tolerable Abweichung für alle Komponenten aneinander angleicht, sein. Das optimale Kraftfeld ist dann ein Minimum der Bewertungsfunktion in Abhängigkeit von allen möglichen Parametern. Auf eine solche „Straffunktion“ können gängige Minimierungsalgorithmen angewendet werden. Die Bestimmung der Gewichtungsfaktoren ist allerdings sehr schwierig. Generell steigt zwar die Vorhersagequalität hoch gewichteter Parameter, ist das Kraftfeld jedoch nicht mehr ausbalanciert, weil eine Sorte der Referenzdaten übermäßig bevorzugt wurde, kann sich das ganze Kraftfeld verschlechtern.

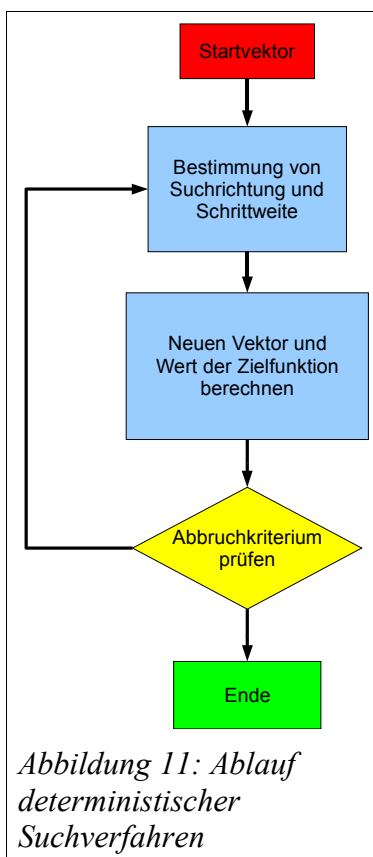
5 Optimierungsverfahren

5.1 Übersicht

Der Anwender hat die Auswahl zwischen globalen, lokalen und heuristischen Suchverfahren. Eine Übersicht findet sich in [Jakob04].

Lokale Verfahren

Die traditionellen lokalen Verfahren eignen sich gut für Parameteroptimierungen ohne kombinatorische Aspekte. Bei geringer Parameterzahl sind Gradienten- und Simplexverfahren etabliert. Sie konvergieren schnell, sind allerdings je nach verwendetem Algorithmus startpunktabhängig. Lokale Verfahren sind dafür bestimmt, um lokale Optima in der Nähe des Startpunktes zu finden. Eine Aussage über größere Gebiete des Suchraums ist mit ihnen nur möglich, wenn man die Algorithmen mehrfach mit unterschiedlichen Startwerten verwendet. Bei großen Suchräumen macht die Startpunktabhängigkeit den Vorteil der schnellen Konvergenz zunichte. Die Arbeitsweise dieser deterministischen Suchverfahren lässt sich, wie in Abb. 11 dargestellt, verallgemeinern.



Ausgehend von einem Startvektor bzw. dem letzten, nicht das Abbruchkriterium erfüllenden Vektor aus der vorhergehenden Iteration, werden Suchrichtung und Schrittweite bestimmt. Wird die Suchrichtung analytisch ermittelt, so spricht man von einem indirekten Verfahren. Hierzu werden partielle Ableitungen der Zielfunktion genutzt. Das Gradientenverfahren [Fletcher1987] und das konjugierte Gradientenverfahren [Fletcher1964] verwenden die Richtung des steilsten Anstiegs der Zielfunktion als Suchrichtung. Die Schrittweite soll dabei in Suchrichtung eine möglichst große Verbesserung der Zielfunktion bringen. Gibt es Probleme bezüglich der Differenzierbarkeit der Zielfunktion, können Quasi-Newton-Verfahren [Fletcher1963] eingesetzt werden. Diese arbeiten mit einem Modell der Zielfunktion, das als Taylorreihe entwickelt wird.

Direkte Verfahren bestimmen Suchrichtung und Schrittweite durch eine Heuristik. Hier genügt es oft, wenn die Zielfunktion abschnittsweise stetig ist. Ein einfacher Vertreter ist der Gauß-

Seidel-Algorithmus [Ortega1996] Hier wird nacheinander in Richtung der Koordinatenachsen mit fester Schrittweite gesucht. Der beste so gefundene Wert dient als Startpunkt für die nächste Iteration.

Einige Ansätze verbessern dieses grundsätzliche Verfahren. So kann entlang einer Richtung nur bis zu einer Verschlechterung gesucht oder die Schrittweite dem Suchraum angepasst werden. Pattern-Strategien passen die Suchrichtung der Zielfunktion an, indem sie Erkundungsschritte durchführen, und sind nicht mehr an die Koordinatenachsen gebunden. Nebenbedingungen müssen durch entsprechende Formulierung der Zielfunktion berücksichtigt werden. Die kann z. B. durch „Straffunktionen“ geschehen.

Die Polyederverfahren arbeiten mit mehreren Suchpunkten. In einem n -dimensionalen Suchraum spannen $n+1$ Punkte einen Polyeder auf, der während der Suche durch die Operationen Expansion, Kontraktion und Reflexion verändert wird. Der bekannteste Vertreter dieser Verfahren, das Simplex-Verfahren von Nelder und Mead [Nelder1965], arbeitet wie folgt:

- Vom Startpunkt aus wird ein Simplex konstruiert. An den Eckpunkten wird der Funktionswert bestimmt und dann der schlechteste Punkt reflektiert.
- Wird hierbei eine Verbesserung erreicht, so wird der Simplex expandiert. Ist das Resultat schlechter als der bisherige zweitschlechteste Punkt, erfolgt eine Kontraktion um den besten Punkt.
- Abbruchkriterium ist die Standardabweichung der Funktionswerte des Simplex vom Mittelpunkt.

Der Polyeder wird in Richtung eines lokalen Optimums bewegt und verkleinert. Kleinere Suboptima können wegen der Ausdehnung des Simplex übersprungen werden.

Neben den in diesem Abschnitt genannten deterministischen Suchverfahren existieren einige stochastische Varianten. Hier wird die Suchrichtung zufällig beeinflusst, der lokale Charakter bleibt jedoch erhalten.

Die direkten lokalen Suchverfahren können allgemeiner angewendet werden als die indirekten Verfahren, da sie geringere Anforderungen an die Zielfunktion stellen. Der Preis dafür ist allerdings die höhere Zahl der nötigen Zielfunktionsberechnungen.

Globale Suchverfahren

Globale Suchverfahren sollen das globale Optimum im Suchraum bzw. bei vergleichbaren Suboptima eines oder mehrere davon identifizieren. Erreichbar ist dieses Ziel allerdings nur mit deutlich mehr Zielfunktionsberechnungen als bei den lokalen Suchverfahren und einer infolgedessen viel langsameren Konvergenz.

Bekannte globale Suchverfahren sind die evolutionären Algorithmen, Branch-and-bound-Algorithmen, die Monte-Carlo-Methode und der Simulated-Annealing-Algorithmus. Auch die nur für überschaubare Suchräume geeignete Rastermethode gehört dazu. Bei dieser wird der Suchraum in konstanten Abständen vollständig abgetastet.

Die Verfahren sollen hier nur kurz vorgestellt werden, auf die für diese Arbeit bedeutsamen evolutionären Algorithmen wird noch im Detail eingegangen.

Branch-and-bound

Der Suchraum muss hier als Baum darstellbar sein, d. h. der Algorithmus [Dakin1965] eignet sich für beschränkte ganzzahlige Probleme. Durch Kombination mit anderen Verfahren können auch gemischte Probleme behandelt werden, wenn sie bestimmte Voraussetzungen erfüllen.

Das Problem wird in Klassen eingeteilt, indem den ganzzahligen Variablen nacheinander die erlaubten Werte zugewiesen werden. Für jede dieser Klassen kann dann der maximal mögliche Wert der Zielfunktion berechnet werden (ob dieser auch erreicht wird, ergibt sich erst bei genauerer Betrachtung der betreffenden Klasse) (branch). In der Klasse mit dem größten potenziellen Zielfunktionswert wird das Optimum vermutet (bond), genau diese wird dann auf die gleiche Art weiter untersucht. Die anderen Klassen werden zurückgestellt. Eliminiert werden zunächst nur Zweige, bei denen Nebenbedingungen verletzt werden. Wird auf diese Weise ein Zweig gefunden, dessen Wert unter dem potenziellen Maximalwert eines zurückgestellten Zweiges liegt, so muss dieser Zweig noch untersucht werden, da sich darin ein besseres Optimum befinden kann. Nach dem Auffinden des globalen Optimums muss unter Umständen also noch ein großer Teil des Suchraums durchsucht werden, um die Optimalität auch nachzuweisen.

Simulated Annealing

Dieser Algorithmus [Kirkpatrick1983] basiert auf dem Metropolis-Verfahren [Metropolis1953], welches Stichproben aus einem Parameterraum unter Ausnutzung der Boltzmannverteilung generiert. Ausgehend von einem Startpunkt wird ein zufälliger neuer Punkt gewählt. Für beide

Punkte wird dann die Zielfunktion berechnet. Hat der neue Punkt einen besseren Wert, so ist er der neue Startpunkt; ist der Wert schlechter, so wird er nur mit einer gewissen von der Differenz zum bisherigen Startpunkt abhängigen Wahrscheinlichkeit zum neuen Startpunkt:

$$P(\Delta E) = \begin{cases} 1 & \text{für } \Delta E < 0 \\ e^{-\frac{\Delta E}{T}} & \text{für } \Delta E \geq 0 \end{cases}$$

$P(\Delta E)$ Wahrscheinlichkeit für die Annahme des neuen Punktes abhängig von der Differenz der Zielfunktion am alten und neuen Punkt

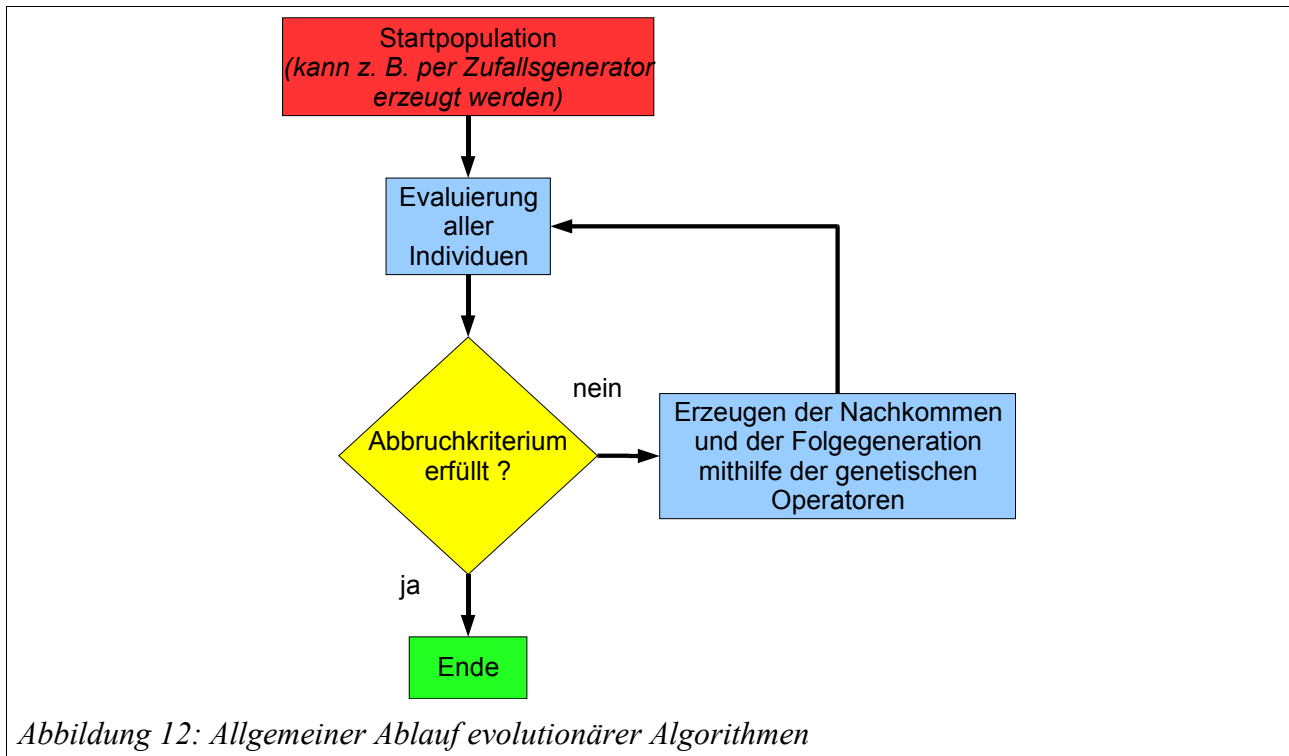
T „Temperatur“

Formel 5.1.1: Annahmewahrscheinlichkeit als neuer Startpunkt beim „simulated annealing“

Die „Temperatur“ dient als Steuerungsparameter für die Annahmewahrscheinlichkeit eines neuen Zielfunktionswertes. Sie ist zu Anfang hoch gewählt und wird später gesenkt. Mit sinkender Temperatur wird es immer unwahrscheinlicher, dass schlechtere Zielfunktionswerte akzeptiert werden. Die Möglichkeit, lokale Optima zu verlassen, sinkt damit stetig. Von dieser typischen Veränderung der Temperatur leitet sich der Name „simulated annealing“ ab. Die Steuerung des Temperaturverlaufs ist kritisch für den Algorithmus. Wird die Temperatur langsam und in kleinen Schritten gesenkt, wird das Verfahren langsam. Senkt man die Temperatur zu schnell, kann man in einem lokalen Optimum stecken bleiben. Die Konvergenz für den Grenzfall einer unendlich langsamen Abkühlung ist nachgewiesen. Ein geeigneter Temperaturverlauf kann jedoch nur empirisch gewählt werden.

5.2 Evolutionäre Algorithmen

Evolutionäre Algorithmen [Fogel1966] [Rechenberg1973] [Holland1995] bilden den Evolutionsprozess nach, womit eine generationsweise Verbesserung vorhandener Lösungen gemeint ist. Diese Algorithmen verwenden Operatoren wie Mutation und Rekombination zur Erzeugung neuer Lösungsvorschläge, arbeiten zumeist mit einer Population von Lösungen (der Suchraum wird parallel von verschiedenen Startpunkten aus durchsucht). Rekombinationsoperatoren sorgen für einen Informationsaustausch zwischen den Lösungen in einer Population, und sie enthalten stochastische Elemente. Die Selektion der Lösungen sorgt für eine intelligente Durchmusterung des Suchraums in Richtung aussichtsreicher Lösungsbereiche. Die



Verfahren lassen sich in die Klassen „Evolutionstrategie, genetische Algorithmen und evolutionäre Programmierung“ einteilen, den in Abb. 12 gezeigten generellen Ablauf haben sie jedoch gemeinsam.

Die Details des jeweils gewählten Verfahrens beeinflussen die Balance zwischen einer eher breiten und groben und einer genauen, auf ein Teilgebiet begrenzten Suche im Lösungsraum. Bei guter Anpassung an die Problemstellung werden die Chancen für das Auffinden des globalen Optimums und die Konvergenzgeschwindigkeit verbessert. Einen mathematischen Beweis für die sichere Identifikation des globalen Optimums nach einer bestimmten Zahl von Iterationen gibt es nicht, in der Praxis funktioniert dies aber meist gut. Nachteilig sind die hohe Zahl der Zielfunktionsberechnungen und die langsame Konvergenz in der Nähe eines Optimums. Andererseits lassen sich diese Algorithmen gut parallelisieren, was die genannten Nachteile teilweise aufwiegt. Eine Verbesserung kann durch die Kombination lokaler Suchverfahren mit den evolutionären Algorithmen erreicht werden. Denkbar sind hier folgende Strategien [Jakob04]:

- Die Teilung der Aufgabe: Ein Teil der Parameter wird mit einem lokalen Verfahren optimiert.
- Mit einem anderen Verfahren wird eine Startpopulation von Lösungen einer gewissen Qualität erzeugt und erst dann der evolutionäre Algorithmus angewendet.

- Die Lösungen des evolutionären Algorithmus werden mit einem lokalen Verfahren nachoptimiert.
- Ein lokales Optimierungsverfahren arbeitet im evolutionären Algorithmus als Reparatur- oder Mutationsmechanismus.

Begriffe

Die im Zusammenhang mit den evolutionären Algorithmen verwendeten Begriffe sind stark vom biologischen Vorbild geprägt.

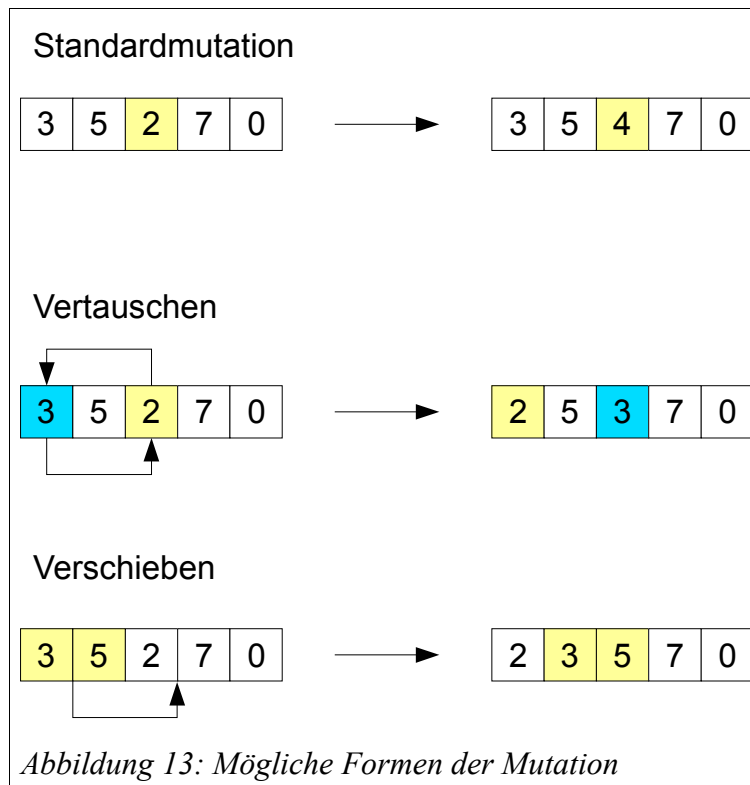
- **Individuum:**
Ein Lösungskandidat. Er enthält alle zur Lösung gehörenden Parameter, die meist in einem **Chromosom** codiert sind.
- **Population:**
Menge von Individuen
- **Chromosom:**
Eine die Eigenschaften eines Individuums codierende Zeichenkette. Meist wird nur ein Chromosom pro Individuum verwendet.
- **Gen:**
Element eines Chromosoms, ein Zeichen. Der Begriff ist fast nur bei den klassischen genetischen Algorithmen gebräuchlich.
- **Allel:**
Ein in der Lösung realisierter konkreter Wert eines Parameters
- **Generation:**
Population zu einem Zeitpunkt
- **Fitness:**
Qualität einer Lösung
- **Eltern:**
An der Reproduktion beteiligte Individuen
- **Nachkommen:**
Aus den Eltern erzeugte Individuen

5.3 Die genetischen Operatoren

Eine Übersicht der gängigen genetischen Operatoren findet sich in [Langer2005].

Der Mutationsoperator

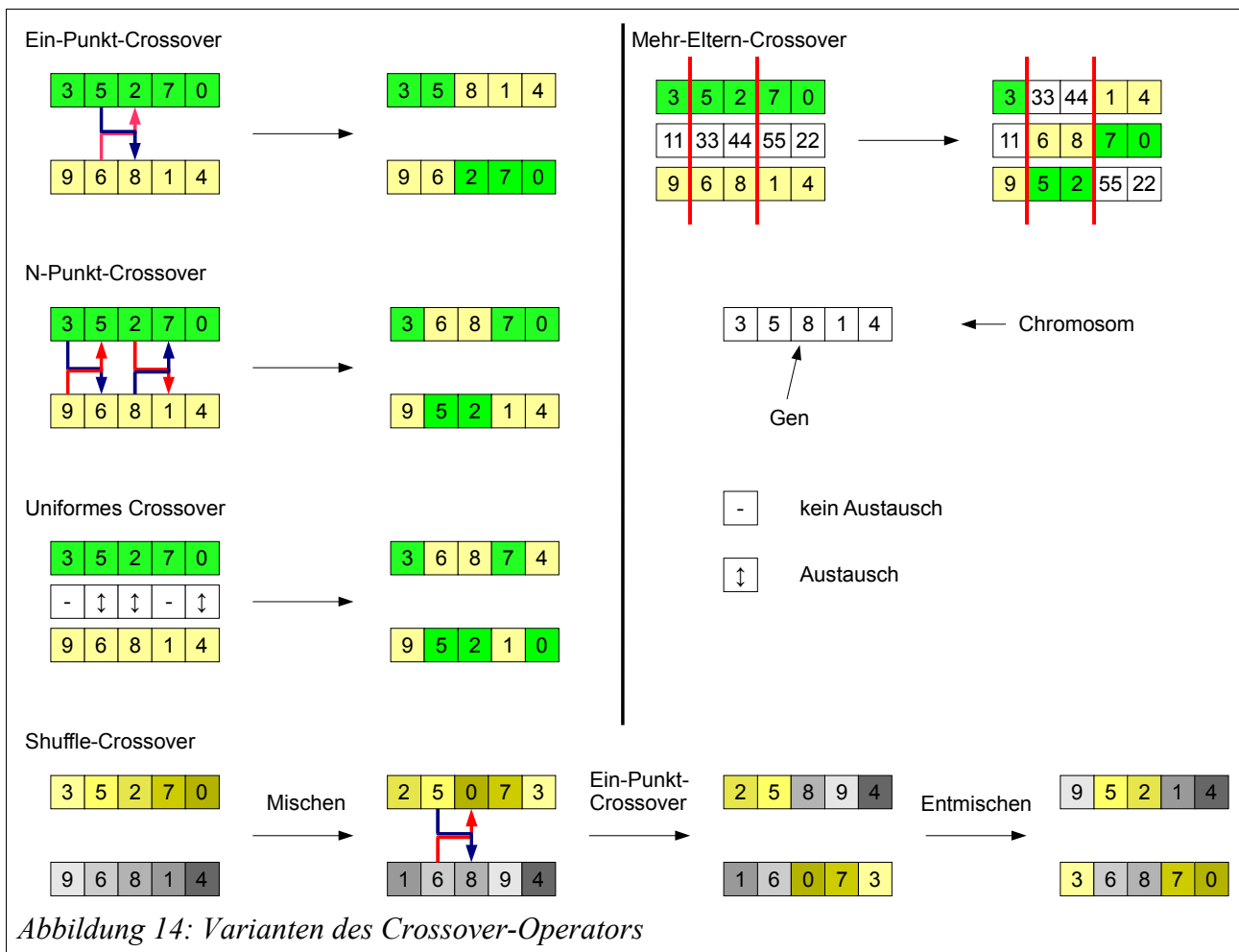
Der Mutationsoperator ändert Gene in zufälliger Weise. Es handelt sich um einen den Suchraum explorierenden Operator, da mit ihm Lösungen erzeugt werden, die so in der Population nicht vorhanden waren. Auf diese Weise wird die Diversität der Population gefördert. Bei binärer Codierung besteht die Mutation einfach in der Inversion zufälliger Bits mit einer gewissen Wahrscheinlichkeit, bei komplexen Datenstrukturen ist die Konstruktion eines speziellen auf das



Problem zugeschnittenen Operators erforderlich. Generell soll durch eine oder mehrere aufeinander folgende Mutationen jeder Punkt des Suchraumes erreicht werden können, andernfalls wäre die Vollständigkeit nicht gegeben. Die Mutation sollte nach Möglichkeit keine ungültigen Lösungen hervorbringen. Entsteht dennoch eine unerlaubte Lösung, so kann dieser Irrläufer entweder korrigiert, eliminiert oder markiert werden. Bei der Markierung wird die Eliminierung durch den Selektionsoperator erledigt, je nach Algorithmus kann die Lösung aber auch mit einer geringen Wahrscheinlichkeit an der Erzeugung von Nachkommen beteiligt sein. Getreu dem Motto „Natura non fecit Saltus“ werden die Operatoren meist so implementiert, dass die Mutationen um einen Punkt im Suchraum normal verteilt sind. Kleine Änderungen finden also häufiger statt als sehr große. Es erscheint logisch, dass eine kleine Änderung an einem guten Individuum eher eine Verbesserung bewirken kann als ein Sprung an eine vollkommen zufällige Stelle im Suchraum. Die Berücksichtigung der Normalverteilung ist jedoch nur möglich, wenn man es mit einem kontinuierlichen Suchraum zu tun hat. Es hat sich als vorteilhaft erwiesen, am Anfang des Suchprozesses größere Mutationen zuzulassen als am Ende.

Der Rekombinationsoperator

Der Rekombinationsoperator, auch „Crossover-Operator“ genannt, bringt Nachkommen durch den Austausch von Genen zwischen den Elternchromosomen hervor. Bei binärer Codierung kann die Rekombination einer Mutation entsprechen, wenn keine Vorkehrungen getroffen werden, um einen Schnitt mitten in einem Gen zu verhindern. Werden Gene durch reelle Zahlen dargestellt, sind auch arithmetische Kombinationen der Elterngene (z. B. als Linearkombination mit zufälligen Gewichten) möglich.



Jedes Individuum in einer Population hat in seinem Genom gute und weniger gute Gene. Die Rekombination nutzt in der Population vorhandene „gute“ Gene aus und beschleunigt damit die Konvergenz, es besteht allerdings die Gefahr, dass die Diversität der Population in diesem Fall nicht mehr gewahrt wird. Rekombination erscheint dann sinnvoll, wenn sich das Optimierungsproblem in voneinander weitgehend unabhängige Unterprobleme zerlegen lässt.

Die gängigen Varianten des Crossover-Operators sind in Abb. 14 gezeigt. Beim Einpunkt-Crossover werden ab einem zufälligen Punkt die Gensequenzen der Elternchromosomen untereinander vertauscht. Beim N-Punkt-Crossover werden die Gensequenzen entsprechend zwischen N Gensequenzen ausgewechselt. Beim uniformen Crossover wird für jedes Gen separat gelost, ob es mit dem entsprechenden Gen aus dem zweiten Elternchromosom substituiert wird oder nicht. Beim Shuffle-Crossover wird die Reihenfolge der Gene auf den Elternchromosomen zufällig verwürfelt und anschließend ein Einpunkt-Crossover durchgeführt. Nach dem Crossover werden die Gene auf den Nachkommenchromosomen wieder sortiert. Uniformes und Shuffle-Crossover sind nicht identisch: während beim Shuffle-Crossover jede Anzahl Genvertauschung gleich wahrscheinlich ist (es wird nur einmal bestimmt, wo der Crossover-Punkt liegt), ist sie beim uniformen Crossover binominal verteilt (für jedes Gen wird die Entscheidung separat getroffen). Beim Mehr-Eltern-Crossover werden Nachkommen aus den Chromosomen von mehr als zwei Eltern erzeugt. Das Beispiel zeigt ein diagonales Crossover zwischen drei Elternchromosomen.

Der Selektionsoperator

Der Selektionsoperator wählt aus einer Population diejenigen Individuen aus, die zur Erzeugung der nächsten Generation herangezogen werden. Die Selektion soll die durchschnittliche Fitness der Population verbessern. Selektionsdruck und Vielfalt müssen dabei gegeneinander abgewogen werden: Ist der Selektionsdruck zu hoch, wird die Population in ein lokales Optimum geraten; ist er zu gering, sinkt die Konvergenzgeschwindigkeit.

Man unterscheidet verschiedene Selektionsverfahren: Hängt die Selektion nur von der Fitness ab, spricht man von harter Selektion, bei der weichen Selektion haben dagegen auch weniger fitte Individuen eine Chance zur Reproduktion. Es folgt eine Beschreibung verbreiteter Selektionsalgorithmen.

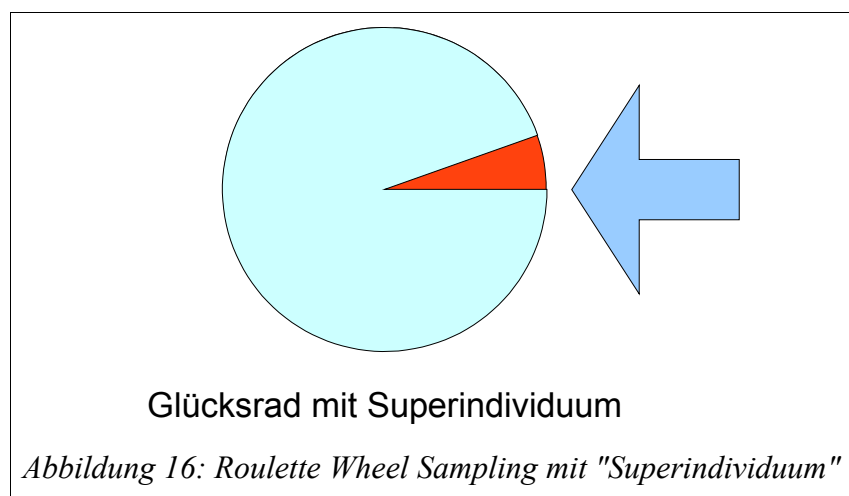
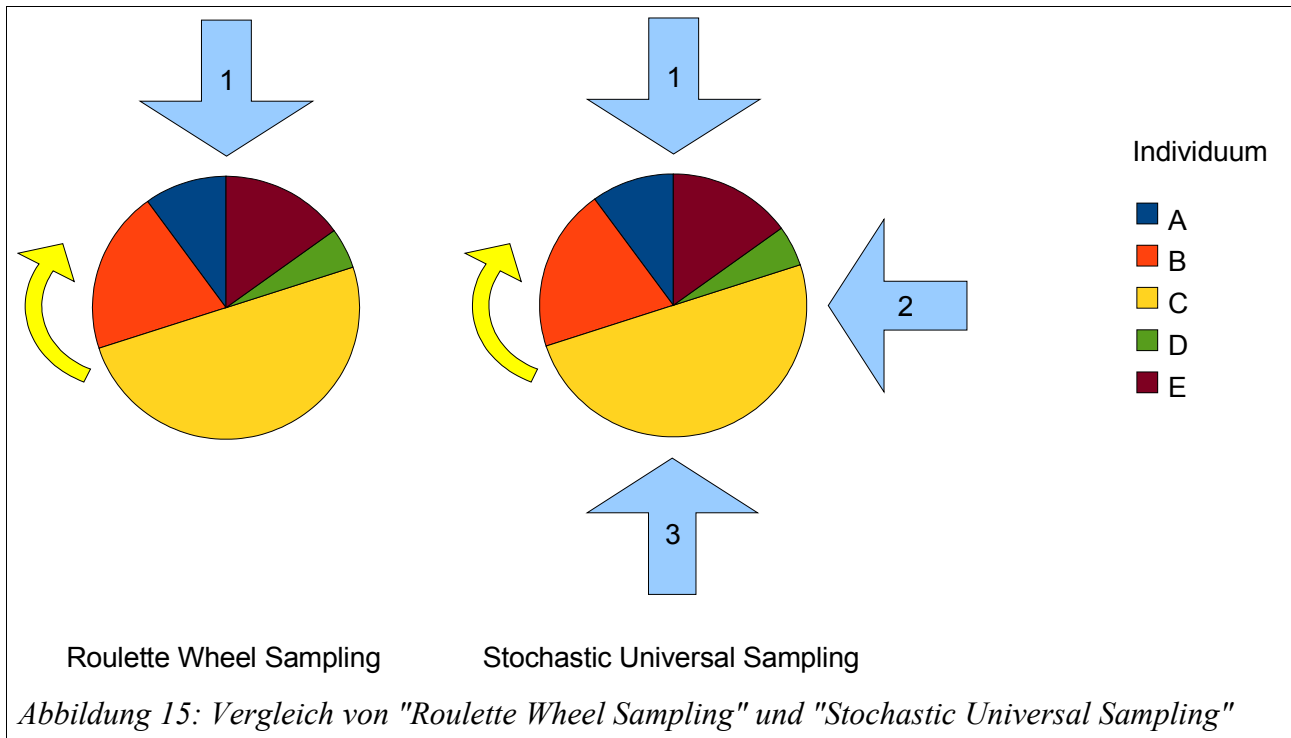
Roulette Wheel Sampling (RWS) und Stochastic Universal Sampling (SUS)

Beide Verfahren ahmen ein Glücksrad nach. Jedem Kandidat wird ein zu seiner relativen Fitness innerhalb der Population proportionales Segment zugeordnet. Beim RWS wird das Glücksrad für jedes zu selektierende Individuum einmal gedreht, beim SUS werden mit nur einer „Drehung“ alle Individuen durch Verwendung entsprechend vieler äquidistanter Zeiger selektiert (Abb. 15).

Die beiden Algorithmen wenden eine fitnessproportionale Selektion an und haben eine problematische Eigenschaft: Anfangs gibt es eine sehr große Bandbreite von Fitnesswerten in der

5 Optimierungsverfahren

Population. Einige wenige Individuen besetzen so fast den gesamten Umfang des Glücksrades und erzeugen so fast alleine die Nachfolgeneration. Im Extremfall kann fast das gesamte Glücksrad mit einem „Superindividuum“ (Abb. 16) besetzt werden. Die Diversität geht auf diese Weise schnell verloren. Nach einigen Generationen haben sich die Fitnesswerte innerhalb der Generation weitgehend angeglichen, und es findet kaum noch eine Selektion statt, da die Individuen ähnlich große Anteile des Glücksradumfangs besetzen.



$$Fitness_{\sigma}(Individuum) = Fitness(Individuum) - (\mu_f(t) - \beta \cdot \sigma_f(t))$$

$\mu_f(t)$	Mittelwert der Fitness von Generation t
$\sigma_f(t)$	Standardabweichung der Fitness von Generation t
β	Empirischer Parameter

Formel 5.3.1: Sigma Scaling

Eine diese Kritikpunkte entschärfende Maßnahme ist das Sigma Scaling (siehe Formel 5.3.1). Hier wird der Selektionsdruck über eine gewisse Zeit konstant gehalten, indem die Fitness mithilfe der Standardabweichung der Fitness in der betreffenden Population skaliert wird.

Die Rangauswahl

Bei der Rangauswahl werden die Individuen innerhalb einer Population nach ihrer Fitness sortiert und so eine Rangfolge aufgestellt. Entscheidend für die Reproduktionswahrscheinlichkeit ist nun der Rang und nicht mehr die absolute Fitness eines Individuums. Der Selektionsdruck bleibt über die Laufzeit konstant, das beste Individuum wird immer mit der gleichen Wahrscheinlichkeit ausgewählt.

Die Turnierauswahl

Einige Individuen werden aus der Population gezogen (mit oder ohne Zurücklegen) und zwischen diesen ein Turnier ausgetragen. Das Gewinnerindividuum erzeugt einen Nachkommen in der Zwischengeneration. Der Algorithmus vermeidet das Dominanzproblem, da die Fitness nicht direkt die Auswahlwahrscheinlichkeit eines Individuums steuert. Der Selektionsdruck kann über die Turniergröße gelenkt werden.

Elitismus

Es besteht die Gefahr, dass gute Individuen durch genetische Operatoren verschlechtert werden. Beim Elitismus werden die besten Individuen unverändert in die nächste Generation übernommen. Von der normalen Auswahl wird die Elite jedoch nicht ausgenommen, da sich durch die genetischen Operatoren noch bessere Individuen ergeben könnten.

Methoden zur Aufrechterhaltung der Diversität in einer Population

Um die frühzeitige Konvergenz in einem lokalen Optimum zu vermeiden, sollen sich die Individuen gut im Lösungsraum verteilen. Die gängigen Methoden hierfür werden kurz vorgestellt.

Fitness Sharing

Beim Fitness Sharing wird die Fitness eines Individuums verringert, wenn es ähnliche Individuen in der Population gibt. Dazu wird allerdings ein für den Suchraum anwendbares Distanzmaß benötigt. Dieses Verfahren wurde 1987 von Goldberg und Richardson [Grefenstette1987] vorgeschlagen.

Crowding

Die von De Jong [De_Jong1975] entwickelte Crowding-Methode ersetzt einen durch den Crowding Faktor bestimmten Anteil der Population durch Nachkommen, dabei ersetzt ein Nachkomme immer das ihm ähnlichste Mitglied der Population. Auf diese Weise haben die Nachkommen von weniger ähnlichen Individuen eine größere Überlebenschance. Ein Zusammenballen der Individuen im Suchraum wird so verhindert. Auch dieses Verfahren ist auf ein Distanzmaß angewiesen.

Immigration

Fleming und Fonseca [Fleming1998] wendeten das Konzept der Immigration an, bei dem ein gewisser Teil der Population durch zufällig generierte Individuen ersetzt wird.

Isolation by Distance

Bei diesen Methoden werden mehrere Subpopulationen entwickelt, zwischen denen es nur eine geringe Migration von Individuen gibt.

Overspecification

Der Ansatz „overspecification“ arbeitet mit redundanten Genen auf den Chromosomen. Eine separate Logik entscheidet, welche Kopie eines Gens gerade aktiv ist. Die inaktive Kopie ist zwar den genetischen Operatoren (Mutation, Rekombination) unterworfen, spielt jedoch bei der Selektion keine Rolle. Da das inaktive Gen jederzeit aktiv werden kann, können neue Individuen in der Population auftauchen, die nicht Resultat des Selektionsprozesses sind.

Behandlung von Einschränkung bei evolutionären Algorithmen

Evolutionäre Algorithmen suchen normalerweise im gesamten Suchraum. Soll dieser eingeschränkt werden, wenn z. B. Randbedingungen zu berücksichtigen sind, müssen diese separat behandelt werden. Im Folgenden werden einige gängige Techniken vorgestellt.

Letales Faktor

Eine sehr einfache Methode besteht darin, Individuen, die Randbedingungen verletzen, aus der Population zu entfernen. Sind die Randbedingungen zu restriktiv, existiert allerdings das Risiko, die komplette Population zu löschen, zumindest aber den größten Teil. Der Algorithmus wird dadurch unbrauchbar für das betreffende Problem.

Straffunktionen

Eine sehr populäre Methode ist die Reduzierung der Fitness eines Individuums in Abhängigkeit zur Verletzung von Randbedingungen. Straf- und Fitnessfunktion werden dazu in einer Funktion zusammengefasst. Möglicherweise dominiert der Einfluss des Strafanteils die Fitnessfunktion, sodass eher die Randbedingungen als die Zielfunktion optimiert werden. Eine mögliche Lösung ist die Anpassung des Strafanteils über die Iterationen hinweg.

Speziell angepasste Operatoren und Datenstrukturen

In einigen Fällen ist es möglich, „verbotene“ Bereiche des Suchraums schon vor der Suche zu entfernen. Der GENOCOP-Algorithmus [Michalewicz1999] startet mit einer Population, in der alle Randbedingungen erfüllt sind. Die Nachkommen werden durch Linearkombinationen erzeugt, die die Einhaltung der Randbedingungen garantieren. Die Randbedingungen selbst müssen allerdings linear sein, was die Anwendbarkeit der Methode sehr einschränkt.

Constraint Tournament Selection

Deb [Deb2000] hat den Ansatz der Constrained Tournament Selection basierend auf dem Prinzip, dass die Randbedingungen verletzende Lösungen immer schlechter als die Randbedingungen erfüllende Lösungen sind, konzipiert. Bei einer Turnierauswahl zwischen zwei Kandidaten werden folgende Regeln angewendet:

- Erfüllt eine Lösung die Randbedingungen nicht, die andere jedoch schon, wird erstere verworfen.

- Erfüllen beide Lösungen die Randbedingungen, wird diejenige mit der besseren Fitness gewählt.
- Verletzen beide Lösungen die Randbedingungen, wird die weniger außerhalb des erlaubten Suchraumes liegende Lösung gewählt.

Der Algorithmus ist auf ein Maß für die Verletzung der Randbedingungen angewiesen, dann allerdings allgemein anwendbar.

5.4 Evolutionsstrategische Algorithmen

Ingo Rechenberg und Hans-Paul Schwefel entwickelten die evolutionsstrategische Optimierung. Eine der ersten Anwendungen war die Optimierung von Flügelformen für Flugzeuge. In den 60er Jahren war man dazu noch darauf angewiesen, dies nach dem Trial-and-Error-Prinzip zu tun. Die Innovation der evolutionsstrategischen Optimierung besteht nun darin, ausgehend von einer bestehenden Form eine zufällige Änderung vorzunehmen. Ergibt sich eine Verbesserung, wiederholt man den Vorgang mit der neuen Form, andernfalls wird die neue Form verworfen und eine andere Änderung versucht. Dies funktioniert allerdings nur unter der Voraussetzung der starken Kausalität. Kleine Änderungen dürfen auch nur geringe Auswirkungen haben. Dies ist die Grundlage für die einfachste Variante des evolutionsstrategischen Algorithmus, dem (1+1) Algorithmus.

5.4.1 Der (1+1) evolutionsstrategische Algorithmus

Der Algorithmus arbeitet nicht mit einer Lösungspopulation und ist damit ein Spezialfall. Aus einem Startindividuum, das auch zufällig erzeugt werden kann, werden so lange durch Mutation Nachkommen erzeugt, bis die Abbruchkriterien erfüllt sind. Ein Nachkomme ersetzt den Vater, wenn er eine höhere Fitness aufweist, andernfalls wird er verworfen.

Im schlimmsten Fall kann der Algorithmus in einem lokalen Optimum stagnieren, das ist z. B. dann der Fall, wenn die Mutationsrate zu klein gewählt wurde. Eine zu groß gewählte Mutationsrate führt dazu, dass der Algorithmus sich von einer Position in der Nähe des globalen Optimums sehr weit von diesem entfernt.

Der Gefahr der Stagnation lässt sich durch das Arbeiten mit einer Population anstelle eines einzelnen Individuums begegnen. Dabei unterscheidet man die (m, n)-Variante, bei der Individuen aus der Elterngeneration nicht Teil der neuen Generation werden können, und der (m+n)-Variante, bei der dies der Fall ist.

5.4.2 (m, n) evolutionsstrategische Algorithmen

Das Verfahren arbeitet mit einer Population von Lösungen. Individuen für die Nachfolgegeneration werden mit den genetischen Operatoren „Mutation“ und „Rekombination“ generiert. Ausgehend von einer n Individuen großen Elterngeneration werden $m > n$ Nachfolgeindividuen erzeugt, von denen die n besten für die nächste Iteration herangezogen werden. Diesen Vorgang nennt man Reduktion, er unterscheidet sich von der Selektion bei den genetischen Algorithmen, die die Auswahl von Kandidaten zur Bildung der Nachfolgegeneration aus der aktuellen Population meint.

5.4.3 (m+n) evolutionsstrategische Algorithmen

Bei dieser Variante werden Individuen der Elterngeneration mit in die Nachfolgegeneration aufgenommen und konkurrieren mit ihren Nachkommen. Anders als bei der (m, n) Variante kann sich die Qualität der besten Lösung nicht verschlechtern, andererseits birgt dieses Vorgehen die Gefahr der Stagnation. Eine Mischform der beiden Algorithmen kann durch Einführen einer maximalen Lebensdauer der Individuen gebildet werden.

5.5 Genetische Algorithmen

Die klassischen genetischen Algorithmen wurden in den 70er Jahren von John Holland entwickelt. Ein Individuum wird durch eine Binärsequenz fester Länge dargestellt. Diese Sequenz beschreibt den Genotyp des Individuums. Jede mögliche Codierung entspricht einem Punkt im Suchraum (Phänotyp). Eine weitere Voraussetzung ist, dass jedem Punkt im Suchraum ein Wert zugewiesen werden kann, der dessen Güte kennzeichnet. Logische, ganzzahlige und reelle Parameter des Phänotyps müssen durch eine entsprechende Codierung berücksichtigt werden, die genetischen Operatoren arbeiten ohne Wissen über diese Zusammenhänge „blind“ mit den Bitstrings. Der Algorithmus kann so unabhängig von einer konkreten Anwendung implementiert werden, eine Anpassung muss dann allerdings bei der Codierung und Decodierung des Genotyps erfolgen.

Die binäre Codierung verursacht eine Reihe von Nachteilen. Zum Beispiel hängt die Wertänderung einer Binärzahl bei einer Mutation erheblich von der Position des geänderten Bits ab. Dieser Nachteil lässt sich durch Verwendung einer angepassten Codierung, dem Gray Code, umgehen. Weiterhin lassen sich die genetischen Operatoren nur schwer an das gestellte Problem anpassen, eine ungünstige Codierung kann die Suche sogar unnötig verkomplizieren.

Die Codierung soll sich möglichst direkt aus der Problemstellung ableiten und Eigenheiten (Regelmäßigkeiten) des Lösungsraumes sollen durch die Codierung erhalten bleiben, sofern sie den Suchprozess erleichtern. Vermieden werden müssen den Algorithmus „irreführende“ Codierungen.

Hollands Schema-Theorem versucht, die Leistungsfähigkeit der genetischen Algorithmen zu erklären. Unter einem Schema versteht man einen Bitstring, bei dem an bestimmten Stellen beliebige Zeichen erlaubt sind. Folgende Begrifflichkeiten werden verwendet:

- Die Ordnung eines Schemas ist die Zahl der definierten Bits im Schema.
- Die definierende Länge eines Schemas ist die Distanz zwischen dem ersten und dem letzten definierten Bit in einem Schema.
- Die Fitness eines Schemas ist der Mittelwert der Fitnesswerte aller Genotypen, die entstehen, wenn man für die beliebigen Zeichen im Bitstring alle möglichen Kombinationen bildet.

Nach dem Schema-Theorem haben bei Anwendung der klassischen genetischen Algorithmen nun solche Schemata eine besonders große Chance, sich in der nächsten Generation zu reproduzieren, die

- eine überdurchschnittliche Fitness aufweisen,
- eine niedrige Ordnung haben,
- und eine kleine definierte Länge besitzen.

Diese Eigenschaften lassen sich erzielen, wenn inhaltlich zusammenhängende Teile eines Chromosoms zusammenhängend und möglichst kompakt codiert werden. Die Codierung selbst sollte über eine möglichst geringe Kardinalität (Anzahl der zur Codierung verwendeten Zeichen) verfügen, was für die Binärcodierung am besten erfüllt ist.

Das Schema-Theorem wird jedoch oft kritisiert, da es streng nur für unendlich große Populationen gilt und die Erwartungswerte der Fitness eines Schemas für endliche Populationen erheblich von der realen Fitness abweichen können. Es wird weiterhin vorausgesetzt, dass die Fitness der zu einem Schema passenden Chromosomen sehr ähnlich ist.

5.6 Verfeinerungstechniken

Die Verfeinerungstechniken haben gemeinsam, dass zunächst alle Datenpunkte mit den neuen Kraftfeldparametern berechnet werden und dann der Wert der Straf- bzw. Bewertungsfunktion

bestimmt wird. Anschließend werden die Parameter variiert und der Zyklus wiederholt, bis Abbruchkriterien erfüllt sind. Das Verfahren ähnelt dem der Geometrieoptimierung. Man kann verschiedene Klassen dieser Techniken unterscheiden:

Methoden, die ohne Gradienten arbeiten

Die systematische Suche testet alle möglichen Parameterkombinationen systematisch durch. Dies ist die einfachste mögliche Implementierung. Nachteil ist der exponentiell mit der Parameterzahl steigende Aufwand, deshalb ist dieses Verfahren nur für kleine Parametersätze sinnvoll.

Daneben sind stochastische Verfahren wie die Monte-Carlo-Suche verbreitet. Hier werden die Parameter zufällig variiert. Dies ist einfach zu implementieren und kann große Suchräume gut abdecken, allerdings ist die Konvergenz recht langsam. Sehr verbreitet ist der Einsatz genetischer Algorithmen. Zunächst werden mehrere Parametersätze abgeschätzt. Zusätzlich werden Regeln benötigt, um die Parametersätze zu kombinieren und zu variieren. Damit werden aus der Parametergeneration Nachkommen erzeugt. Die Auswahl der Nachkommen kann nach der Bewertungsfunktion und anderen Kriterien erfolgen.

Methoden, die mit Gradienten der Bewertungsfunktion arbeiten

Der Gradient der Bewertungsfunktion in Abhängigkeit von den Parametern gibt die optimale Richtung für eine Verbesserung an. Stehen zusätzlich die zweiten Ableitungen zur Verfügung (Hesse-Matrix), kann die Konvergenz sogar noch beschleunigt werden. Die Ableitung der Bewertungsfunktion bereitet allerdings spezifische Probleme. Viele experimentell bestimmte Werte basieren auf Energieminima. Für jede Parameteränderung ist deshalb eine erneute Geometrieoptimierung des kompletten Referenzdatensatzes erforderlich. In dieser Situation ist eine analytische Ableitung nach den Parametern kaum möglich, die daher notwendige numerische Differenzierung vervielfacht die Geometrieoptimierung. Vorteil dieser Methode ist dagegen die besonders genaue Wiedergabe der Kraftfeldeigenschaften. Es existieren Näherungen für die Ableitungen der Bewertungsfunktion, die die erwarteten Geometrieänderungen für kleine Parameteränderungen enthalten.

5.7 Probleme bei der automatischen Parametrisierung

Parameter in einem Kraftfeld hängen mehr oder weniger voneinander ab. Zwei Parameter sind voneinander abhängig, wenn sich eine Kombination von Parameteränderungen finden lässt, die keine Änderung bei den berechneten Datenpunkten hervorruft. Ein einfaches Beispiel hierfür sind

die Winkel um ein Zentralatom. Betrachtet man Formaldehyd, kann die Erhöhung des H-C-H Referenzwinkels immer durch die Erhöhung des O=C-H-Referenzwinkels kompensiert werden.

Wendet man bei einer Verfeinerung die Methode der kleinsten Fehlerquadrate an, ergeben sich wegen der Abhängigkeit Singularitäten. Löst man das Problem durch Anwendung der Singulärwertzerlegung (SVD), so sorgt man lediglich dafür, dass voneinander abhängige Parameter in der Nähe ihrer Startwerte bleiben. Abhängige und schlecht definierte Parameter können mit einer harmonischen „Straffunktion“ in der Nähe von sinnvoll per Hand gewählten Werten gehalten werden.

Validierung der Parameter

Zunächst muss überprüft werden, ob das Kraftfeld innerhalb der zur Parametrisierung herangezogenen Referenzdaten (Trainingsdatensatz) eine befriedigende Genauigkeit erreicht. Dazu können die bei der Optimierung verwendete Bewertungsfunktion und andere statistische Größen wie der RMSD- oder MAD-Wert dienen. Einzelne besonders große Abweichungen sollten individuell begutachtet werden. Abweichungen können durch eine ungeeignete mathematische Form des Kraftfeldes, Fehler im Parametrisierungsprozess oder Fehler in den Referenzdaten zustande kommen. Wichtig ist auch die Betrachtung des Einflusses einzelner Parameter auf die Qualität des Kraftfeldes. Dabei sollten auch Abhängigkeiten zwischen verschiedenen Parametern und den Fehlern der Referenzdaten mit einbezogen werden. Die Übertragbarkeit des Parametersatzes kann durch Evaluierung des Kraftfeldes mit Referenzdaten, die nicht für den Parametrisierungsprozess verwendet wurden, geschehen. Man unterscheidet zwischen der Kreuzvalidierung und der externen Validierung. Bei der Kreuzvalidierung teilt man die verfügbaren Referenzdaten in mehrere Gruppen auf und schließt wechselseitig eine Gruppe aus und verfeinert die Parameter mit den verbleibenden Gruppen. Nachteilig wirkt sich hier aus, dass die mathematische Form im Hinblick auf alle Referenzdaten beurteilt worden ist, bevor die Validierung durchgeführt wird. Eine zuverlässige Beurteilung ist nur bei der externen Validierung mit einem separaten Testdatensatz möglich.

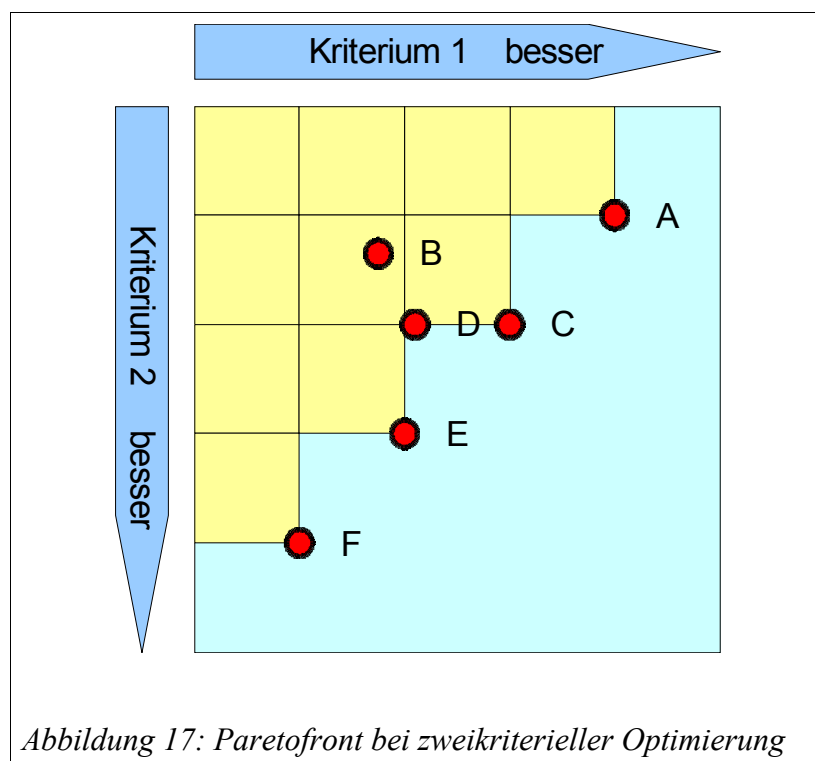
5.8 Mehrkriterielle Optimierung

Bei der Behandlung von Problemen, bei denen mehrere Kriterien optimiert werden müssen, die sich unter Umständen widersprechen, gilt es einen guten Kompromiss zu finden. Keine Lösung kann von vornherein als optimal angesehen werden, es existiert aber eine Menge von Lösungen, die in der Hinsicht optimal sind, dass sich keine Eigenschaft der Lösung verbessern lässt, ohne dies auf

5 Optimierungsverfahren

Kosten einer anderen Eigenschaft zu tun. Diese Menge stellt eine Grenze im Suchraum dar - die sogenannte Paretofront. Optimierungsalgorithmen sollen nun Lösungen finden, die möglichst nahe an dieser Grenze liegen.

Voraussetzung für ein Optimierungsverfahren ist ein Maß für die Lösungsqualität. Im einfachsten Fall ist dies eine gewichtete Summe über alle zu optimierenden Kriterien. Das Problem wird so auf eine einkriterielle Aufgabenstellung zurückgeführt und kann mit den bereits besprochenen Verfahren gelöst werden. In den Gewichten steckt dabei schon die Auswahl einer Lösung auf der Paretofront. Um eine fundierte Entscheidung treffen zu können, erscheint aber eine nachgelagerte Entscheidungsmöglichkeit zwischen verschiedenen paretooptimalen Lösungen wünschenswert. Eine Möglichkeit, um die Lösungen miteinander zu vergleichen, ist das Konzept der Dominanz. Dies soll anhand der Optimierung nach zwei Kriterien, wie in Abb. 17 gezeigt, demonstriert werden:



Lösung C ist in beiden Kriterien besser als Lösung B ($C > B$), C dominiert also B.

Lösung C ist nur bei einem Kriterium besser als Lösung D, das zweite Kriterium ist gleich ($C \geq D$). Hier dominiert C schwach über D.

5 Optimierungsverfahren

Beim Vergleich zwischen den Lösungen A und F ist mal A und mal F besser, man bezeichnet diese Lösungen als invariant zueinander ($A \sim F$).

Die nicht-dominierten Lösungen A, C, E und F spannen den dominierten Bereich des Suchraums auf, der gelb markiert ist.

6 Das MOMO-Programmsystem

MOMO wird seit 1986 in unserem Arbeitskreis entwickelt. Es basiert auf dem PIMM85-Kraftfeld und dient als Umgebung, um verschiedene Erweiterungen implementieren zu können. Abb. 18 veranschaulicht den logischen Programmaufbau.

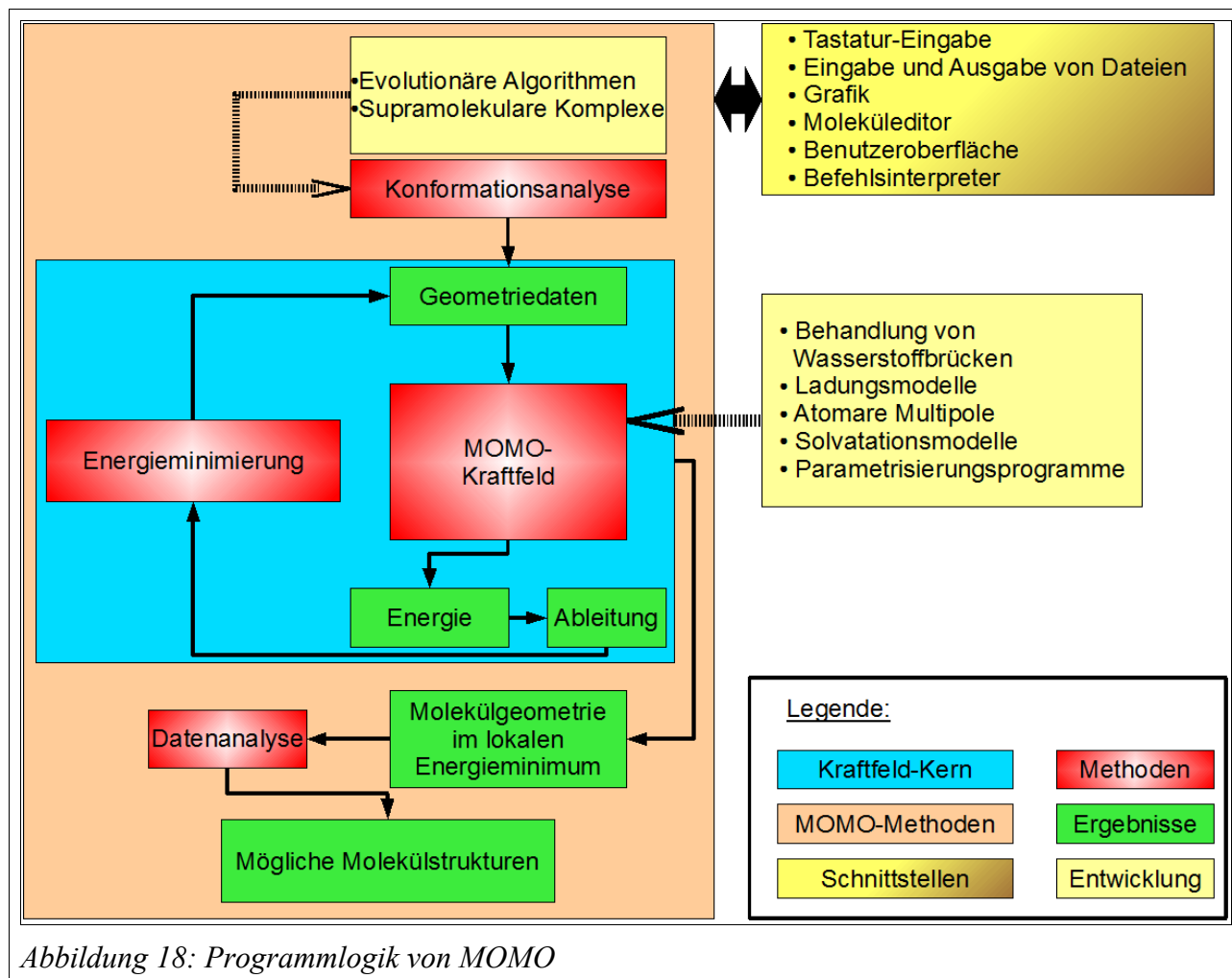


Abbildung 18: Programmlogik von MOMO

Kern des Programms ist das MOMO-Kraftfeld. Hier ist das mathematische Modell für die Energieberechnung aus der Molekülstruktur durch die verschiedenen Kraftfeldterme implementiert. Für von den Potenzialfunktionen benötigte Hilfsgrößen existieren separate Module, die etwa Parameter einlesen oder die Zerlegung des Moleküls im Sinne der vorhandenen Potenzialfunktionen übernehmen. Methoden zur Energieminimierung von Strukturen nutzen direkt das Kraftfeld und die von ihm gelieferten partiellen Ableitungen. Die Programmteile kommunizieren über einen großen gemeinsam benutzten Speicherbereich, der für die Steuerung des Programmablaufs benötigte

Informationen, die Daten der aktuellen Berechnung und die Repräsentation des betrachteten Molekülsystems enthält. Das Kraftfeldmodell wird von einer zweiten Programmebene, den MOMO-Methoden, genutzt. Dazu gehören Programme zur Konformationsanalyse von einzelnen Molekülen und supramolekularen Komplexen sowie zur Datenanalyse. Das Programm verfügt über Schnittstellen zur Steuerung über eine grafische Benutzeroberfläche und eine Kommandosprache. Die Steuerung über Kommandos ist auch durch Textdateien möglich. Die Kommandosprache erlaubt die Änderung von Kraftfeldparametern, die Auswahl alternativer Berechnungsverfahren im Programm, den Aufbau von Molekülsystemen und das Starten von Programmteilen für spezielle Aufgaben mit entsprechenden Parametern. Der Befehlsinterpreter ist innerhalb der grafischen Benutzeroberfläche über ein Shell-Fenster zugänglich.

6.1 Die MOMO-Versionen

Immer wieder ist es erforderlich gewesen, MOMO an die technischen Gegebenheiten und Anforderungen der bearbeiteten Aufgabenstellung anzupassen. Die ursprüngliche Version wurde ohne grafische Benutzeroberfläche als Fortran-Programm für Unix Workstations realisiert. Im Zuge der Verbreitung von MS-DOS und Windows erhielt MOMO durch Egon Gemmel [Gemmel1999] eine grafische Benutzeroberfläche und wurde für DOS portiert. Dabei gelangte eine von der Firma Lahey vertriebene Grafikbibliothek zum Einsatz. Aufgrund der hohen Verbreitung von Windows rückte die leichte Übertragbarkeit zwischen verschiedenen Systemen in den Hintergrund. Zwar ist das Programm prinzipiell weiter übertragbar, man ist jedoch auf die Verfügbarkeit einer entsprechenden Version der Grafikbibliothek angewiesen. Im Rahmen meiner Diplomarbeit erfolgte eine weitere Überarbeitung des MOMO-Programms. Im Laufe der Jahre hatten sich einige unterschiedliche, für die Projekte verschiedener Mitarbeiter erweiterte Versionen des Programms angesammelt, die nicht in die veröffentlichte Version aufgenommen worden waren. Da die Problemstellung meiner damaligen Arbeit, die Evaluierung der in MOMO enthaltenen empirischen Punktladungsmodelle, sich auf das gesamte Kraftfeld auswirkt, wurde eine Sammelversion von MOMO erstellt. Dazu ergaben sich einige technische und organisatorische Probleme wie eine fehlende Dokumentation der Programmquellen, undokumentierte Befehle und unter speziellen Bedingungen auftretende Programmfehler, die auch noch im Laufe dieser Arbeit durch viele überarbeitete Versionen korrigiert werden mussten. Die Verfügbarkeit von DOS und Windows 98 im Kontext mit modernen Rechnern und Entwicklungswerkzeugen verschlechterte sich ebenfalls. Da von der Grafikbibliothek eine Version für Linux existierte, wurde MOMO auf dieses System portiert. Nebenbei ergab sich so ein großer Geschwindigkeitsvorteil und vor allem die detaillierte

Kenntnis der Programminterna. Auch dieser Schritt löste nicht alle Probleme. Die Grafikbibliothek wird nicht mehr vom Hersteller gepflegt, und deren Kompatibilität mit neueren Versionen von Linux verschlechtert sich zunehmend. MOMO eignet sich zwar prinzipiell als Grundlage für spezielle Erweiterungen, erzwingt aber dennoch die Pflege eines kompletten Programmpaketes inklusive aller nicht-wissenschaftlichen Aspekte. Eine Lösung der Problematik bietet die Verwendung von auf das Molecular Modelling zugeschnittenen und auch allgemeinen Programmbibliotheken. Hiervon wurde in der vorliegenden Arbeit ausgiebig Gebrauch gemacht. Die hierbei vorgenommenen Arbeiten werden in den nächsten Kapiteln ausführlich beschrieben.

6.2 Anforderungen an Bibliotheken zum Molecular Modelling

Der Einsatz von Computersimulationen in der Chemie steigt immer mehr an, ständig werden neue Techniken entwickelt, die meistens jedoch viel Zeit brauchen, bis sie für die Allgemeinheit in einer einfach zu verwendenden Implementierung vorliegen, wenn dies überhaupt der Fall ist. Sowohl technische als auch rechtliche Gründe verhindern oft die Implementierung und Modifizierung von Simulationsprogrammen durch den Wissenschaftler:

- Simulationsprogramme sind häufig umfangreich und sehr komplex, eine Modifikation durch nicht an deren Entwicklung Beteiligte ist daher schwierig.
- Die Lizenzbestimmungen verbieten oftmals die Weitergabe modifizierter Versionen an andere Benutzer.

Eine Lösung für die rechtlichen Probleme bietet Open Source Software; die technischen Hindernisse, die vor allem aus Umfang und Komplexität erwachsen, können durch Arbeitsteilung und moderne Entwicklungsmethoden überwunden werden. Von großer Bedeutung sind in diesem Zusammenhang das objektorientierte Programmdesign und der Einsatz von high-level-Programmiersprachen. Beim objektorientierten Programmdesign wird Software um die Datenstrukturen herum entwickelt und nicht wie sonst üblich um die Reihenfolge der Datenverarbeitung; high-level-Programmiersprachen bieten an typische Problemstellungen angepasste Datenstrukturen, low-level-Sprachen haben an Rechner angepasste Datenstrukturen. Programme in high-level-Sprachen sind kürzer, was zu kürzeren Entwicklungszeiten, weniger Programmierfehlern und leichter lesbarem Code führt. Wünschenswert wäre folglich eine Bibliothek, welche

- die gängigsten Simulationsmethoden implementiert,

- moderne Programmier Techniken nutzt (high-level-Sprachen und objektorientiertes Programmdesign),
- leicht zu erweitern und mit anderen Bibliotheken kombinierbar ist,
- die Verwendung einer gemeinsamen high-level-Sprache für Erweiterungen und Anwendungen erlaubt,
- gut dokumentierte und geräteunabhängige Dateiformate nutzt,
- Schnittstellen zu anderen Simulations- und Visualisierungsprogrammen mitbringt,
- Lizenzbestimmungen unterliegt, die eine problemlose Weitergabe an andere Benutzer ermöglichen.

Diese Voraussetzungen werden vom Molecular Modelling Toolkit von Konrad Hinsen, MMTK abgekürzt [Hinsen2000], erfüllt:

MMTK wurde in der high-level-Sprache Python programmiert. Python bietet eine gute Unterstützung für ein modulares und objektorientiertes Programmdesign und hat flexible und systemübergreifend nutzbare Schnittstellen zu low-level-Programmiersprachen wie C oder Fortran; low-level-Sprachen können dazu dienen, um zeitkritische Programmteile zu implementieren. Der modulare Aufbau ermöglicht theoretisch die Programmierung neuer Methoden und sogar ganzer Kraftfelder, ohne den vorhandenen Quelltext kennen zu müssen, wie so oft hängt dies aber von der Verfügbarkeit einer entsprechend ausführlichen Dokumentation ab.

Für die vorliegende Arbeit war eine Analyse des MMTK-Quellcodes von Hand erforderlich.

6.3 Das Molecular Modelling Toolkit

Die erste MMTK Version wurde 1997 veröffentlicht [Hinsen1997] und bereits für einige Forschungsprojekte eingesetzt [Hinsen1998] [Thomas1999] [Hinsen1999]. Auch ein interaktives Analyseprogramm nutzt MMTK.

Chemische Systeme werden in MMTK als Objekte abgebildet. Mit Objekt ist hier eine Datenstruktur gemeint, die Daten und die dazu gehörenden Operatoren in Form von angepassten Unterprogrammen zusammenfasst. Die Objekte können reale Objekte wie Moleküle oder abstrakte Dinge wie Listen darstellen. Die Definition einer Objektart nennt man Klasse. MMTK hält eine Reihe von Klassen bereit, um Moleküle entsprechend den Gewohnheiten des Chemikers zu beschreiben: Atomobjekte werden durch Bindungen verknüpft und bilden Molekülobjekte. Diese

können zu Komplexen kombiniert werden. Objekte können allgemein Gruppenobjekten zugeordnet werden, z. B. Atome zu einer funktionellen Gruppe. Die zentrale Objektklasse in MMTK ist das „universe“. Es beinhaltet ein komplettes chemisches System aus Molekülen, einer Geometrie, einem Kraftfeld und einer Umgebung (Druck, Temperatur).

Der objektorientierte Aufbau vereinfacht den Zugriff auf Objekte in allen Ebenen. Atome können z. B. durch den Weg entlang der Hierarchie Universe → Molekül → Atom angesprochen werden. Da Daten und Unterprogramme zu deren Bearbeitung und Abfrage immer Teil eines Objektes sind, ist automatisch für Kompatibilität gesorgt. Eine Abfrage der Energie für ein Universe nutzt so automatisch das richtige Berechnungsverfahren, egal, was für ein Typ (unendlich oder periodisch) vorliegt.

MMTK bietet die Möglichkeit, alle Objekte oder Kombinationen von Objekten in Dateien zu speichern. Da alle Daten in Python-Objekten vorhanden sind, können die entsprechenden zu Python gehörenden Methoden benutzt werden. Trajektorien, die den Verlauf einer Simulation Schritt für Schritt speichern, können zu umfangreich werden, um sie im Arbeitsspeicher zu erzeugen und dann in eine Datei zu schreiben. Hier müssen direkt Dateien angelegt werden. Zudem ist es notwendig, einen effizienten Zugriff auf Teile der möglicherweise sehr umfangreichen Daten zu realisieren. Glücklicherweise gibt es Bibliotheken zum effizienten Umgang mit großen Datenmengen. MMTK gebraucht das netCDF-Format [Rew1990] für Trajektorien. Die Dateien sind systemunabhängig und können eine Beschreibung der enthaltenen Daten mit aufnehmen. Daneben werden das PDB-Format [Berman2000] und das DCD-Format [Brooks1983] für Trajektorien unterstützt. Andere Dateiformate lassen sich leicht hinzufügen.

Python wurde nicht nur genutzt, um große Teile von MMTK zu implementieren, es dient auch dazu, das Programm zu bedienen. Alle Anwendungen werden als Python-Skript programmiert. Durch die Nutzung von Bibliotheken können diese Programme sehr einfach ausfallen, sodass gegenüber einem Modellingprogramm mit eigener Kommandosprache kein Nachteil im Hinblick auf die Benutzerfreundlichkeit entstehen muss. Auch eine Nutzung im „interaktiven Modus“ (Python Shell) ist möglich.

Kraftfelder sind die Basis molekularer Simulationen. MMTK unterstützt die Verwendung mehrerer Kraftfelder und ermöglicht das einfache Hinzufügen von Energiebeiträgen (z. B. harmonische Potenziale, um einen bestimmten Abstand zu fixieren). Auch Kraftfelder werden als Klassen realisiert. Ein neues Kraftfeld kann also als zusätzliche Klasse implementiert werden. Dieser

Aufbau macht es leicht, Kraftfelder zu kombinieren oder zu modifizieren. Ein Amber94-Kraftfeld und einige „Einfachst-Kraftfelder“ sind Teil des Paketes.

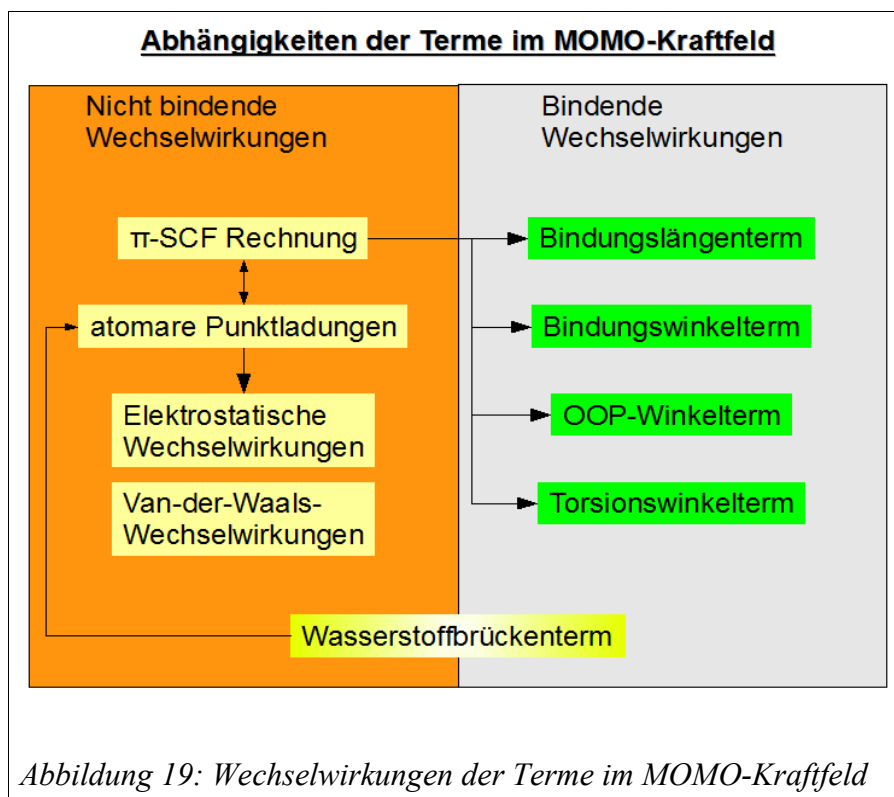
MMTK beinhaltet eine Reihe von Standardalgorithmen, dazu gehören:

- grundlegende Operationen für den Aufbau, die Analyse und die Modifizierung molekularer Systeme und Koordinatenmanipulation
- Energieminimierung (steepest descent und conjugate gradients)
- molekulare Dynamik (Velocity-Verlet)
- Fit atomarer Punktladungen
- Berechnung molekularer Oberflächen

Visualisierungsaufgaben werden über externe Programme realisiert. Dies ist über PDB- oder VRML-Dateien möglich, direkt unterstützt werden die Programme PyMol [DeLano] und VMD [Humphrey1996], mit denen Trajektorien animiert und molekulare Grafiken mit anderen Grafikobjekten kombiniert werden können.

6.4 Die Implementierung von MOMO für MMTK

6.4.1 Vorüberlegungen

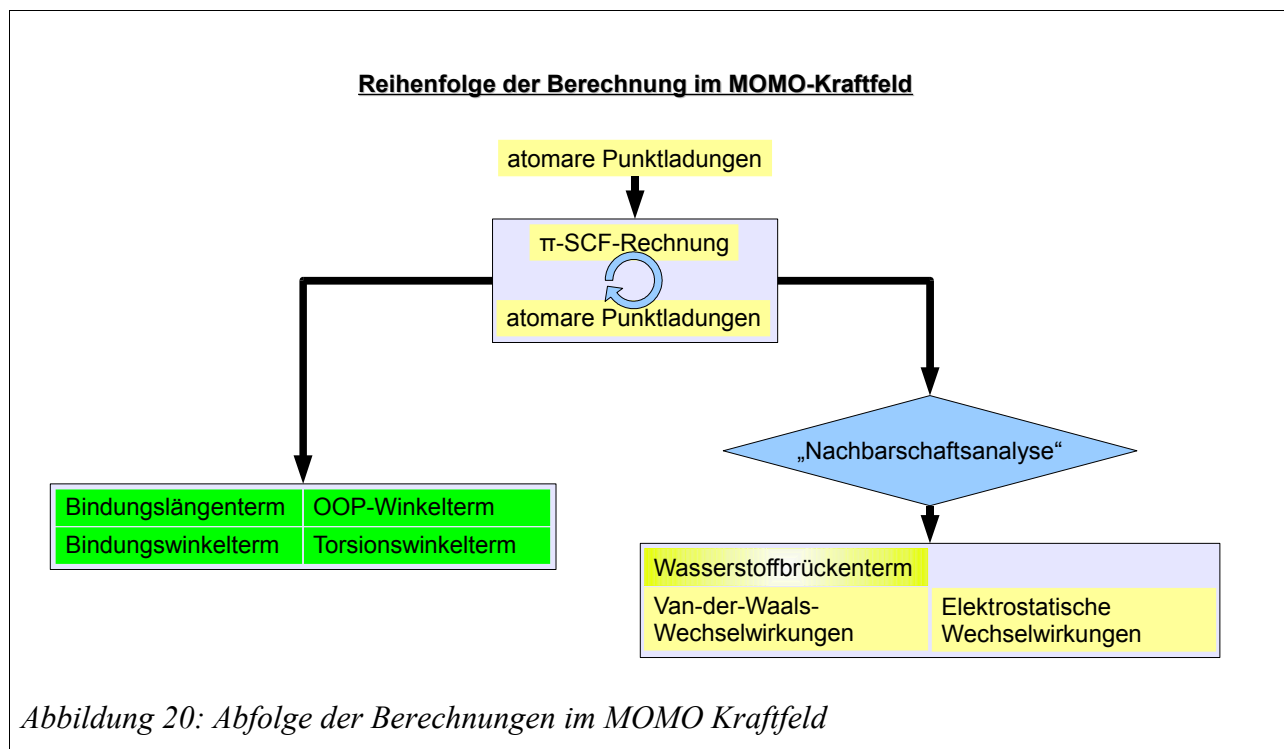


Das MOMO-Kraftfeld kommt mit vergleichsweise wenigen Parametern aus, da Potenzialfunktionen in Abhängigkeit von den Ergebnissen der π -SCF-Rechnung und von der Umgebung eines Atoms modifiziert werden. Abb. 19 zeigt die Einflüsse der Terme in MOMO aufeinander. Die einzelnen Terme von MOMO müssen daher in einer bestimmten Reihenfolge berechnet werden, damit Zwischenergebnisse rechtzeitig vorliegen. Die Prioritäten im Kraftfeld sind in Abb. 20 ersichtlich. Die Ergebnisse der π -SCF-Rechnung werden für fast alle anderen Terme im MOMO-Kraftfeld gebraucht. Voraussetzung für die π -SCF-Rechnung ist wiederum ein initialer Satz atomarer Punktladungen. Da sich beides gegenseitig beeinflusst, muss eine iterative Berechnung der zwei Terme erfolgen. Der Wasserstoffbrückenterm bewirkt eine kleine Ladungsverschiebung von 0.0125 Elementarladungen vom Wasserstoff auf den Akzeptor, auf weitere Iterationen der Ladungsmodelle wird jedoch wegen des kleinen Beitrags verzichtet. Liegen diese Ergebnisse vor, so können die elektrostatischen Wechselwirkungen berechnet werden. Die Van-der-Waals-Wechselwirkungen sind nicht von den Ergebnissen anderer Kraftfeldterme abhängig. Mithilfe der durch die π -SCF-Rechnung erhaltenen Bindungsordnungen können nun die Parameter für die bindenden

Wechselwirkungen modifiziert werden. Daraus ergeben sich folgende Prioritäten für die Berechnungen in einem Kraftfeldzyklus:

Zuerst muss eine Ladungsberechnung vorgenommen werden, daraufhin dann eine π -SCF-Rechnung, die ihrerseits wieder die Ladungsberechnung starten kann, bis π - und σ -Ladungen konvergieren. Jetzt können die bindenden Wechselwirkungen, die van-der-Waals-Wechselwirkungen und die Wasserstoffbrücken berechnet werden. Ist die Ladungsübertragung durch den Wasserstoffbrückenterm abgeschlossen, können schließlich auch die elektrostatischen Wechselwirkungen ermittelt werden.

Nichtbindende Potenzialfunktionen und die Wasserstoffbrückenfunktion erfordern technisch alle die Suche nach geeigneten Wechselwirkungspartnern in ihrer näheren Umgebung. Es ist daher vorteilhaft, diesen Suchprozess gesammelt durchzuführen.

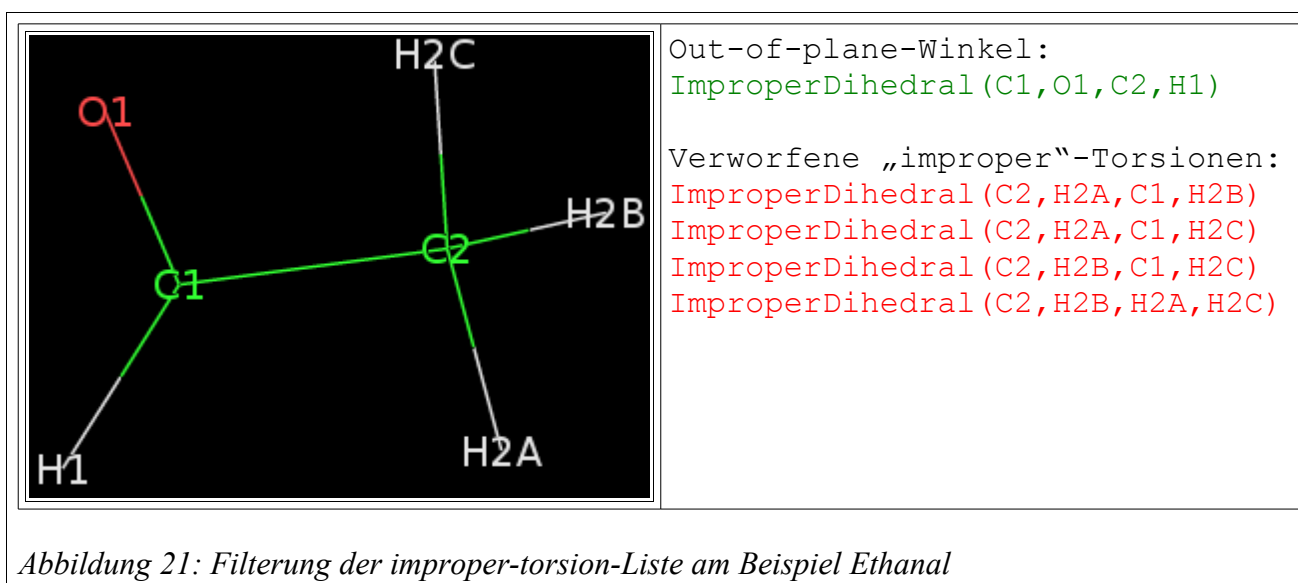


6.4.2 Aufbereitung des molekularen Systems für das MOMO-Kraftfeld

Neben der gedanklichen Zerlegung des Molekülsystems in Atome, Bindungen, Bindungswinkel, Torsionswinkel und out-of-plane-Winkel sind noch weitergehende Zerlegungen erforderlich, für die effiziente Algorithmen entwickelt wurden. Für die π -SCF-Rechnung müssen π -Systeme im Molekül gefunden werden, für 3-, 4- und 5-Ringe gibt es eine spezielle Behandlung der Bindungswinkel.

Eine weitere Besonderheit ist eine Matrix, in der einzelne nichtbindende Wechselwirkungen zwischen Atompaaaren ausgeschlossen werden können. So werden z. B. zwischen an derselben Bindung beteiligten Atomen keine van-der-Waals- oder Coulomb-Kräfte berechnet. Nach dieser Vorarbeit können schließlich die einzelnen Potenzialfunktionen mit Parametern versorgt und in einer Warteschlange zur Abarbeitung abgelegt werden. Die Aufbereitung wird in den nächsten Unterkapiteln genauer beschrieben.

Von MMTK werden Listen über Bindungen, Bindungswinkel, Torsionswinkel und „improper“-Torsionswinkel geliefert. Letztere können so nicht von MOMO übernommen werden, da hier mit einem out-of-plane-Potenzial gearbeitet wird. Eine Liste der out-of-plane-Winkel lässt sich jedoch durch Filterung der improper-Torsion-Liste erhalten: Das Zentralatom muss genau drei Substituenten haben. Die Liste von MMTK enthält auch nicht-trigonale Zentren (Abb. 21).



Je nach den paarweise zwischen den Atomen wirkenden Kraftfeldtermen wird nun die Wechselwirkungsmatrix befüllt. Die Werte aus Tabelle 4 werden jeweils addiert bzw. subtrahiert. Da diese Werte Zweierpotenzen entsprechen, kann im Programm einfach das entsprechende Bit abgefragt werden, um zu prüfen, ob ein Attribut gesetzt ist. Aus praktischen Gründen werden auch einige häufig benötigte Bitkombinationen in der Tabelle angeboten.

Neben den paarweisen Wechselwirkungen zwischen Atomen werden die einzelnen Atome mit speziellen Attributen markiert. Tabelle 5 zeigt die möglichen Attribute, Tabelle 6 listet die Standardattribute für die in MOMO parametrisierten Atomtypen auf.

```
IRC      = 1   # DFIX-Restriction
IBC      = 2   # Bindung
IAC      = 4   # Endatom eines Bindungswinkels
ITC      = 8   # Endatom eines Torsionswinkels
INC      = 16  # VdW-Wechselwirkung
ICC      = 32  # Coulomb-Wechselwirkungen
INFC     = 64  # Erzwungene nichtbindende Wechselwirkungen
IHBC     = 128 # Wasserstoffbrücke
IBRC     = 3   # Bindung oder Restriction
IBAC     = 6   # Bindung oder Bindungswinkel
IBATC    = 15  # Bindung, Bindungswinkel oder Torsionswinkel
IBBC     = 34  # Bindung und Coulomb-Wechselwirkungen
IACC     = 36  # Bindungswinkel und Coulomb-Wechselwirkungen
ITCC     = 40  # Torsionswinkel und Coulomb-Wechselwirkungen
IDC      = 48  # Default-Code
IDFC     = 112 # Default-Code INF-Set
INNFC    = 80  # VdW-Wechselwirkungen und erzwungene nichtbindende
                Wechselwirkungen
IOPENC   = 256 # Für Ring-Torsionstreiber geöffnete Bindung
IMUPOL   = 512 # Multipol-Wechselwirkungen zwischen 2 Atomen
```

Folgende Attribute werden gesetzt bzw. gelöscht:

- Bindungen: IBBC setzen, INC und ICC löschen
- Bindungswinkel: IAC setzen, INC löschen
- Torsionswinkel: ITC setzen, INC löschen.

Tabelle 4: Wechselwirkungsattribute im MOMO-Kraftfeld

```
IFPI     = 1   # Pi Atom
IFHYDR   = 2   # Wasserstoffatom
IFHET    = 4   # Heteroatom
IFHBAC   = 8   # H-Brücken-Akzeptor
IFHBDO   = 16  # H-Brücken-Donor
IFTERM   = 32  # Terminales Atom
IFCINC   = 64  # Zentralatom für Gruppeninkremente ( $\sigma$ -Bindungsenergie)
IFSINC   = 128 # Substituent für Gruppeninkremente ( $\sigma$ -Bindungsenergie)
IFEREP   = 256 # Substituent mit (repulsiven) freien Elektronenpaaren
IFDISP   = 512 # MOMO.für Atom im GFX Modus anzeigen
IFLABL   = 1024 # MOMO.für Atomlabel im GFX Modus
IFXFIX   = 2048 # Fix x-Koordinate
IFYFIX   = 4096 # Fix y-Koordinate
IFZFIX   = 8192 # Fix z-Koordinate
IFSEL    = 16384 # MOMO.für Atom mit Maus ausgewählt
IFSRING  = 32768 # Atom in einem 3-, 4- oder 5-Ring
```

Tabelle 5: Atomattribute im MOMO-Kraftfeld

Atome in 3-, 4- und 5-Ringen bekommen das „IFSRING“-Attribut. Für die Suche von Ringen existieren verschiedene Algorithmen in der Literatur, für den hier vorliegenden Spezialfall mit drei

festen Ringgrößen wurde ein eigener Ringsuchalgorithmus entwickelt, der nach einer kurzen Einführung in die bekannten Methoden erörtert wird.

Atomtyp	Nr.	Attribute
C(sp ₃)	1	Zentralatom für Gruppen- und Substituenteninkremente
C(sp ₂)	2	π-Atom
N(3)	3	π-Atom, H-Brücken-Donor und -Akzeptor, Heteroatom
N(2)	4	π-Atom, H-Brücken-Donor und -Akzeptor, Heteroatom
O(2)	5	π-Atom, H-Brücken-Donor und -Akzeptor, Heteroatom
O(1)	6	π-Atom, H-Brücken-Donor und -Akzeptor, Heteroatom
S(2)	7	π-Atom, H-Brücken-Donor und -Akzeptor, Heteroatom, repulsive freie Elektronenpaare
Cl	8	π-Atom, Heteroatom, terminales Atom
F	9	π-Atom, Heteroatom, terminales Atom
H	10	Wasserstoff, terminales Atom
C(sp)	11	π-Atom, repulsive freie Elektronenpaare
N(1)	12	π-Atom, Heteroatom, terminales Atom, repulsive freie Elektronenpaare
Mg	13	Heteroatom, repulsive freie Elektronenpaare
Ca	14	Heteroatom, repulsive freie Elektronenpaare
Si	15	π-Atom
P	16	π-Atom, Heteroatom
S(4)	17	π-Atom, Heteroatom
Br	18	π-Atom, Heteroatom, terminales Atom
I	19	π-Atom, Heteroatom, terminales Atom
-H	20	Wasserstoff, terminales Atom

Tabelle 6: Atomtypen und deren Standardattribute im MOMO-Kraftfeld

Ringsuche in Molekülen

Ringe beeinflussen die Eigenschaften eines Moleküls derart, dass eine besondere Behandlung im Kraftfeld nötig ist. Die Betrachtung von Molekülen als Graphen ermöglicht die Ringsuche per Computer. Verschiedene Ansätze sollen je nach Anwendungsfall den Aufwand für die Suche minimieren. Für die Beschreibung eines Moleküls ist es oft ausreichend, nur nach Teilmengen aller

vorhandenen Ringe zu suchen. Um den Sinn gängiger Teilmengen von Ringen zu verstehen, muss kurz auf die Eigenschaften eines Molekülgraphen eingegangen werden.

In einem Molekülgraphen entsprechen Atome den Knoten und kovalente Bindungen den Kanten. Die Kanten sind ungerichtet, zwischen verschiedenen Typen von Bindungen (Einfach- und Mehrfachbindungen) wird nicht unterschieden. Ringe werden als Zyklen bezeichnet.

Folgende Begriffe aus der Graphentheorie sind hierbei von Bedeutung:

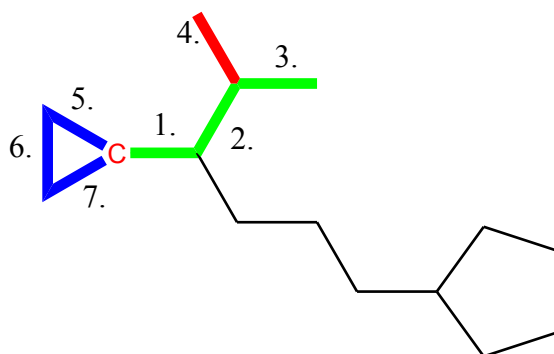
- Zwei durch eine Kante verbundene Knoten gelten als benachbart.
- Die Anzahl der Kanten an einem Knoten gibt den Grad des Knotens an.
- Eine Teilmenge von Knoten und Kanten aus einem Graphen ist ein Teilgraph.
- Ein Weg oder Pfad ist die Durchquerung einer Folge benachbarter Knoten.
- Können von jedem Knoten aus alle anderen Knoten der Graphen durch einen Weg erreicht werden, ist der Graph zusammenhängend.
- Ist ein Graph nicht zusammenhängend, so heißen die zusammenhängenden Teilgraphen „Komponenten“.
- Ein „Polytope“ ist die begrenzende Hülle einer Menge von Punkten, die eine endliche Menge von Punkten enthält.
- Planare Graphen teilen die Ebene in Regionen ein, die durch Kanten und Knoten begrenzt werden. Innerhalb einer Region können alle Knoten verbunden werden, ohne eine Kante dabei zu schneiden.
- Zwei Knoten in einem Zyklus, die durch eine nicht zum Zyklus gehörende Kante verbunden sind, heißen „Nachbarpunkte“.
- Zyklen ohne Nachbarpunkte nennt man „Simple Cycles“.
- Regionen, die Simple Cycles sind, heißen „Simple Faces“.
- Simple Cycles, die keiner Region entsprechen, nennt man „Cut Faces“.
- Zerfällt ein zusammenhängender Graph durch Entfernen eines Knotens in Komponenten, so heißt der betreffende Knoten „Cut Vertex“.
- Eine Komponente ohne „Cut Vertex“ ist ein Block.

Bei der Ringsuche werden nun Zyklen aus dem molekularen Graphen extrahiert. Eine Möglichkeit, nicht aufwendig nach allen Zyklen zu suchen, ist, sich auf die minimale Anzahl zu beschränken. Damit ist die minimale Anzahl linear unabhängiger Zyklen gemeint, die alle Knoten und Kanten abdecken. Daneben existieren noch weitere Mengen von Ringen. Ziel dieser Ringmengen ist u. a., das Molekül für Datenbankanwendungen möglichst eindeutig oder unter speziellen Gesichtspunkten zu beschreiben.

Die Ringsuche in MOMO

Die Ringsuche in MOMO hat nur den Zweck, Atome in kleinen Ringen zu markieren, und kann deshalb sehr einfach ausfallen. Ein trivialer Ansatz ist, von jedem Atom aus den molekularen Graphen in maximal fünf Schritten zu durchsuchen, bis man wieder auf das Ausgangsatom trifft. Dies lässt sich als Tiefensuche oder als Breitensuche verwirklichen. Bei der Tiefensuche (Abb. 22) wird immer zuerst weiter weg vom Startknoten gesucht, bevor eine alternative Verzweigung vom Startknoten aus inspiziert wird, wohingegen bei der Breitensuche zuerst alle Nachbarknoten kontrolliert werden, bevor man sich weiter vom Startknoten entfernt.

Für ungerichtete Graphen, wie es Molekülgraphen sind, sieht das Verfahren wie folgt aus:



Die Suche startet beim rot markierten Kohlenstoffatom, die Zahlen geben die Reihenfolge der abgesuchten Kanten an. Eine Farbe markiert immer den Weg zu einem Knoten, an dem der Algorithmus „umkehren“ muss, weil es keine weiteren Nachbarknoten gibt.

Abbildung 22: Illustration der Tiefensuche

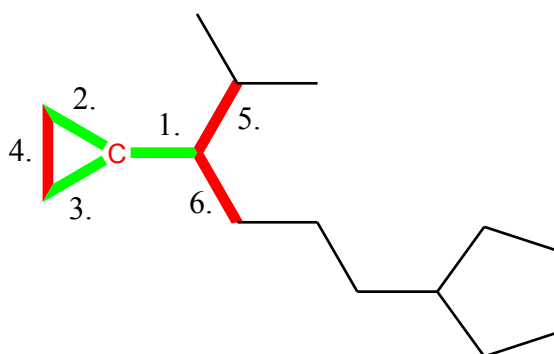
Ausgehend von einem Startknoten (Atom), wird die erste Kante (Bindung) gewählt und getestet, ob der Nachbarknoten das gesuchte Atom ist oder schon besucht wurde. Ist dies nicht der Fall, so wird

von diesem Nachbarknoten aus eine neue Suche gestartet. Dieser Schritt ist eine Rekursion, da sich das Suchverfahren selbst aufruft. Dies wird so lange wiederholt, bis entweder

- das Ausgangsatom gefunden ist
- oder alle benachbarten Atome besucht sind (z. B. in einer „Sackgasse“),
- die maximale Schrittweite (hier 5 Bindungen) ausgeschöpft ist.

Die Breitensuche (Abb. 23) läuft folgendermaßen ab:

Das Startatom wird in einer Warteschlange abgelegt. Dann wird ein Atom aus der Warteschlange entnommen und als besucht markiert. Ist das Atom das gesuchte und wurde mindestens über eine Kante vom Startatom aus erreicht, war die Suche erfolgreich, andernfalls werden alle noch nicht besuchten Nachbarknoten an die Warteschlange angehängt. Ist die Warteschlange leer, ist die Suche beendet.



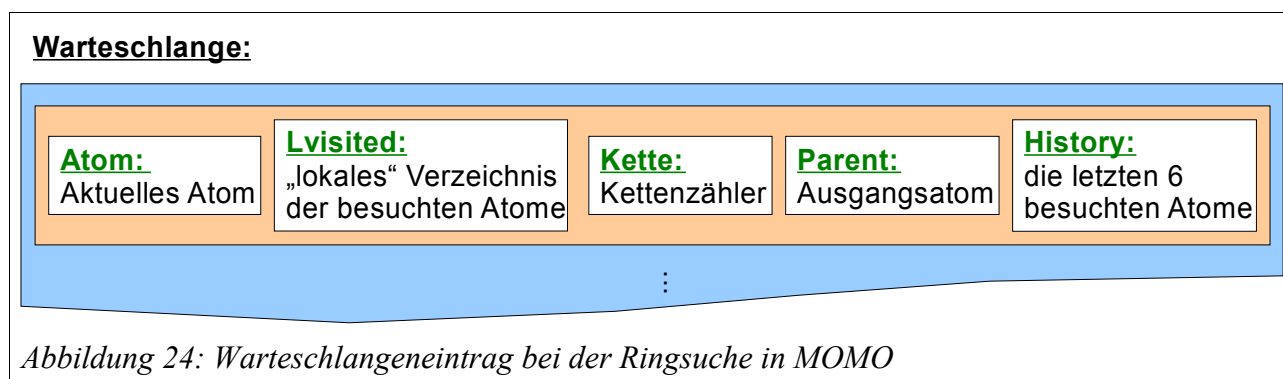
Die Suche startet beim rot markierten Kohlenstoffatom, die Zahlen geben die Reihenfolge der abgesuchten Kanten an. Eine Farbe markiert immer eine Ebene der sich „ausbreitenden“ Suche.

Abbildung 23: Illustration der Breitensuche

Um nicht von jedem Atom aus eine neue Suche starten zu müssen, wurde ein spezieller Algorithmus entwickelt, der auf der Breitensuche basiert. Hier muss allerdings mehr gespeichert werden als nur die bereits besuchten Knoten im Molekülgraphen.

Ausgehend von der Breitensuche spart dieser Algorithmus Zeit: Teile des Moleküls, die keine 3-, 4- und 5-Ringe enthalten können (Ketten), werden markiert und müssen dann nicht mehrfach untersucht werden.

Der Algorithmus benötigt eine Warteschlange und eine Markierung für bereits besuchte Atome. Jedes Element der Warteschlange enthält ein aktuelles Atom, eine Liste der sechs zuletzt besuchten Atome, ein Verzeichnis aller bereits von dieser Verzweigung besuchten Atome, einen Zähler für die „Kettenlänge“ der Verzweigung und das Atom, von dem die Verzweigung ausging. Abb. 24 zeigt ein Element der Warteschlange.



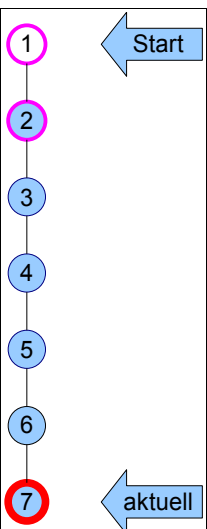
Zunächst wird ein beliebiges Atom im Molekülgraphen in die Warteschlange eingetragen; die weiteren Felder bleiben noch leer. Aus der Warteschlange werden nun so lange Elemente entnommen, bis diese leer ist, dann ist die Suche beendet. Für jedes entnommene Element werden die folgenden Schritte durchgeführt, bevor das nächste Element entnommen wird:

- 1) Das aktuelle Atom wird an die Liste „History“ angehängt. Wenn History mehr als sechs Einträge hat, wird der älteste Eintrag an die Liste „Lvisited“ angefügt. Steht der Zähler „Kette“ auf einem Wert größer vier, wird das Atom aus diesem Eintrag zusätzlich als „besucht“ markiert. Anschließend wird überprüft, ob das aktuelle Atom in der History-Liste mindestens zweimal vorkommt, wenn ja, gehört es zu einem kleinen Ring und wird entsprechend gekennzeichnet.
- 2) Jetzt müssen die Nachbarknoten des aktuellen Atoms untersucht und in die Warteschlange gelegt werden, um die Breitensuche fortzusetzen. Die Nachbarknoten werden verworfen,
 - falls das Nachbaratom das einen Schritt zurückliegende aktuelle Atom ist oder es bereits besucht wurde (Markierung oder lvisited),
 - falls es sich beim Nachbaratom um Wasserstoff handelt,
 - falls das Nachbaratom das Ende einer Sackgasse ist (also keine weiteren Bindungen von ihm ausgehen).

- 3) Bleiben mindestens zwei Nachbarn übrig, wird der Zähler „Kette“ auf null zurückgesetzt, andernfalls um eins erhöht.
- 4) Für jeden Nachbarn wird nun ein neuer Eintrag in der Warteschlange erzeugt und anschließend mit Schritt zwei fortgefahren.

Abb. 25 soll die Arbeitsweise des Algorithmus verdeutlichen.

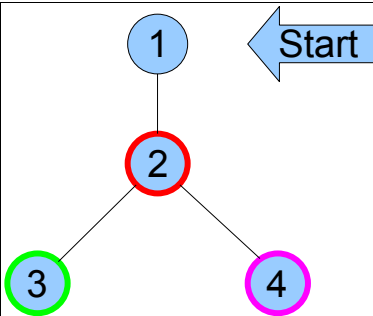
Der Algorithmus startet in diesem Beispiel am Anfang einer Kette, nach sechs Schritten sieht der Datensatz für das aktuelle Atom (Nr. 7) wie folgt aus:



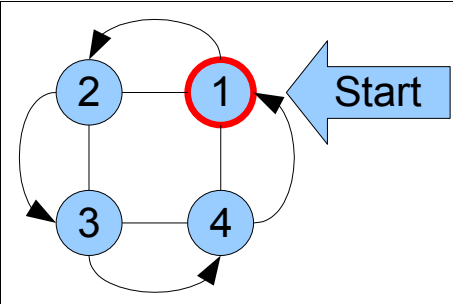
„History“ enthält die Atome 2, 3, 4, 5, 6 und 7. Atom 1 wird aus „History“ gelöscht, da die Maximallänge der Liste überschritten wird.

- Atom 1 wird in das lokale Verzeichnis der besuchten Atome übernommen.
- Die Atome 1 und 2 wurden als Kettenatome markiert. Sie können nicht Teil eines 3-, 4-, oder 5-Rings sein, da sie schon mehr als vier Schritte, die sich alle entlang einer Kette bewegen, zurückliegen.

Bei einer Verzweigung werden nach der Verarbeitung von Atom Nr. 2 zwei Einträge in der Warteschlange erzeugt. Die Einträge für Atom Nr. 3 und Nr. 4 haben die Liste History, das Verzeichnis Lvisited und den wegen der Verzweigung auf null zurückgesetzten Zähler „Kette“ gemeinsam.



Bei der Erkennung eines Rings ergibt sich Folgendes:

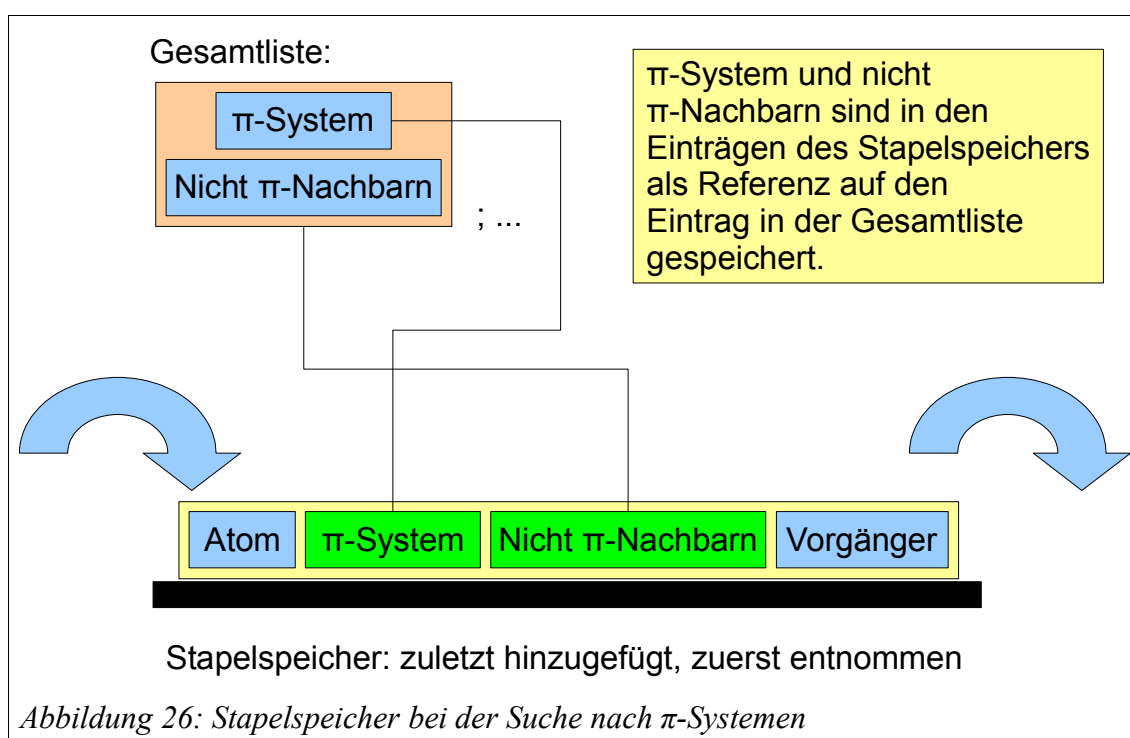


Ausgehend von Atom Nr. 1 wird entlang des Rings gesucht. Kommt der Algorithmus wieder bei Atom Nr. 1 an, so ist dieses zweimal in der Liste „History“ vorhanden und damit der Ring erkannt. Die maximale Ringgröße wird durch die begrenzte Länge der „History“-Liste eingehalten.

Abbildung 25: Beispiel zur Ringsuche in MOMO

Die Erkennung von π -Systemen im Molekül

In MOMO werden π -Systeme in Molekülen besonders behandelt. Ein π -System ist ein zusammenhängender Bereich von mit dem Merkmal π -Atom markierten Atomen. Ein Molekül kann mehrere π -Systeme enthalten. Die π -SCF-Rechnung erfolgt für jedes dieser Systeme separat. Zum Schluss werden die Ergebnisse für das gesamte Molekül zusammengefasst, wobei Wechselwirkungen zwischen Atomen aus verschiedenen π -Systemen vernachlässigt werden. Für die Suche nach solchen zusammenhängenden π -Bereichen wurde ein ebenfalls auf der Breitensuche basierender Suchalgorithmus entwickelt. Benötigt werden für jeden zusammenhängenden π -Bereich eine Liste der π -Atome und eine Liste der an den Bereich angrenzenden Nicht- π -Atome.



Der Algorithmus basiert auf einem Stapelspeicher (Stack, siehe Abb. 26), aus dem immer das zuletzt dort abgespeicherte Element entnommen wird. Daneben wird ein Verzeichnis der bereits untersuchten Atome, eine Liste der π -Systeme und der angrenzenden Nicht- π -Atome geführt.

Zunächst wird ein beliebiges Startatom des zu analysierenden Moleküls auf den Stapel gelegt. Die anderen Felder sind zunächst leer. Für jedes Element werden die unten aufgelisteten Schritte durchgeführt, bis der Stack leer ist:

- Nehme das oberste Element vom Stack und:

- Prüfe, ob das Atom schon bearbeitet wurde. Wenn ja, verwirfe das Element und beginne von vorne.
- Markiere das aktuelle Atom als bearbeitet.
- Erstelle eine Liste der dem aktuellen Atom benachbarten Atome. Die π -Atome müssen zuletzt in der Liste stehen.
- Falls das aktuelle Atom ein π -Atom ist:
 - Sofern noch keine Liste für ein π -System und dessen angrenzende Atome vorhanden ist, wird diese angelegt und eine Referenz auf diese lokale Liste in die Liste der π -Systeme abgelegt.
 - Das aktuelle Atom wird an die Liste des π -Systems angehängt.
 - Ein eventuelles Vorgängeratom (ist im Stack-Element eingetragen) wird als angrenzendes Atom gespeichert, wenn es kein π -Atom ist.
 - Die benachbarten Atome werden auf den Stack gelegt, die oben durchgeführte Sortierung gewährleistet, dass die π -Atome zuletzt dort abgelegt werden und damit oben liegen. Wäre dies nicht der Fall, könnte die Atomliste eines π -Systems nicht korrekt fortgeführt werden.
- Wenn nicht:
 - Die Liste der an das π -System angrenzenden Atome wird um das aktuelle Atom ergänzt, falls diese schon existiert. In diesem Fall muss ein π -Bereich gerade verlassen worden sein.
 - Die Nachbaratome des aktuellen Atoms werden wie oben auf den Stack gelegt. Die Liste für ein π -System wird aber nicht weitergegeben.

6.5 *Verwendete Hilfsprogramme*

Neben der MMTK-Bibliothek wurden für diese Arbeit weitere Hilfsprogramme verwendet, die hier kurz vorgestellt werden. Die Programme dienen der Visualisierung von Molekülen (PyMol, VMD), stellen mathematische Funktionen bereit (Scientific Python), generieren Diagramme (Matplotlib) und vereinfachen die Datei Ein- und Ausgabe (cElementTree) im XML-Dateiformat.

PyMOL

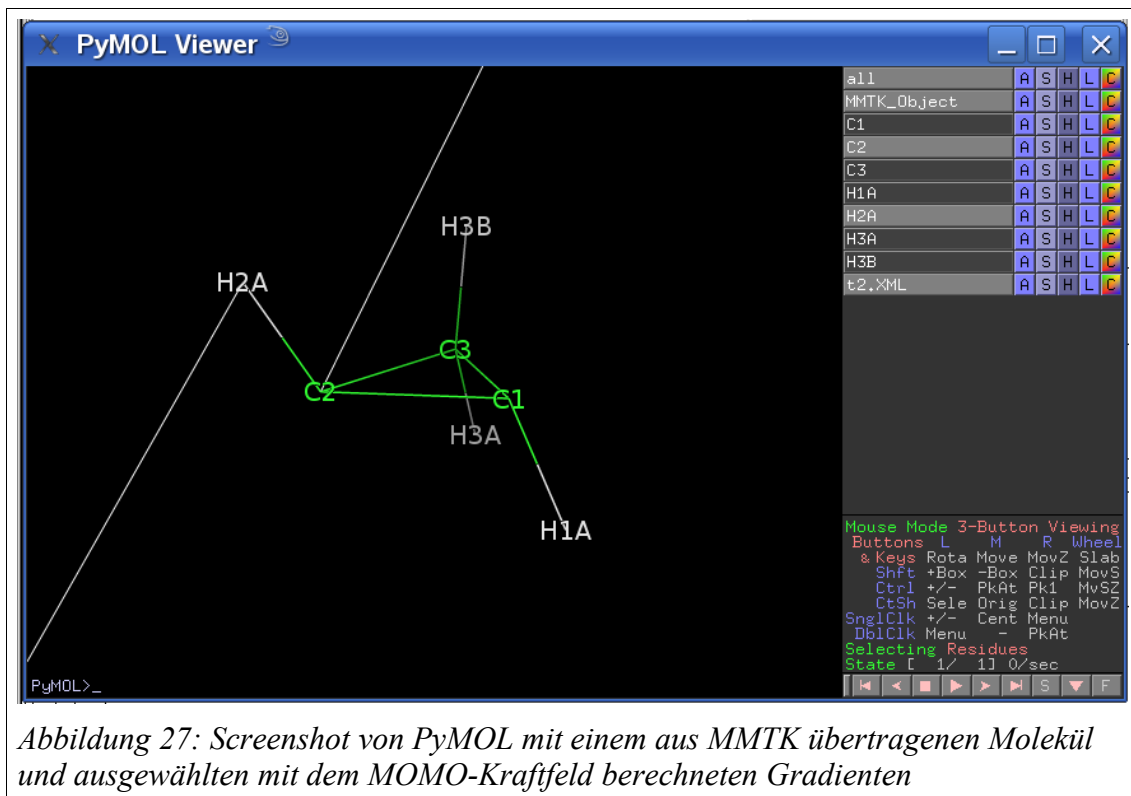
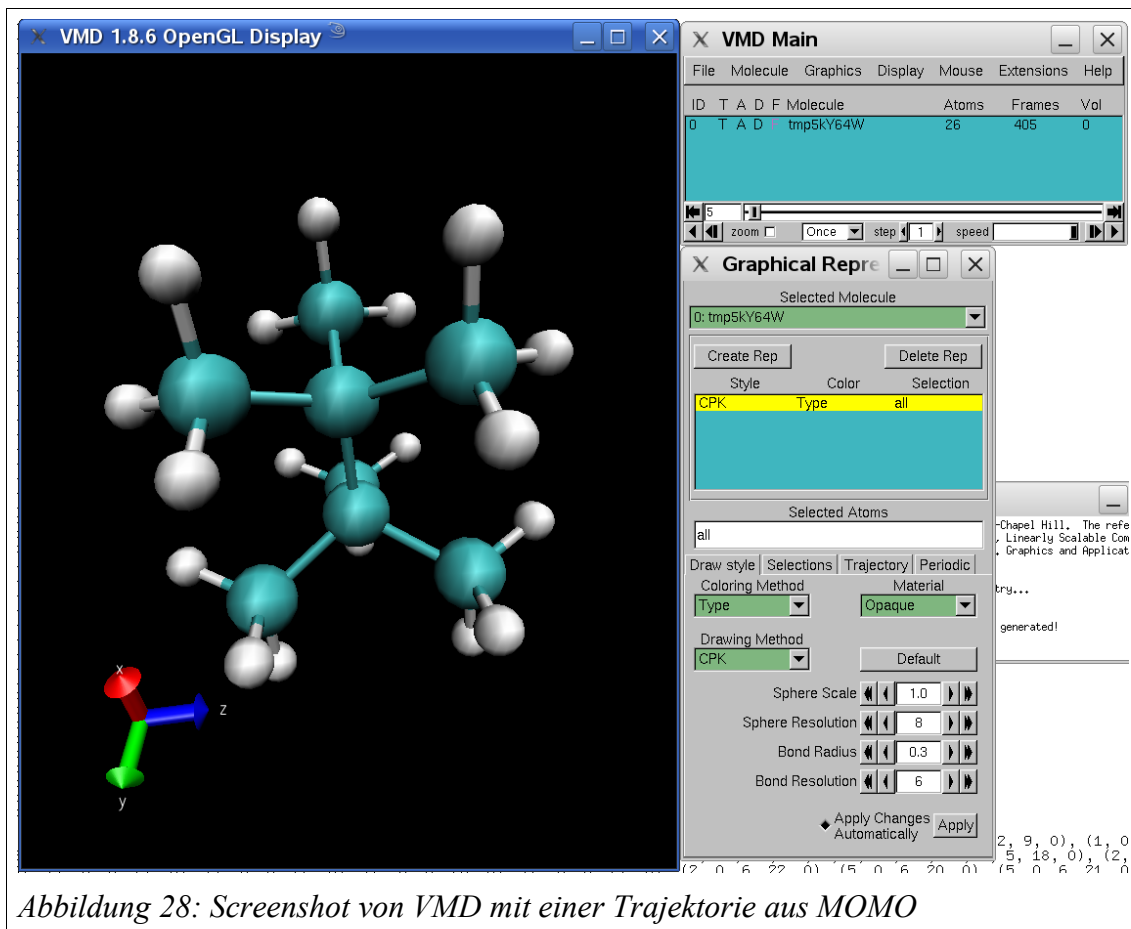


Abbildung 27: Screenshot von PyMOL mit einem aus MMTK übertragenen Molekül und ausgewählten mit dem MOMO-Kraftfeld berechneten Gradienten

PyMOL [DeLano] ist ein 3D-Visualisierungsprogramm für Moleküle. Es benutzt in der Strukturbiologie gebräuchliche Dateiformate und erlaubt auf verschiedenen Betriebssystemen die Berechnung hochwertiger Grafiken. PyMOL hat einen eingebundenen Interpreter für die Programmiersprache Python. Mit ihm können beliebige Python-Programme geladen werden und dann die Funktionen von PyMOL nutzen. Neben molekularen Grafiken können auch individuelle Grafikobjekte und Animationen erzeugt werden. PyMOL wird in MOMO als Standardprogramm zur Anzeige von Molekülgrafiken verwendet, durch benutzerdefinierte Grafikobjekte können z. B. die auf ein Atom wirkenden Kräfte dargestellt werden (Abb. 27). Die Trajektorie (der Verlauf) kann als Animation visualisiert werden.

VMD

VMD [Humphrey1996] steht für „Visual Molecular Dynamics“. Es handelt sich um ein molekulares Grafikprogramm, das auf die Darstellung von Biopolymeren optimiert ist. Daneben können insbesondere Trajektorien z. B. von MD (molecular dynamics) Simulationen animiert



werden. MOMO nutzt VMD zur Darstellung von Trajektorien und Molekülen (Abb. 28), da die Trajektorienanstellung mit PyMOL auf einigen Systemen Probleme bereitet.

SciPy

SciPy [Eads2008] ist eine Sammlung wissenschaftlicher Software für Python. Darin enthalten ist Open-Source-Software für Mathematik, Natur- und Ingenieurwissenschaften. Die Bibliothek stellt benutzerfreundliche und effiziente Routinen für gängige numerische Probleme wie Integration, die Lösung von Gleichungssystemen, Optimierungsalgorithmen und die Manipulation großer Zahlenfelder, Clusteranalyse, Statistik, Interpolation usw. zur Verfügung.

Matplotlib

Matplotlib [Hunter2007] ist eine Python-Bibliothek zum Erstellen vielfältiger Diagramme, die sowohl in gängige Grafikdateiformate exportiert als auch interaktiv zusammen mit verbreiteten Bibliotheken zur Erzeugung grafischer Benutzeroberflächen dienen können. MOMO nutzt diese

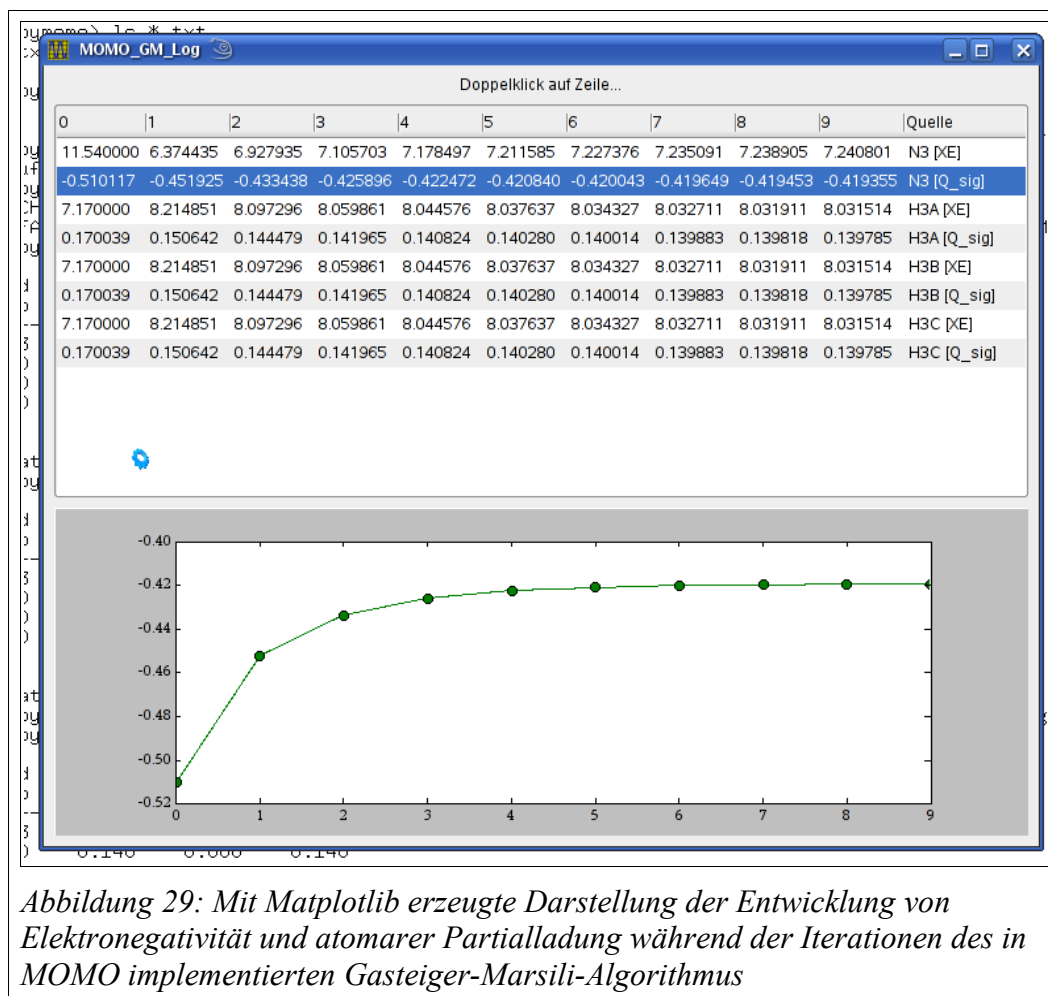


Abbildung 29: Mit Matplotlib erzeugte Darstellung der Entwicklung von Elektronegativität und atomarer Partialladung während der Iterationen des in MOMO implementierten Gasteiger-Marsili-Algorithmus

Bibliothek sowohl zur Erzeugung von Diagrammen in Protokolldateien als auch interaktiv (Abb. 29).

cElementTree

Um Daten über Moleküle sinnvoll für Parametrisierungsprogramme einsetzen zu können, ist es nicht mehr ausreichend, nur eine Liste von Atomkoordinaten zu speichern. Hier werden Zusatzinformationen wie eine Liste der Bindungen und verschiedene Referenzen wie experimentelle Daten oder Ergebnisse aus quantenmechanischen Rechnungen benötigt. Ein weitverbreitetes universelles Dateiformat ist XML [Bray], eine erweiterbare Auszeichnungssprache für hierarchisch strukturierte Daten in Form von Textdateien, für die Problemstellung also genau das Richtige. Hier können alle Daten sowohl für den Menschen als auch für den Rechner lesbar abgespeichert werden und Programmbibliotheken für XML ermöglichen eine einfache und effiziente Implementierung. Eine XML-basierte Sprache ist die CML (Chemical Markup Language)

[Murray-Rust1999], für die Bedürfnisse von MOMO sind allerdings so viele Erweiterungen und Änderungen nötig, dass das Dateiformat den Namen CML nicht mehr verdient.

cElementTree [Lundh2005] ist eine durch Verwendung eines in der Programmiersprache C geschriebenen Kerns besonders schnelle Bibliothek zur Handhabung von XML-Dateien.

Die Abkürzung XML steht für Extensible Markup Language. Ein XML-Dokument besteht aus Textzeichen und ist daher für Menschen lesbar, Binärdaten sind per Definition nicht enthalten. Der logische Aufbau einer XML-Datei ist eine Baumstruktur, womit sich die hierarchische Struktur ergibt. Mögliche Knoten im Baum sind

- Paare aus Start- und End-Tag (<TagName> </TagName>)
- Empty Element Tags (<TagName />)
- Attribute in Form von Schlüsselworten beim Start-Tag oder einem Empty-Element-Tag (<Tag-Name Attribut="wert">)
- Verarbeitungsanweisungen
- Kommentare

Ein XML-Dokument enthält genau einen Wurzelknoten, darunter können weitere Ebenen verschachtelt werden (Abb. 30).

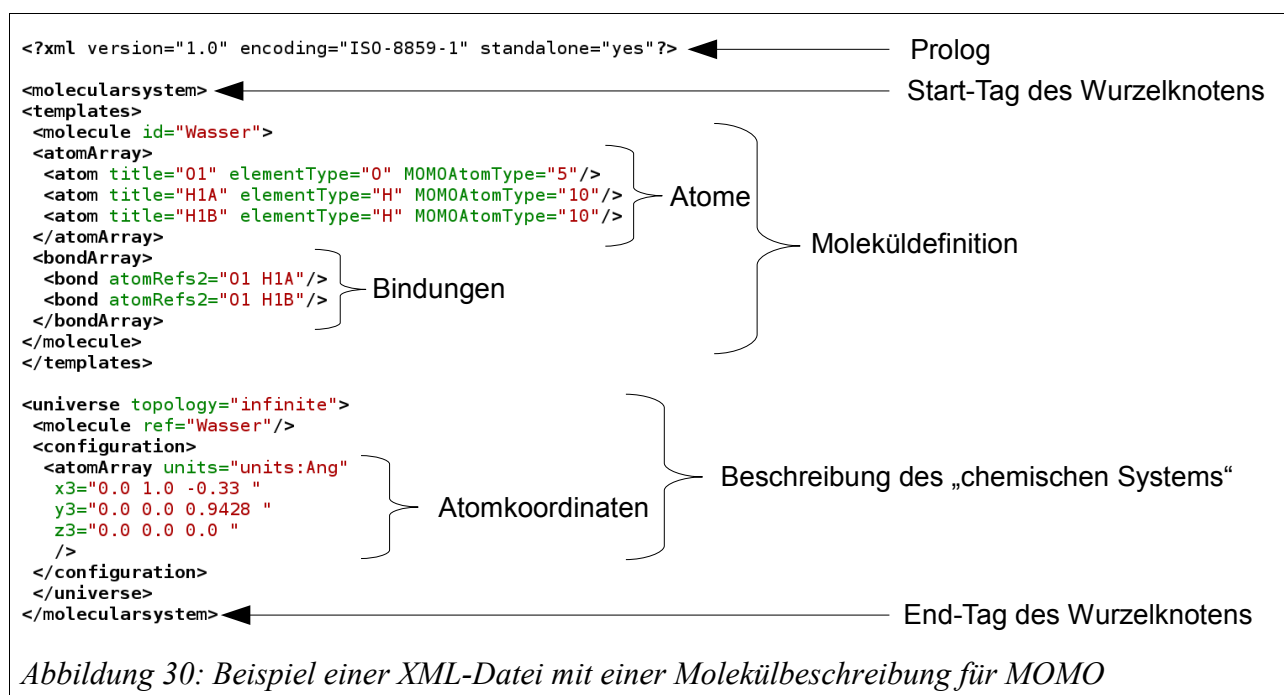


Abbildung 30: Beispiel einer XML-Datei mit einer Molekülbeschreibung für MOMO

XML ist keine Sprache zur Definition von Inhalten, sondern die Grundlage für die Definition einer solchen. Kein einziger Tag ist bezüglich seiner Bedeutung definiert, nur der Aufbau des Dokuments ist vorgegeben. Die Namen der Elementtypen werden von der Anwendung definiert.

Jedes XML-Dokument beginnt mit einem Prolog:

```
<?xml version="1.0" encoding="iso-8859-1" standalone="yes" ?>
```

Er kennzeichnet eine XML-Datei und macht Angaben bezüglich der Codierung.

Die Elemente der Datei bestehen aus einem Inhalt (die eigentliche Information), der von Tags eingerahmt wird.

MOMO nutzt XML, um Moleküle und Referenzdaten für das Parametrisierungsprogramm zu speichern.

Parallel Python

Parallel Python [Vanovschi2005] ist eine Bibliothek für die Programmiersprache Python, welche auf einfache Weise die Abarbeitung von Aufgaben (Jobs) aus einer Warteschlange unterstützt. Die Jobs werden automatisch auf mehrere Prozessoren in einem Rechner und ggf. auf mehrere vernetzte Rechner verteilt. MOMO selbst ist wegen der Abhängigkeiten von Zwischenergebnissen schlecht parallelisierbar. Für die massenweise Berechnung von Molekülen bei der Parametrisierung oder für Konformationsanalysen zerfällt die Aufgabe in viele unabhängige Teilprobleme und kann so von der parallelen Berechnung profitieren.

7 Angewandter Teil

7.1 Parametrisierung des MOMO-Wasserstoffbrückenpotenzials

Das Fehlen von Parametersätzen für bestimmte Substanzklassen schränkt die Verwendung von Kraftfeldern stark ein. Eine wichtige Anwendung von MOMO in unserem Arbeitskreis ist die Berechnung von wasserstoffbrückegebundenen Komplexen. Hier bestand konkreter Bedarf für neue Parameter, weshalb mit der Parametrisierung von Wasserstoffbrücken begonnen wurde.

Der Begriff „Wasserstoffbrückenbindung“ [Steiner2002] wurde im Jahr 1923 von Maurice Huggins geprägt und bezog sich damals nur auf Wechselwirkungen zwischen Sauerstoff und Wasserstoff zweier verschiedener Wassermoleküle. Inzwischen ist der Begriff jedoch sehr viel weiter gefasst und bezeichnet heute die Wechselwirkung zwischen einem Protonendonor und einem oder mehreren räumlich benachbarten Protonenakzeptoren. Es handelt sich hierbei um gerichtete intra- oder intermolekulare Wechselwirkungen, die breite Übergangsbereiche zu kovalenten, van-der-Waals-, ionischen und Kation- π -Bindungen aufweisen.

Die Orientierung der Wasserstoffbrücken ist zum Teil auf elektrostatische Kräfte zwischen komplementären Partialladungen zurückzuführen, kann jedoch nicht darauf beschränkt werden. Polarisation, Ladungstransfer und Dispersion spielen ebenfalls eine Rolle, wenngleich die elektrostatischen Wechselwirkungen angesichts eines Donor-Akzeptorabstandes von 3-4 Å die übrigen Wechselwirkungen überwiegen dürften. Wechselwirkt ein Donor mit mehreren Akzeptoren, spricht man von einer „gegabelten“ Wasserstoffbrücke.

Die Dissoziationsenergien der Wasserstoffbrücken decken einen weiten Bereich ab. Man unterteilt grob in starke und schwache Wasserstoffbrücken. Unter starken Wasserstoffbrücken versteht man solche mit einer Dissoziationsenergie über 15 kJ/mol. Dazu zählen u. a. O-H...O- (z. B. in Wasser), N-H...O- oder O-H...N-Wasserstoffbrücken. Die Wechselwirkungen werden als kovalent beschrieben, weisen eine Bindungslänge von 1.2 bis 1.5 Å und einen Bindungswinkel von 170°- 180° auf und haben eine starke Richtungspräferenz. Schwache Wasserstoffbrücken haben eine Dissoziationsenergie von unter 4 kJ/mol. Die H...Akzeptor-Bindungslänge ist größer als 2.2 Å und der Bindungswinkel $> 90^\circ$. Beispiele sind C-H...O-, C-H...N- und C-H...Halogen-Brücken. Solche schwachen Wechselwirkungen als Wasserstoffbrücken zu bezeichnen, ist zwar umstritten, dennoch gibt es kristallografische Ergebnisse, die dafür sprechen [Taylor1981].

7.1.1 Referenzdaten für Wasserstoffbrücken

Denkbar sind hier kristallografische Daten und ab-initio-Rechnungen. In Kristallstrukturen lassen sich Wasserstoffbrücken statistisch nachweisen, indem man Datenbanken wie die CSD [Allen2002] nach den charakteristischen Molekülfragmenten durchforstet. Hier treten bestimmte räumliche Anordnungen gehäuft auf, aus denen sich klar eine Vorzugsrichtung und ein Vorzugsabstand zwischen Donor und Akzeptor ergeben. Die Bestimmung von Referenzdaten wurde in zwei Schritten vorgenommen. Zunächst wurde in der CSD nach den für einen Wasserstoffbrückenbindungstyp typischen Molekülfragmenten gesucht und eine Statistik über die geometrischen Parameter des in MOMO verwendeten Dunitz-Potenzials [Vedani1985] erstellt. Hieraus wurde jeweils der Modus bestimmt. Die Potenzialfunktion wurde dann anhand von ab-initio-Rechnungen eines Modellkomplexes parametrisiert.

7.1.2 Energieberechnung von Wasserstoffbrückenbindungen

Da die Energie einer Wasserstoffbrücke nicht direkt erhältlich ist, muss sie als Differenz bestimmt werden. Nahe liegend ist es hier, die Differenz zwischen der Energie des Komplexes und der Summe der Energien der Einzelmoleküle anzusetzen.

$$\Delta E = E(AB)_{ab}^{Komplex} - E(A)_a^{Monomer} - E(B)_b^{Monomer} \quad (7.1.1)$$

Die Dissoziationsenergie ist die Energie des Komplexes AB im Basissatz beider Monomere $E(AB)_{ab}^{Komplex}$ abzüglich der Monomere A und B in ihren jeweiligen Basissätzen ($E(A)_a^{Monomer}$ und $E(B)_b^{Monomer}$). Die Geometrie, auf die sich die Energie bezieht, ist hochgestellt, der Basissatz tiefgestellt.

Da die verwendeten Basissätze nicht vollständig sind, kommt es dazu, dass die Monomere im Komplex den Basissatz des Bindungspartners „mitbenutzen“, was zu einer zu geringen Energie des Komplexes führt. Die Dissoziationsenergie wird deshalb oft überbewertet. Dieser Effekt ist unter dem Namen „basis set superposition error“ (BSSE) bekannt [Lendvay1998].

Rechnerisch kann dem BSSE durch die sogenannte Counterpoise-Korrektur begegnet werden. Hier werden die Monomere im Basissatz des Dimers berechnet:

$$\Delta E_{counterpoise} = E(AB)_{ab}^{Komplex} - E(A)_{ab}^{Komplex} - E(B)_{ab}^{Komplex} \quad (7.1.2)$$

Die Berechnung wird in der Literatur sehr unterschiedlich gehandhabt. Im einfachsten Fall werden Einzelmoleküle und Komplexe mit der Hartree-Fock-Methode berechnet und die Energiedifferenz

gebildet. Empfohlen [Yongho_Kim2002] wird die Verwendung einer Methode, die Elektronenkorrelation berücksichtigt (MP2 oder B3LYP), und einer Korrektur für den BSSE sowie die Berücksichtigung der Geometrieänderung der Einzelmoleküle im Komplex. Diese Fragment-Relaxationsenergie sollte bei der Berechnung der Komplexdissoziationsenergie mit einfließen.

$$E_{Relaxation} = [E(A)_a^{Monomer} - E(A)_a^{Komplex}] + [E(B)_b^{Monomer} - E(B)_b^{Komplex}]$$

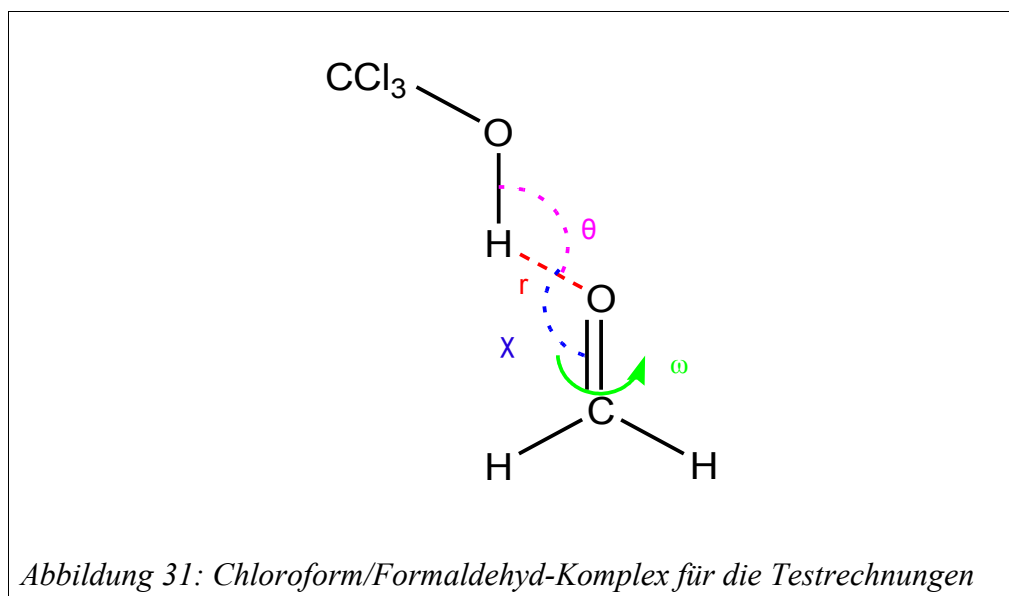
Die Energie der Wasserstoffbrücke berechnet sich demnach wie folgt:

$$\Delta E_{HB} = \Delta E_{counterpoise} + E_{Relaxation}$$

Da für die Parametrisierung des Dunitz-Potenzials viele Referenzpunkte nötig sind und sich der Aufwand bei den genaueren quantenchemischen Rechnungen stark erhöht, wurden zunächst einige Testrechnungen mit den verschiedenen Methoden vorgenommen und verglichen.

7.1.3 Testrechnungen mit verschiedenen Ab-initio-Methoden

Die Testrechnungen erfolgten am Beispiel des Komplexes Formaldehyd-Chloroform (Abb. 31). Chloroform besitzt aufgrund der drei elektronenziehenden Chlor-Atome am Kohlenstoff ein azides H-Atom und eignet sich daher gut als H-Brücken-Donor.



Zunächst wurden die Energien der Einzelmoleküle und des Komplexes mit verschiedenen Methoden, jeweils nach einer Geometrieoptimierung, berechnet und dann auf den verschiedenen oben beschriebenen Wegen die Dissoziationsenergie bestimmt.

Die ab-initio-Rechnungen wurden mit dem Programm Gaussian03 [Gaussian03] durchgeführt, welches vom Programm „MOMOGauSlavekJ.py“ gesteuert wurde. Das Programm arbeitet wie folgt: Zunächst wird eine im MOMO-Kraftfeld mit Standard-Parametern vorminimierte Struktur des Komplexes eingelesen. Aus dieser werden drei Eingabedateien für Gaussian erzeugt. Jeweils eine zur Geometrieoptimierung der Einzelmoleküle und des Komplexes. Schließlich werden noch die Geometrien der Einzelmoleküle aus dem optimierten Komplex extrahiert und deren Energien ohne weitere Optimierung (single-point-Rechnung) berechnet. Durch Vergleich mit den optimierten Einzelmolekülen wird eine Relaxationsenergie (wie oben beschrieben) ermittelt. Bei der Berechnung der Wasserstoffbrückenbindungsenergie gelangt man mit dem Basissatz 6-31G** zu folgenden Ergebnissen:

- Berechnung der Energie nur nach der Hartree-Fock-Methode:

$$E(HB)^{HF} = E_{Komplex} - E_{Molekül A}^{HF} - E_{Molekül B}^{HF} = -16,1 \text{ kJ/mol}$$

- Berechnung mit Møller-Plesset 2. Ordnung:

$$E(HB)^{HF/MP2} = E_{Komplex}^{MP2/HF} - E_{Molekül A}^{MP2/HF} - E_{Molekül B}^{MP2/HF} = -21,0 \text{ kJ/mol}$$

- Berechnung mit Møller-Plesset 2. Ordnung und Counterpoise-Korrektur:

$$E(HB)^{HF/MP2/Counterpoise} = E_{Komplex}^{MP2/HF+Counterpoise} - E_{Molekül A}^{MP2/HF} - E_{Molekül B}^{MP2/HF} = -13,7 \text{ kJ/mol}$$

- Berechnung mit Møller-Plesset 2. Ordnung, Counterpoise-Korrektur und Anrechnung der Fragmentrelaxationsenergie:

$$E(HB)^{HF/MP2/Counterpoise/Relax} = E(HB)^{HF/MP2/Counterpoise} + E_{Relaxation} = -13,5 \text{ kJ/mol}$$

Die Relaxationsenergie fällt, da kleiner als 0.1 kJ/mol, kaum ins Gewicht.

Bevor nun verschiedene Konstellationen der Wasserstoffbrücken getestet wurden, musste wegen der großen Zahl der Berechnungen noch die schnellste akzeptable Methode gefunden werden. Hierzu wurden die beschriebenen Schritte mit verschiedenen Basissätzen wiederholt. Die Ergebnisse sind in Tabelle 7 zusammengefasst.

Es fällt auf, dass die Fragmentrelaxationsenergie praktisch vernachlässigt werden kann, ohne Korrektur für den BSSE wird die Energie des Komplexes zu niedrig und damit die der

Basissatz:	STO-3G	3-21G	3-21+G**	6-31G**
$E(HB)^{HF}$	-11.2 kJ/mol	-32.5 kJ/mol	-26.4 kJ/mol	-16.1 kJ/mol
$E(HB)^{HF/MP2}$	-9.6 kJ/mol	-35.3 kJ/mol	-28.0 kJ/mol	-21.0 kJ/mol
$E(HB)^{HF/MP2/Counterpoise}$	-6.3 kJ/mol	-16.8 kJ/mol	-16.0 kJ/mol	-13.7 kJ/mol
$E(HB)^{HF/MP2/Counterpoise/Relax}$	-6.3 kJ/mol	-16.7 kJ/mol	-15.8 kJ/mol	-13.5 kJ/mol

Basissatz:	B3LYP/6-31G**	B3LYP/6-311G**	B3LYP/6-311++G**
$E(HB)^{DFT}$	-18.7 kJ/mol	-19.2 kJ/mol	-14.4 kJ/mol
$E(HB)^{DFT/Counterpoise}$	-13.1 kJ/mol	-13.0 kJ/mol	-12.5 kJ/mol
$E(HB)^{DFT/Counterpoise/Relax}$	-12.9 kJ/mol	-13.0 kJ/mol	-12.3 kJ/mol

Tabelle 7: Berechnung der Dissoziationsenergie des Chloroform/Formaldehyd-Komplexes mit verschiedenen Methoden im Vergleich

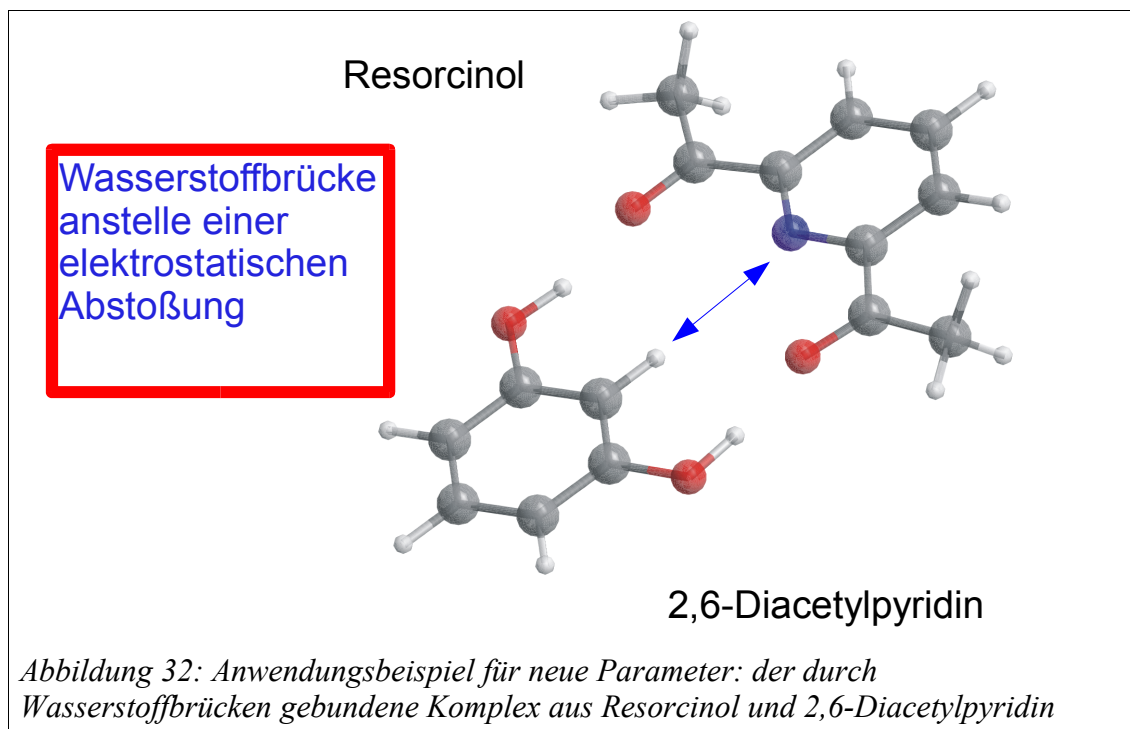
Wasserstoffbrückenbindung zu hoch bewertet. Ein guter Kompromiss zwischen Rechengenauigkeit und Rechenzeit ist die Methode B3LYP/6-31G**, die Counterpoise-Korrektur sollte auf jeden Fall stattfinden.

Die weiteren Berechnungen für Wasserstoffbrücken wurden mit dem Basissatz 6-31G**, der Counterpoise-Korrektur und der Berücksichtigung der Fragmentrelaxationsenergie durchgeführt, da deren Einfluss bei weiter vom Minimum entfernten Komplexkonstellationen noch geklärt werden muss.

7.1.4 Die untersuchten Wasserstoffbrücken-Typen

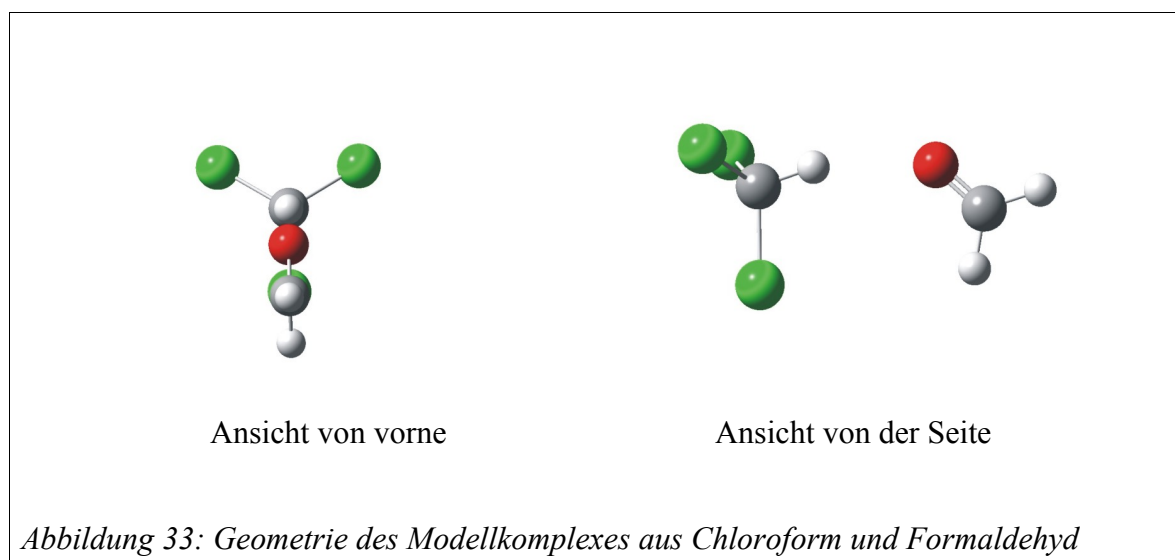
Ziel der Parametrisierung war es, die Berechnung wasserstoffbrückengebundener supramolekularer Komplexe zu verbessern bzw. zu ermöglichen. Damit sollten Kandidaten zur Kokristallisation kleiner organischer Moleküle identifiziert werden.

Vor allem für die schwachen Wasserstoffbrücken waren keine Parameter vorhanden, was dazu führen konnte, dass fälschlicherweise eine abstoßende anstelle einer anziehenden Wechselwirkung angenommen wurde, wie das beim Komplex aus Resorcinol und 2,6-Diacetylpyridin (Abb. 32) der Fall war.

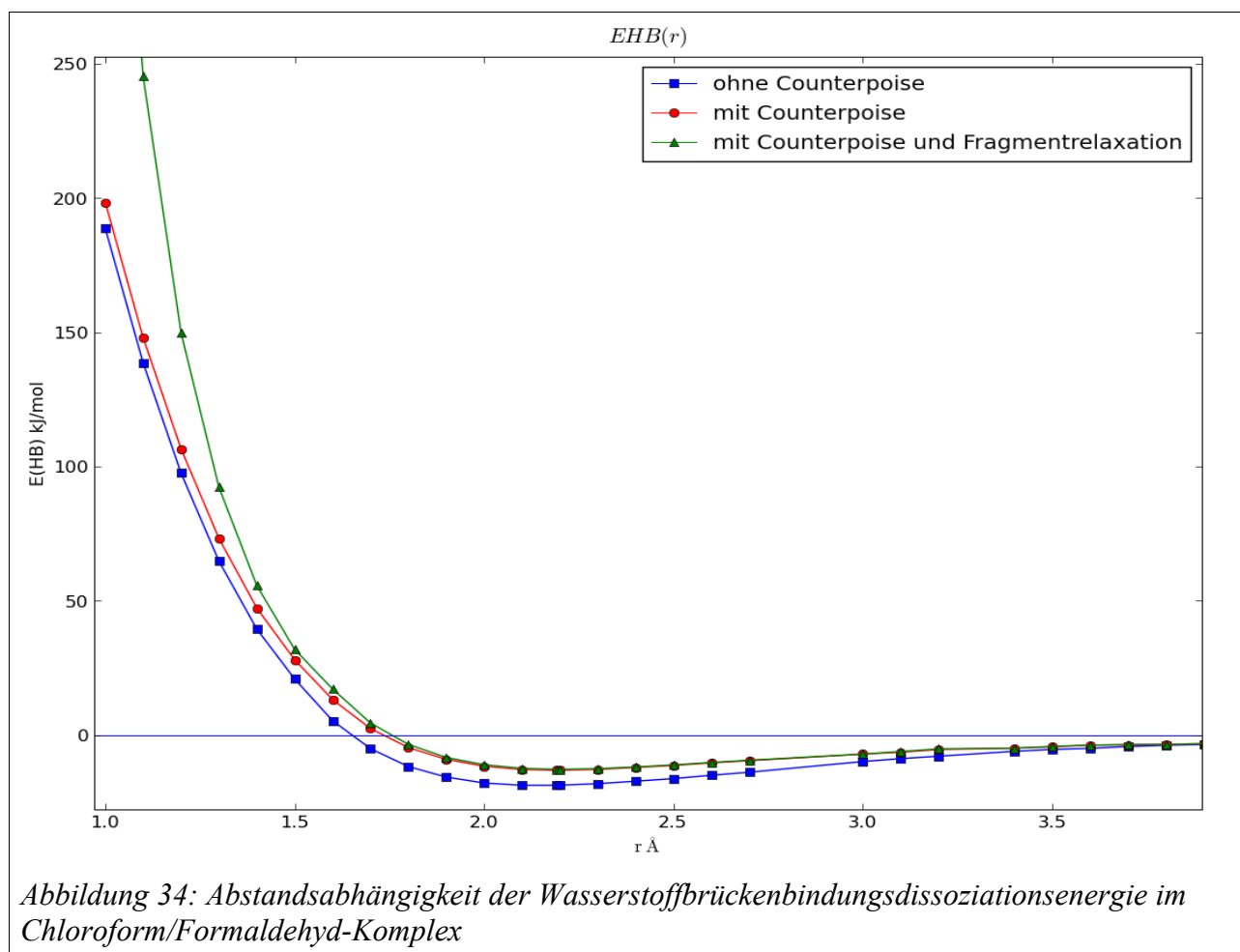


7.1.4.1 C–H...O= - Wasserstoffbrücken

Dieser Wasserstoffbrückentyp wurde anhand des Modellsystems Chloroform/Formaldehyd



untersucht. Das elektronegative Sauerstoffatom fungiert als Wasserstoffbrückenakzeptor. Chloroform hat ein azides H-Atom und eignet sich deshalb gut als Wasserstoffbrückendonator. Mit dem oben beschriebenen Verfahren wird zunächst eine freie Optimierung des Komplexes durchgeführt. Dabei ergibt sich die in Abb. 33 gezeigte Geometrie.



Es folgen systematisch partielle Optimierungen, um Stützwerte für das Dunitz-Potenzial zu erhalten. Hierbei wird jeweils einer der vier Parameter des Dunitz-Potenzials variiert und bei der Optimierung festgehalten. Die Dissoziationsenergie wird mit der eingangs beschriebenen Methode zu 12.9 kJ/mol bestimmt. Abb. 33 zeigt die Struktur des frei optimierten Komplexes von vorne und von der Seite. In der Ansicht von der Seite ist gut die Ausrichtung des Wasserstoffatoms zum freien Elektronenpaar am Carbonylsauerstoff zu erkennen. Im Anschluss an die Optimierung wurde eine Frequenzanalyse vorgenommen. Diese ergab für die abgebildete Struktur keine imaginären Frequenzen, was bedeutet, dass diese Anordnung einem Minimum auf der Energiehyperfläche entspricht. In der Ansicht von vorne stehen die Wasserstoffatome des Formaldehydes auf Deckung mit einem Chlor-Atom des Chloroforms. Unter anderen Bedingungen als bei den für die Berechnung geltenden 0 K im Vakuum geht diese Deckungskonformation wahrscheinlich verloren. Den Verlauf der H-Brückenbindungsenergie in Abhängigkeit vom Donor-Akzeptor-Abstand zeigt Abb. 34.

Der berechnete Gleichgewichtsabstand beträgt 2.19 Å. Die Werte für Abstände < 1.8 Å wachsen sehr schnell, da der Molekülabstand in und unter den Bereich der Van-der-Waals Radien fällt. Hier findet eine unnatürliche Verformung der berechneten Komplexe unter den drastischen Randbedingungen (eingefrorene Koordinaten) statt. In der Tabelle 8 ist zu erkennen, dass sich die Geometrie der Wasserstoffbrücke oberhalb des Gleichgewichtsabstandes (rot markiert) kaum ändert und mit sinkendem Abstand immer stärkere Verformungen auftreten.

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
188.63	198.35	430.70	1.00	159.56	118.33	59.22
138.54	147.93	245.26	1.10	166.12	121.54	93.48
97.53	106.56	149.79	1.20	168.34	123.52	88.44
64.74	73.20	92.28	1.30	170.09	126.06	87.14
39.44	47.25	55.65	1.40	171.46	129.16	91.41
20.83	27.99	31.84	1.50	173.15	133.53	88.41
5.31	13.18	17.12	1.60	177.48	125.13	-1.62
-5.08	2.39	4.46	1.70	176.95	125.63	-0.05
-11.79	-4.69	-3.52	1.80	174.94	125.61	2.05
-15.67	-9.09	-8.37	1.90	174.97	126.89	-2.95
-17.90	-11.67	-11.10	2.00	171.89	125.59	0.27
-18.76	-12.87	-12.45	2.10	169.29	125.04	0.00
-18.76	-13.15	-12.81	2.19	167.03	124.74	-0.04
-18.72	-13.14	-12.83	2.20	166.88	124.77	-0.21
-18.17	-12.84	-12.56	2.30	164.21	124.35	0.02
-17.23	-12.11	-11.93	2.40	160.01	123.43	-0.55
-16.25	-11.38	-11.16	2.50	157.91	123.34	-0.05
-15.01	-10.42	-10.25	2.60	154.74	123.15	-3.72
-13.89	-9.59	-9.43	2.70	152.65	122.75	0.05
-9.92	-7.13	-7.11	3.00	155.83	129.98	1.20
-8.83	-6.43	-6.18	3.10	153.43	130.27	1.92
-7.91	-5.45	-5.19	3.20	130.02	116.96	8.72
-6.10	-4.91	-4.96	3.40	165.22	143.09	-1.19
-5.41	-4.48	-4.26	3.50	160.65	142.43	4.25
-5.01	-3.79	-3.85	3.60	129.27	115.37	3.77
-4.31	-3.78	-3.56	3.70	155.71	142.49	4.52
-3.87	-3.50	-3.56	3.80	162.42	149.94	-0.57
-3.53	-3.27	-3.29	3.90	163.80	151.55	7.74

Tabelle 8: Abstandsabhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Formaldehyd-Komplex. Der Gleichgewichtsabstand ist rot markiert.

Der in Abb. 34 veranschaulichte Potenzialverlauf ähnelt dem Morse-Potenzial, die eingezeichneten Punkte entsprechen Rechenergebnissen, die Linien sind durch Matplotlib zwischen diesen Punkten interpoliert. Lücken sind durch nicht konvergierte Berechnungen aufgetreten. Der Potenzialverlauf wurde einmal ohne jede Korrektur, mit der Counterpoise-Korrektur und zusätzlich mit Korrektur der Fragmentrelaxationsenergie berechnet. Der gefundene Gleichgewichtsabstand unterscheidet sich kaum (< 0.1 Å), der Einfluss der Fragmentrelaxation beim Gleichgewichtsabstand und bei größeren Abständen ist gering. Mit kleiner werdenden Abständen steigt die Fragmentrelaxationsenergie mit der zunehmenden Deformierung der Moleküle erwartungsgemäß

stark an. Der gefundene Potenzialverlauf sollte sich durch den dem Lennard-Jones-Potenzial ähnlichen Anteil des Dunitz-Potenzials gut wiedergeben lassen.

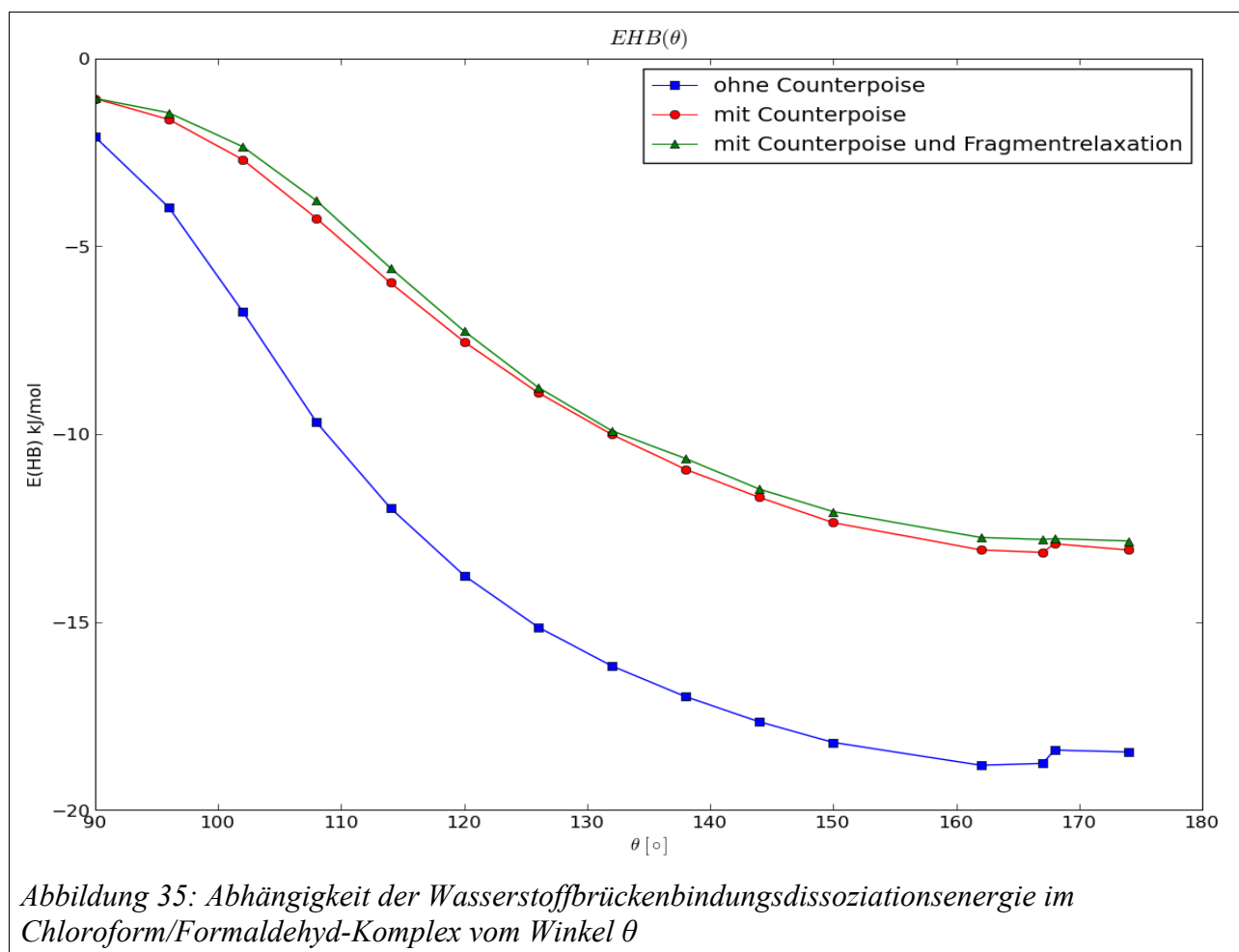


Abbildung 35: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Formaldehyd-Komplex vom Winkel θ

Als Nächstes wurden die Winkelabhängigkeiten untersucht. In Abb. 35 und Tabelle 9 wurde der Winkel θ systematisch variiert. Aus Symmetriegründen ist eine Abdeckung des Bereiches von 90° - 180° ausreichend. Die Berechnung für 180° ist nicht konvergiert. Offensichtlich hat Gaussian ein Problem mit der Fixierung von Winkeln auf genau 180° , da Berechnungen in diesem Fall nie konvergieren.

Für die Winkelabhängigkeit ergeben sich keine extremen Energiewerte, wie dies für geringe Wasserstoffbrückenbindungsabstände der Fall ist. In der deshalb größer skalierten Darstellung ist der Einfluss der Counterpoise-Korrektur zu sehen. In den Kurven ohne Fragmentrelaxationskorrektur ist ein Sprung vorhanden, der in der korrigierten Kurve fehlt, was auf eine leichte Deformation hindeutet. In der Wertetabelle findet sich an dieser Stelle ein Sprung in der

7 Angewandter Teil

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-2.09	-1.07	-1.06	4.59	90.00	101.07	13.23
-3.97	-1.63	-1.45	3.84	96.00	116.96	-10.33
-6.74	-2.69	-2.35	3.31	102.00	126.56	8.65
-9.69	-4.26	-3.79	2.92	108.00	131.91	6.86
-11.97	-5.97	-5.58	2.71	114.00	132.90	5.68
-13.76	-7.54	-7.26	2.57	120.00	132.20	5.23
-15.14	-8.89	-8.75	2.47	126.00	130.65	5.12
-16.17	-10.01	-9.91	2.39	132.00	129.12	4.68
-16.99	-10.94	-10.65	2.33	138.00	127.74	4.22
-17.66	-11.68	-11.46	2.29	144.00	126.05	3.89
-18.21	-12.36	-12.06	2.26	150.00	126.03	1.96
-18.81	-13.08	-12.75	2.20	162.00	124.18	0.01
-18.76	-13.15	-12.81	2.19	167.03	124.74	-0.04
-18.41	-12.92	-12.78	2.19	168.00	125.75	4.40
-18.46	-13.08	-12.84	2.19	174.00	127.88	0.16

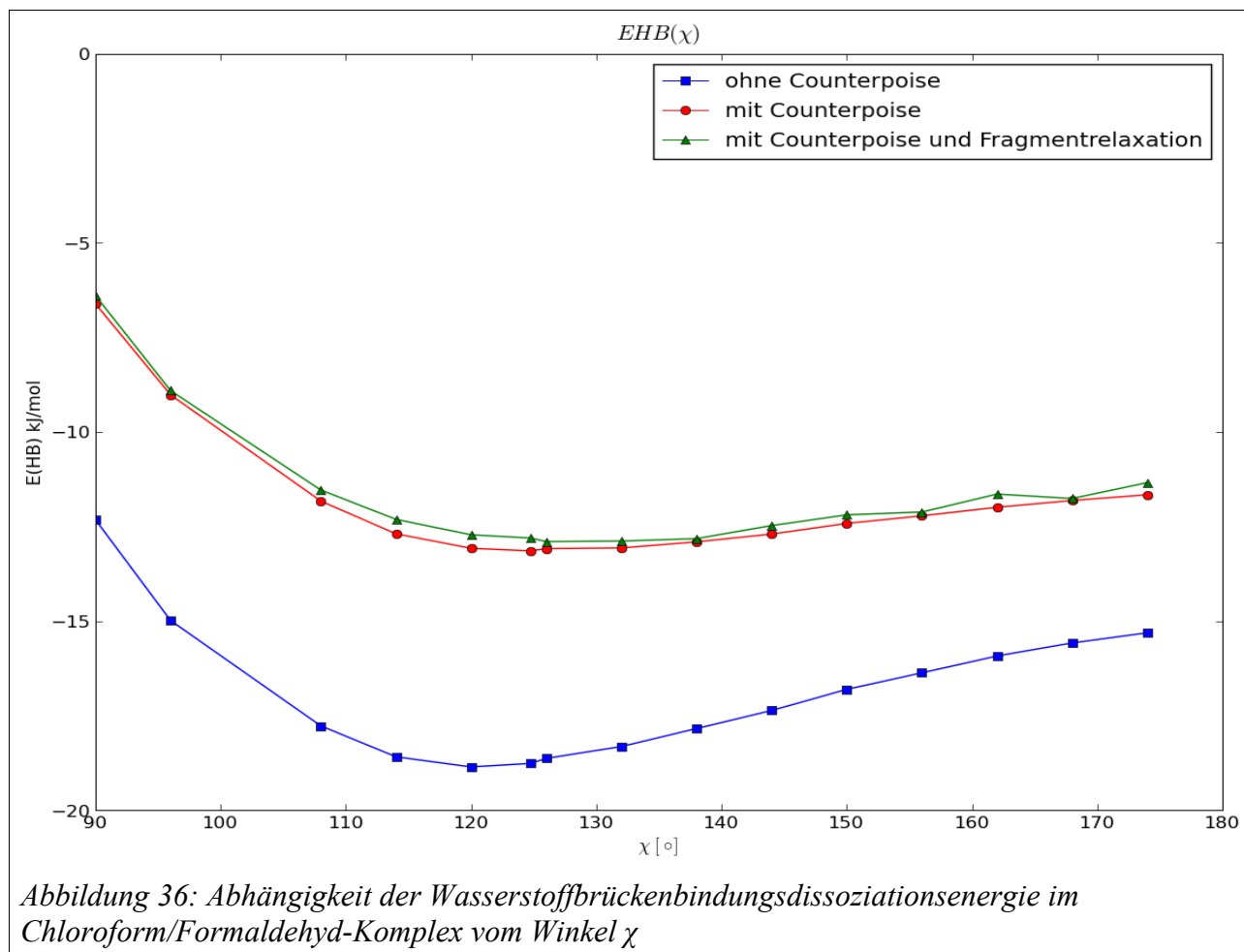
Tabelle 9: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Formaldehyd-Komplex vom Winkel θ

Größe des Torsionswinkels ω . Die Verwendung von Cosinus-Funktionen als Näherung erscheint dem Kurvenverlauf angemessen.

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-12.32	-6.61	-6.40	2.36	169.57	90.00	-0.67
-14.99	-9.02	-8.91	2.27	168.21	96.00	-0.23
-17.77	-11.83	-11.54	2.22	168.15	108.00	-0.98
-18.59	-12.69	-12.31	2.20	167.27	114.00	0.21
-18.86	-13.08	-12.72	2.19	166.38	120.00	-0.11
-18.76	-13.15	-12.81	2.19	167.03	124.74	-0.04
-18.63	-13.08	-12.90	2.19	167.59	126.00	1.62
-18.32	-13.07	-12.88	2.19	171.51	132.00	-0.10
-17.84	-12.91	-12.82	2.19	175.02	138.00	0.23
-17.37	-12.70	-12.48	2.20	177.12	144.00	0.42
-16.81	-12.42	-12.19	2.20	177.30	150.00	-5.44
-16.37	-12.22	-12.12	2.20	177.94	156.00	0.43
-15.92	-11.99	-11.64	2.21	179.15	162.00	2.87
-15.58	-11.81	-11.76	2.21	177.86	168.00	4.20
-15.31	-11.66	-11.34	2.21	177.87	174.00	34.21

Tabelle 10: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Formaldehyd-Komplex vom Winkel χ

In Abb. 36 und Tabelle 10 wurde der Winkel χ systematisch variiert. Aus Symmetriegründen ist eine Abdeckung des Bereiches von 90° - 180° ausreichend.



Der Kurvenverlauf in Abhängigkeit zum Winkel χ kann nur deutlich schlechter durch eine Cosinus-Funktion angenähert werden. Auch hier sind Unregelmäßigkeiten im Verlauf zu erkennen, die sich in der Wertetabelle als Deformationen widerspiegeln.

Eine Abhängigkeit vom Torsionswinkel ω existiert, wie Abb. 37 und Tabelle 11 zu entnehmen, bei diesem Wasserstoffbrückenbindungstyp kaum. Auffällig ist die Beeinflussung des Winkels χ . Durch das Fixieren von ω geht die Ausrichtung der Wasserstoffbrücke nach den freien Elektronenpaaren des Akzeptors Sauerstoff zum Teil verloren. Der entsprechende Teil des Dunitz-Potenzials sollte hier nicht genutzt werden. Besonders auffällig erscheint der große Unterschied in der Bandbreite der Werte mit und ohne Counterpoise-Korrektor. Der BSSE hat hier offenbar großen Einfluss auf das Ergebnis. Insgesamt wirken die Resultate mit Counterpoise-Korrektur und Korrektur der Fragmentrelaxation am geeignetsten für die nachfolgende Parametrisierung.

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-18.76	-13.15	-12.81	2.19	167.03	124.74	-0.04
-18.75	-13.13	-12.93	2.19	167.07	124.77	0.00
-18.71	-13.10	-12.75	2.19	166.68	124.74	6.00
-18.43	-12.96	-12.67	2.19	168.08	126.16	12.00
-18.01	-12.78	-12.67	2.19	170.29	128.91	18.00
-17.53	-12.55	-12.46	2.20	171.80	132.50	24.00
-17.14	-12.33	-12.17	2.20	170.56	134.64	30.00
-15.99	-11.82	-11.72	2.20	174.47	146.35	42.00
-15.67	-11.69	-11.63	2.21	174.05	151.11	48.00
-15.27	-11.65	-11.62	2.21	178.92	179.69	60.01
-15.27	-11.64	-11.50	2.21	179.23	178.78	66.00
-15.27	-11.65	-11.57	2.21	179.00	179.84	72.02
-15.29	-11.62	-11.47	2.20	178.73	179.97	78.05
-15.23	-11.60	-11.26	2.21	178.95	179.71	84.00
-15.27	-11.65	-11.33	2.21	179.14	179.99	89.70
-15.08	-11.46	-11.39	2.23	171.88	175.65	102.00
-15.27	-11.64	-11.58	2.21	179.77	179.69	113.99
-15.26	-11.63	-11.59	2.21	178.71	179.04	120.00
-15.40	-11.56	-11.36	2.21	173.30	155.89	126.00
-15.42	-11.65	-11.62	2.21	175.51	161.77	132.00
-16.00	-11.82	-11.65	2.20	174.00	145.55	138.00
-15.28	-11.65	-11.42	2.21	178.82	179.82	150.01
-15.27	-11.64	-11.42	2.21	179.21	179.74	167.99

Tabelle 11: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Formaldehyd-Komplex vom Winkel ω

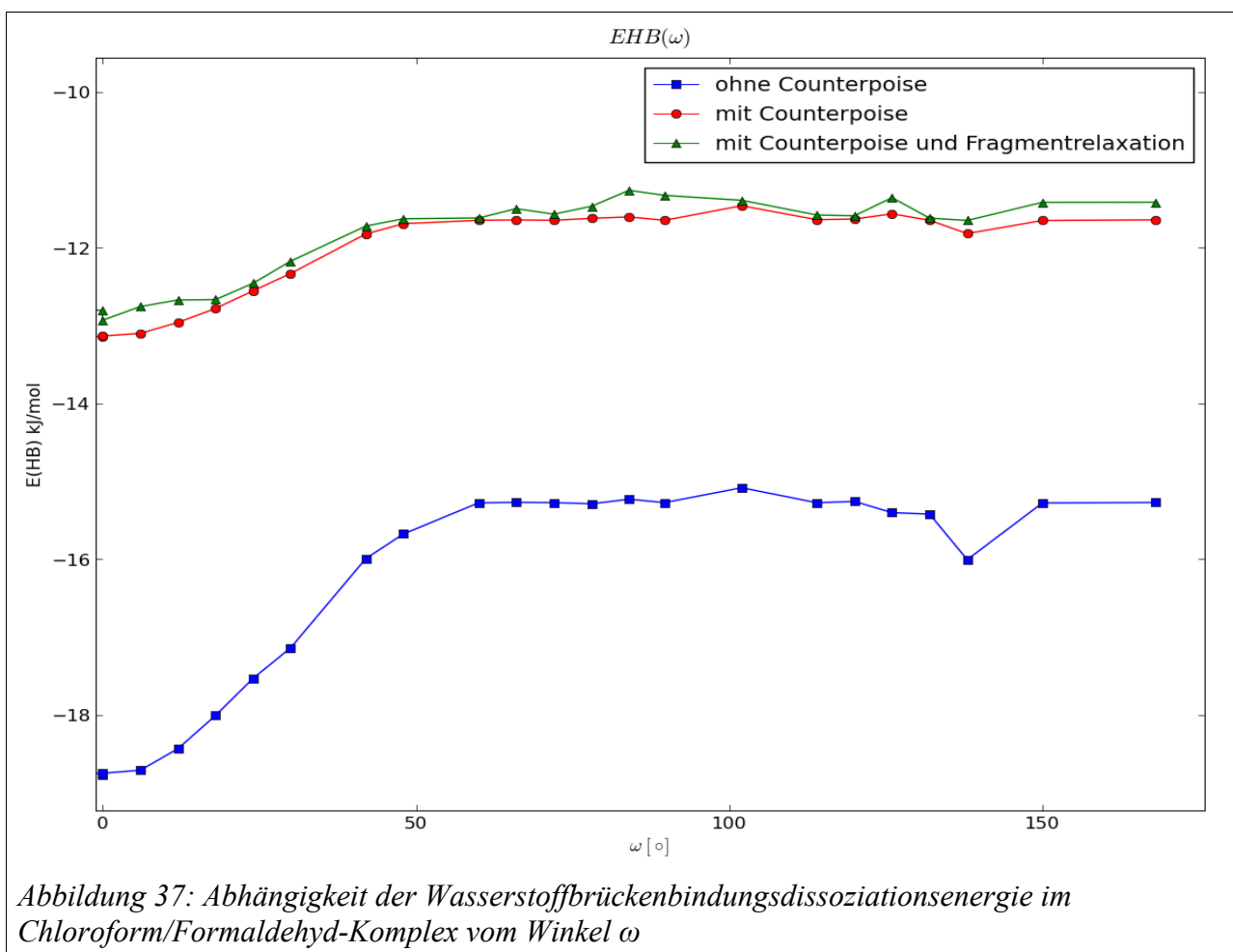
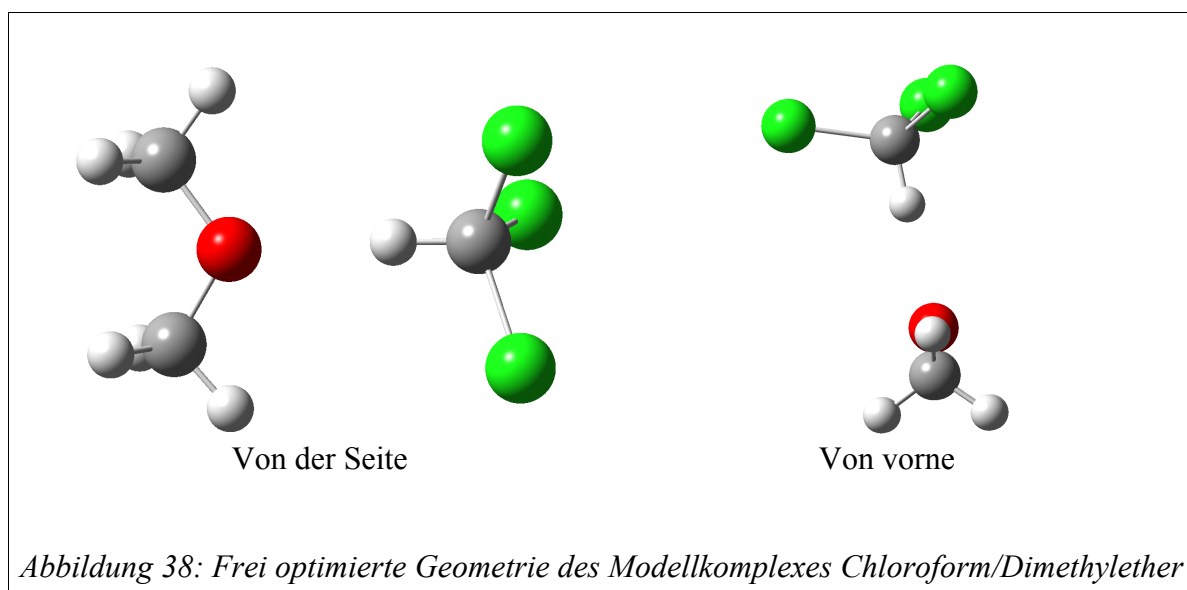


Abbildung 37: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Formaldehyd-Komplex vom Winkel ω

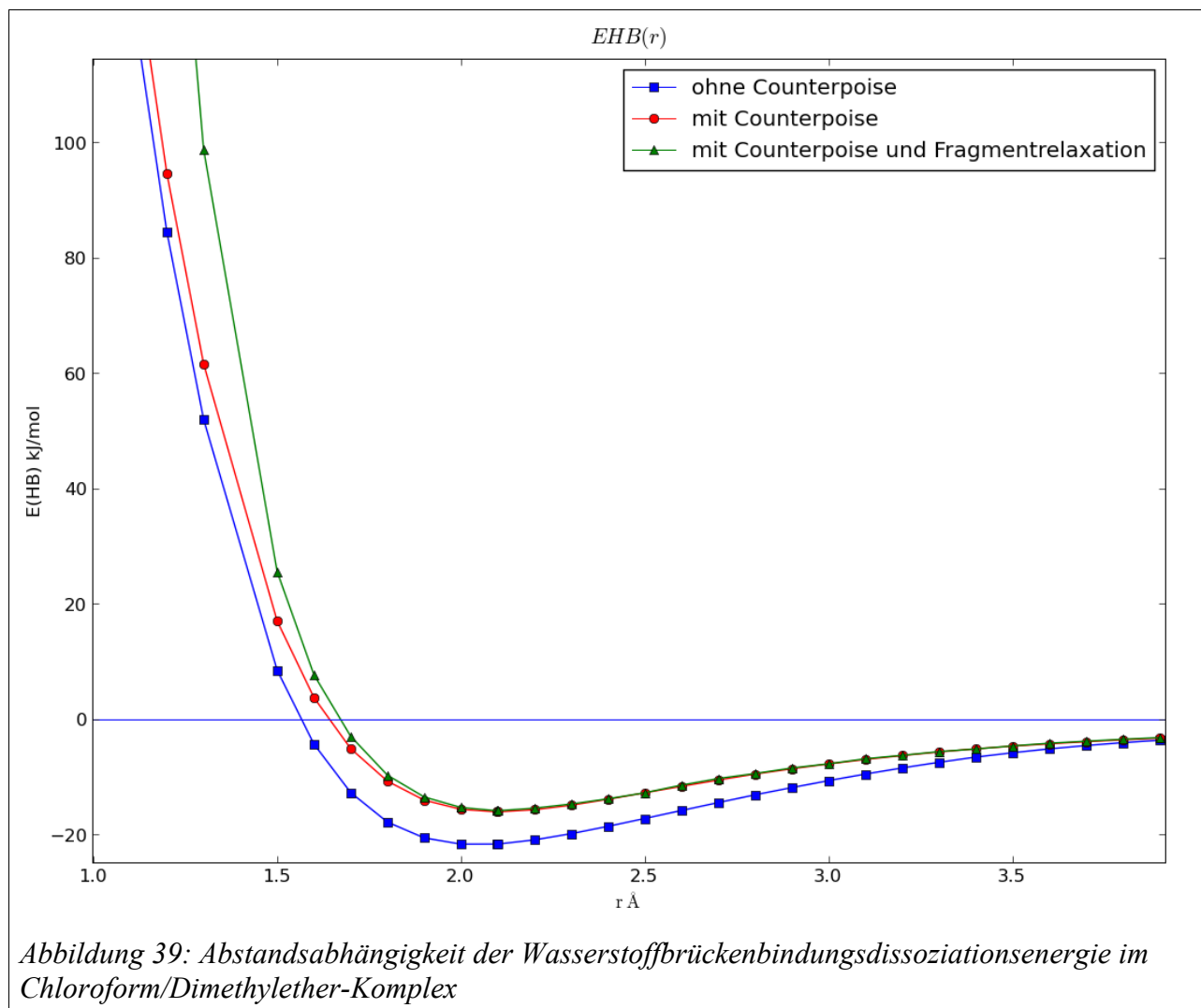
7.1.4.2 C-H...O- -Wasserstoffbrücken

Dieser Wasserstoffbrückenbindungstyp wurde anhand des Modellkomplexes Chloroform/Dimethylether untersucht. Die Equilibriumsstruktur demonstriert Abb. 38. Die Ausrichtung des Donors zu einem der freien Elektronenpaare am Akzeptor Sauerstoff hin ist deutlich zu erkennen. Die Überprüfung dieser Struktur mittels einer Frequenzanalyse offenbart, dass es sich aufgrund der Abwesenheit imaginärer Frequenzen um ein Minimum handeln muss. Die Energie dieser Wasserstoffbrücke beträgt in der Equilibriumsstruktur 15.9 kJ/mol.



Es folgen systematisch partielle Optimierungen, um Stützwerte für das Dunitz-Potenzial zu erhalten. Für den Verlauf der H-Brückenbindungsenergie ergibt sich der in Abb. 39 gezeigte Graph. Der berechnete Gleichgewichtsabstand beträgt 2.1 Å, die Werte für $r < 1.8$ Å erscheinen unrealistisch und ergeben, wie in den anderen Fällen, eine unnatürliche Verformung aufgrund der bei der Optimierung festgehaltenen Abstandscoordinate. Dies ist auch an der sich vergrößernden Differenz der Kurve mit Korrektur der Fragmentrelaxation und denen ohne zu sehen. Der Kurvenverlauf ähnelt dem Morse-Potenzial, die berechneten Werte sind in Tabelle 12 aufgelistet. Auch hier fällt eine deutliche Veränderung der Geometrie der Wasserstoffbrücke bei extrem kleinen Abständen r auf. Der mit den verschiedenen Berechnungsmethoden gefundene Gleichgewichtsabstand unterscheidet sich praktisch nicht.

In Abhängigkeit vom Donor-Wasserstoff-Akzeptor-Winkel (θ) ergibt sich der in Abb. 40 veranschaulichte Potenzialverlauf. In der Gleichgewichtsstruktur ist $\theta = 177.9^\circ$. Im Bereich von 168°



bis 180° entsteht jedoch bei partieller Optimierung mit festgehaltenem Winkel θ nur eine geringe Energiedifferenz von < 0.5 kJ/mol. Zu kleineren Winkeln hin steigt die Energiedifferenz schnell an und erreicht bei 90° schon 15.6 kJ/mol. Die genauen Werte sind Tabelle 13 zu entnehmen. Die „Dellen“ in der Potenzialkurve um 90° und 180° sind in der Wertetabelle als Verformung des Komplexes manifestiert und haben ihre Ursache in sterischen Problemen aufgrund der Methylsubstituenten am Dimethylether.

7 Angewandter Teil

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
170.22	181.16	492.71	1.00	175.87	112.31	-39.36
84.51	94.67	173.11	1.20	179.30	123.18	-10.33
51.99	61.62	98.80	1.30	179.62	124.17	3.40
8.50	17.03	25.60	1.50	179.12	124.29	5.27
-4.31	3.76	7.72	1.60	179.64	124.15	-1.75
-12.72	-5.10	-2.96	1.70	179.50	123.85	-6.17
-17.85	-10.75	-9.71	1.80	179.03	123.46	-3.48
-20.58	-14.03	-13.49	1.90	178.51	122.99	-5.95
-21.67	-15.64	-15.31	2.00	177.87	122.51	10.03
-21.66	-16.07	-15.88	2.10	177.75	122.33	12.72
-21.66	-16.07	-15.87	2.10	177.85	122.32	13.49
-20.93	-15.70	-15.43	2.20	178.09	122.45	-14.34
-19.85	-14.91	-14.71	2.30	177.98	122.36	-15.13
-18.58	-13.89	-13.80	2.40	178.02	122.37	-17.57
-17.22	-12.76	-12.76	2.50	177.75	122.32	-18.92
-15.83	-11.64	-11.43	2.60	177.32	122.17	-19.46
-14.46	-10.55	-10.29	2.70	177.07	122.11	-19.61
-13.11	-9.52	-9.42	2.80	177.80	122.26	-19.70
-11.86	-8.60	-8.46	2.90	176.29	122.44	-19.35
-10.65	-7.76	-7.73	3.00	176.01	122.51	-18.75
-9.52	-6.99	-6.86	3.10	171.09	122.23	-20.08
-8.45	-6.30	-6.27	3.20	175.11	122.59	-17.72
-7.48	-5.69	-5.63	3.30	174.24	122.67	-17.52
-6.57	-5.13	-5.17	3.40	176.36	123.00	-16.34
-5.82	-4.68	-4.63	3.50	172.19	122.83	-16.91
-5.15	-4.27	-4.17	3.60	171.35	122.86	-16.55
-4.57	-3.90	-3.82	3.70	171.25	122.93	-16.21
-4.07	-3.58	-3.48	3.80	170.72	122.91	-16.09
-3.65	-3.30	-3.19	3.90	170.59	122.95	-15.89

Tabelle 12: Abstandsabhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Dimethylether-Komplex. Der Gleichgewichtsabstand ist rot markiert.

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-0.97	0.35	0.32	4.48	90.00	111.53	-51.40
-1.34	-0.52	-0.40	4.68	96.00	100.28	-17.86
-5.24	-1.18	-0.72	3.33	102.00	123.90	-0.75
-8.90	-2.97	-2.28	2.87	108.00	123.06	15.26
-11.68	-5.04	-4.42	2.65	114.00	123.39	12.68
-14.04	-7.08	-6.59	2.49	120.00	123.16	14.31
-15.90	-8.93	-8.54	2.38	126.00	123.49	10.61
-17.38	-10.56	-10.23	2.30	132.00	123.28	12.52
-18.53	-11.93	-11.63	2.24	138.00	123.66	7.13
-19.35	-13.05	-12.89	2.20	144.00	123.17	7.88
-20.07	-14.02	-13.80	2.17	150.00	122.70	6.82
-20.61	-14.79	-14.59	2.14	156.00	122.29	3.11
-21.16	-15.39	-15.13	2.12	162.00	124.61	8.40
-21.41	-15.81	-15.41	2.11	168.00	121.32	-5.53
-21.64	-16.03	-15.85	2.10	174.00	121.62	-11.38
-21.66	-16.07	-15.88	2.10	177.75	122.33	12.72
-21.70	-15.85	-15.47	2.10	180.00	120.51	-30.59

Tabelle 13: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Dimethylether-Komplex vom Winkel θ

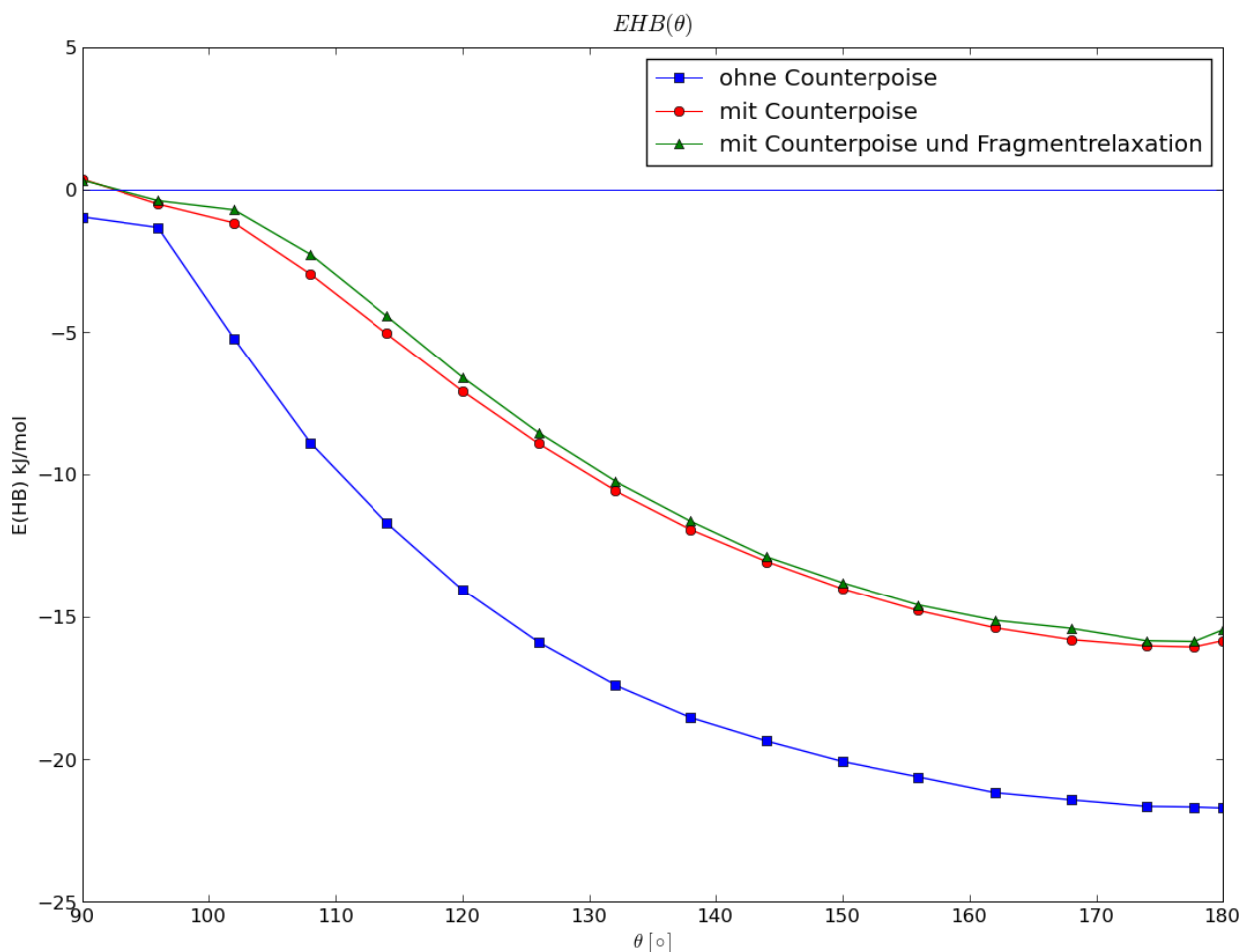


Abbildung 40: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Dimethylether-Komplex vom Winkel θ

In den Wertetabellen für die partiellen Optimierungen bei festgehaltenem Winkel χ (Tabelle 14) und ω (Tabelle 15) ist eine deutliche gegenseitige Beeinflussung dieser beiden Winkel auszumachen, die letztendlich durch sterische, auf die voluminösen Methylsubstituenten zurückzuführende Gründe verursacht wird. Viele partielle Optimierungen, bei denen einer dieser Winkel festgehalten wurde, sind nicht konvergiert, weshalb die Wertetabelle vergleichsweise kurz ausfällt. In der Darstellung zur Variation des Winkels χ (Abb. 41) ist der typische Potenzialverlauf noch zu sehen, beim Winkel ω wurde jedoch aufgrund zu starker Beeinflussung der Komplexgeometrie (Tabelle 15) auf eine grafische Auftragung verzichtet.

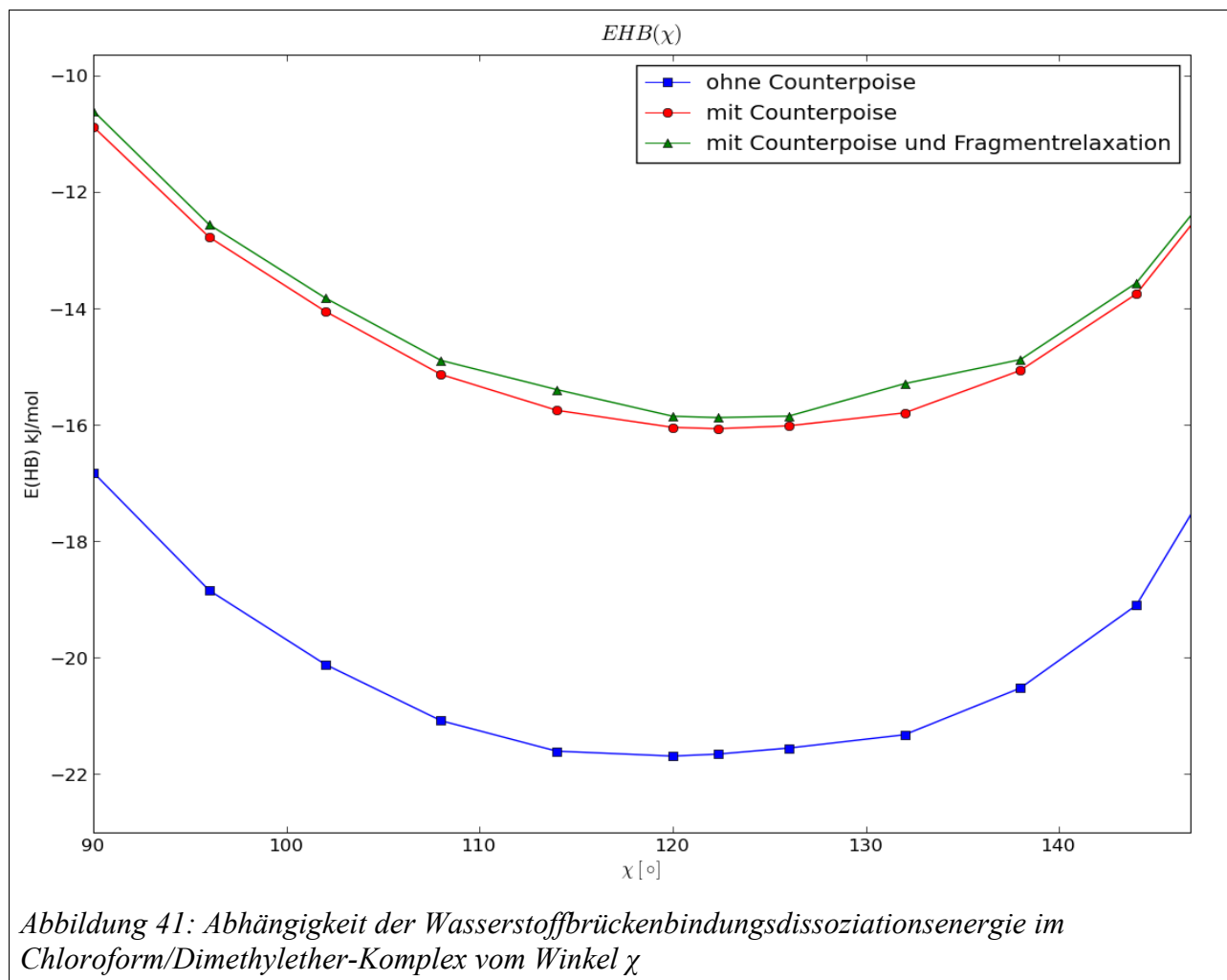
7 Angewandter Teil

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-16.82	-10.88	-10.61	2.23	176.82	90.00	-53.66
-18.84	-12.78	-12.56	2.18	176.42	96.00	-47.88
-20.11	-14.05	-13.82	2.15	176.74	102.00	-42.71
-21.08	-15.14	-14.89	2.12	176.48	108.00	-33.51
-21.61	-15.75	-15.40	2.11	176.36	114.00	-27.28
-21.69	-16.04	-15.85	2.10	177.22	120.00	-16.05
-21.66	-16.07	-15.88	2.10	177.75	122.33	12.72
-21.55	-16.02	-15.85	2.10	178.78	126.00	-6.81
-21.33	-15.80	-15.29	2.10	175.80	132.00	2.91
-20.52	-15.07	-14.88	2.12	176.06	138.00	2.07
-19.10	-13.75	-13.56	2.13	175.35	144.00	-1.41
-2.55	-1.25	-1.25	3.49	163.34	174.00	-13.00

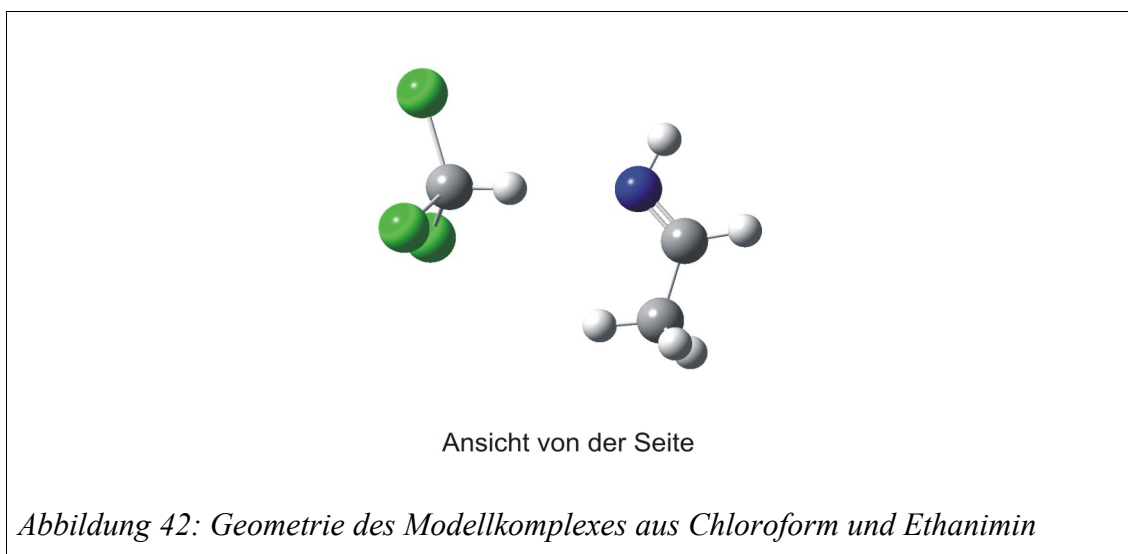
Tabelle 14: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Dimethylether-Komplex vom Winkel χ

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-20.77	-14.50	-14.01	2.13	177.00	115.66	-180.00
-21.66	-16.07	-15.88	2.10	177.75	122.33	12.72
-21.67	-15.85	-15.50	2.11	178.58	121.03	90.00
-21.69	-15.96	-15.61	2.10	178.68	122.09	96.00
-10.20	-3.76	-3.33	2.19	174.89	108.01	126.00
-11.80	-5.45	-5.07	2.21	175.72	107.64	132.00
-13.48	-7.09	-6.74	2.21	174.85	108.31	138.00
-17.67	-11.26	-10.92	2.17	175.33	109.25	156.00
-19.56	-13.16	-13.01	2.15	176.11	110.67	168.00
-20.23	-13.90	-13.62	2.14	176.34	113.83	174.00

Tabelle 15: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Dimethylether-Komplex vom Winkel ω



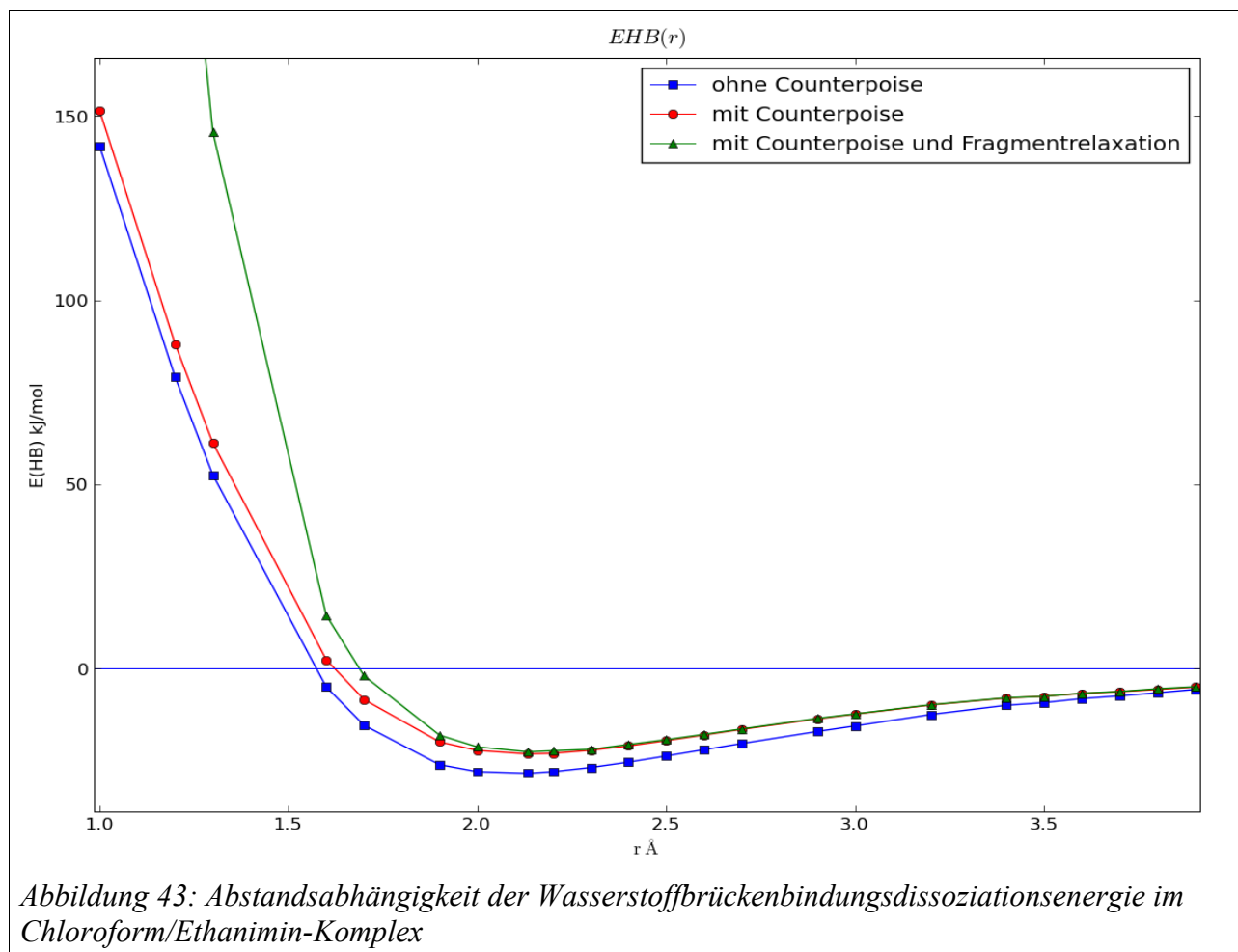
7.1.4.3 C-H...N= -Wasserstoffbrücken



Dieser Wasserstoffbrückenbindungstyp wurde anhand des Modellsystems Chloroform/Ethanimin untersucht. Abb. 42 skizziert die Struktur des frei optimierten Komplexes von der Seite. Man erkennt die Ausrichtung des Chloroform-Wasserstoffs (Donor) zum Stickstoffatom (Akzeptor).

Die Energie dieser Wasserstoffbrücke wurde zu 22.6 kJ/mol berechnet. Man nimmt in der Equilibriumsstruktur eine Deckungskonformation zwischen dem Wasserstoff am Stickstoff und einem Chloratom des Chloroforms wahr. Man kann davon ausgehen, dass diese Cl-H-Wechselwirkung viel schwächer ausgeprägt ist als die H-N-Wechselwirkung, dennoch geht sie mit in die ermittelte Energie ein. Man kann wie beim Chloroform/Formaldehyd-Komplex (Kap. 7.1.4.1) antizipieren, dass die Deckungskonformation bei höheren Temperaturen verloren geht. Im Anschluss an die Optimierung fand eine Frequenzanalyse statt. Diese ergab für die abgebildete Struktur keine imaginären Frequenzen, was bedeutet, dass diese Anordnung einem Minimum auf der Energiehyperfläche entspricht.

Es folgen systematisch partielle Optimierungen, um Stützwerte für das Dunitz-Potenzial zu erhalten. Hierbei wird jeweils einer der vier Parameter des Dunitz-Potenzials variiert und bei der Optimierung festgehalten. Für den Verlauf der H-Brückenbindungsenergie in Abhängigkeit vom Donor-Akzeptor-Abstand ergibt sich der in Abb. 43 gezeigte Verlauf.



Der berechnete Gleichgewichtsabstand beträgt 2.13 \AA . Die Werte für $< 1.9 \text{ \AA}$ erscheinen unrealistisch und sind vermutlich der unnatürlichen Verformung der berechneten Komplexe unter den drastischen Randbedingungen (eingefrorene Koordinaten) geschuldet.

Der in Abb. 43 illustrierte Potenzialverlauf ähnelt dem Morse-Potenzial, die detaillierten Ergebnisse sind in Tabelle 16 aufgeführt.

7 Angewandter Teil

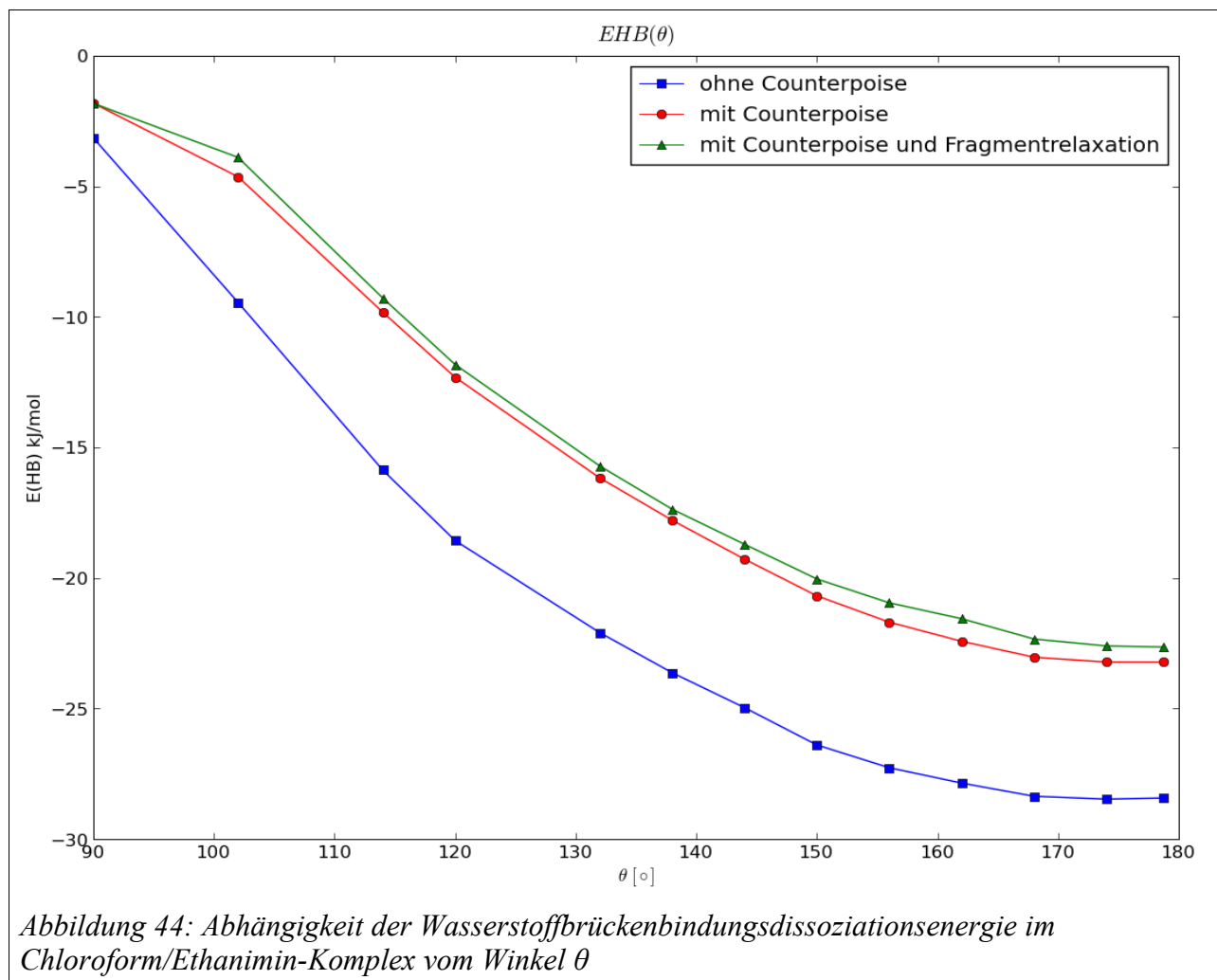
EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
141.86	151.52	645.40	1.00	157.75	119.84	154.27
79.12	88.12	248.24	1.20	169.72	122.05	147.05
52.54	61.22	145.58	1.30	174.37	122.25	151.83
-5.00	2.38	14.43	1.60	176.25	122.08	172.78
-15.41	-8.42	-1.98	1.70	178.42	123.73	176.22
-26.10	-19.92	-18.11	1.90	177.82	126.20	178.09
-28.00	-22.26	-21.29	2.00	180.00	127.54	-179.77
-28.42	-23.22	-22.64	2.13	178.71	128.02	178.96
-28.00	-23.04	-22.34	2.20	179.99	128.50	179.45
-26.88	-22.22	-21.96	2.30	177.17	128.29	178.11
-25.41	-21.02	-20.62	2.40	179.18	129.13	178.27
-23.75	-19.57	-19.32	2.50	178.06	129.66	178.19
-22.06	-18.07	-17.90	2.60	178.98	129.23	178.39
-20.35	-16.53	-16.42	2.70	174.80	127.90	177.77
-17.11	-13.68	-13.53	2.90	173.70	127.60	177.69
-15.60	-12.40	-12.33	3.00	174.70	128.03	176.71
-12.50	-9.91	-9.84	3.20	162.17	123.20	165.04
-9.99	-8.02	-7.96	3.40	161.24	119.04	165.84
-9.26	-7.56	-7.59	3.50	171.57	127.66	172.68
-8.19	-6.80	-6.70	3.60	168.39	125.96	166.82
-7.45	-6.31	-6.25	3.70	172.54	127.13	177.88
-6.59	-5.70	-5.50	3.80	165.23	124.05	169.91
-5.69	-5.04	-4.94	3.90	154.14	117.20	163.43

Tabelle 16: Abstandsabhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Ethanimin-Komplex. Der Gleichgewichtsabstand ist rot markiert.

In Abhängigkeit vom Donor-Wasserstoff-Akzeptor-Winkel (θ) ergibt sich der in Abb. 44 aufgetragene Potenzialverlauf.

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-3.15	-1.82	-1.83	4.24	90.00	167.72	48.45
-9.45	-4.64	-3.90	3.10	102.00	154.25	175.58
-15.87	-9.83	-9.28	2.63	114.00	149.88	179.08
-18.57	-12.32	-11.83	2.50	120.00	144.69	178.13
-22.09	-16.17	-15.71	2.34	132.00	142.86	165.20
-23.63	-17.79	-17.37	2.27	138.00	142.01	165.82
-24.97	-19.28	-18.71	2.24	144.00	138.01	166.54
-26.39	-20.68	-20.04	2.21	150.00	123.48	173.64
-27.26	-21.69	-20.95	2.18	156.00	122.62	173.29
-27.85	-22.42	-21.56	2.15	162.00	122.78	173.35
-28.36	-23.03	-22.34	2.14	168.00	124.14	178.97
-28.47	-23.22	-22.60	2.13	174.00	126.11	178.95
-28.42	-23.22	-22.64	2.13	178.71	128.02	178.96

Tabelle 17: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Ethanimin-Komplex vom Winkel θ



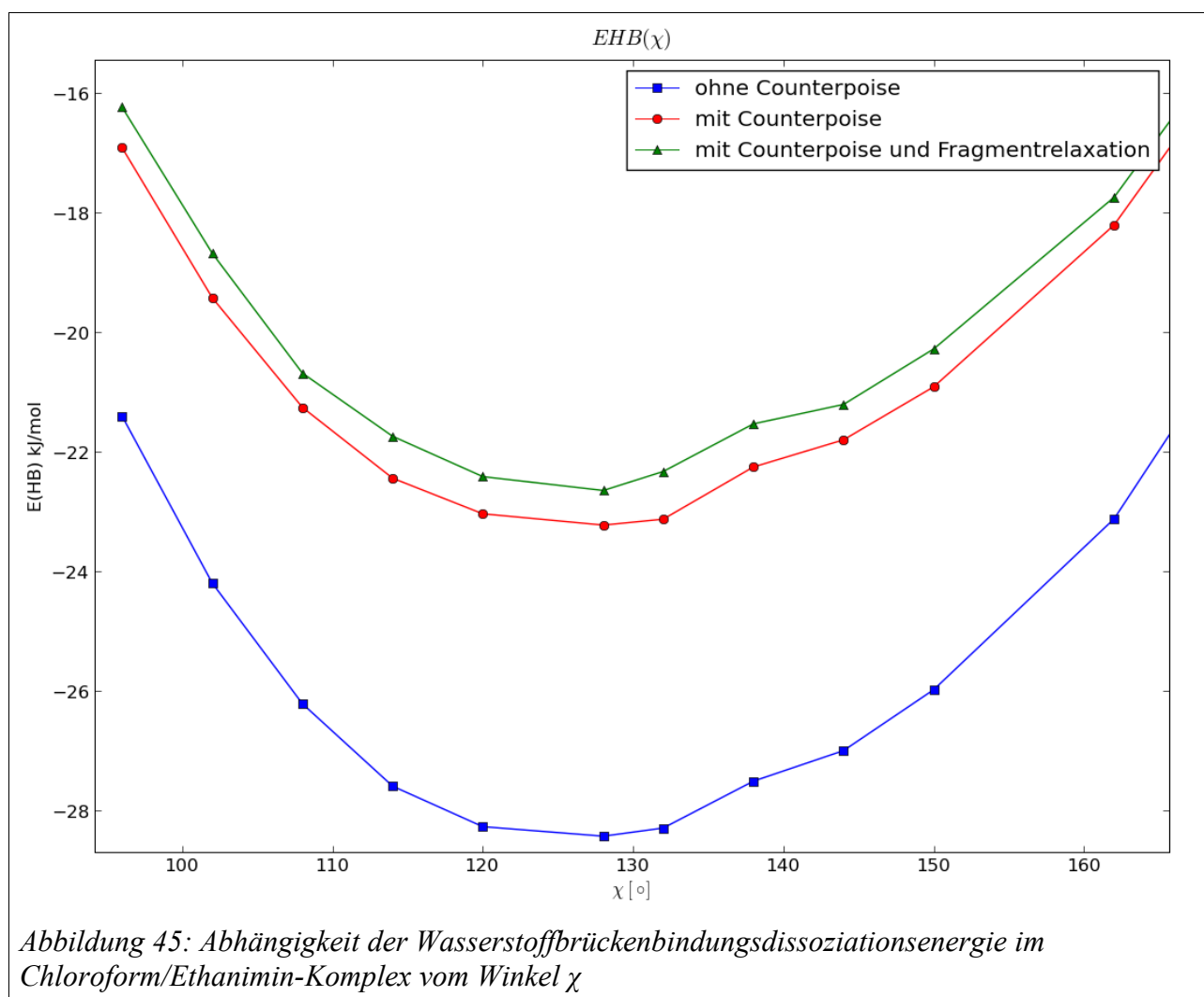
Der entsprechende Winkel in der Gleichgewichtsstruktur beträgt 179° . Im Bereich zwischen 168° und 179° ist die Energiedifferenz der mit festgehaltenem Winkel θ partiell optimierten Struktur zur frei optimierten Struktur am geringsten, der Unterschied zwischen der Equilibriumsstruktur und der mit $\theta = 168^\circ$ optimierten beträgt gerade einmal 0.3 kJ/mol.

Die Wasserstoffbrücke wird durch einen etwas kleineren Donor-Wasserstoff-Akzeptor Winkel also kaum geschwächt. Zu kleineren Winkeln steigt die Energiedifferenz schnell an und erreicht für 90° bereits 20.8 kJ/mol. Die genauen Werte können aus Tabelle 17 entnommen werden.

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-21.40	-16.91	-16.23	2.22	173.70	96.00	174.64
-24.19	-19.42	-18.68	2.17	172.21	102.00	176.83
-26.21	-21.25	-20.68	2.15	172.80	108.00	178.45
-27.59	-22.44	-21.74	2.14	171.76	114.00	179.80
-28.26	-23.03	-22.41	2.13	172.88	120.00	179.90
-28.42	-23.22	-22.64	2.13	178.71	128.02	178.96
-28.29	-23.12	-22.33	2.14	177.89	132.00	179.61
-27.50	-22.25	-21.53	2.15	164.62	138.00	176.45
-26.99	-21.80	-21.20	2.15	166.74	144.00	176.25
-25.97	-20.91	-20.27	2.16	168.96	150.00	-179.52
-23.12	-18.20	-17.74	2.20	160.96	162.00	-179.83
-16.24	-11.91	-11.57	2.36	153.69	180.00	171.66

Tabelle 18: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Ethanimin-Komplex vom Winkel χ

Abb. 45 und die Tabelle 18 präsentieren die berechnete Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie vom Winkel χ .



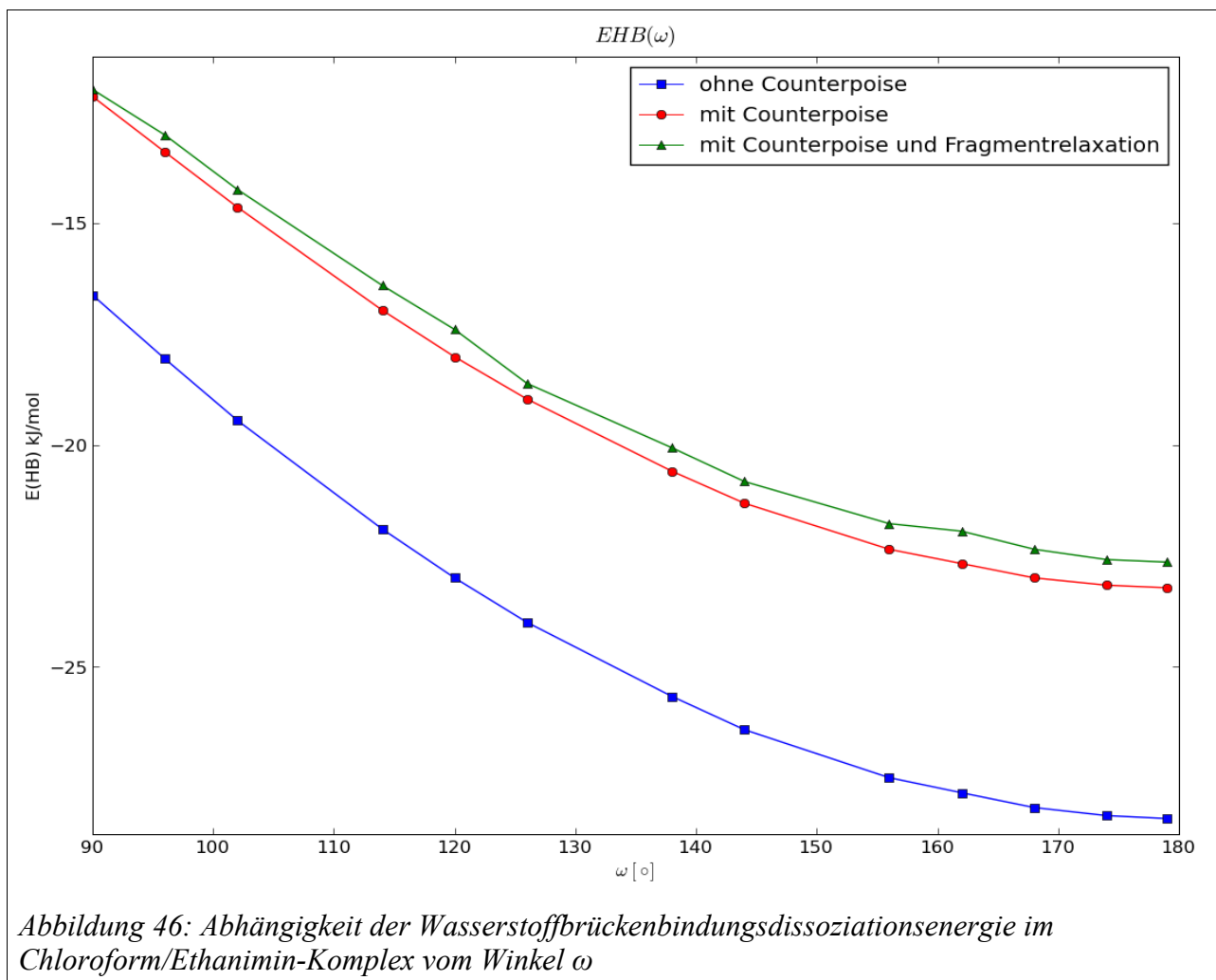
Der Winkel hat in der Equilibriumsstruktur den Wert 128° . Im Bereich von 120° bis 132° wird die Wasserstoffbrücke durch eine Abweichung von diesem Wert jedoch nur geringfügig geschwächt. Bei 132° beträgt der Energieunterschied zur Gleichgewichtsstruktur nur 0.3 kJ/mol. Außerhalb dieses Fensters wächst die Energiedifferenz recht schnell bis auf 10.9 kJ/mol für $\chi=180^\circ$.

Die partiellen Optimierungen mit fixiertem Winkel ω ergeben den in Abb. 46 und Tabelle 19 wiedergegebenen Potenzialverlauf.

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [$^\circ$]	Chi [$^\circ$]	Omega [$^\circ$]
-16.62	-12.15	-11.99	2.31	155.38	143.97	90.00
-18.06	-13.40	-13.01	2.27	157.24	138.00	96.00
-19.45	-14.64	-14.24	2.23	159.76	133.98	102.00
-21.90	-16.96	-16.40	2.19	163.96	131.91	114.00
-23.00	-18.01	-17.40	2.18	165.26	131.59	120.00
-24.00	-18.97	-18.61	2.17	167.12	130.28	126.00
-25.67	-20.59	-20.07	2.15	171.51	129.16	138.00
-26.42	-21.31	-20.83	2.15	172.59	128.94	144.00
-27.50	-22.35	-21.78	2.14	175.39	128.17	156.00
-27.84	-22.68	-21.95	2.13	177.10	127.12	162.00
-28.18	-23.00	-22.35	2.14	176.96	127.41	168.00
-28.36	-23.16	-22.59	2.13	179.33	127.87	174.00
-28.42	-23.22	-22.64	2.13	178.71	128.02	178.96

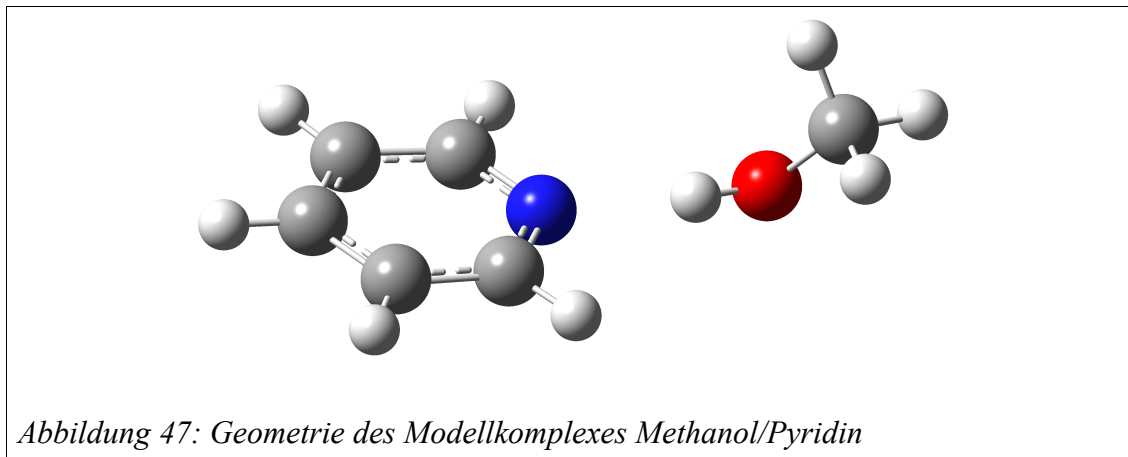
Tabelle 19: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Chloroform/Ethanimin-Komplex vom Winkel ω

In der frei optimierten Struktur beträgt der Torsionswinkel $\omega = 179^\circ$. Die Wasserstoffbrücke wird mit abweichendem Torsionswinkel kontinuierlich geschwächt. Im Bereich von 168° bis 179° wird die Wasserstoffbrücke durch eine Abweichung von diesem Wert jedoch nur geringfügig beeinflusst. Bei 168° beträgt der Energieunterschied zur Gleichgewichtsstruktur nur 0.3 kJ/mol. Außerhalb dieses Fensters wächst die Energiedifferenz recht schnell bis auf 10.9 kJ/mol für $\chi=180^\circ$. Anders als beim Chloroform/Formaldehyd-Komplex wirkt sich ω deutlich auf die Stärke der Wasserstoffbrücke aus und muss in der Potentialfunktion entsprechend berücksichtigt werden.

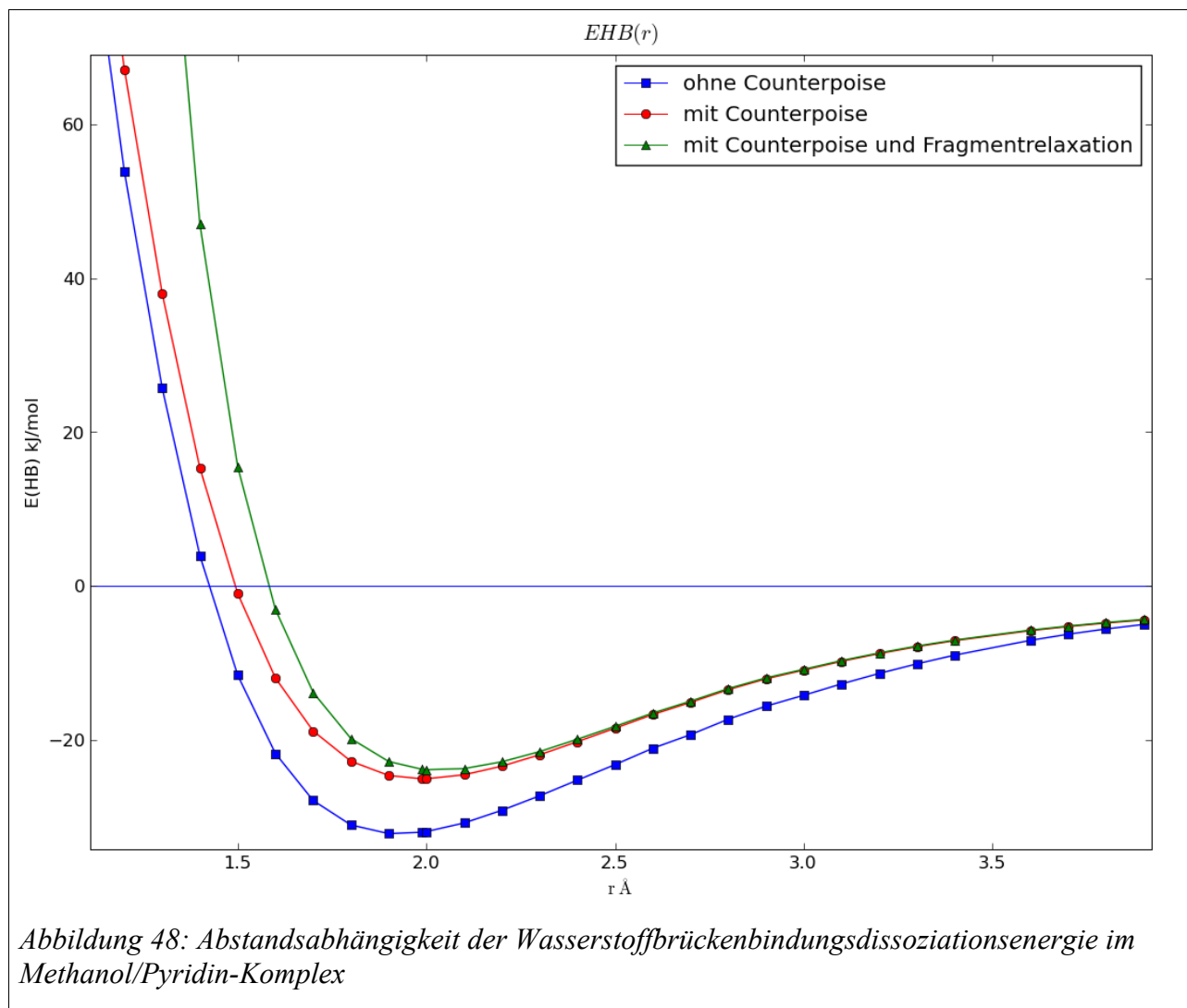


7.1.4.4 O–H...N(Aromat)-Wasserstoffbrücken

Dieser Wasserstoffbrückenbindungstyp wurde anhand des Modellsystems Methanol/Pyridin bestimmt. Abb. 47 zeigt die Struktur des frei optimierten Komplexes.



Die Überprüfung dieser Struktur mittels einer Frequenzanalyse zeigt, dass es sich aufgrund der Abwesenheit imaginärer Frequenzen um ein Minimum handeln muss. Die Energie der Wasserstoffbrücke beträgt im Gleichgewicht 23.9 kJ/mol. Durch systematische partielle Optimierung wurden Stützwerte für die Parametrisierung des Dunitz-Potenzials erhalten. Die Abhängigkeit der Wasserstoffbrückenbindungsenergie zum Abstand ist aus Abb. 48 und der Tabelle 20 ersichtlich. Der Wasserstoff-Akzeptor Abstand in der Equilibriumsstruktur beträgt 1.99 Å, die Lage des Minimums unterscheidet sich leicht, je nach Energieberechnungsmethode. Für Abstände $r < 1.8$ Å sinkt die Wasserstoffbrückenbindungsenergie sehr schnell und ändert schließlich ihr Vorzeichen. Dies ist wie in den anderen Fällen einer unrealistischen Verformung des Komplexes aufgrund des kurzen, festgehaltenen Molekülabstandes geschuldet.



In Abhängigkeit vom Donor-Wasserstoff-Akzeptor-Winkel (θ) ergibt sich der in Abb. 49 aufgetragene Potenzialverlauf. In der Gleichgewichtsstruktur ist $\theta = 172.6^\circ$. Im Bereich von 162° bis 174° entsteht bei partieller Optimierung mit festgehaltenem Winkel θ nur ein geringer Energieunterschied von < 0.3 kJ/mol. Die genauen Zahlen sind der Tabelle 21 zu entnehmen. Hier sind deutliche Auswirkungen der festgehaltenen θ -Koordinate auf die Komplexgeometrie, vor allem was den Winkel χ angeht, sichtbar. Umgekehrt ist dies auch in Tabelle 22 bei fixierter χ -Koordinate bezüglich des Winkels θ zu beobachten.

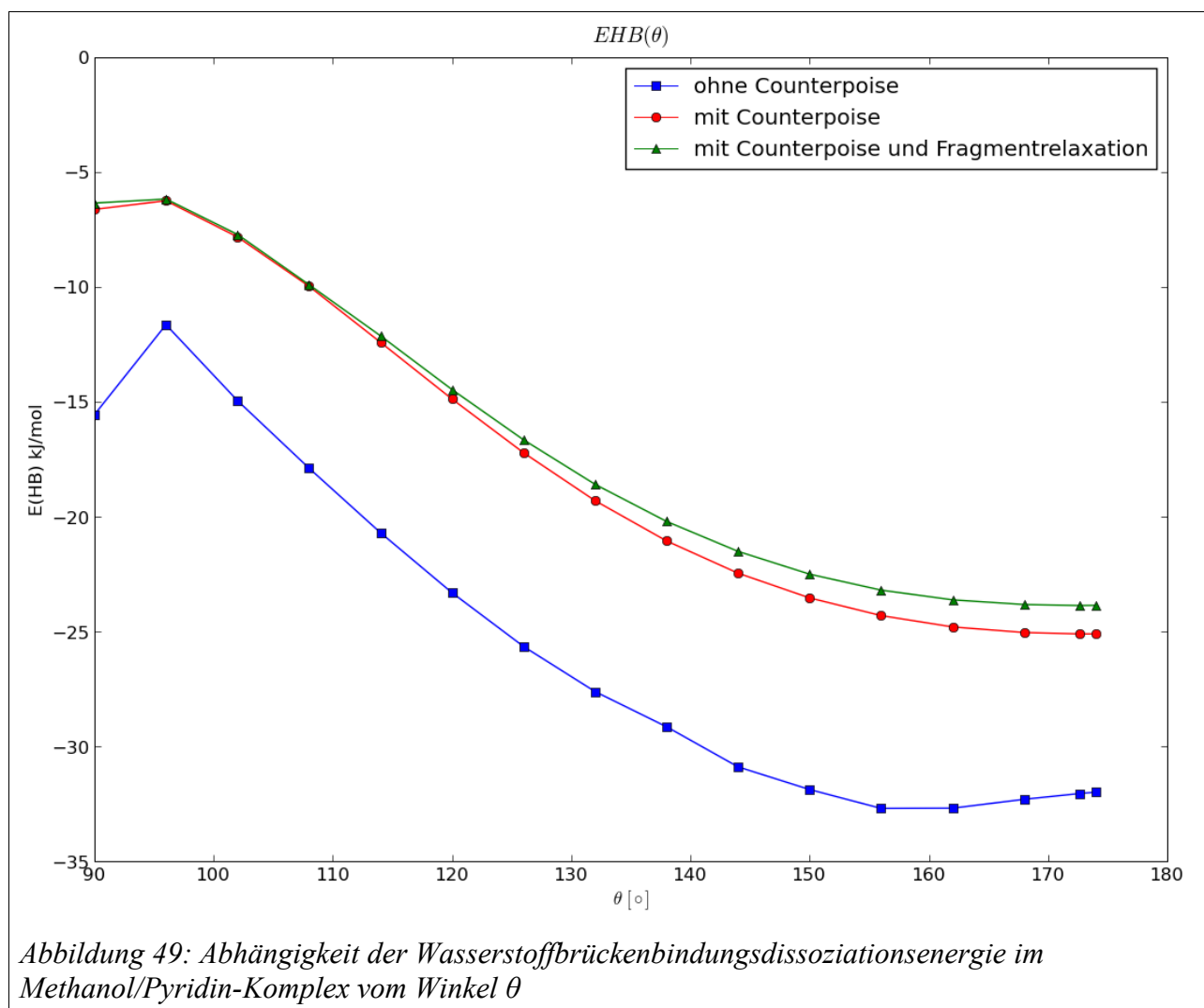
7 Angewandter Teil

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
135.64	148.66	542.44	1.00	164.68	119.88	-18.23
90.02	103.24	340.41	1.10	171.12	122.79	-12.96
53.79	67.12	198.68	1.20	173.25	124.98	4.44
25.71	37.99	102.34	1.30	171.77	126.62	2.80
3.89	15.30	46.97	1.40	171.84	127.71	3.44
-11.63	-1.02	15.48	1.50	171.62	128.60	3.92
-21.81	-12.00	-3.08	1.60	171.34	129.27	4.07
-27.92	-18.89	-13.88	1.70	171.17	129.65	4.09
-31.11	-22.82	-19.87	1.80	171.24	129.67	4.06
-32.23	-24.67	-22.85	1.90	171.76	129.22	4.03
-32.05	-25.10	-23.87	1.99	172.64	128.41	4.01
-31.97	-25.10	-23.92	2.00	172.80	128.26	4.01
-30.84	-24.58	-23.78	2.10	174.15	127.00	4.05
-29.20	-23.46	-22.89	2.20	176.44	124.83	4.39
-27.31	-21.98	-21.57	2.30	179.25	120.80	4.53
-25.28	-20.28	-19.96	2.40	176.80	118.35	4.61
-23.26	-18.53	-18.28	2.50	176.45	123.53	4.50
-21.13	-16.75	-16.55	2.60	178.44	123.73	5.58
-19.36	-15.17	-15.01	2.70	175.41	121.26	4.95
-17.35	-13.51	-13.37	2.80	179.74	124.09	5.08
-15.66	-12.10	-11.98	2.90	180.00	124.25	5.09
-14.24	-10.97	-10.87	3.00	172.61	121.20	5.38
-12.76	-9.83	-9.73	3.10	170.64	122.37	4.20
-11.40	-8.82	-8.73	3.20	169.84	121.33	5.25
-10.15	-7.93	-7.85	3.30	168.06	121.40	4.99
-9.02	-7.14	-7.06	3.40	166.74	121.28	4.95
-7.11	-5.85	-5.77	3.60	164.85	121.30	5.30
-6.31	-5.32	-5.25	3.70	164.05	122.09	5.87
-5.63	-4.86	-4.79	3.80	163.23	121.58	6.20
-5.04	-4.46	-4.39	3.90	162.32	121.56	6.45

Tabelle 20: Abstandsabhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Methanol/Pyridin-Komplex. Der Gleichgewichtsabstand ist rot markiert.

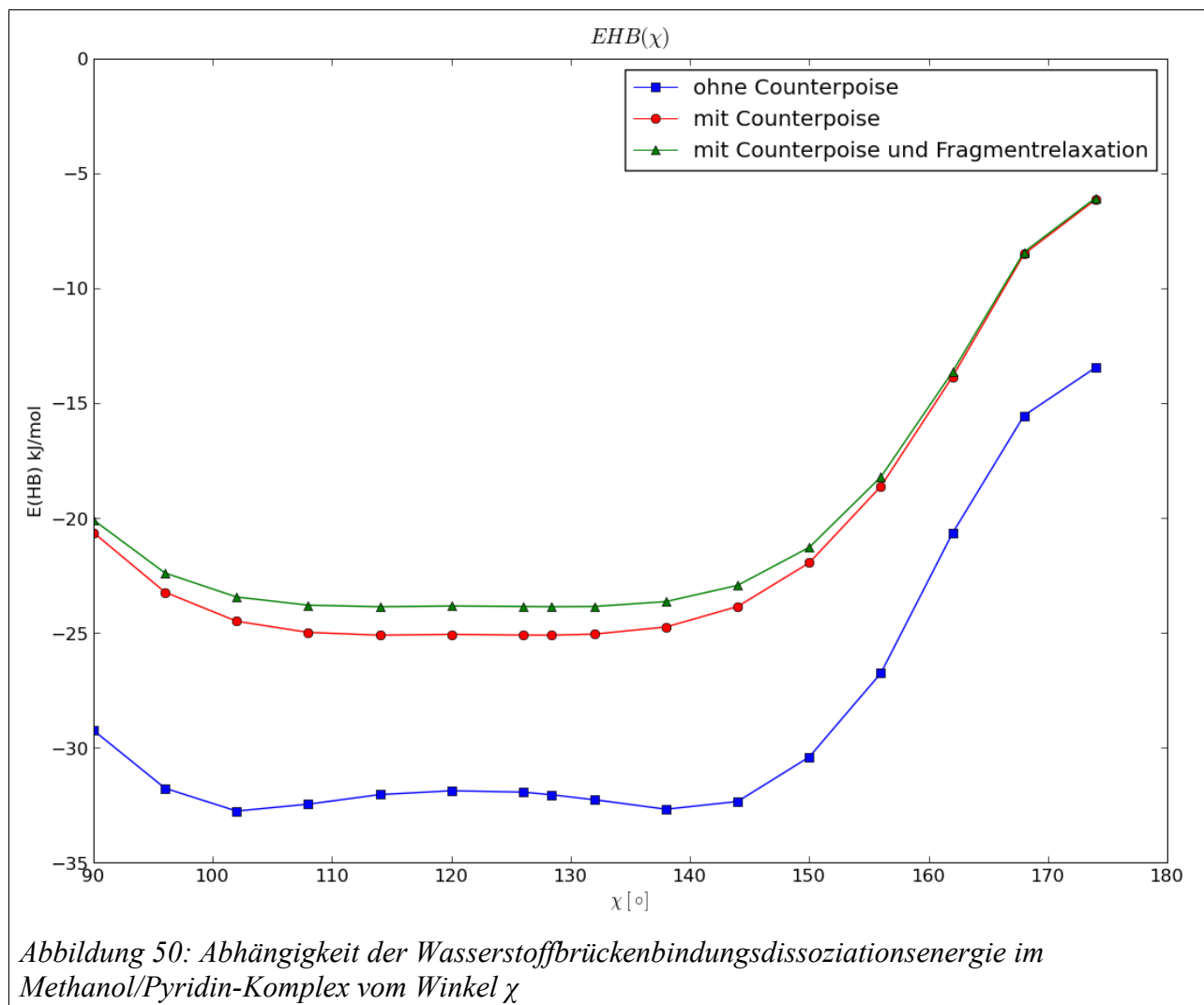
EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-15.56	-6.63	-6.35	3.58	90.00	161.03	47.92
-11.65	-6.25	-6.18	3.33	96.00	159.68	0.08
-14.95	-7.83	-7.72	2.84	102.00	148.84	0.01
-17.89	-9.97	-9.91	2.61	108.00	145.45	-0.01
-20.70	-12.41	-12.13	2.45	114.00	143.79	-0.04
-23.31	-14.88	-14.47	2.33	120.00	143.04	-0.01
-25.65	-17.22	-16.66	2.23	126.00	142.53	-0.01
-27.61	-19.31	-18.59	2.16	132.00	142.08	-0.00
-29.15	-21.06	-20.21	2.10	138.00	141.49	0.82
-30.89	-22.46	-21.51	2.05	144.00	140.54	3.71
-31.88	-23.54	-22.50	2.02	150.00	139.29	4.19
-32.70	-24.31	-23.20	2.00	156.00	137.40	3.92
-32.69	-24.80	-23.62	1.99	162.00	134.93	3.18
-32.30	-25.04	-23.82	1.99	168.00	131.57	3.44
-32.05	-25.10	-23.87	1.99	172.64	128.41	4.01
-31.99	-25.10	-23.86	1.99	174.00	127.43	4.07

Tabelle 21: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Methanol/Pyridin-Komplex vom Winkel θ

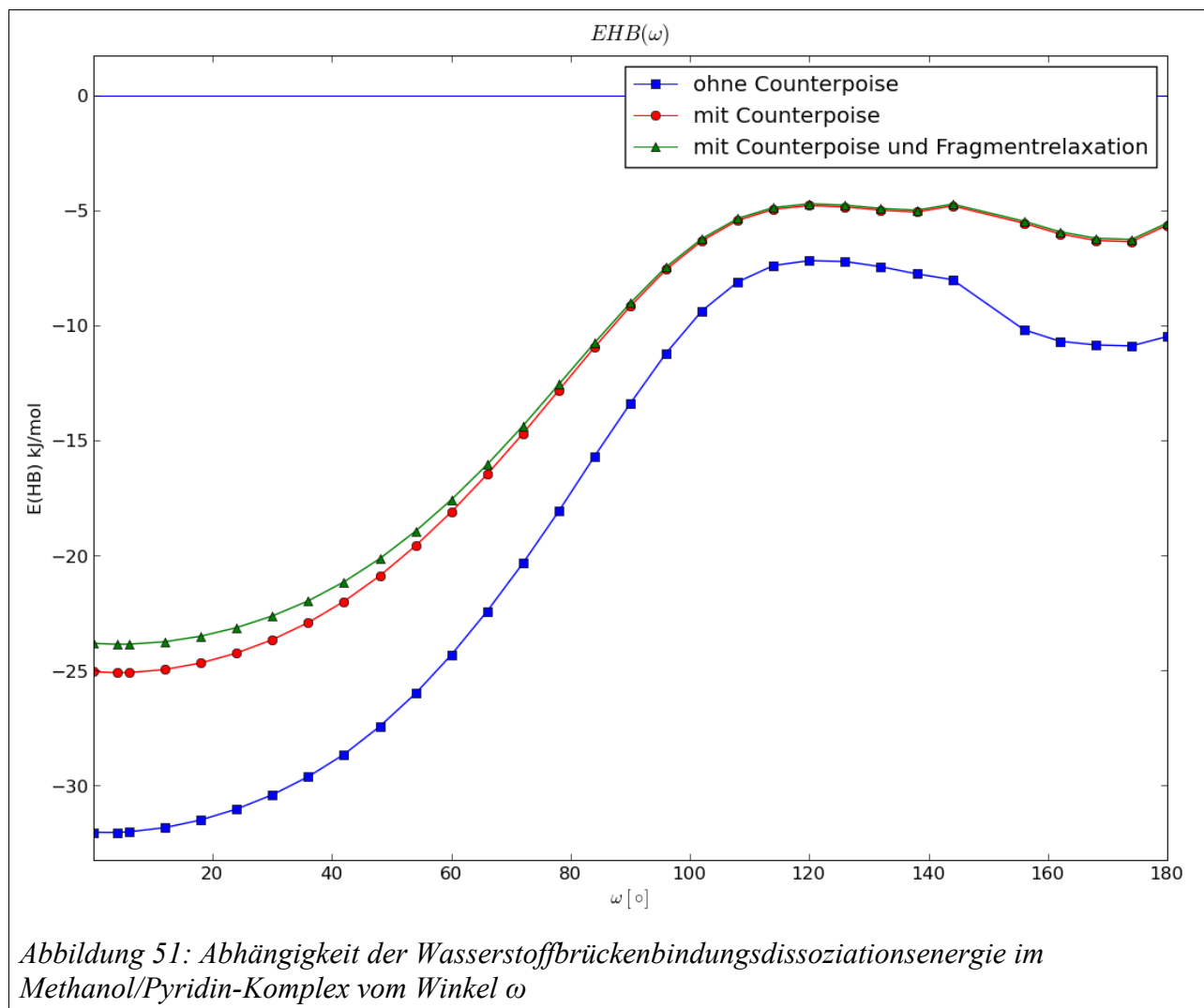


EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-29.24	-20.64	-20.09	2.09	158.21	90.00	4.37
-31.77	-23.23	-22.40	2.02	159.58	96.00	3.76
-32.76	-24.50	-23.44	1.99	163.08	102.00	2.95
-32.46	-24.98	-23.80	1.99	167.67	108.00	2.45
-32.04	-25.10	-23.87	1.99	172.80	114.00	3.42
-31.88	-25.07	-23.83	1.99	179.98	120.00	3.93
-31.94	-25.10	-23.85	1.99	174.70	126.00	4.34
-32.05	-25.10	-23.87	1.99	172.64	128.41	4.01
-32.27	-25.06	-23.85	1.99	169.69	132.00	3.41
-32.68	-24.75	-23.64	1.99	164.82	138.00	2.90
-32.34	-23.86	-22.93	2.01	160.87	144.00	3.22
-30.41	-21.95	-21.28	2.05	158.49	150.00	2.61
-26.76	-18.63	-18.22	2.13	158.07	156.00	1.94
-20.67	-13.86	-13.65	2.27	158.28	162.00	1.16
-15.54	-8.52	-8.43	2.61	153.27	168.00	2.49
-13.46	-6.14	-6.07	3.60	109.99	174.00	10.35

Tabelle 22: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Methanol/Pyridin-Komplex vom Winkel χ



Der typische Potenzialverlauf für die partielle Optimierung bei festgehaltenem Winkel χ ist — wie in Abb. 50 zu sehen — aufgrund der θ/χ – Wechselwirkungen deutlich beeinträchtigt. In der frei optimierten Struktur des Komplexes ist $\chi = 128.4^\circ$.



Die partielle Optimierung mit fixiertem Winkel ω ergibt den in Abb. 51 und Tabelle 23 wiedergegebenen Potenzialverlauf. In der Equilibriumsstruktur beträgt der Torsionswinkel $\omega = 4.0^\circ$. Im Bereich von 0° bis 24° wird die Wasserstoffbrücke bei einer Abweichung von diesem Wert jedoch nur geringfügig um weniger als 0.8 kJ/mol geschwächt. Außerhalb dieses Fensters wächst die Energiedifferenz bis auf 18.3 kJ/mol an. Ab ca. $\omega = 90^\circ$ sind in der Wertetabelle deutliche Geometrieänderungen bezüglich aller anderen Parameter der Wasserstoffbrücke zu identifizieren, die grafische Auftragung stimmt entsprechend nicht mehr mit dem typischen cosinusartigen Potenzialverlauf überein.

7 Angewandter Teil

EHB (DFT) [kJ/mol]	EHB (DFT+CP) [kJ/mol]	EHB (DFT+CP+RELAX) [kJ/mol]	r [Angstroem]	Theta [°]	Chi [°]	Omega [°]
-32.05	-25.07	-23.83	1.99	172.37	128.80	0.00
-32.05	-25.10	-23.87	1.99	172.64	128.41	4.01
-32.02	-25.10	-23.86	1.99	172.71	128.24	6.00
-31.84	-24.97	-23.76	1.99	172.46	128.04	12.00
-31.51	-24.68	-23.52	1.99	171.69	128.16	18.00
-31.03	-24.25	-23.14	1.99	170.57	128.44	24.00
-30.41	-23.67	-22.64	2.00	169.36	128.63	30.00
-29.63	-22.93	-21.98	2.01	168.43	128.47	36.00
-28.66	-22.01	-21.16	2.02	167.68	128.13	42.00
-27.44	-20.89	-20.15	2.04	167.15	127.46	48.00
-26.01	-19.59	-18.95	2.06	166.60	126.70	54.00
-24.33	-18.11	-17.58	2.08	166.14	125.65	60.00
-22.43	-16.48	-16.06	2.12	166.07	124.28	66.00
-20.34	-14.71	-14.37	2.16	166.54	122.49	72.00
-18.08	-12.83	-12.57	2.22	167.50	120.19	78.00
-15.70	-10.95	-10.76	2.29	169.01	116.86	84.00
-13.40	-9.17	-9.03	2.38	171.58	113.53	90.00
-11.21	-7.57	-7.46	2.49	174.82	109.83	96.00
-9.37	-6.32	-6.23	2.62	179.29	106.49	102.00
-8.11	-5.44	-5.35	2.81	175.44	106.72	108.00
-7.40	-4.95	-4.88	3.00	171.27	106.60	114.00
-7.19	-4.79	-4.72	3.17	167.59	104.00	120.00
-7.22	-4.85	-4.77	3.36	162.82	97.82	126.00
-7.45	-4.99	-4.92	3.63	163.17	87.68	132.00
-7.76	-5.07	-4.99	3.88	167.38	83.76	138.00
-8.01	-4.81	-4.73	4.14	175.10	83.14	144.00
-10.19	-5.56	-5.47	5.62	79.27	110.20	156.00
-10.69	-6.02	-5.93	5.87	69.78	109.56	162.00
-10.85	-6.31	-6.21	6.10	59.31	108.60	168.00
-10.89	-6.36	-6.27	6.25	49.21	109.48	174.00
-10.49	-5.64	-5.55	5.88	70.72	118.71	180.00

Tabelle 23: Abhängigkeit der Wasserstoffbrückenbindungsdissoziationsenergie im Methanol/Pyridin-Komplex vom Winkel ω

7.1.5 Parametrisierungsprozedur für Wasserstoffbrücken

Nach Berechnung der Stützwerte erfolgte die Parametrisierung zunächst von Hand, indem die Stützwerte, die Dunitz-Potenzialfunktion und deren Parameter in ein Tabellenkalkulationsprogramm eingegeben und die Parameter schließlich von Hand variiert wurden. Die Abweichung des Potentials von den Stützwerten wurde dabei vom Benutzer beurteilt. Wegen der zu diesem Zeitpunkt noch nicht vorhandenen Software wurden Wechselwirkungen des übrigen Kraftfeldes zunächst nicht betrachtet. Eine erste Version des Parametrisierungsprogrammes las die per Hand zusammengestellten Referenzdaten aus einer Datei und ermittelte Parameter mittels eines evolutionsstrategischen Algorithmus. Auch hier wurden zunächst isoliert vom übrigen Kraftfeld nur Werte in die Dunitz-Funktion eingesetzt und mit der Referenz verglichen. Da die Parameter immer nur für einen Modellkomplex optimiert wurden, gelangte ein klassischer evolutionsstrategischer Algorithmus zum Einsatz. Optimierungsziel war eine möglichst geringe Abweichung von allen Stützwerten, weitere Einschränkungen erfolgten nicht. Mit der Fertigstellung der MOMO-Implementierung für MMTK konnte auch die Parametrisierung weiter entwickelt und die Berechnung der Stützwerte automatisiert werden. Der neue Algorithmus wird im Folgenden genauer beschrieben.

7.1.6 Automatische Generierung der Referenzdaten

Die Stützwerte für die Parametrisierung wurden, wie in den vorhergehenden Kapiteln beschrieben, anhand von möglichst einfachen Modellkomplexen bestimmt, um die Rechenzeiten für die ab-initio-Berechnungen in einem handhabbaren Rahmen zu halten. Zunächst wurde jeder Komplex mit dem MOMO-eigenen Editor mit Standardgeometrien aufgebaut und in das MOMO-XML-Dateiformat übertragen. Zu diesem wurden dann Tags hinzugefügt, die die nachfolgend aufgelisteten Zusatzinformationen beinhalten. Abb. 52 zeigt die Hierarchie der Daten.

```
<ParameterizationReference>
<hbond IA="O1" IAA="C2" IAB="H1A" ID="C1" IH="H1">
</hbond>
  <HBEquilibrium EHB="-12.8059070099" chi="124.744133007"
  omega="-0.0366551941558" r="2.18972656781" theta="167.027313016" />
  <HBEnergy EHB="-12.8059070099" chi="124.744133007" omega="-0.036655194155"
  plot="r" r="2.18972656781" theta="167.027313016" />
</ParameterizationReference>
```

Abbildung 52: XML-Tags zur Aufnahme von Informationen über Wasserstoffbrücken

Ausgehend von der Wurzel wird der Tag `<ParameterisationReference>` eingeführt. Unterhalb von diesem sollen immer Referenzdaten für die Parametrisierung angelegt werden. Für die Wasserstoffbrücken werden folgende Informationen benötigt:

- Angabe der Atomnummern/Namen für Donor, Donorwasserstoff, Akzeptor, Nachbar des Akzeptors und Nachbar des Nachbarn des Akzeptors
- Die Stützwerte für die Parametrisierung
- Angaben zur frei optimierten Struktur des Modellkomplexes

Für die Berechnung der Referenzdaten ist lediglich der erste Punkt entscheidend, für die eigentliche Parametrisierung werden dann auch die Stützwerte und die Equilibriumsstruktur benötigt.

Der Tag `<hbond>` deklariert die an der Wasserstoffbrücke beteiligten (bezogen auf das Dunitz-Potenzial) Atome in den Attributen `„IH“` (Wasserstoffatom), `„ID“` (Wasserstoffbrückendonor), `„IA“` (Wasserstoffbrückenakzeptor), `„IAA“` (Nachbar des Akzeptors) und `„IAB“` (Nachbar des Nachbarn des Akzeptors). Im Tag `<HBEnergy>` signalisieren die Attribute `„r“` [Å], `„theta“` [°], `„chi“` [°] und `„omega“` [°] die Geometrie der Wasserstoffbrücke, und `„EHB“` [kJ/mol] symbolisiert die Wasserstoffbrückenbindungsenergie. Mit dem Attribut `„plot“` wird ein Parameter als bei der partiellen Optimierung festgehalten gekennzeichnet. Der Tag `<HBEquilibrium>` ist vom Aufbau her identisch mit `<HBEnergy>`, markiert jedoch die frei optimierte Komplexstruktur. Der Stützwert wird in einem `<HBEnergy>`-Tag noch einmal wiederholt.

Die Steuerung des Ablaufs zur Berechnung der Stützwerte übernimmt das Programm „GauSlaveDFTScan.py“. Der bei den ab-initio-Rechnungen zu verwendende Basissatz sowie Schrittweite und Wertebereich der zu variierenden Geometrien der Wasserstoffbrücke müssen beim Programmstart übergeben werden. Zunächst erfolgt eine freie Energieminimierung des Komplexes ausgehend von der in der XML-Datei gespeicherten Geometrie, hier mit dem im Editor aufgebauten Komplex. Anschließend werden die zu variierenden Werte eingefroren und eine Geometrieoptimierung der verbleibenden Freiheitsgrade durchgeführt. Die ab-initio-Rechnung erfolgt durch den Aufruf des Gaussian-Paketes mit der erzeugten Eingabedatei. Das Programm kann eine Wertetabelle ausgeben und den Potenzialverlauf grafisch darstellen.

Das Programm „GauSlaveXML.py“ ist eine Variante des Programmes, die ermittelten Stützwerte werden in eine vom Programm angelegte Kopie der XML-Datei des zu berechnenden Komplexes geschrieben.

Die Parametrisierung erfolgt nun durch das Programm „HBOptimizerEQ.py“. Die Berechnung der Parameter erfolgt in zwei Abschnitten. Zunächst wird der abstandsabhängige Term 7.1.3 mithilfe der Methode der kleinsten Fehlerquadrate optimiert.

$$E_{HB}^{Radius} = \left(\frac{a}{r_{HA}^i} - \frac{c}{r_{HA}^j} \right) \quad 7.1.3$$

Ergebnis sind Werte für die Parameter a, c, i und j. Da der Abstand r nicht negativ werden kann, sind die so möglichen nicht-ganzzahligen Ergebnisse für die Exponenten i und j akzeptabel. Schließlich müssen noch Werte für den winkelabhängigen Teil (Formel 7.1.4) des Potentials gefunden werden.

$$E_{HB}^{Gesamt} = E_{HB}^{Radius} \cdot \cos^k(\theta) \cos^m(\chi - \chi_0) \cos^n(\omega - \omega_0) \quad 7.1.4$$

Da hier auch negative Zahlen auftreten können, kommen nur ganzzahlige Exponenten in Frage. Die Methode der kleinsten Fehlerquadrate ist deshalb nicht mehr anwendbar. Die Exponenten k, m und n werden deshalb mittels eines evolutionsstrategischen Algorithmus bestimmt. Um gegenseitige Beeinflussungen zu vermeiden, werden für die Parametrisierung des abstandsabhängigen Teils nur solche Stützwerte verwendet, bei deren Berechnung der Radius fixiert und die übrigen Koordinaten frei optimiert worden sind. In die Optimierung des winkelabhängigen Teils fließen dann alle Stützwerte mit ein.

Es hat sich gezeigt, dass die Energie für unter dem Gleichgewichtsabstand liegende Radien durch die Potenzialfunktion nur schlecht wiedergegeben werden kann, deshalb werden nur solche Stützwerte verwendet, deren Wasserstoff-Akzeptor-Abstand den Gleichgewichtsabstand um maximal 0.2 Å unterschreitet. Hier kann der Kurvenverlauf fast perfekt an die Referenzdaten angepasst werden. Durch die Beschränkung auf ganzzahlige Exponenten ergeben sich für den winkelabhängigen Teil recht grobe Schritte bei der Anpassung der Funktion. Zusätzlich ist eine Begrenzung der für die Stützwerte akzeptierten Winkel in Abhängigkeit zum Optimum erforderlich, um eine Fehlleitung der Optimierung durch unnatürliche Verformungen des Komplexes aufgrund der erzwungenen Koordinaten bei der partiellen Optimierung zu vermeiden.

7.1.7 Die Parametrisierung des Wasserstoffbrückenpotenzials

Nicht alle berechneten Stützwerte sollten auch mit in die Parametrisierung einbezogen werden. Viele der berechneten Werte müssen aufgrund unnatürlicher Verformungen des Komplexes als „unphysikalisch“ angesehen werden und eignen sich daher nicht zur Beschreibung eines Wasserstoffbrückenbindungspotenzials. Das Parametrisierungsprogramm „HBOptimizerEQ.py“ ermöglicht aus diesem Grund die Angabe der erlaubten Abweichung der Energie und der Parameter der Dunitz-Funktion in Bezug auf die Gleichgewichtsstruktur. Gegenüber der Einbeziehung aller Stützwerte in die Parametrisierung lässt sich so eine deutliche Verbesserung erzielen. Geeignete Werte liegen zumindest bei den hier behandelten Wasserstoffbrücken immer recht nahe an der frei optimierten Struktur.

Aufgrund der beschriebenen Auswahl der in die Parametrisierung zu integrierenden Werte erscheint die Angabe statistischer Kennzahlen für die Qualität der Anpassung nicht angebracht. Zu weit vom Minimum entfernte Stützwerte beschreiben die eigentliche Wasserstoffbrücke nicht und sollten deshalb auch nicht in die Bewertung einfließen, auf der anderen Seite liegt das Ziehen dieser Grenze im Ermessen des Anwenders und erschwert so die Vergleichbarkeit der Kennzahlen bei verschiedenen Parametrisierungen. Zur Beurteilung des Parametersatzes dienen deshalb Korrelationsdiagramme, in denen die bei der Optimierung berücksichtigten Stützwerte hervorgehoben werden. Im Folgenden werden die Ergebnisse der Parametrisierungen mit den in den vorangegangenen Kapiteln ermittelten Stützwerten aufgeführt.

Modellkomplex Chloroform/Formaldehyd

Eingabe

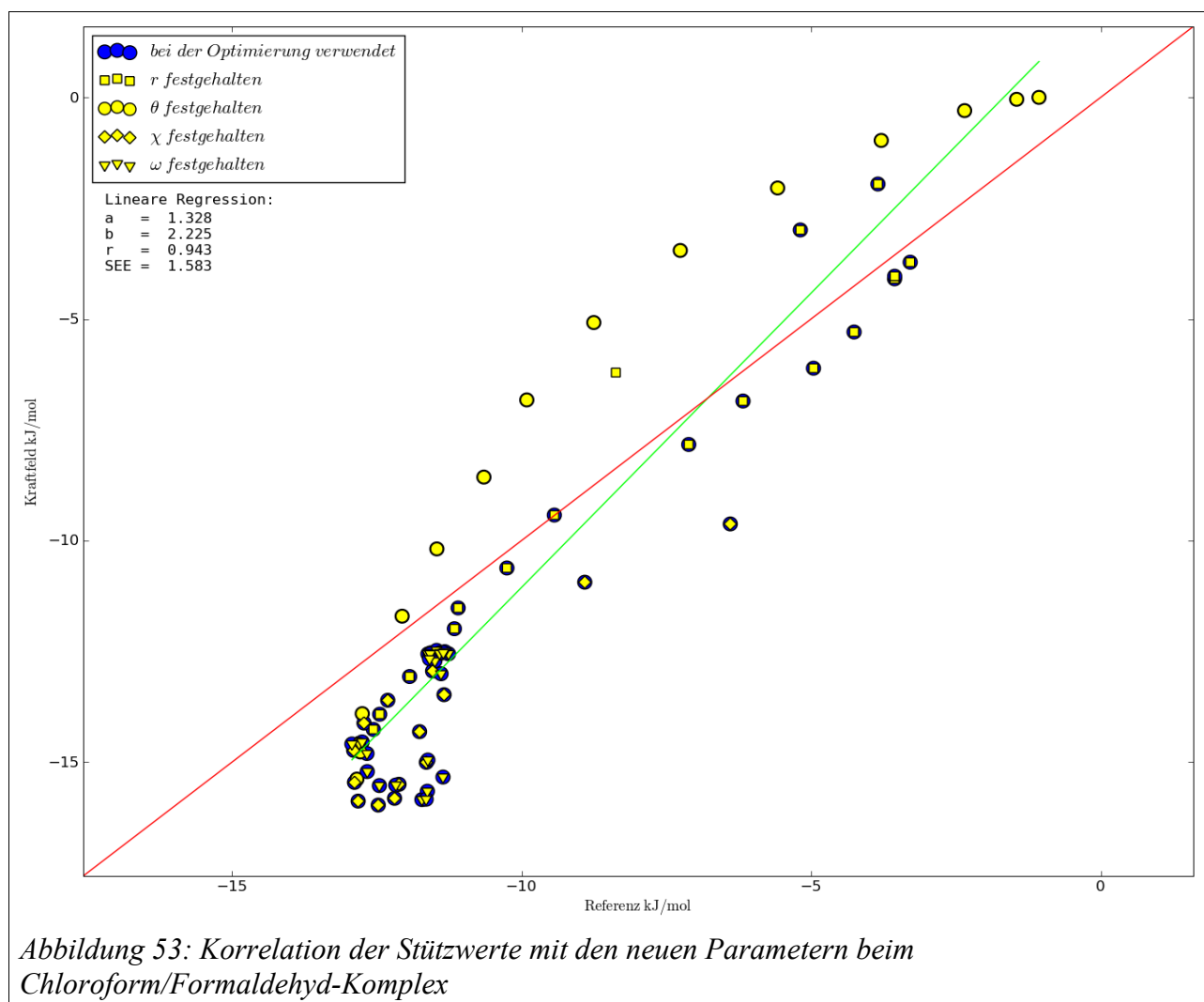
Erlaubte Abweichungen der verwendeten Stützwerte von der Gleichgewichtsstruktur:

$$\Delta E = 100 \text{ kJ/mol} ; \Delta r = \text{unbegrenzt} ; \Delta \theta = \text{unbegrenzt} ; \Delta \chi = \text{unbegrenzt} ; \\ \Delta \omega = \text{unbegrenzt}$$

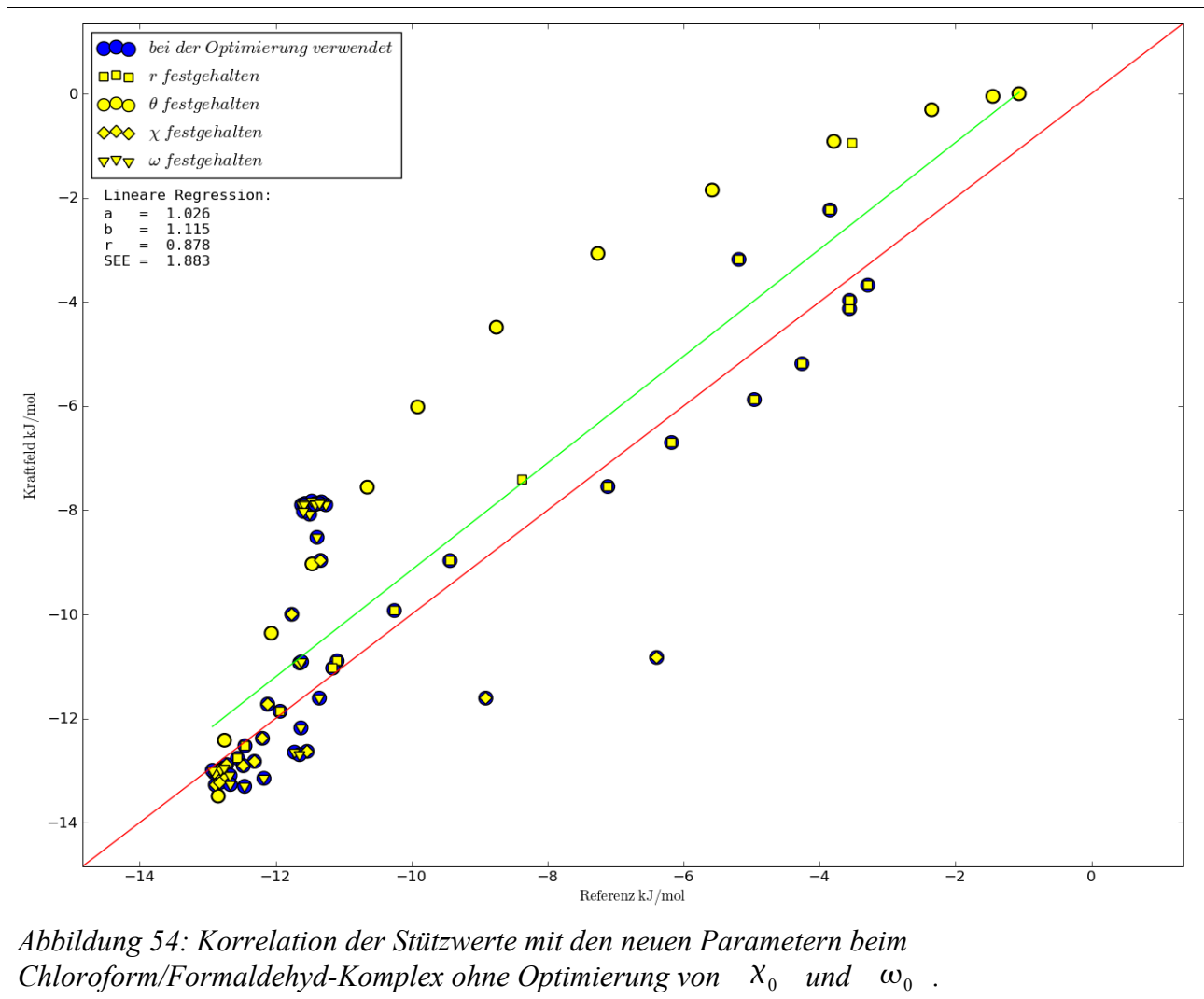
Ergebnisse

Parameter für das Dunitz-Potential nach 10 000 Generationen:

$$A = 115158.6 ; C = 112697.37 ; i = 4.88 ; j = 4.84 ; k = 2 ; m = 1 ; n = 0 ; \\ \chi_0 = 141.0^\circ ; \omega_0 = 0^\circ$$



Die Grenzen für die Selektierung der Stützwerte wurden durch manuelles Probieren und Beurteilen des Korrelationsdiagramms (Abb. 53) ermittelt. Die Herkunft der Stützwerte ist im Diagramm durch die Form der Marker gekennzeichnet und bezieht sich auf die bei der partiellen Optimierung der Komplexstruktur festgehaltenen Koordinaten bezüglich des Dunitz-Potenzials. Es fällt auf, dass sich die Parameter für χ_0 und ω_0 deutlich von den betreffenden Winkeln der Gleichgewichtsstruktur unterscheiden. Da der Optimierungsalgorithmus sich nur an der Übereinstimmung der mit dem Kraftfeld berechneten Energie mit den gegebenen Stützwerten orientiert, ist dies jedoch zu erklären. Die Bandbreite der Energien bei festgehaltenen Winkeln χ und ω ist nach Tabelle 10 und 11 geringer als in vielen Fällen die Abweichung der mit dem Kraftfeld berechneten Energie von der Referenz. Die Optimierung erreicht ihr Ziel daher bei freier Wahl dieser Parameter eher. Für das Kraftfeld ist die Lage des Energieminimums wichtiger als die



genauen Energiewerte, eine freie Optimierung dieser beiden Parameter sollte deshalb unterbleiben und stattdessen die Zahlen der Equilibriumsstruktur übernommen werden.

Die Parametrisierung wurde wiederholt, ohne χ_0 und ω_0 mit zu optimieren. So sind bedeutend weniger Zyklen des evolutionstrategischen Algorithmus erforderlich, da der Wertebereich der Exponenten k , m und n sehr klein ist. Dabei ergab sich Folgendes:

Parameter für das Dunitz-Potenzial nach 1000 Generationen:

$$A=56625.78 \ ; \ C=55501.73 \ ; \ i=4.29 \ ; \ j=4.26 \ ; \ k=2 \ ; \ m=1 \ ; \ n=0 \ ;$$

$$\chi_0=\chi_{eq}=124.74^\circ \ ; \ \omega_0=\omega_{eq}=0.04^\circ$$

Die Eingabe entsprach dem 1. Versuch. Die sich nun ergebende Korrelation ist in Abb. 54 gezeigt.

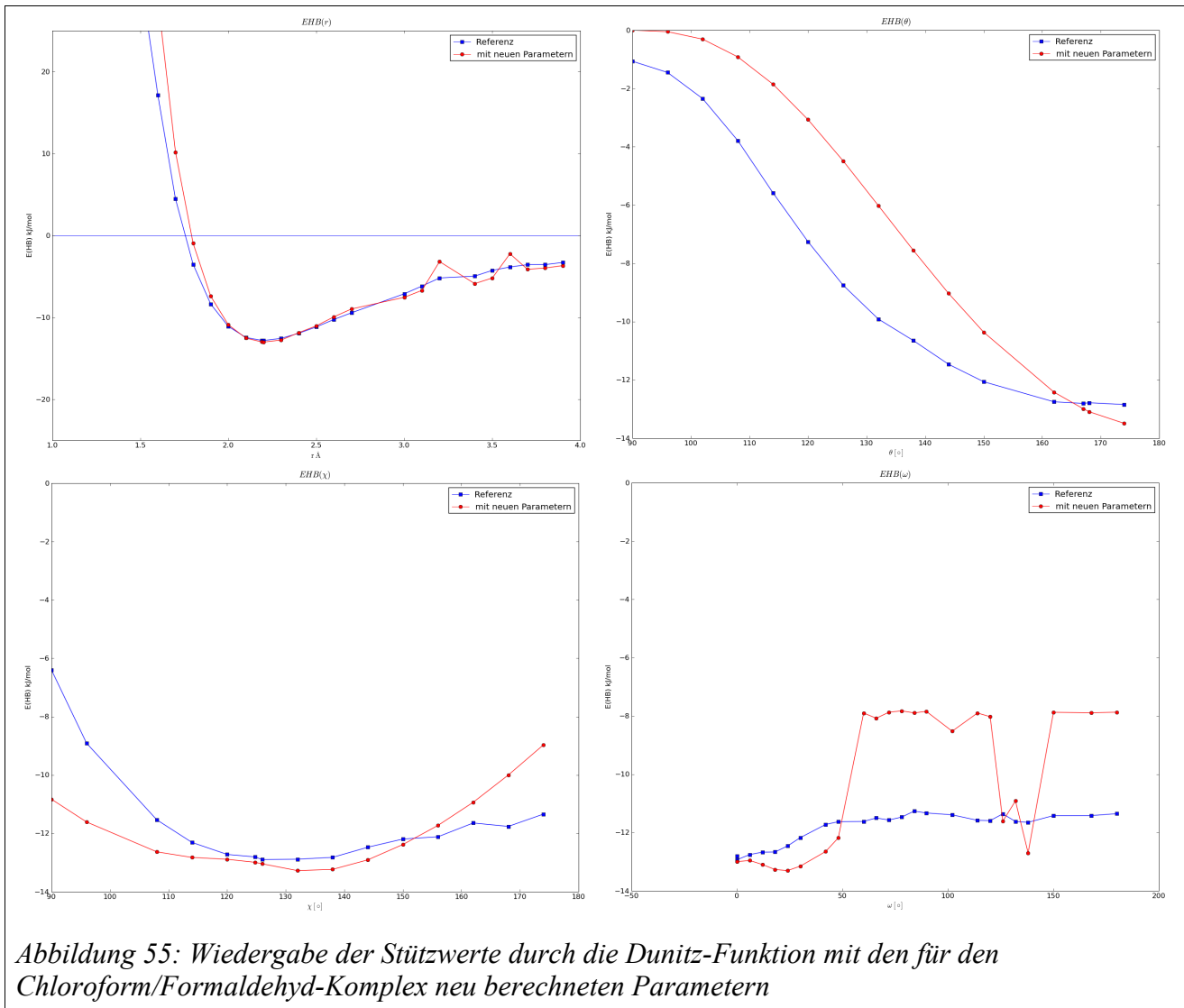


Abbildung 55: Wiedergabe der Stützpunkte durch die Dunitz-Funktion mit den für den Chloroform/Formaldehyd-Komplex neu berechneten Parametern

In Abb. 55 ist zu sehen, dass die Anpassung des abstandsabhängigen Terms gut gelingt, die winkelabhängigen Teile jedoch Probleme bereiten. Der Einfluss letzterer Terme macht sich in Form von Ausreißern im abstandsabhängigen Potenzial bemerkbar. Der Exponent für den ω -abhängigen Term ist auf „null“ gesetzt, was einer Abschaltung entspricht. Die hier berechneten Werte ergeben sich ausschließlich aus den Einflüssen der übrigen Terme (Graph unten rechts). Diese „Abschaltung“ erscheint gerechtfertigt, da der Torsionswinkel ω die Dissoziationsenergie des Komplexes offenbar kaum beeinflusst.

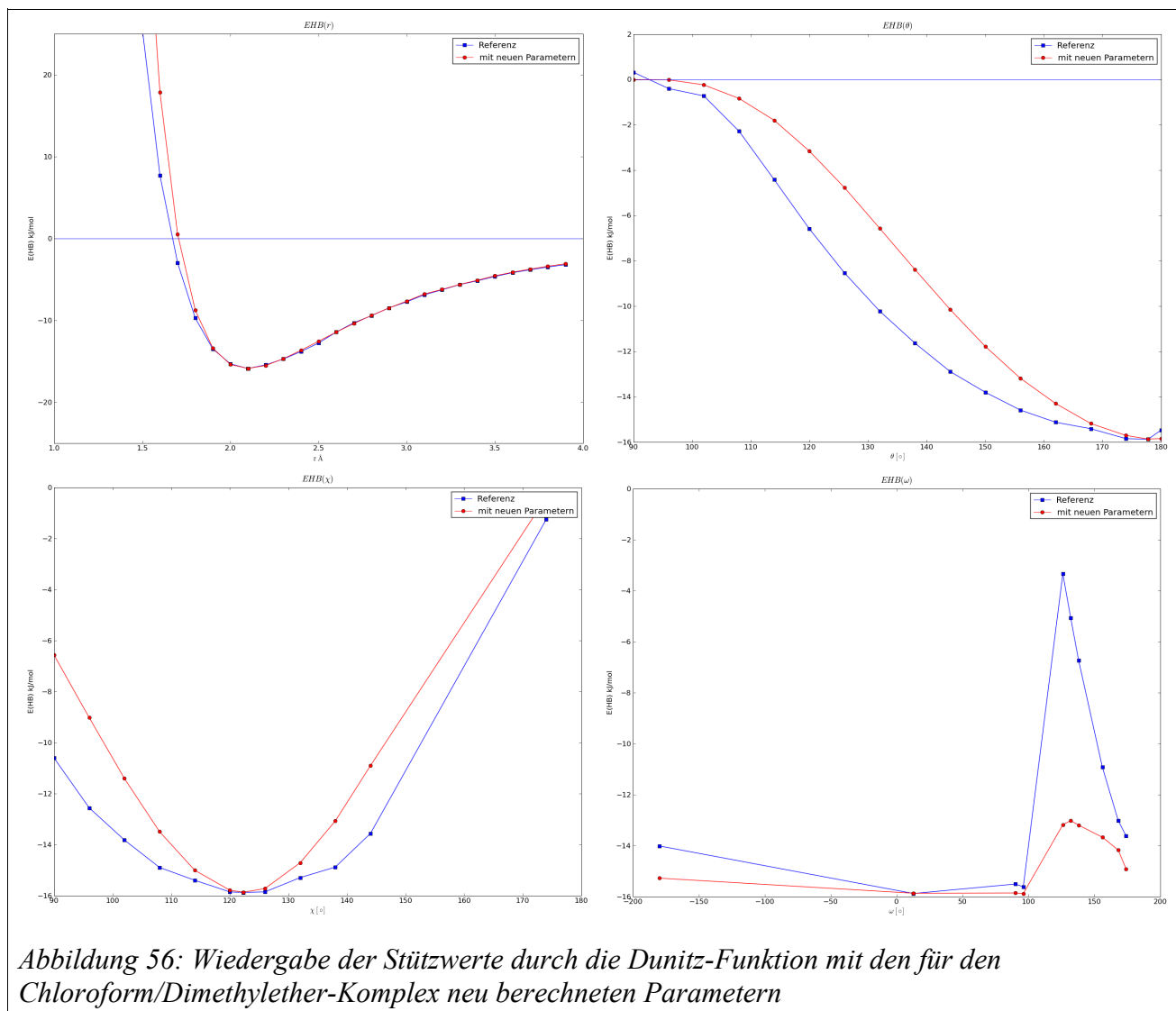
Modellkomplex Chloroform/Dimethylether

Eingabe

Erlaubte Abweichungen der verwendeten Stützwerte von der Gleichgewichtsstruktur:

$$\Delta E = 100 \text{ kJ/mol} ; \Delta r = \text{unbegrenzt} ; \Delta \theta = \text{unbegrenzt} ; \Delta \chi = \text{unbegrenzt} ;$$

$$\Delta \omega = \text{unbegrenzt}$$

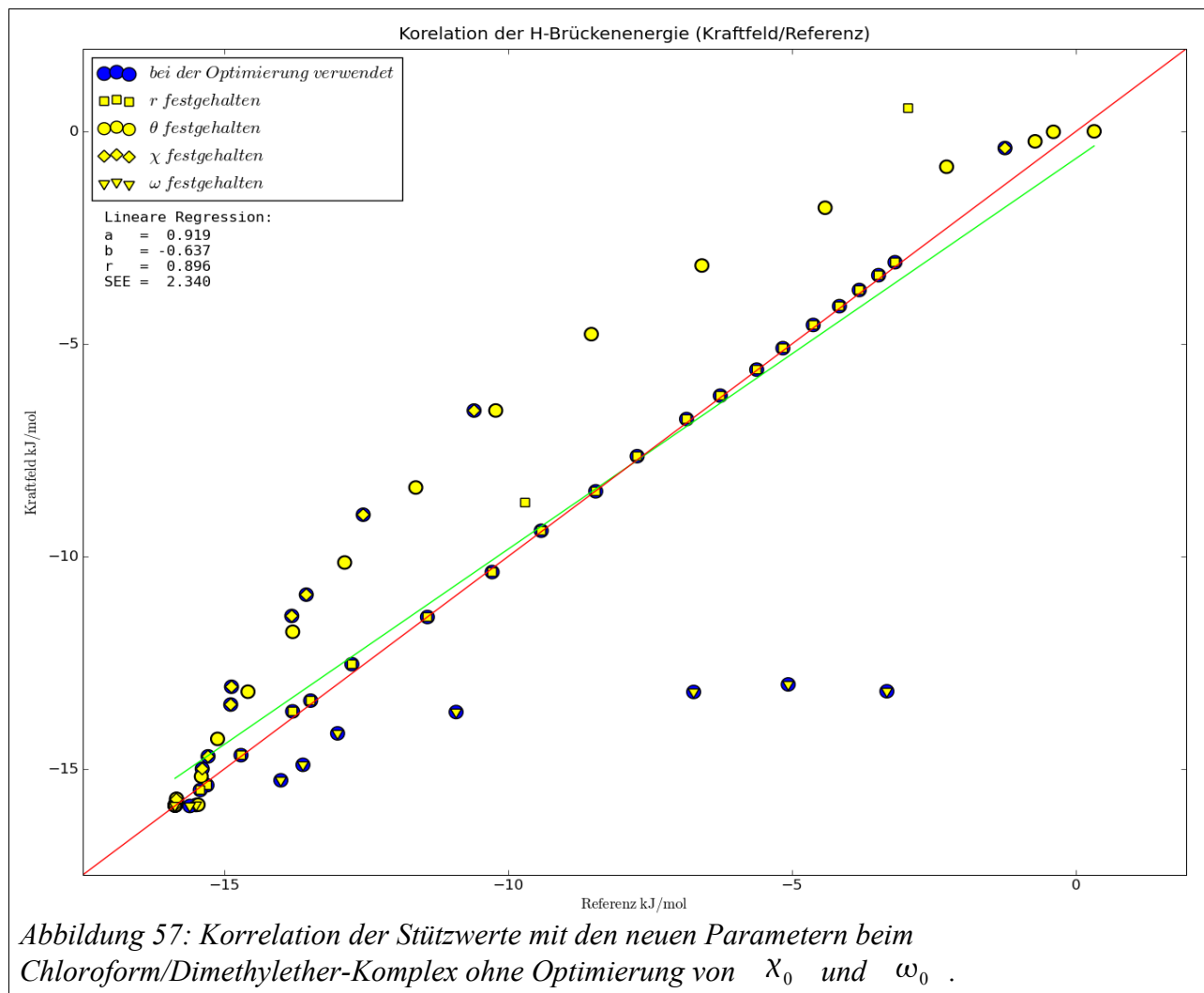


Ergebnisse

Parameter für das Dunitz-Potenzial nach 1000 Generationen:

$$A=90241.06 ; C=88778.69 ; i=4.85 ; j=4.82 ; k=2 ; m=5 ; n=0 ;$$

$$\chi_0=\chi_{eq}=122.33^\circ ; \omega_0=\omega_{eq}=12.72^\circ$$



Die Korrelation der mit den neuen Parametern berechneten Energien zu den Referenzdaten kann in den Abb. 56 und 57 abgelesen werden. Auch hier kann der abstandsabhängige Term besser wiedergegeben werden, als die winkelabhängigen Terme. Der ω -abhängige Term wurde vom Optimierungsalgorithmus aufgrund seines geringen Einflusses deaktiviert.

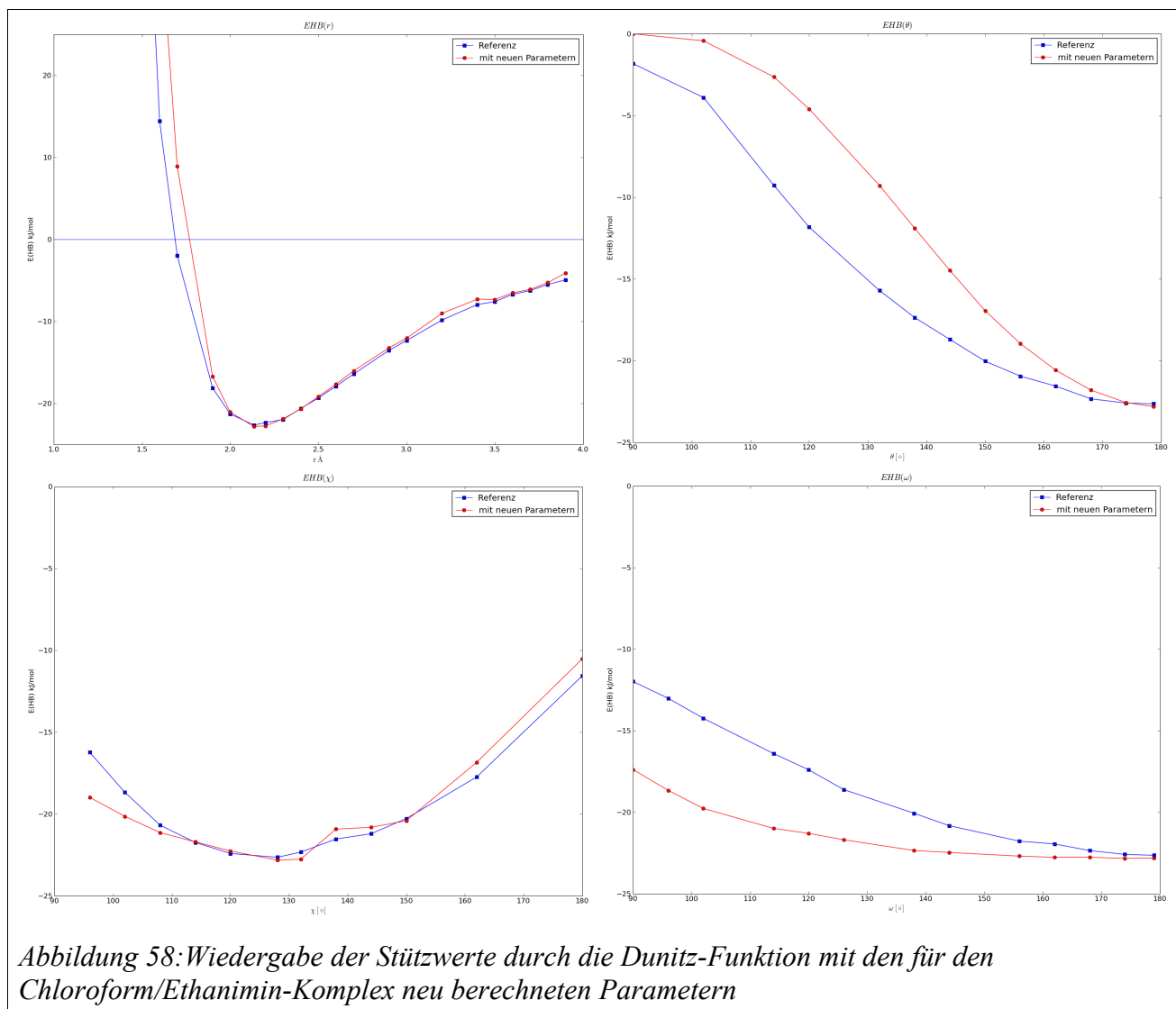
Modellkomplex Chloroform/Ethanamin

Eingabe

Erlaubte Abweichungen der verwendeten Stützwerte von der Gleichgewichtsstruktur:

$$\Delta E = 100 \text{ kJ/mol} ; \Delta r = \text{unbegrenzt} ; \Delta \theta = \text{unbegrenzt} ; \Delta \chi = \text{unbegrenzt} ;$$

$$\Delta \omega = \text{unbegrenzt}$$

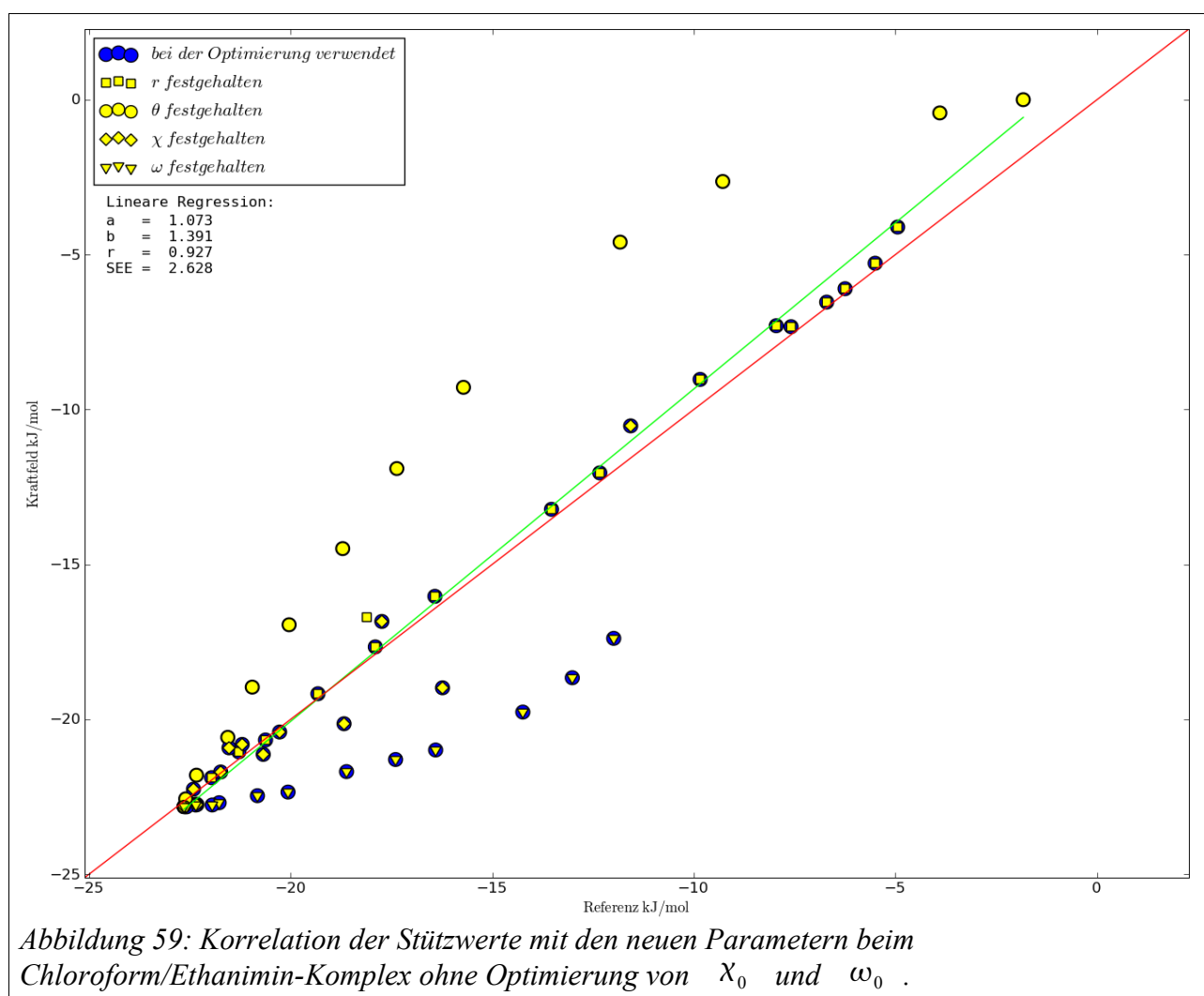


Ergebnisse

Parameter für das Dunitz-Potenzial nach 10 000 Generationen:

$$A=179563.7 ; C=177219.82 ; i=4.78 ; j=4.76 ; k=2 ; m=1 ; n=0 ; \\ \chi_0=\chi_{eq}=128.02^\circ ; \omega_0=\omega_{eq}=178.96^\circ$$

Die Korrelation der mit den neuen Parametern berechneten Energien zu den Referenzdaten kann in den Abb. 58 und 59 entnommen werden. Das Parametrierungsprogramm kommt mit dem Modellkomplex gut zurecht, wie man an der vom Energieminimum ausgehenden recht lange und dicht auf der Diagonalen liegenden Punktwolke erkennen kann. Der ω -abhängige Term wurde vom Optimierungsalgorithmus aufgrund seines geringen Einflusses deaktiviert.



Modellkomplex Methanol/Pyridin

Eingabe

Erlaubte Abweichungen der verwendeten Stützwerte von der Gleichgewichtsstruktur:

$$\Delta E = 100 \text{ kJ/mol} ; \Delta r = \text{unbegrenzt} ; \Delta \theta = \text{unbegrenzt} ; \Delta \chi = \text{unbegrenzt} ; \\ \Delta \omega = \text{unbegrenzt}$$

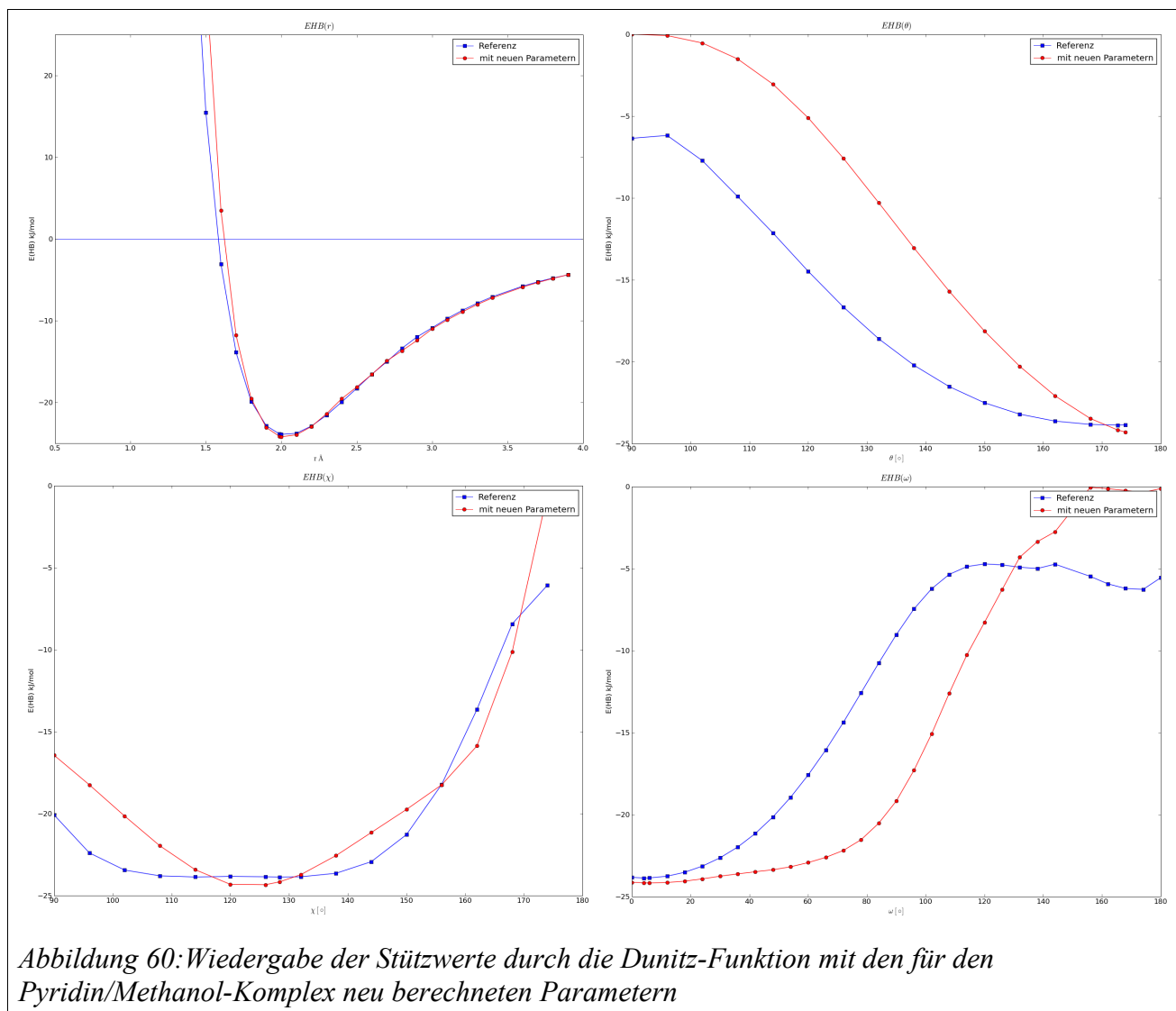


Abbildung 60: Wiedergabe der Stützwerte durch die Dunitz-Funktion mit den für den Pyridin/Methanol-Komplex neu berechneten Parametern

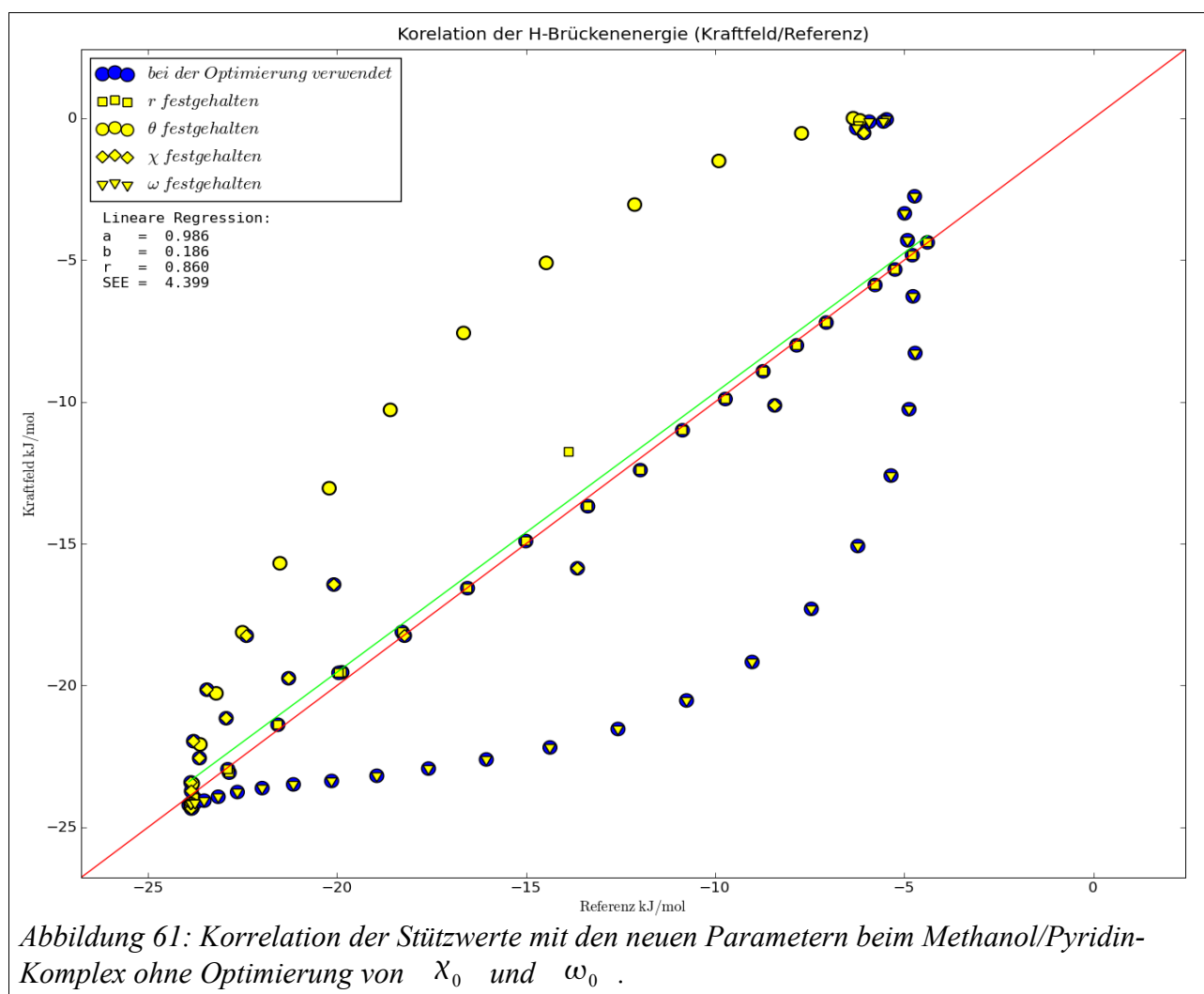
Ergebnisse

Parameter für das Dunitz-Potenzial nach 1000 Generationen:

$$A=107330.29 ; C=106014.02 ; i=4.57 ; j=4.55 ; k=2 ; m=1 ; n=0 ;$$

$$\chi_0=\chi_{eq}=128.41^\circ ; \omega_0=\omega_{eq}=4.01^\circ$$

Die Korrelation der mit den neuen Parametern berechneten Energien zu den Referenzdaten kann in den Abb. 60 und 61 abgelesen werden. Das Parametrierungsprogramm kommt mit dem Modellkomplex gut zurecht, wie man an der vom Energieminimum ausgehenden recht lange und dicht auf der Diagonalen liegenden Punktwolke erkennen kann. Der ω -abhängige Term wurde vom Optimierungsalgorithmus aufgrund seines geringen Einflusses deaktiviert. Die bei der Berechnung der Stützwerte zu diesem Modellkomplex aufgefallene starke Wechselwirkung zwischen den Winkeln θ und χ wird durch die optimierten Parameter offenbar gut wiedergegeben.



7.2 Parametrisierung von atomaren Punktladungsmodellen

7.2.1 Atomare Punktladungsmodelle

Die Berechnung von Partialladungen und den daraus resultierenden Coulomb-Kräften nimmt eine zentrale Rolle im MOMO-Kraftfeld ein. Alle anderen Kraftfeldterme offenbaren mehr oder weniger starke Abhängigkeiten von den berechneten Partialladungen oder den in der π -SCF-Rechnung ermittelten Bindungsordnungen, die ebenfalls wieder im Zusammenhang mit den Partialladungen stehen. In meiner Diplomarbeit [GW2004] habe ich die in MOMO implementierten Ladungsmodelle einem Test unterzogen und die Ergebnisse mit verschiedenen quantenchemischen Methoden zur Ladungsberechnung verglichen.

Da Partialladungen keine physikalischen Eigenschaften, sondern ihrerseits wieder Modelle für die Ladungsdichteverteilung bzw. deren Wirkung repräsentieren, sind je nach Anwendungsfall verschiedene Ladungsmodelle geeignet.

In meiner Diplomarbeit habe ich die berechneten Partialladungen u. a. anhand von chemischen Gesichtspunkten wie der Wiedergabe des induktiven Effektes und der systematischen Entwicklung der Partialladungen bei verschiedenen substituierten Molekülen im Hinblick auf die Elektronegativitäten bewertet. Unter diesen Gesichtspunkten hatte das Ladungsmodell von Merz und Kollman besonders gut abgeschnitten. Für die Parametrierung sollte zunächst der bewährte, schon in der Diplomarbeit verwendete Testdatensatz von Sabine Monz [Monz1997] eingesetzt werden. Neben den reinen Zahlenwerten für die Partialladungen stehen hier auch Betrachtungen zur systematischen Entwicklung der Ladungen unter chemischen Gesichtspunkten zur Verfügung, die zur Parametrisierung herangezogen werden sollen.

Zur Bestimmung von Partialladungen aus der Molekülwellenfunktion existieren verschiedene Verfahren, die sich in drei Klassen einteilen lassen. Alle müssen die Elektronendichte, die sich als Quadrat der Molekülwellenfunktion ergibt, auf die Atome eines Moleküls aufteilen und dann aus dem Elektronendichteanteil eines Atoms dessen elektrostatische Eigenschaften ableiten. Populationsanalysen analysieren die Besetzung der Atomorbitale, die den Basissatz der Molekülorbitale bilden (LCAO). An der Topologie der Elektronendichte orientierte Methoden spalten die Ladungsdichte in einem Molekül in den einzelnen Atomen zugeordnete Subräume auf und integrieren innerhalb dieser Gebiete über die Ladungsdichte. ESP-Methoden passen Partialladungen so an, dass sie das elektrostatische Potenzial eines Moleküls möglichst gut wiedergeben.

Thomas Kleinöder hat in seiner Dissertation [Kleinoeder2005] ein dem Gasteiger-Marsili-Verfahren verwandtes Partialladungsmodell entwickelt und parametrisiert. Dazu hat er verschiedene auf der Molekülwellenfunktion basierende Modelle als mögliche Quelle von Referenzdaten evaluiert. Neben den auch in meiner Diplomarbeit eingesetzten Verfahren gehört dazu auch die „natural bond orbital“-Methode, die von ihm favorisiert wird. Die Beschreibung der Verfahren aus beiden Arbeiten wird im Folgenden kurz zusammengefasst, Details finden sich in der zitierten Literatur.

Populationsanalysen

Die einfachste Form der Populationsanalyse ist die Coulson-Populationsanalyse [Coulson1947]. Voraussetzung ist hier die Orthogonalität aller Basisorbitale im Molekül. Die Ladungsdichteverteilung wird durch Aufsummieren der Besetzungen der Atomorbitale, die zu den Molekülorbitalen beitragen, erhalten. Da orthogonale Atomorbitale in der Praxis kaum anzutreffen sind, muss die Überlappungspopulation zwischen den Bindungspartnern aufgeteilt werden. Bei der Mulliken-Populationsanalyse erfolgt diese Aufteilung gleichmäßig, ungeachtet der Unterschiede in der Elektronegativität und der Art des betrachteten Atomorbitals [Mulliken1955]. Die Probleme der Mulliken-Populationsanalyse sollen durch die „natürliche“ Populationsanalyse [Reed1985] überwunden werden. Hier wird die Elektronendichtematrix in Unterblöcke zerlegt, von denen jeder einem Atom zugeordnet ist. Diese lokalisierten atomaren Basissätze werden „natürliche“ Atomorbitale (NAOs) genannt und sind orthogonal zueinander, die Überlappungspopulation muss nicht mehr aufgeteilt werden. NAOs sind spezifisch für die Wellenfunktion, weniger abhängig vom Basissatz und konvergieren mit sich verbessernder Wellenfunktion. Die Ergebnisse verbessern sich also mit der Qualität der Wellenfunktion, was bei der Mulliken-Populationsanalyse nicht der Fall ist. Der Transformation zu den NAOs folgen noch weitere Schritte, die zu „natürlichen Hybridorbitalen“ und „natürlichen Bindungsorbitalen“ führen. Die Methoden sind unter dem Begriff „natural bond orbital“-Methoden bekannt und wurden entwickelt, um die in der Molekülwellenfunktion enthaltenen Informationen in klassische Konzepte wie Hybridisierung oder Lewis-Formeln übertragen zu können.

Auswertung der Elektronendichtetopologie

Die Methoden dieser Klasse definieren eine räumliche Besetzung mit Atomen innerhalb der Elektronendichte eines Moleküls. Dazu müssen die Grenzen der einem Atom zugeordneten Elektronendichte gefunden werden. Die Ladung kann dann durch Integration bestimmt werden.

Besonders bekannt ist das „atoms in molecules“-Verfahren von Bader [Bader1991]. Hierbei werden zunächst bindungskritische Punkte bestimmt. Ein bindungskritischer Punkt ist jener mit der geringsten Elektronendichte zwischen den Bindungspfaden. Eine Bindung ist der Pfad mit der größten Elektronendichte zwischen zwei Atomen. Bindungskritische Punkte besitzen also ein Minimum in einer Richtung und zwei Maxima in der anderen Richtung. Folgt man der Richtung der größten Abnahme der Elektronendichte in Richtung der Maxima, erhält man sogenannte zero-flux Flächen, die atomare „Bassins“ begrenzen.

Hirshfeld schlug ein anderes Schema zur Definition der Bassins vor [Hirshfeld1977]: Die Partialladung ist hier die Differenz des Elektronendichteintegrals eines Atoms mit jenem über ein in das zu untersuchende Molekül an die betreffende Atomposition projizierten Referenzatoms. Hirshfeld-Ladungen sind auch unter dem Namen „Stockholder-Ladungen“ bekannt.

Fit des molekularen elektrostatischen Potentials

Das elektrostatische Potential eines Moleküls kann direkt als Observable aus der Molekülwellenfunktion bestimmt werden. Für den Fit des Potentials wird das Molekül in einem Punktgitter platziert und das Potential an diesen Punkten berechnet. Die atomaren Punktladungen werden nun dahingehend optimiert, bis an den Messpunkten ein ähnliches Potential auftritt. Die Varianten dieser Methoden unterscheiden sich in der Wahl des Punktgitters und der Optimierungsmethode. Die bekanntesten Methoden sind CHELP [CHELP1987], CHELPG [CHELPG1989] und das Merz-Kollman-Verfahren [Kollman1984]. Bei CHELP werden die Gitterpunkte auf kugelförmigen Schalen um die Atome herum positioniert. CHELPG ist eine Weiterentwicklung, die die Abhängigkeit von der Molekülorientierung bei CHELP vermeiden soll. Beim Merz-Kollman-Verfahren befinden sich die Gitterpunkte auf verschachtelten Connolly-Oberflächen (mit einer definierten Kugel abrollbare Oberfläche) des Moleküls.

7.2.2 Anforderungen an Ladungsmodelle als Referenz

Verschiedene Anwendungsgebiete richten unterschiedliche Anforderungen an die Ladungsmodelle. Sollen die Ladungen z. B. für Moleküldynamik-Rechnungen eingesetzt werden, so müssen die elektrostatischen Eigenschaften für alle dabei möglichen Koordinaten im Phasenraum der Berechnung beschrieben werden. Bei Berechnungen mit fixierten Molekülen reicht eine Beschreibung dieses Zustandes aus. Offensichtliche Qualitätskriterien für quantenchemisch berechnete Ladungen sind deren Unabhängigkeit vom Basissatz und der Orientierung des Moleküls. Daneben sollten die Ladungen möglichst unempfindlich gegenüber der internen Geometrie sein und

Eigenschaften wie das elektrostatische Potenzial, Dipolmoment und höhere Multipolmomente richtig wiedergeben. Auch chemische Trends, wie der induktive Effekt oder die Einflüsse von Elektronegativitäten, sollten wiedergegeben werden. Dazu kommt die Fähigkeit, experimentelle Resultate wie die chemische Verschiebung in der NMR-Spektroskopie, zu beschreiben. Die verschiedenen Ansätze der Ladungsmodelle machen es schwer, alle Kriterien zugleich zu erfüllen. So beschreiben potenzialbasierte Methoden die elektrostatischen Eigenschaften eines Moleküls als Ganzes am besten, während die Populationsanalysen bei den auf Elektronegativitätsunterschieden und der Topologie beruhenden Effekten einen prinzipiellen Vorteil haben.

Bei der Parametrisierung wird letztendlich ein empirisches Ladungsmodell an ein auf der Molekülwellenfunktion basierendes angepasst. Am erfolgreichsten wird dies sein, wenn die Ansätze der Modelle miteinander kompatibel sind. Ein auf Elektronegativitäten und der Topologie basierendes Modell, wie jenes von Gasteiger und Marsili, sollte am besten mit einer zu den Populationsanalysen zählenden Methode korrelieren. Polarisierbare Ladungsmodelle, wie das von Sabine Monz, sind wahrscheinlich eher in der Lage, das elektrostatische Potenzial eines Moleküls zu reproduzieren.

Für die Eigenschaften des Referenzmodells lassen sich so folgende Forderungen aufstellen:

- Die Bindungspolaritäten sollten wiedergegeben werden.
- Ladungen sollten möglichst nur von der Molekültopologie abhängen.
- Ladungen sollten für den Chemiker sinnvolle Werte haben.

Die auf der Elektronendichtetopologie basierenden Verfahren disqualifizieren sich leider durch das Fehlen einer verfügbaren und ausreichend robusten Implementierung. Am interessantesten erscheinen daher das Merz-Kollman-Verfahren und die NBO-Ladungen.

7.2.3 Zusammenstellung der Referenzdaten

Aus dem Testdatensatz meiner Diplomarbeit wurden 36 Moleküle in 20 Reihen gruppiert, die eine systematische Entwicklung der Partialladungen im Hinblick auf die Elektronegativitäten aufweisen. Zu jeder dieser Reihen wurden Referenzdaten als absoluter Wert der Partialladungen und empirische Regeln in Form von Vergleichsanweisungen gespeichert.

Immer wieder ergaben sich Probleme bei der Parametrisierung, die auf Eingabefehler bei den Referenzreihen zurückzuführen waren. Da hier auch durch mehrfache Korrektur von Hand zusammen mit verschiedenen Praktikanten keine endgültige Abhilfe zu schaffen war, wurde ein

Prüfprogramm implementiert. Das Programm „ReihenCheckPro.py“ inspiziert die Reihen anhand von folgenden Kriterien:

- Sind die Bindungslängen der gespeicherten Moleküle plausibel?
- Stimmt die Gesamtladung mit der sich aus Atomtypen und Bindungsliste ergebenden Formalladung überein?
- Erfüllen die Referenzladungen die mitgespeicherten empirischen Vergleichsregeln?
- Stimmen die Referenzladungen mit denen aus einer Neuberechnung mithilfe des Gaussian-Paketes überein? (ReihenCheckPro.py erzeugt die dazu nötigen Eingabedateien automatisch und liest das Ergebnis zurück.)

Mithilfe dieses Programms konnten die Eingabedateien erfolgreich korrigiert werden. Auch zur Neuberechnung von Referenzladungen nach verschiedenen in Gaussian implementierten Methoden und dem automatischen Erzeugen entsprechend geänderter Reihendateien erwies es sich als sehr nützlich.

Zur Bestimmung der Formalladung eines Moleküls aus Atomtypen des MOMO-Kraftfeldes und der Bindungsliste wurde ein spezieller Algorithmus entwickelt. Hier wurde versucht, die Intuition des Chemikers nachzuahmen. Damit dies funktionieren kann, sind nämlich Annahmen und Zusatzinformationen erforderlich, da die Anzahl der über das Molekül zu verteilenden Elektronen nicht von vornherein bekannt ist.

Die Angabe der Ladung eines Moleküls ist Voraussetzung für die Berechnung mit ab-initio-Programmen. Ein Algorithmus wie der hier beschriebene kann dort nicht eingesetzt werden, da die hier gemachten Annahmen nicht allgemein gelten müssen. Für den Zweck des hier vorgestellten Programms funktionieren sie jedoch sehr gut.

Der Algorithmus arbeitet wie folgt:

Zunächst werden für alle Atome „Valenzschalen“ aus 2 bzw. 4 gleichwertigen Orbitalen definiert. Diese werden gemäß der Hund'schen Regel mit Valenzelektronen befüllt. Bei diesem Vorgang wird mitgezählt, wie viele Orbitale mit nur einem Elektron besetzt sind. Nun wird über die Bindungsliste iteriert und versucht, die Bindungen durch den Aufbau bindender Elektronenpaare zu definieren. Pro Zyklus werden jeweils nur Einfachbindungen aufgebaut. Wenn alle Orbitale vollständig besetzt

sind, wird die Rechnung beendet. Zur Bindungsknüpfung wird für jedes Atom ein einfach besetztes Orbital gesucht. Dabei können folgende Fälle auftreten:

1. Bei beiden Bindungspartnern befindet sich in einem Orbital ein einzelnes Elektron, sodass eine Bindung ausgebildet werden kann.
2. Bei beiden Bindungspartnern existieren keine einfach besetzten Orbitale, sodass keine Bindung entsteht.
3. Ein Bindungspartner besitzt ein einfach besetztes Orbital und der andere Bindungspartner ein freies Elektronenpaar. Der Fall wird zurückgestellt, da vielleicht ein anderer Nachbar ein freies Elektron zur Verfügung hat. Falls dies nicht der Fall ist, wird ein Elektron verworfen, welches durch die erfolgte Bindung überzählig ist und eine positive Ladung registriert (Beispiel mit NH_4^+ siehe Abb. 62).

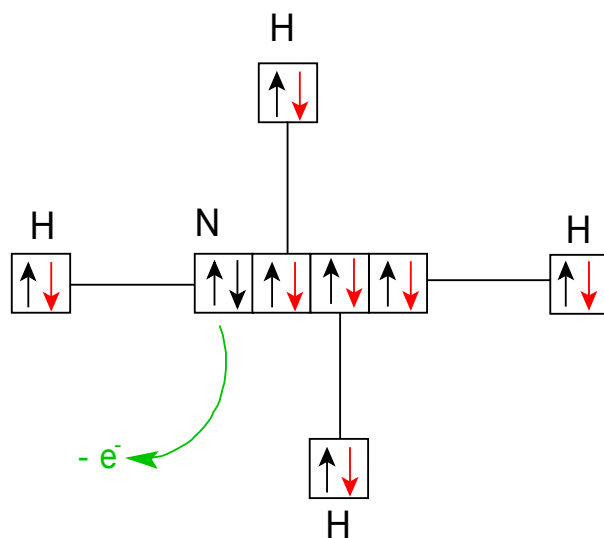
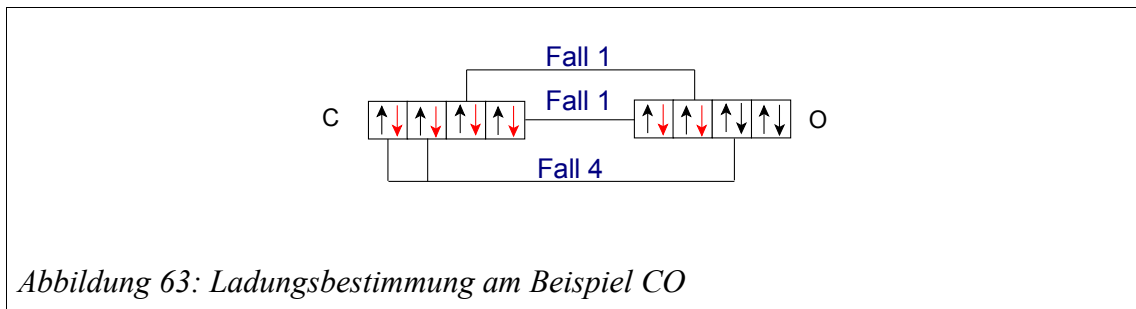
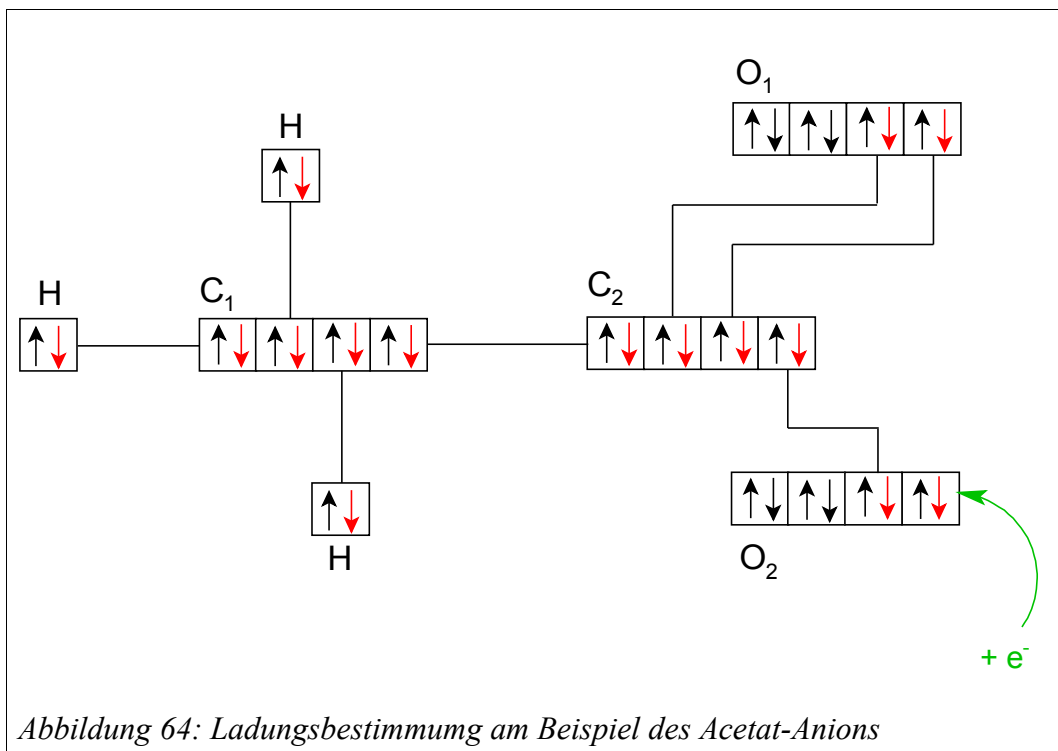


Abbildung 62: Ladungsbestimmung am Beispiel NH_4^+

In schwarz ist die Besetzung der „Valenzorbitale“ vor dem Start des Algorithmus eingezeichnet, die hinzugefügten Elektronen sind rot markiert.

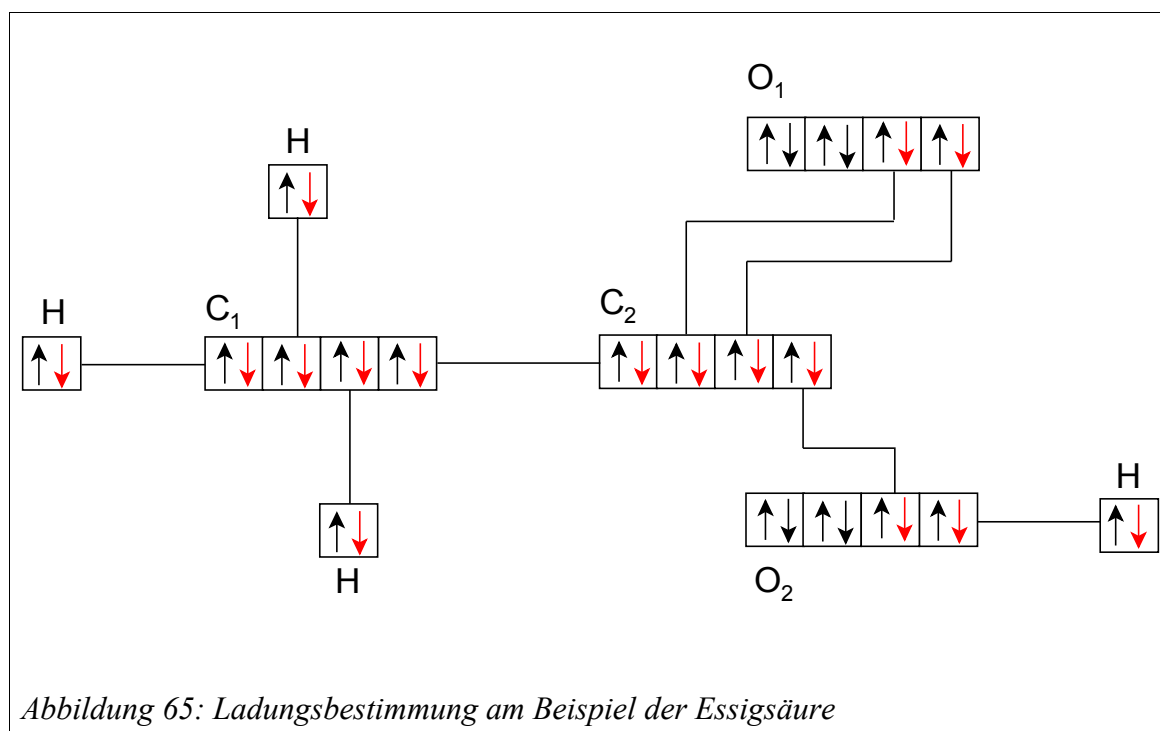


4. Ein Bindungspartner verfügt über zwei einfach besetzte Orbitale und der andere Bindungspartner über ein freies Elektronenpaar. Hier wird, wie im Fall 3, eine Bindung zwischen dem Elektronenpaar und einem der einzelnen Elektronen ausgebildet. Dabei bleibt ein Elektron übrig, welches in das zweite einfach besetzte Orbital platziert wird (Beispiel mit CO siehe Abb. 63).



5. Nach Ausbildung aller möglichen Bindungen bleibt noch ein einzelnes Elektron übrig. Um die Edelgasschale zu erreichen, wird ein Elektron ergänzt, woraus eine negative Ladung resultiert (Beispiel Acetat, siehe Abb. 64. Zum Vergleich Essigsäure in Abb. 65).

Die Bindungen werden durch Iteration über die Bindungsliste in Zyklen erstellt. Falls innerhalb eines Zyklus ein Fall zurückgestellt wird, wird vor dem nächsten Zyklus die Bindungsliste auf die Originallänge gekürzt. Ist ein Zyklus beendet und konnte in dessen Verlauf keine neue Bindung mehr ausgebildet werden, so werden die zurückgestellten Fälle bearbeitet. Die Summe der registrierten Formalladungen ergibt die Gesamtladung des Moleküls.

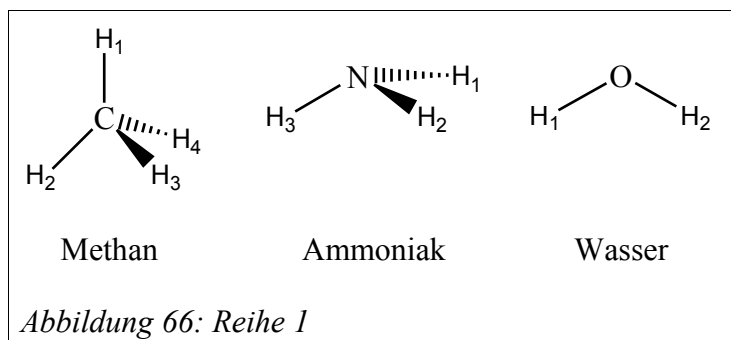


7.2.4 Der Referenzdatensatz

Es wurden 36 Moleküle in 20 Reihen zusammengefasst, die eine systematische Entwicklung aufwiesen. Zu jeder Reihe wurden jeweils die Referenzdaten und Vergleichsanweisungen gespeichert, die später vom Parametrisierungsprogramm genutzt werden können. Diese empirischen Regeln bilden systematische Entwicklungen der atomaren Monopole nach chemischen Gesichtspunkten ab. Im Folgenden sind die so erstellten Reihen und die mit ihnen zusammen gespeicherten empirischen Regeln aufgeführt.

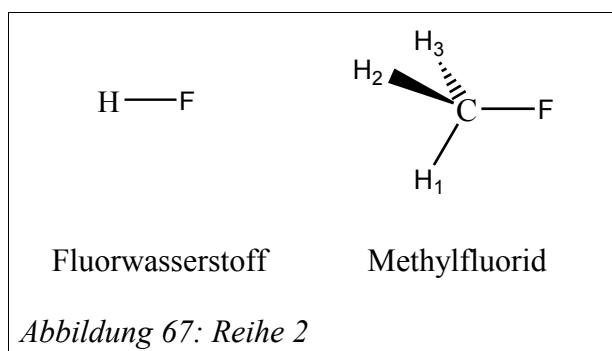
Die Numerierung der Atome in den folgenden Abbildungen erfolgt nicht nach IUPAC, sondern orientiert sich an den von MOMO verwendeten Atomnummern.

Reihe 1



Die Partialladungen am Wasserstoff sollten aufgrund dessen geringer Elektronegativität immer positiv sein. Die negative Ladung am Zentralatom bzw. die positive Ladung am Wasserstoffs sollte in der Reihenfolge Methan, Ammoniak, Wasser gemäß der Reihenfolge der Elektronegativitäten von C, N und O, steigen, was als Vergleichsregel gespeichert wurde.

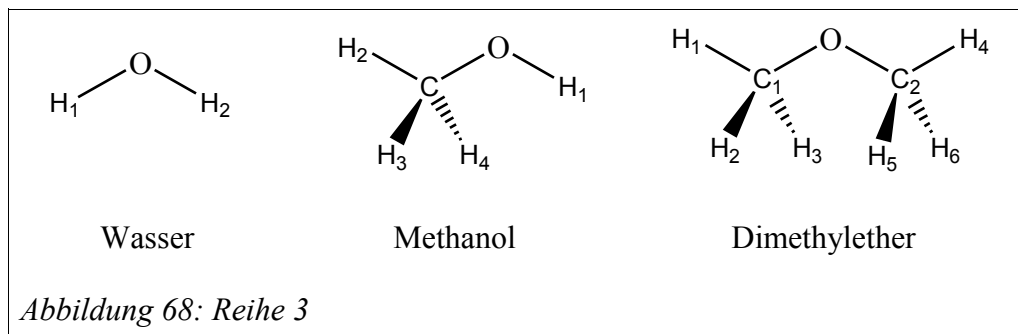
Reihe 2



Methylgruppen sind elektronenziehende Substituenten. Aufgrund der größeren Elektronegativität des Kohlenstoffs gegenüber dem Wasserstoff sollte das Zentralatom bei einem H/CH₃-Austausch positiviert werden. Beim Methylfluorid wird die negative Ladung am Fluor dementsprechend abgeschwächt, was in dieser Reihe als Vergleichsregel dient.

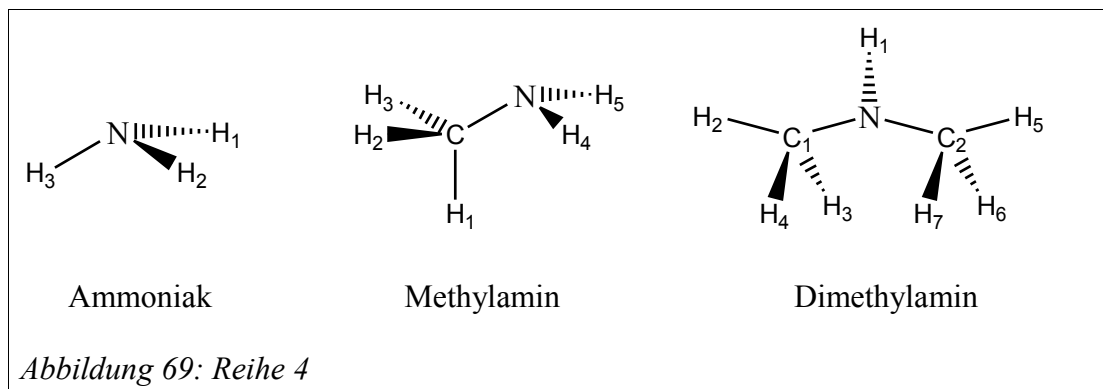
Reihe 3

Durch den Einfluss der Methylsubstituenten wird der zentrale Sauerstoff zunehmend vom Wasser zum Dimethylether positiviert. Dies ist als Vergleichsregel hinterlegt.



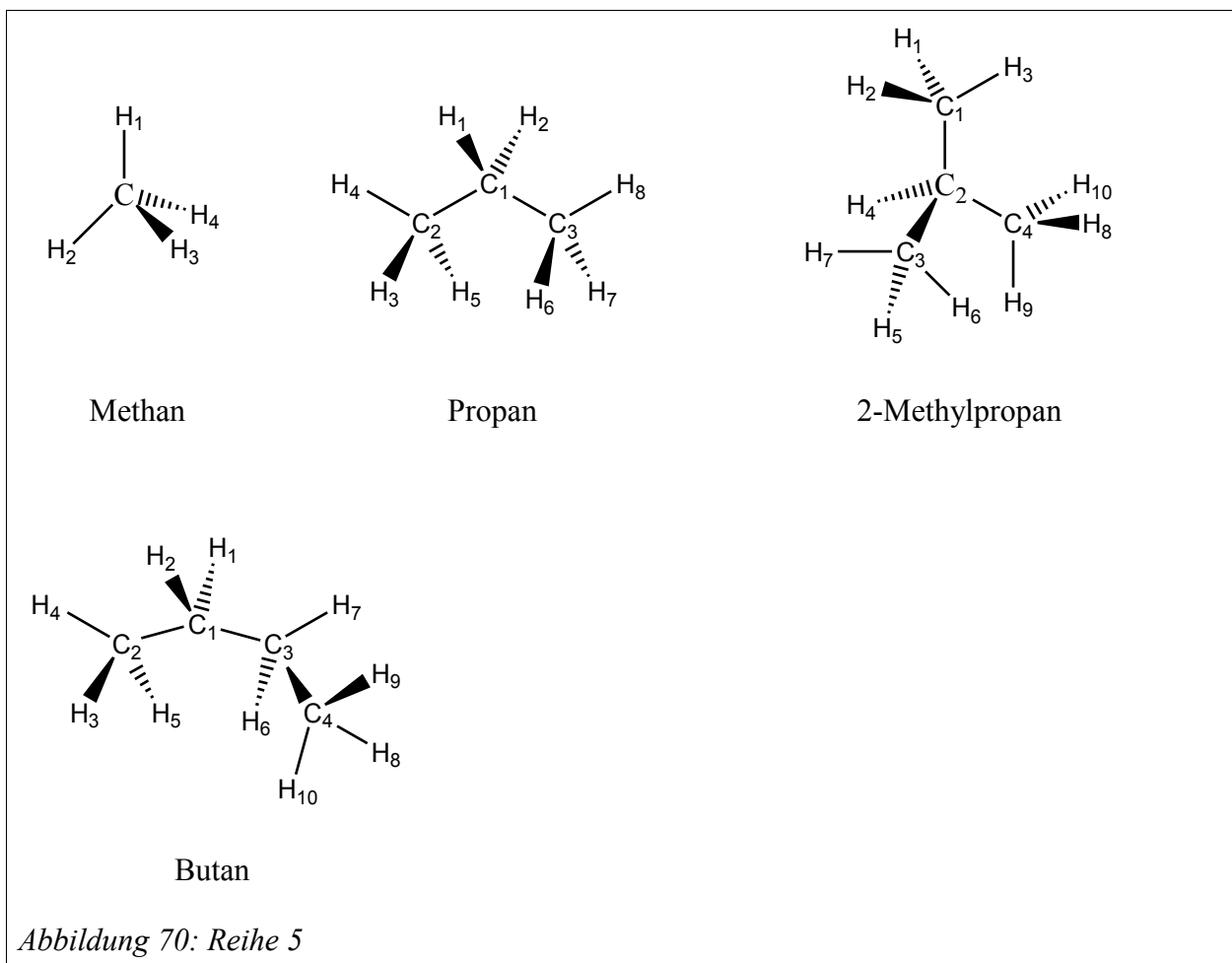
Reihe 4

Wie in Reihe 3 ergibt sich durch die zunehmende Substituierung durch Methylgruppen eine Positivierung des Zentralatoms, was auch hier als Vergleichsanweisung dient.

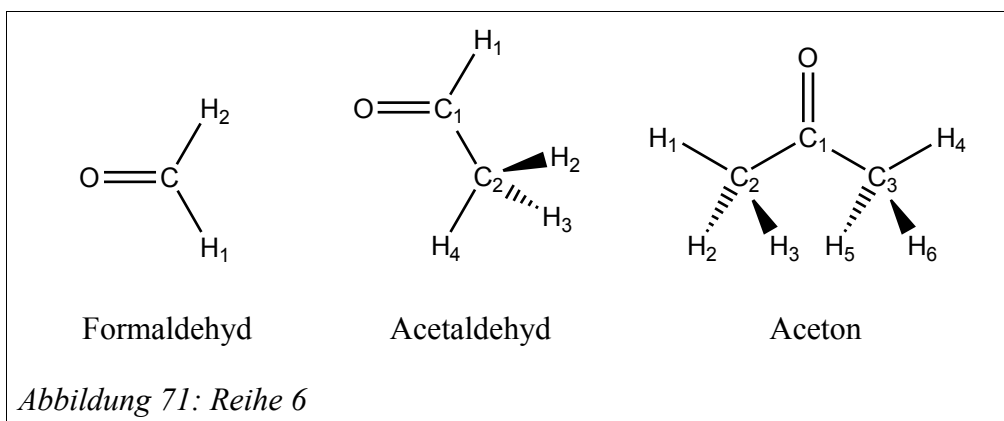


Reihe 5

Ähnlich wie in den vorhergehenden Reihen ist auch bei reinen Kohlenwasserstoffbindungen ein induktiver Effekt zu beobachten. Der zentrale Kohlenstoff wird mit zunehmender Substituierung immer positiver. Wird die Kette vom Propan zum Butan verlängert, ist eine Positivierung der terminalen Methylgruppe zu erwarten. Dies sind die Vergleichsregeln für diese Reihe.

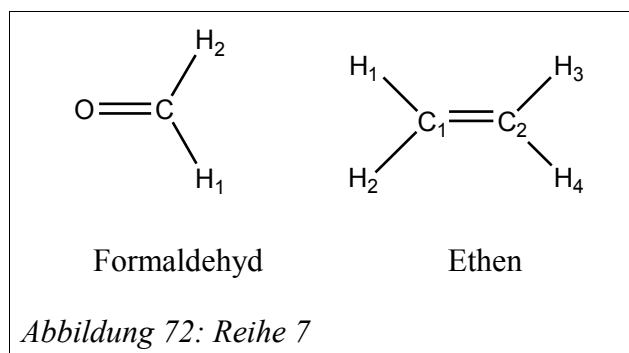


Reihe 6



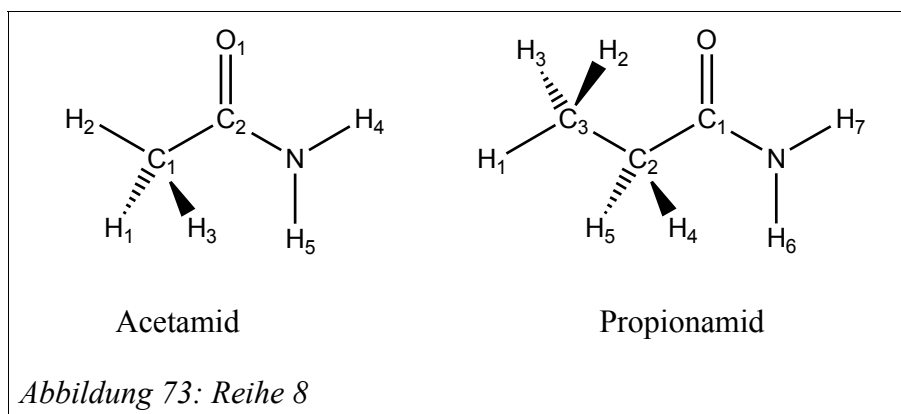
Die Positivierung durch Methylsubstituenten ist auch bei den Carbonylverbindungen zu beobachten. Durch die Induktion der Substituenten sollte sich die negative Ladung des Carbonylsauerstoffs erhöhen. Ersteres wird als Vergleichsregel gespeichert.

Reihe 7



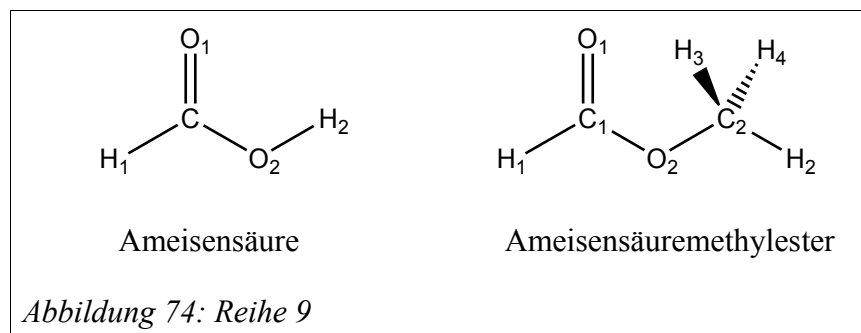
Ethen ist isoelektronisch zum Formaldehyd. Die größere Elektronegativität des Sauerstoffs resultiert in einer Reduktion der Ladungsdichte am Kohlenstoff, durch den induktiven Effekt ist gleichzeitig eine Negativierung der Wasserstoffatome zu erwarten. Beide Effekte werden als Vergleichsregeln der Reihe genutzt.

Reihe 8



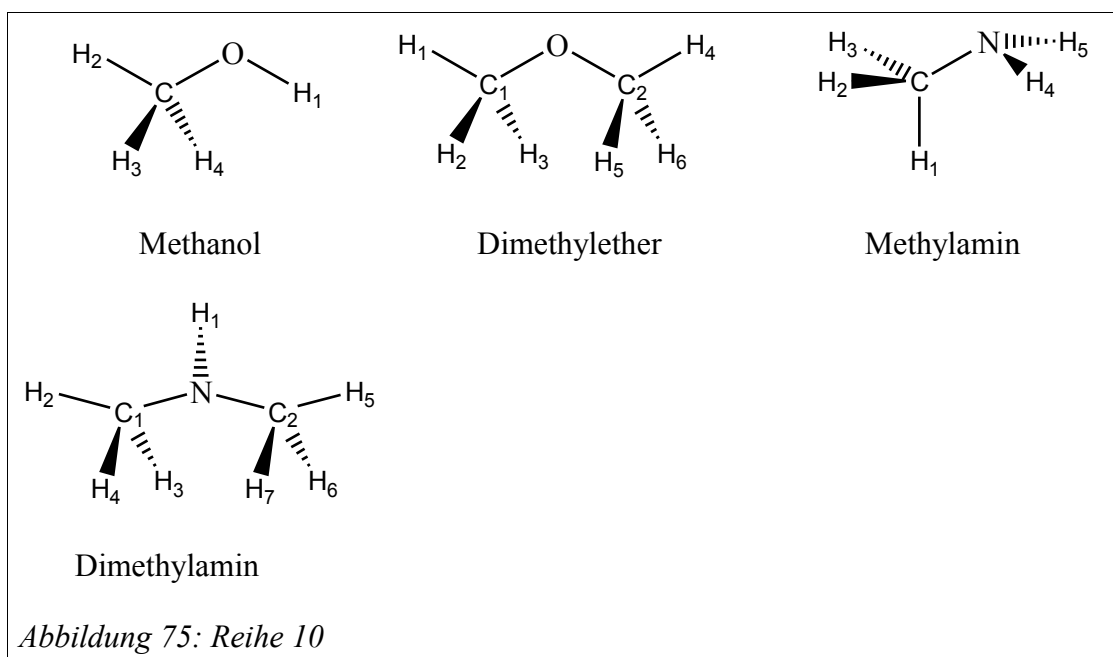
Vergleicht man die Partiaalladungen von Acetamid und Propionamid, so wirkt der Einfluss des Methylsubstituenten in γ -Stellung sehr klein. Das Abspeichern einer Vergleichsregel erscheint hier nicht gerechtfertigt. Dies gilt sowohl für den Carbonyl-Sauerstoff als auch für den Stickstoff.

Reihe 9



Ein ähnlich geringer Einfluss des γ -Substituenten, wie in Reihe 8, ist auch hier zu beobachten. Signifikant erscheint lediglich eine leichte negative Induktion am H1, die hier als Vergleichsregel dient.

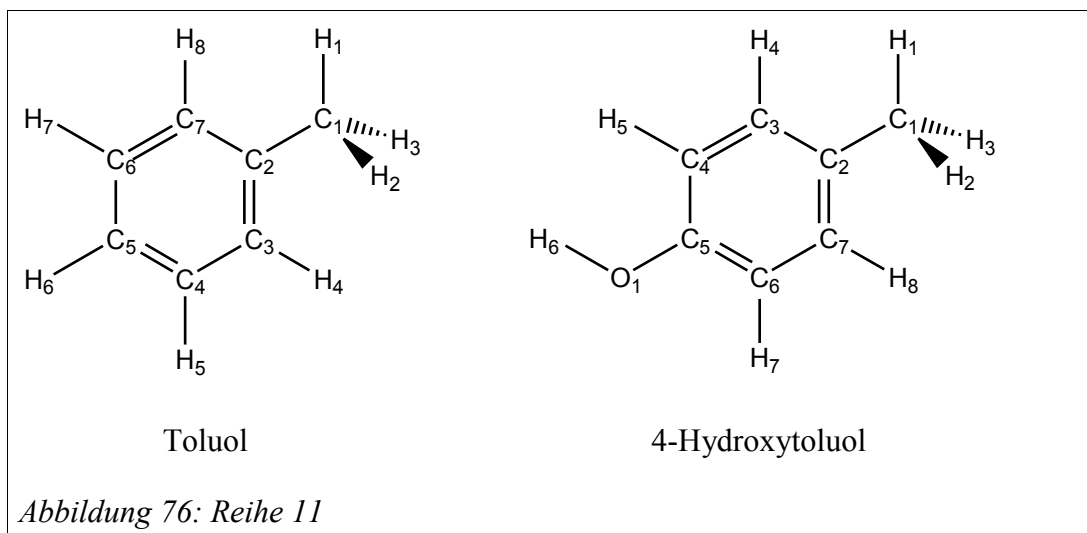
Reihe 10



Beim Vergleich von Methanol mit Dimethylether und Methylamin mit Dimethylamin kann der Einfluss der Methylgruppen gut beobachtet werden. Als Vergleichsregeln dienen hier die Positivierung des Sauerstoffs vom Methanol zum Dimethylether, des Stickstoffs vom Methylamin zum Dimethylamin und die Positivierung der Methyl-Wasserstoffatome vom Methylamin zum Dimethylether. Beim Übergang zum jeweils höheren Homologen verringert sich die

Elektronendichte an den Wasserstoffen. Um eine Konformationsabhängigkeit auszumitteln, werden die Wasserstoffpartialladungen vor dem Vergleich aufsummiert.

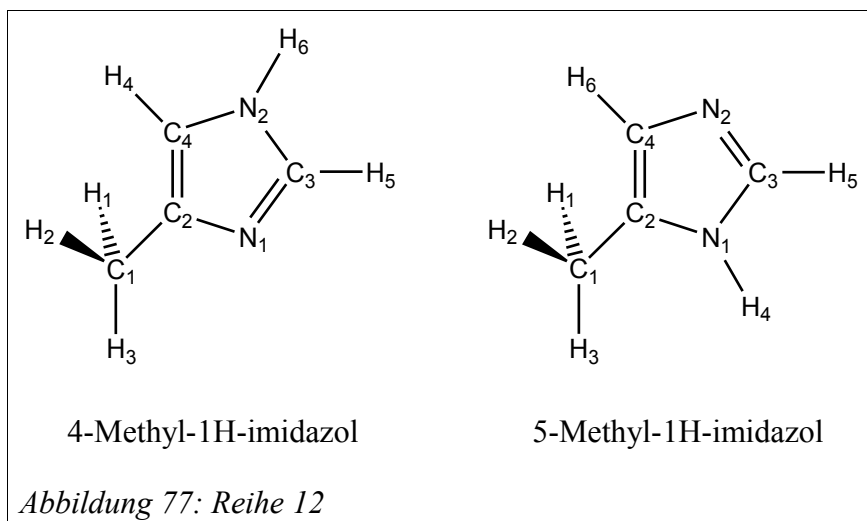
Reihe 11



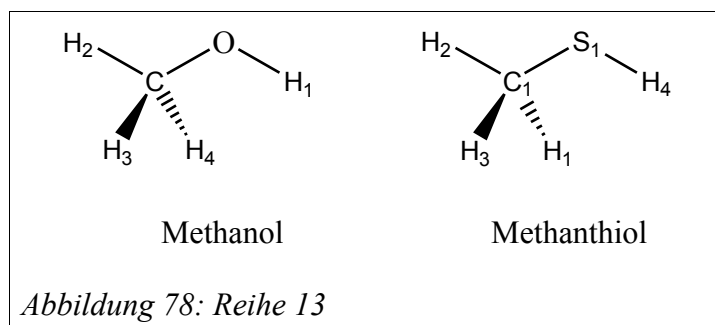
In Aromaten führen induktive und mesomere Effekte zur Veränderung der Ladungsdichte am aromatischen System. Im Toluol wird die Elektronendichte an der Verknüpfungsstelle (C2) abgesenkt, in Ortho- und Para-Stellung kommt es zu einer Negativierung gegenüber der Meta-Position. Dies ist innerhalb des Toluols als Vergleichsregel angelegt. Die Substitution mit einer Hydroxyl-Gruppe zum 4-Hydroxytoluol mündet in eine Reduktion der Ladungsdichte am C5. In Ortho- und Meta Stellung dagegen (C2, C4, C6) erhöht sich die Ladungsdichte gegenüber der im Toluol. In der Meta-Position (C7,C3) erfolgt eine Positivierung. Diese Vergleiche werden als empirische Regeln gespeichert.

Reihe 12

Zweifach koordinierter Stickstoff zieht mehr Elektronendichte von seinen Nachbarn ab als dreifach koordinierter. Dementsprechend muss die Partiaalladung am zweifach koordinierten Stickstoff negativer sein. Dieser Vergleich ist in den Referenzdaten intramolekular definiert.



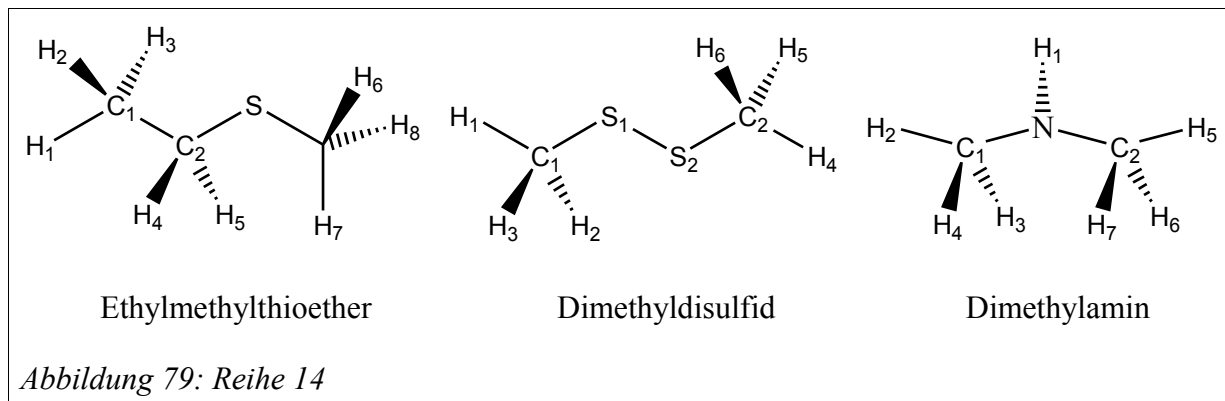
Reihe 13



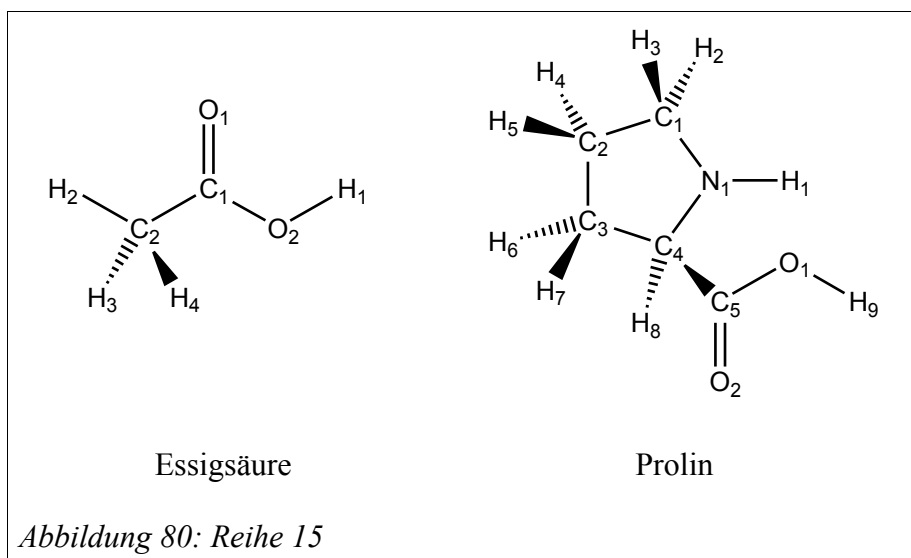
Innerhalb einer Hauptgruppe nimmt die Elektronegativität der Elemente mit steigender Ordnungszahl ab. Die Partialladung der Hydroxylwasserstoffe sollten deshalb positiver ausfallen als die in den entsprechenden Schwefelanaloga, was hier als empirische Regel dient.

Reihe 14

Im Ethylmethylthioether wird die Ladung der beiden α -Substituenten im Kontrast zu Methanethiol deutlich reduziert, eine Teilkompensation kann durch die negativen Methyl-Kohlenstoffatome in β -Stellung angenommen werden. Vergleichsregeln werden in dieser Reihe nicht angewendet.

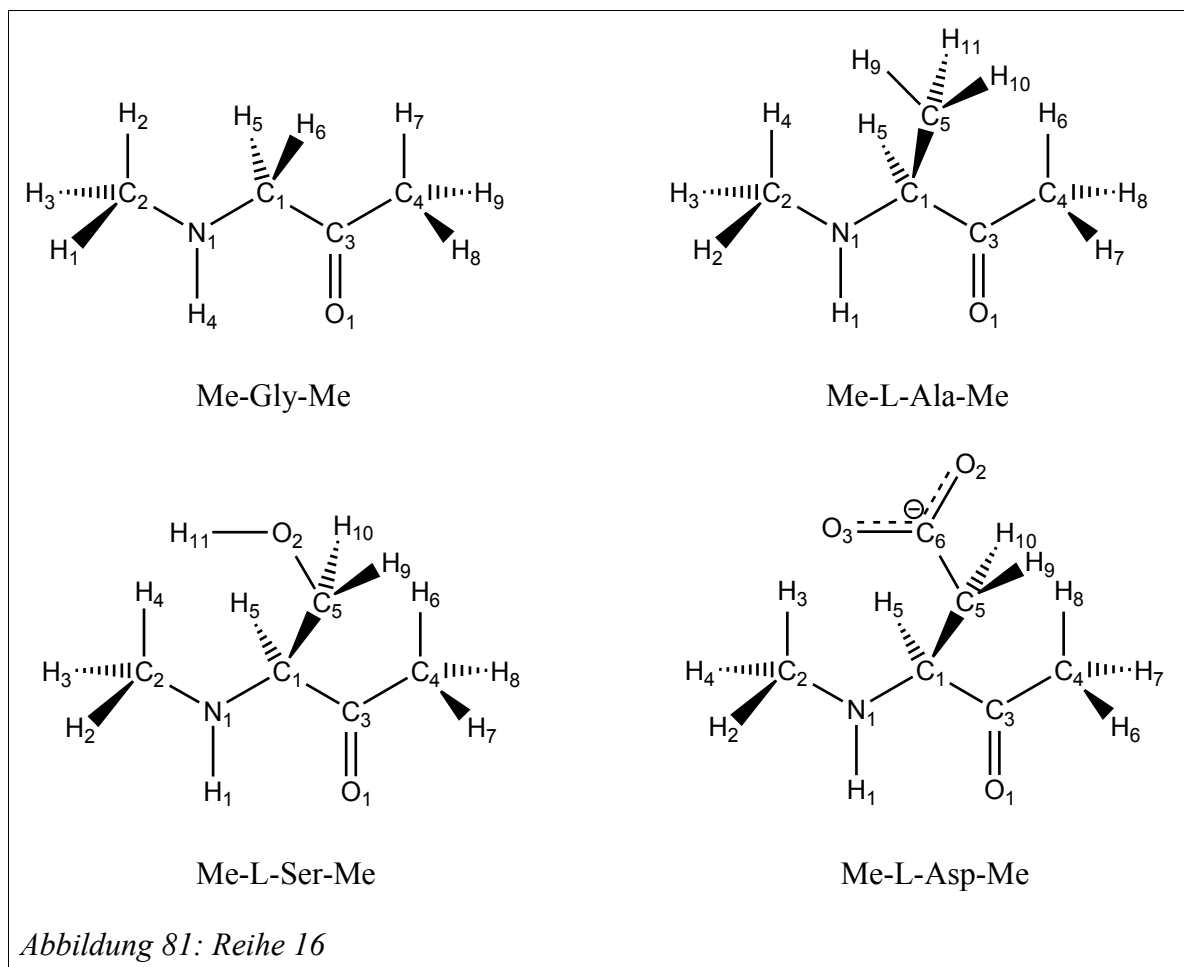


Reihe 15



In dieser Reihe werden die Carbonylkohlenstoffe miteinander verglichen. Im Prolin sollte die Ladungsdichte wegen des induktiven Effekts des Stickstoffs in γ -Position gesteigert werden. Dies ist die einzige empirische Regel in dieser Reihe.

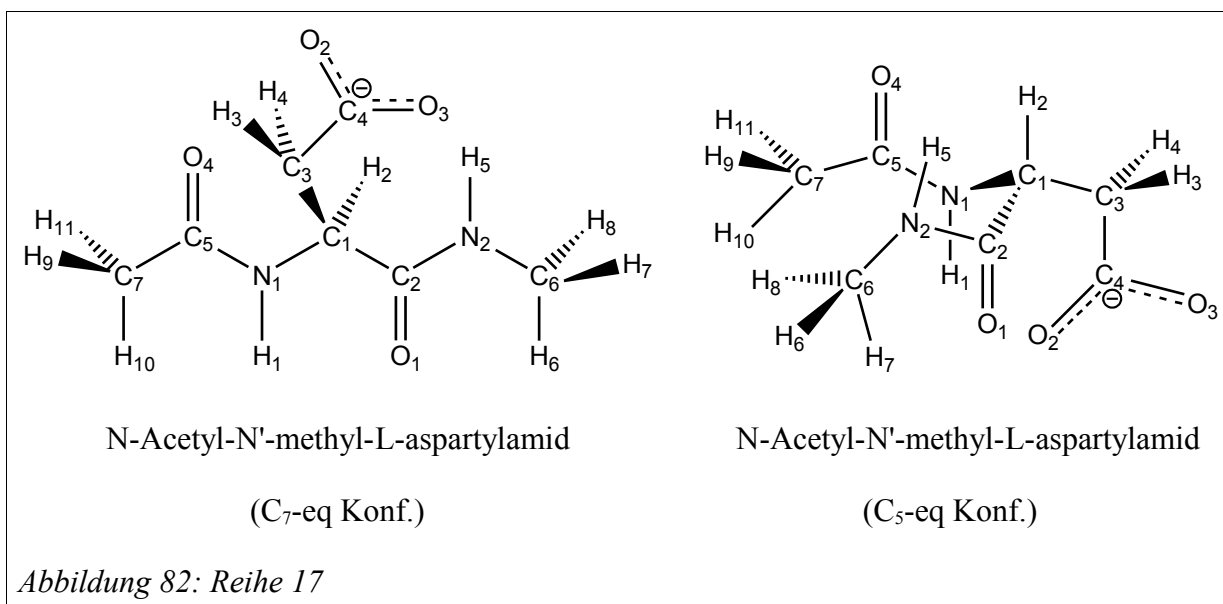
Reihe 16



Diese Reihe enthält einige Modellverbindungen für Aminosäureseitenketten. Bei der Substitution des zentralen Kohlenstoffatoms mit einer Methylgruppe wird dieses positiviert. Geht man weiter zu Me-L-Ser-Me, wird C1 durch die Induktion des Hydroxyl-Sauerstoffs negativer und der vormalige Methyl-Kohlenstoff C5 positiver. Zwischen N1 und der Hydroxylgruppe (H11–O2-) bildet sich eine Wasserstoffbrücke, die die Ladungsdichte am Stickstoff erniedrigt und gleichzeitig das Carbonyl-C3 durch den Einfluss der Hydroxylgruppe bzw. der reduzierten induktiven Kraft des N1 positiviert. Im Me-L-Asp-Me ist das C5 gegenüber dem Me-L-Ala-Me durch den induktiven Effekt der COO⁻-Gruppe positiviert. Die genannten Vergleiche sind als Regeln in der Reihe abgespeichert.

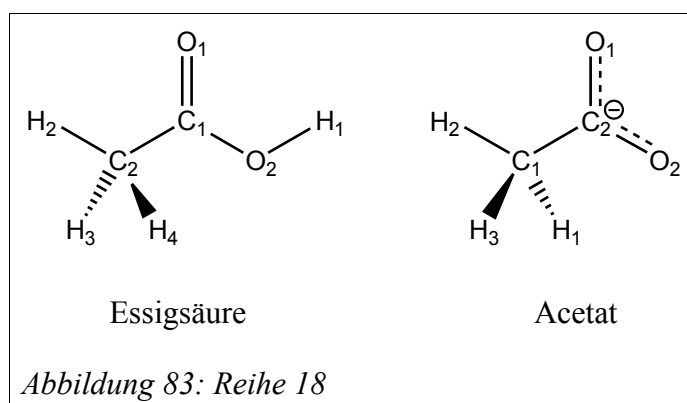
Reihe 17

In dieser Reihe soll der Einfluss der Konformation des Rückgrats auf die Partiaalladungen getestet werden. Die beiden Konformationen divergieren im Wasserstoffbrückenbindungsmuster. In der



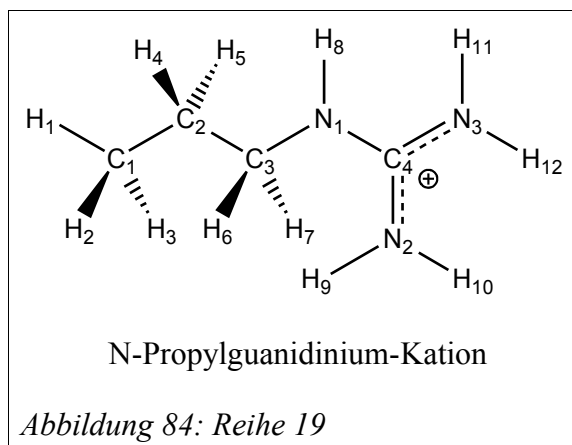
C₅-eq Konformation bilden sich zwischen O₃ und H₅ sowie zwischen O₁ und H₁ Wasserstoffbrücken und in der C₇-eq Konformation zwischen O₂ und H₁ sowie zwischen O₄ und H₅. Die Partiaalladung am N₂ ist deshalb in der C₅-eq Konformation positiver als in der C₇-eq Konformation, was als Vergleich codiert ist.

Reihe 18



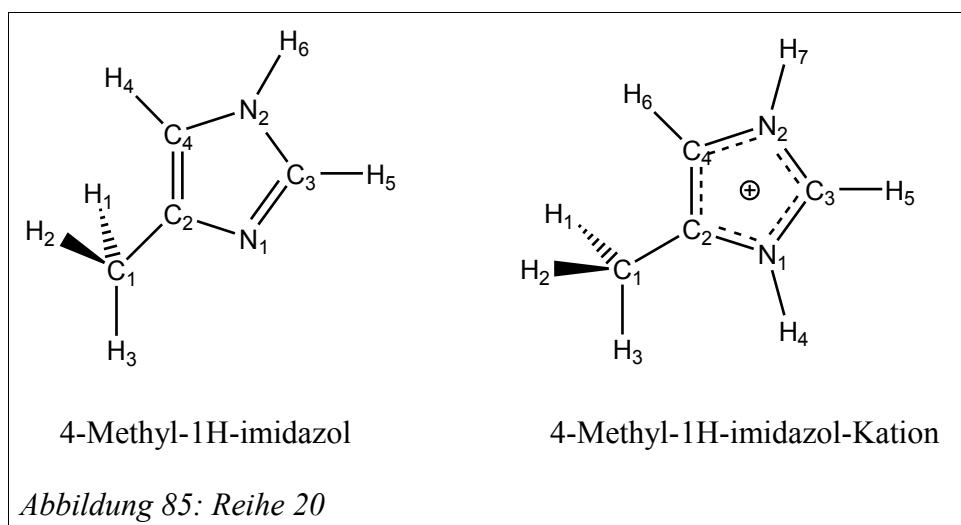
Durch die Deprotonierung der Essigsäure zum Acetat wird die Ladung am zentralen Carbonyl-Kohlenstoff deutlich positiviert. Die Methylgruppe (Summe der Partiaalladungen) wird negativiert. Beide Vergleiche wurden für die Reihe codiert.

Reihe 19



Im N-Propylguanidinium-Kation ist die positive Formalladung über N2, C4 und N3 delokalisiert. Eine empirische Regel für die Partialladungen wurde nicht aufgestellt.

Reihe 20



Im 4-Methyl-1H-imidazol-Kation sind die Ringatome gegenüber der Neutralverbindung positiviert, einzig am C3 lässt sich, da beide umliegende Stickstoffe in trigonal koordinierter Form vorliegen, eine Negativierung feststellen, welche in dieser Reihe als empirische Regel genutzt wird.

7.2.5 Die Codierung (XML-Tags und Vergleichsanweisungen)

Für die Speicherung der Referenzdaten werden die bisher verwendeten XML-Dateien um zusätzliche Tags erweitert.

```

<ParameterizationReference>
<PointCharges ChargeMethod="MK">
  <universe ref="Methan universe"/>
  <AtomArray e="-0.495 0.124 0.124 0.124 0.124"/>
</PointCharges>
<PointCharges ChargeMethod="MK">
  <universe ref="Ammoniak universe"/>
  <AtomArray e="-1.082 0.361 0.360 0.361"/>
</PointCharges>
<PointCharges ChargeMethod="MK">
  <universe ref="Wasser universe"/>
  <AtomArray e="-0.806 0.403 0.403"/>
</PointCharges>
<QualityRules>
  <compareCharge operator="gt">
    <operand1 refUniverse="Wasser universe" refMolecule="Wasser" proc="mean" refAtoms="H1A H1B"/>
    <operand2 refUniverse="Ammoniak universe" refMolecule="Ammoniak" proc="mean" refAtoms="H1A H1B H1C"/>
  </compareCharge>
  ...
  <compareCharge operator="gt">
    <operand1 refUniverse="Ammoniak universe" refMolecule="Ammoniak" proc="mean" refAtoms="H1A H1B H1C"/>
    <operand2 refUniverse="Methan universe" refMolecule="Methan" proc="mean" refAtoms="H1A H1B H1C H1D"/>
  </compareCharge>
  ...
</QualityRules>

</ParameterizationReference>

```

Abbildung 86: XML Tags für die Parametrisierung von Ladungsmodellen

Unterhalb des Tags `<ParameterizationReference>` werden die Tags `<PointCharges>` und `<QualityRules>` aufgenommen. Unter `<PointCharges>` werden die Referenzladungen und unter `<QualityRules>` die im letzten Abschnitt angesprochenen empirischen Vergleichsregeln abgelegt.

<PointCharges> Dieser Tag hat das Attribut „ChargeMethod“, dessen Wert einen Text enthält, der die verwendete Methode zur Berechnung der gespeicherten Partiaalladungen kennzeichnet. Das Kürzel „MK“ steht hier für das in Gaussian implementierte Merz-Kollman-Verfahren. Untertags sind `<universe>` und `<AtomArray>`. `<Universe>` enthält ein Attribut „ref“, dessen Wert der Name des Universe-Objektes, also der Name, unter dem die Daten des chemischen Systems, auf das sich die Referenzladungen beziehen, angesprochen wird. `<AtomArray>` enthält das Attribut „e“, das die Partiaalladungen in der Einheit „Elementarladungen“ für alle Atome innerhalb des referenzierten Universe-Objektes in der Reihenfolge von dessen Atomliste aufführt.

<QualityRules> Dieser Tag fasst Untertags zusammen, die empirische Regeln über Partialladungen von Molekülen beschreiben. Der Tag `<CompareCharge>` hat als Attribut den „operator“, der die Werte „eq“ (gleich), „gt“ (größer) oder „lt“ (kleiner) annehmen kann. Die entsprechenden Operanden sind in Untertags `<operand1>` und `<operand2>` angegeben. Die Attribute „refUniverse“, „refMolecule“ und „refAtoms“ adressieren die für den Vergleich heranzuziehenden Atome. „refUniverse“ beinhaltet den Namen des Universe-Objekts, „refMolecule“ bezeichnet das Molekül innerhalb dieses Universe-Objektes und „refAtoms“ schließlich Atome innerhalb des Moleküls. Es können ein oder mehrere Atome angegeben werden, verglichen wird letztendlich aber nur eine Zahl. Das Attribut „proc“ gibt an, was mit den Werten bei Angabe mehrerer Atome geschehen soll: „sum“ summiert die Partialladungen auf, „mean“ berechnet deren arithmetisches Mittel.

7.2.6 Analyse des Testdatensatzes mit den verfügbaren Ladungsmodellen

Als Kandidaten für Referenzladungen wurden die mit dem Merz-Kollman-Verfahren und der „natural bond orbital“-Methode berechneten ausgewählt. Beide können mit Gaussian03 berechnet werden. Die so ermittelten Ladungen werden untereinander und mit den beiden in MOMO implementierten Ladungsmodellen GM (Gasteiger-Marsili) und ABR27 (Sabine Monz) verglichen. Für diese Analyse wurde das Programm „CCP.py“ geschrieben. Eingabe ist hier ein Verzeichnis mit den im letzten Kapitel beschriebenen Reihendateien, und der Name zweier zu vergleichender Ladungsmodelle. Die Berechnungen erfolgen für die Ladungsmodelle GM und ABR27 über die MOMO-Bibliothek, MK- und NBO-Ladungen werden durch Aufruf von Gaussian03 berechnet. Ein- und Ausgabedateien können archiviert werden, sodass bereits mit Gaussian berechnete Ergebnisse (es wird auf Übereinstimmung der Eingabedaten geprüft) einfach gelesen werden können. Die Ausgabe erfolgt als HTML-Seite und enthält Histogramme zur Verteilung der Partialladungen nach MOMO-Atomtyp und Ladungsmodell, die Testergebnisse der in den Reihendateien hinterlegten empirischen Regeln für beide Ladungsmodelle sowie Korrelationsdiagramme der Ladungen insgesamt und nach MOMO-Atomtyp. Für jede dieser Korrelationen findet eine lineare Regression statt. Schließlich wird für jeden Atomtyp noch eine Aufstellung der Ladungen in Abhängigkeit zu den Nachbaratomen erzeugt, um die Histogramme zur Ladungsverteilung besser interpretieren zu können.

Die Analyse wurde mit den zuvor beschriebenen 20 Reihen durchgeführt. Für die Berechnungen mit Gaussian03 wurde die Methode „HF/6-31G**“ verwendet. Zunächst werden die beiden quantenchemischen Methoden MK und NBO miteinander verglichen.

7.2.7 Vergleich der Ladungsmodelle MK und NBO

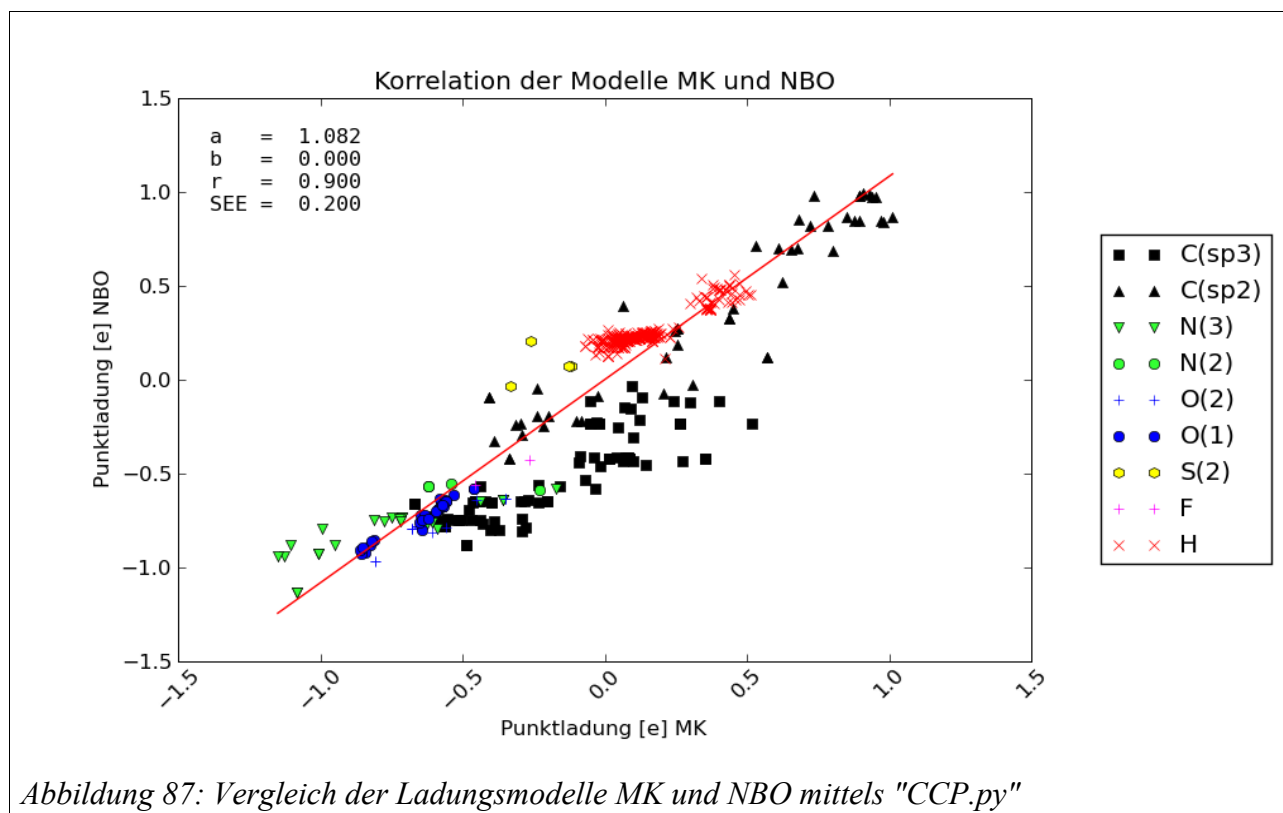


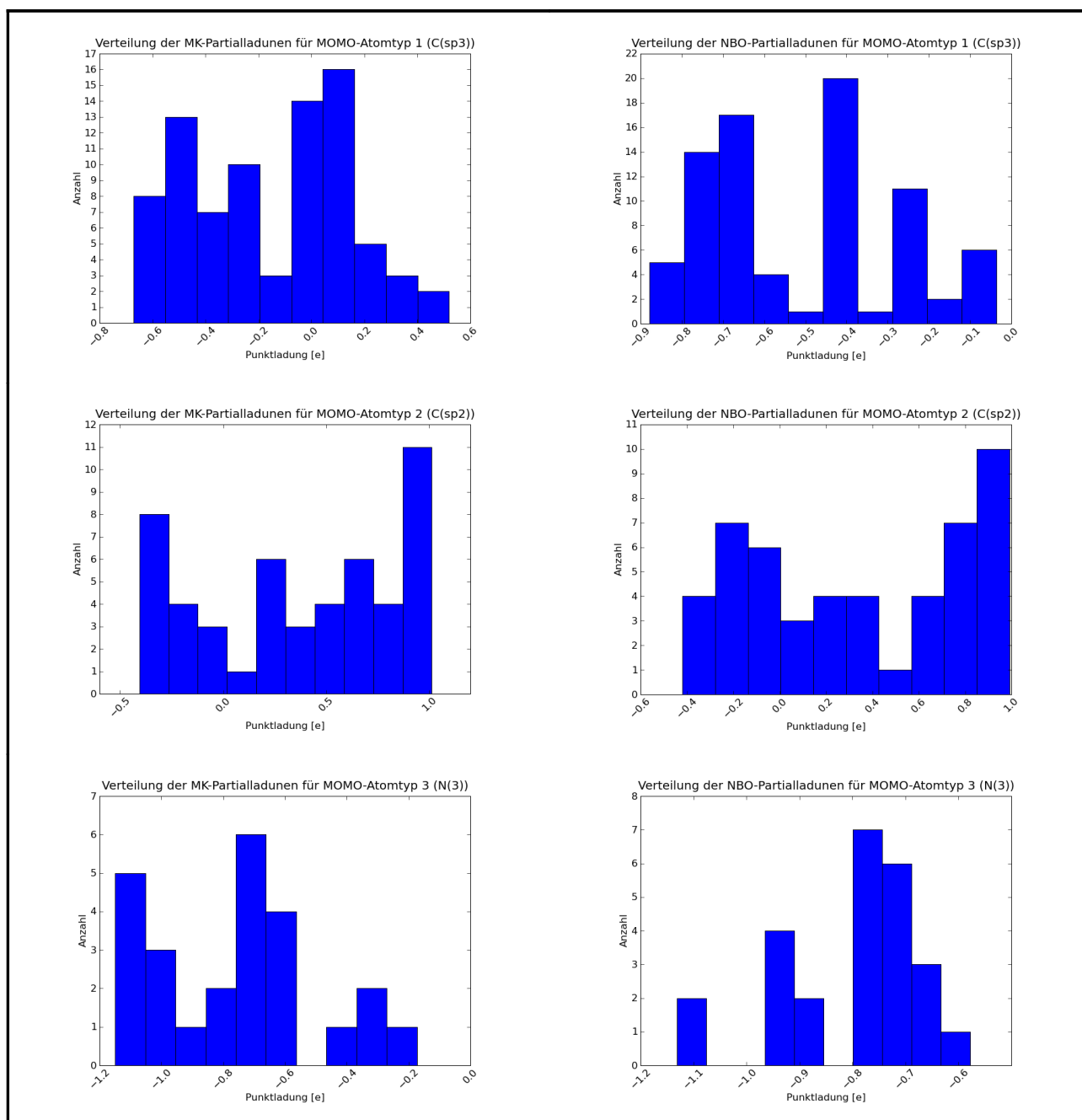
Abbildung 87: Vergleich der Ladungsmodelle MK und NBO mittels "CCP.py"

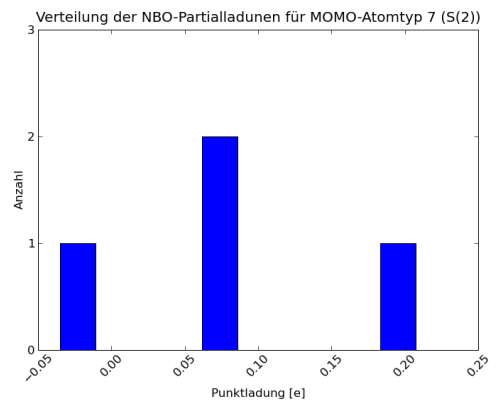
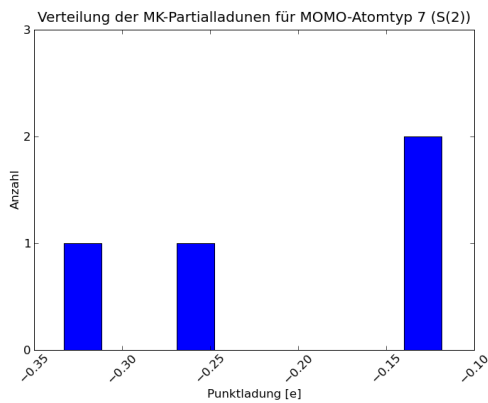
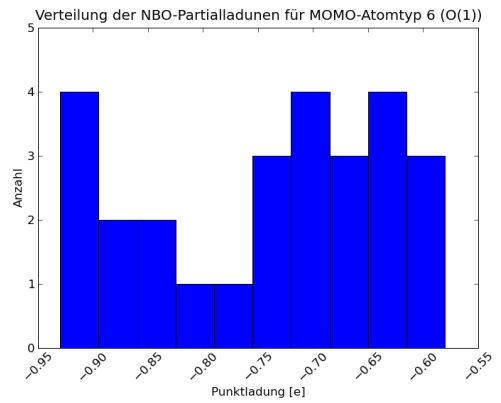
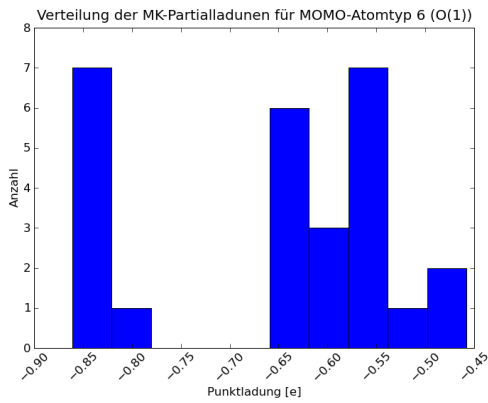
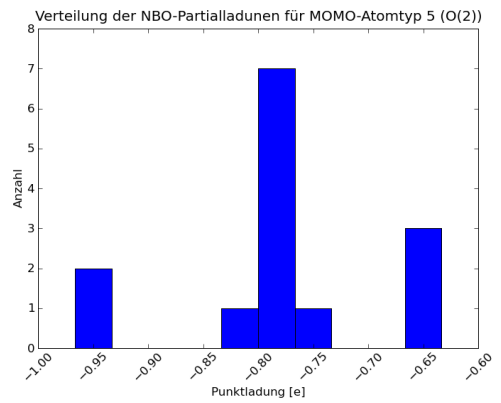
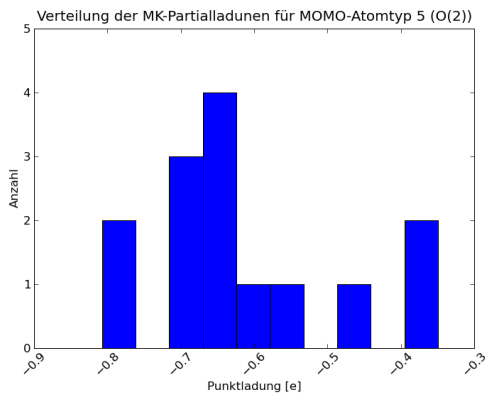
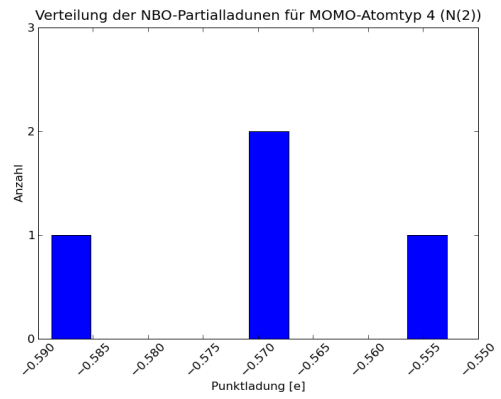
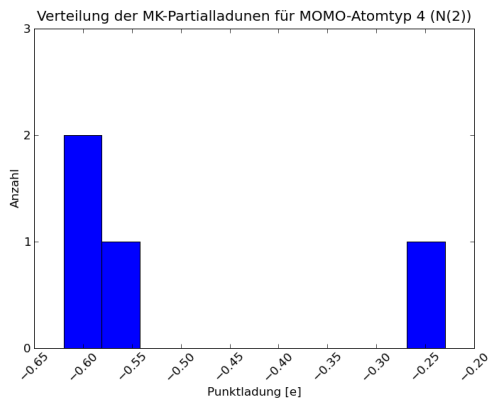
Abb. 87 zeigt die Korrelation der MK- und NBO-Ladungen, links oben sind die Ergebnisse der linearen Regression ($NBO = a \cdot MK + b$) angegeben. Der Korrelationskoeffizient ist r . Die Korrelation ist nicht sehr gut, obwohl beide Methoden dieselbe Eigenschaft bestimmen sollen. Der Achsenabschnitt von 0.0 ist bei den den überwiegend neutralen Verbindungen im Testdatensatz zu erwarten. Eine Skalierung der Werte zwischen den beiden Ladungsmodellen ist angesichts einer Steigung der Ausgleichsgerade von 1.082 nicht nötig. Die Standardabweichung des Schätzwertes beträgt 0.2 e. Sie ist in der Abbildung mit SEE bezeichnet und wird nach Formel 7.1 berechnet. Betrachtet man die Bandbreite der Ladungen, ist dieser Fehler recht groß.

$$SEE = \sigma_{est} = \sqrt{\frac{\sum (NBO - a \cdot MK + b)^2}{n - 2}} \quad (7.1)$$

7 Angewandter Teil

Besonders schlechte Korrelationskoeffizienten ergeben sich für die Atomtypen N(2) und S(2), allerdings liegen für diese nur drei bzw. vier Wertepaare vor, die zudem noch aus verschiedenen Verbindungsklassen stammen. Die Ladungen werden hier durch das MK-Verfahren meist deutlich negativer bewertet als durch die NBO-Methode. Bei S(2) gelangen beide Verfahren nur für den Fall des Ethylmethylthioethers zu ähnlichen Ergebnissen, bei den drei Werten für Atomtyp N(2) weicht der für die Ladung im 4-Methyl-1H-imidazol stark ab.





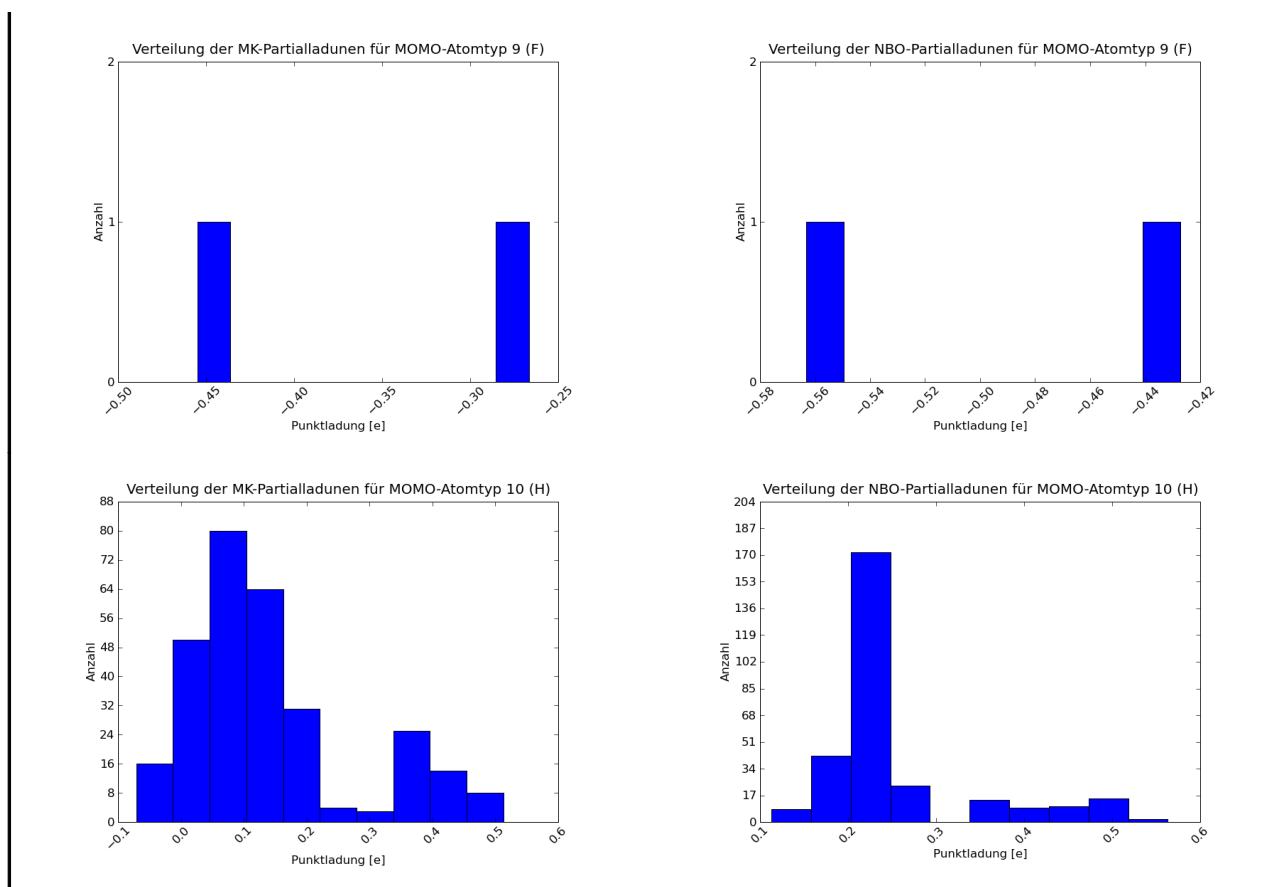


Tabelle 24: Verteilung der Atomladungen für die Modelle MK und NBO nach Atomtyp

Betrachtet man die Verteilung der Ladungen (Tabelle 24), so fällt auf, dass die MK-Ladungen zu einer breiten Verteilung neigen, wohingegen sich die NBO-Ladungen eher auf einige ausgeprägte Peaks konzentrieren. Besonders prägnant ist dies beim Kohlenstoff. Der Effekt lässt sich vermutlich damit erklären, dass die MK-Ladungen relativ stark von der Geometrie des betreffenden Moleküls abhängen und so leicht mit variierender Konformation schwanken, wohingegen die NBO-Ladungen hauptsächlich von der Topologie abhängen.

Die sp^3 -Kohlenstoffladungen ($C(sp^3)$) decken einen großen Bereich von -0.671 e bis 0.130 e ab, wobei die nach MK berechneten Zahlen durchweg positiver ausfallen. Die negativsten Kohlenstoffatome sind solche in Nachbarschaft von Carbonylgruppen, der positivste sp^3 -Kohlenstoff findet sich im Methylfluorid. Aus Tabelle 25 ist die Verteilung der Partialladung in Abhängigkeit zu den Nachbaratomen ersichtlich. Es zeigt sich auch hier deutlich, dass die Bandbreite der mit dem NBO-Verfahren berechneten Werte sehr viel geringer ist als die der nach MK bestimmten.

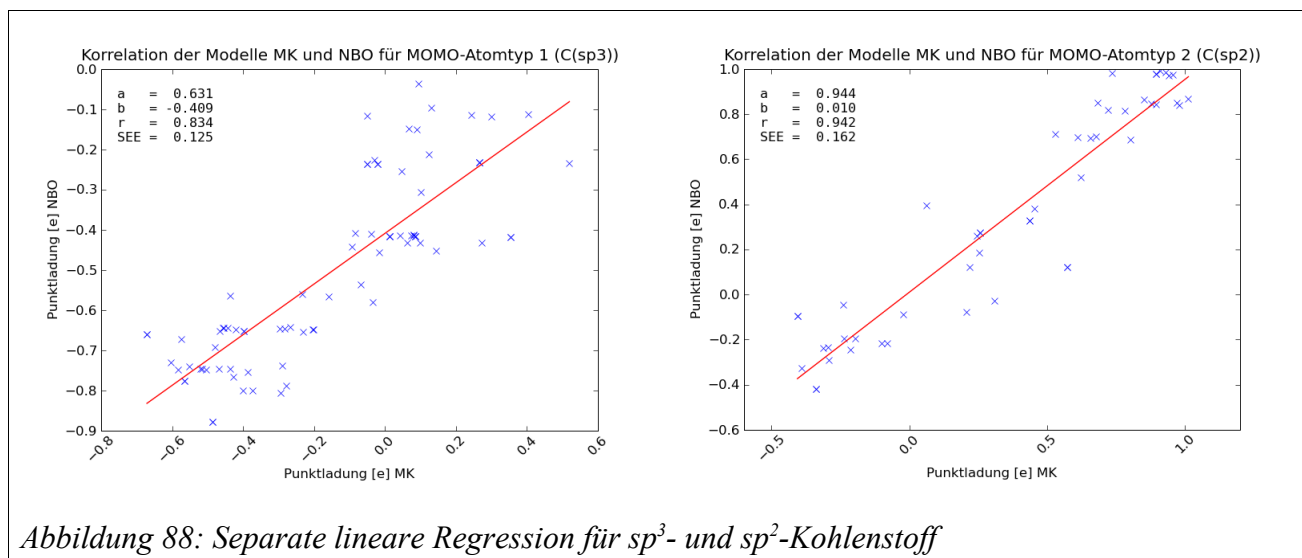


Abbildung 88: Separate lineare Regression für sp^3 - und sp^2 -Kohlenstoff

Sehr viel besser ist die Übereinstimmung bei den sp^2 -Kohlenstoffatomen. Die bei den sp^3 -Kohlenstoffen gefundenen Abweichungen treten zwar in gleicher Weise auf, sind jedoch längst nicht so stark ausgeprägt. Dies offenbart sich deutlich, wenn für diese beiden Atomtypen eine separate lineare Regression getätigt wird (Abb. 88).

Bei den Wasserstoffatomen gibt es einen großen Peak bei 0.1 e (MK) bzw. 0.2 e (NBO) der, wie Tabelle 26 zeigt, zu den Wasserstoffen der Alkylgruppen im Testdatensatz gehört. Dieser Peak fällt in beiden Ladungsmodellen mit den an sp^2 -Kohlenstoffzentren gebundenen Wasserstoffatomen zusammen. Die negativsten Wasserstoffatome treten in Alkylgruppen auf, die positivsten in Alkoholen. Die Ladungen des NBO-Modells sind insgesamt deutlich positiver und weniger breit gestreut.

Carbonylsauerstoffatome (O(1)) werden vom MK-Verfahren negativer (ca. 0.086 e) als von der NBO-Methode bewertet, die auch hier eine geringere Streuung der Werte aufweist. Dies passt zur, im Vergleich zum NBO-Modell, zu stark positiven Bewertung der sp^2 -Kohlenstoffatome, da dieser Atomtyp in jedem Fall Nachbaratom von Atomen des Typs O(1) ist. Dies ist zwingend, weil die Summe der Partialladungen bei allen Ladungsmodellen der Gesamtladung des Moleküls entspricht.

Auch die Ladungen des Atomtyps O(2) werden vom NBO-Modell negativer bewertet; betrachtet man die Bindungspartner, so sind dies solche, die gegenüber der MK-Methode positiver eingestuft wurden. Gleiches kann auch über den Atomtyp N(3) gesagt werden, allerdings bildet hier die Amino-Gruppe eine Ausnahme.

C(sp ³)	Ladungsmodelle:	MK			NBO		
		min	max	mean	min	max	mean
19	C(sp ²); H; H; H;	-0.671	-0.290	-0.507	-0.777	-0.649	-0.724
14	N(3); H; H; H;	-0.085	0.354	0.083	-0.418	-0.409	-0.415
11	C(sp ³); H; H; H;	-0.466	-0.201	-0.336	-0.653	-0.642	-0.647
8	O(2); H; H; H;	-0.050	0.265	0.088	-0.254	-0.232	-0.236
6	C(sp ³); C(sp ²); N(3); H;	-0.051	0.404	0.176	-0.151	-0.112	-0.127
6	C(sp ³); C(sp ³); H; H;	-0.092	0.273	0.078	-0.457	-0.431	-0.441
4	S(2); H; H; H;	-0.400	-0.278	-0.336	-0.805	-0.789	-0.799
4	C(sp ³); C(sp ²); H; H;	-0.437	-0.068	-0.224	-0.566	-0.535	-0.556
2	H; H; H; H;	-0.487	-0.487	-0.487	-0.879	-0.879	-0.879
2	C(sp ³); N(3); H; H;	-0.030	0.123	0.047	-0.226	-0.212	-0.219
1	C(sp ³); O(2); H; H;	0.094	0.094	0.094	-0.036	-0.036	-0.036
1	C(sp ²); N(3); H; H;	0.101	0.101	0.101	-0.306	-0.306	-0.306
1	C(sp ³); C(sp ³); C(sp ³); H;	0.520	0.520	0.520	-0.234	-0.234	-0.234
1	C(sp ³); S(2); H; H;	-0.035	-0.035	-0.035	-0.581	-0.581	-0.581
1	F; H; H; H;	0.130	0.130	0.130	-0.096	-0.096	-0.096

Tabelle 25: Ladungsverteilung am sp³-Kohlenstoff in Abhängigkeit von den Nachbaratomen
Alle Werte in [e].Erzeugt mit CCP.py

H	Ladungsmodelle:	MK			NBO		
		min	max	mean	min	max	mean
216	C(sp ³);	-0.071	0.195	0.080	0.172	0.270	0.218
35	N(3);	0.302	0.513	0.391	0.373	0.490	0.415
28	C(sp ²);	-0.033	0.259	0.129	0.126	0.279	0.204
13	O(2);	0.339	0.472	0.415	0.483	0.538	0.496
1	F;	0.455	0.455	0.455	0.563	0.563	0.563
1	S(2);	0.211	0.211	0.211	0.113	0.113	0.113
1	N(2);	0.379	0.379	0.379	0.487	0.487	0.487

Tabelle 26: Ladungsverteilung am Wasserstoff in Abhängigkeit von den Nachbaratomen
Alle Werte in [e].Erzeugt mit CCP.py

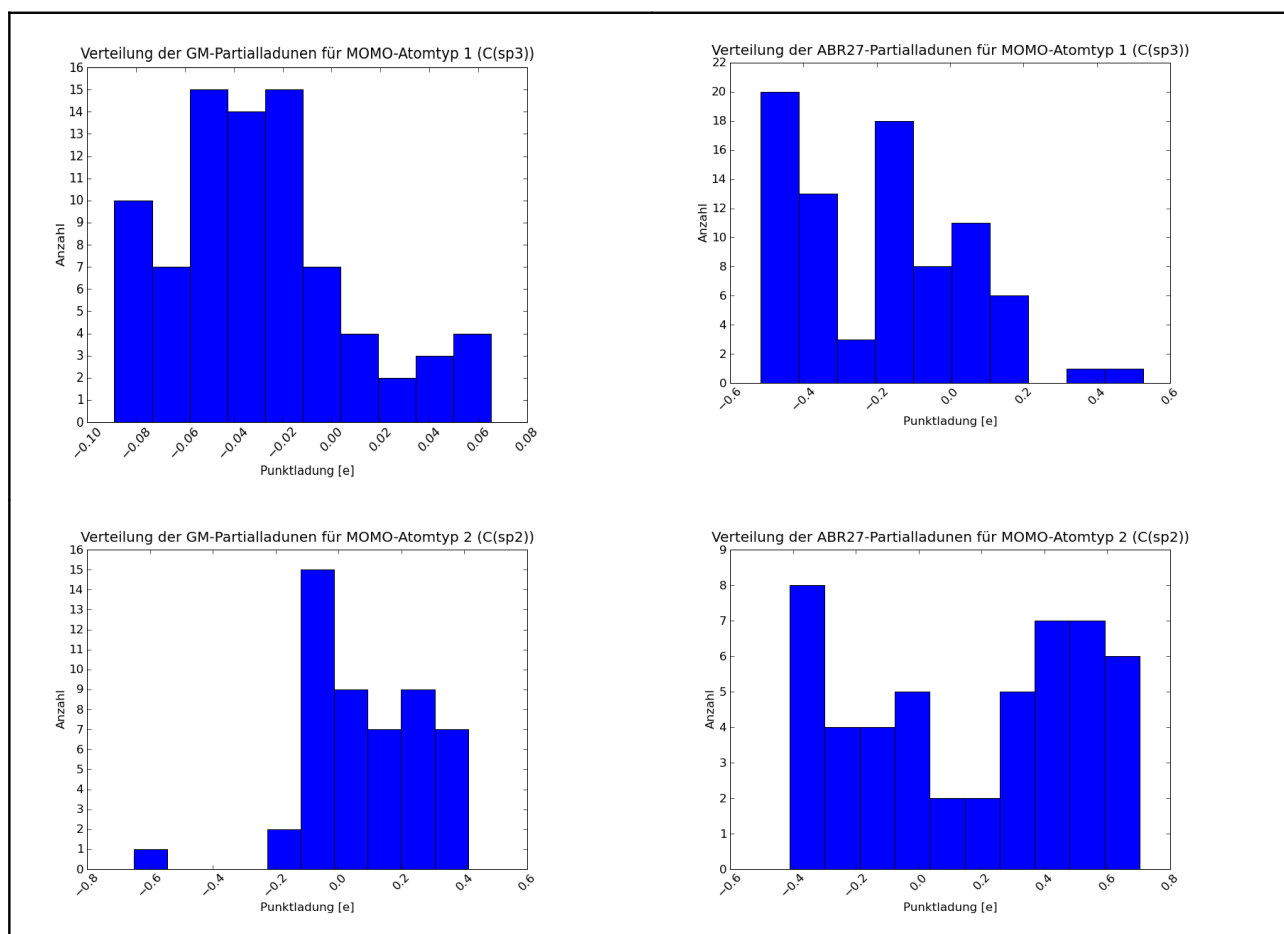
Der vorgestellte Satz empirischer Regeln wird durch das MK-Ladungsmodell vollständig erfüllt, was allerdings auch daran liegt, dass er anhand von Ergebnissen eines ähnlichen Modells entwickelt wurde. Leider ist dies bei den NBO-Ladungen nicht der Fall. In Reihe 5 sind die terminalen Methylgruppen im Propan und Butan neutral, deren Positivierung vom Propan zum Butan kann deshalb nicht nachvollzogen werden. In Reihe 9 werden für H1 praktisch identische Partialladungen berechnet, die Regel zur negativen Induktion wird daher mit diesem Verfahren wahrscheinlich nur zufällig erfüllt oder verletzt. Die Positivierung der Methylwasserstoffatome in Reihe 10 vom Methylamin zum Dimethylether wird umgedreht, der Unterschied der Partialladungen ist im MK-Modell zusätzlich um ein Vielfaches höher. In Reihe 12 wird bei den Molekülen 4-Methyl-1H-imidazol und 5-Methyl-1H-imidazol zweifach koordinierter Stickstoff positiver bewertet als dreifach koordinierter, was eigentlich umgekehrt sein sollte. Wieder ist die Bandbreite der Partialladungen viel geringer als beim MK-Modell. Der Vergleich der Carbonyl-Kohlenstoffe in Reihe 15 scheitert am zu geringen Unterschied der ermittelten Ladungen. Die Positivierung des Stickstoffs im Peptidmodell Me-L-Ser-Me (Reihe 16) durch eine intramolekulare Wasserstoffbrücke beeinflusst dessen Partialladung im NBO-Modell kaum und wird als Negativierung bewertet. Dasselbe Phänomen lässt sich bei der intramolekularen Wasserstoffbrücke in Reihe 17 identifizieren. In Reihe 18 ist praktisch keine Differenz der Partialladung der Methylgruppe zwischen Essigsäure und dem Acetat-Ion festzustellen, wenn man nur den Kohlenstoff betrachtet. Summiert man alle Atomladungen der Methylgruppe auf, erhält man, wie erwartet, eine Negativierung beim Acetat. Anstelle einer Negativierung am C3 im 4-Methyl-1H-imidazol-Kation (Reihe 20) gegenüber der Neutralverbindung ergibt sich beim NBO-Modell eine Positivierung.

7.2.8 Vergleich der Ladungsmodelle GM und ABR27

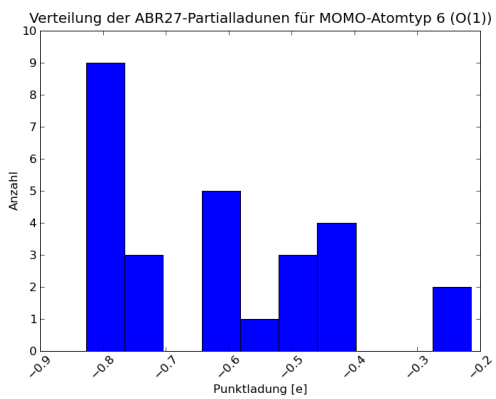
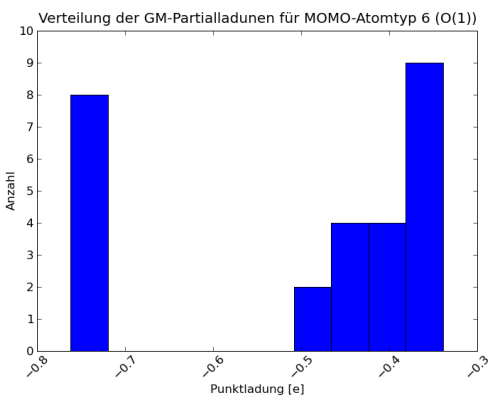
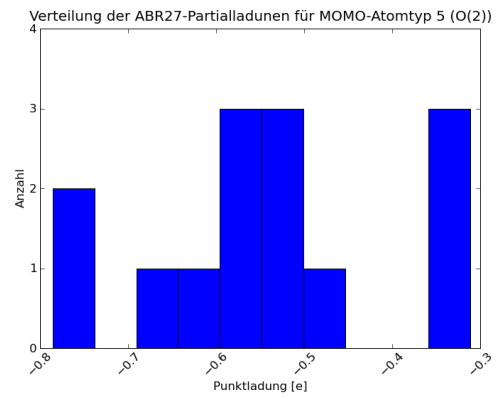
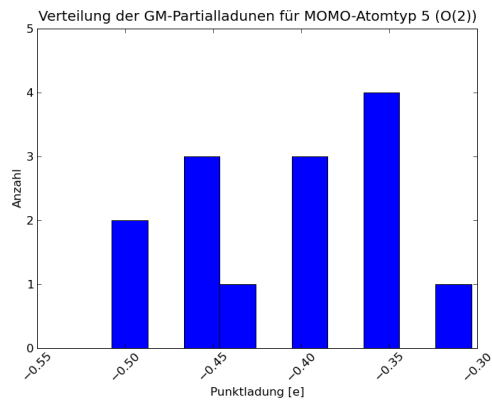
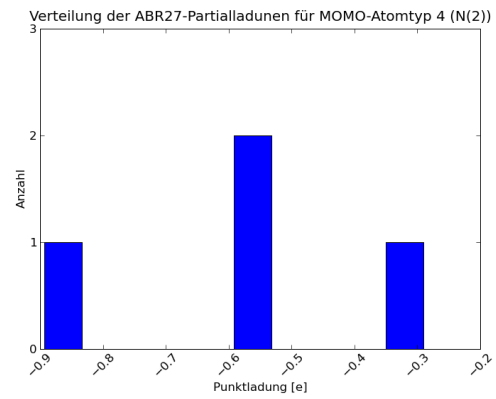
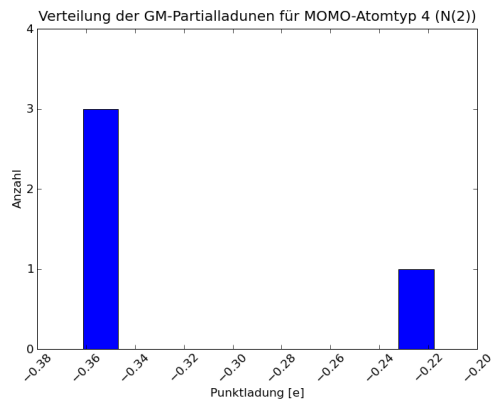
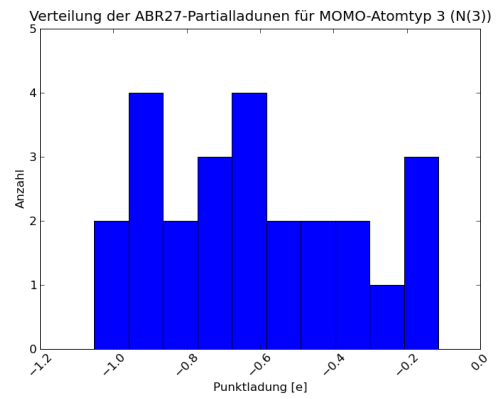
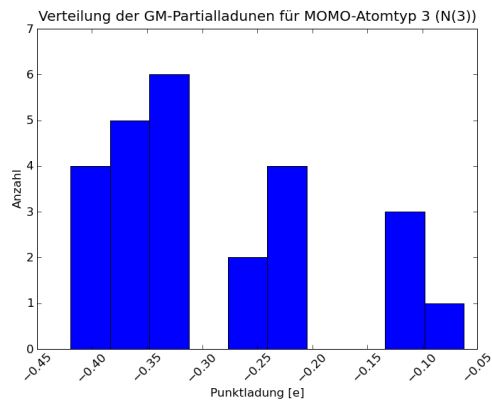
Als Nächstes werden die Ladungsmodelle GM und ABR27 mit ihren bisherigen Parametern untersucht. Interessant ist zunächst, inwieweit die beiden Ladungsmodelle die empirischen Regeln erfüllen können, bevor sie dann mit den quantenchemischen Verfahren verglichen werden.

In Reihe 5 werden die terminalen Methylgruppen durch GM fast gleich und annähernd neutral bewertet, die Einhaltung der Positivierung vom Propan zum Butan ist deshalb nur Zufall, ABR27 ermittelt fälschlicherweise eine Negativierung. Beim Übergang von Ethen zum isoelektronischen Formaldehyd (Reihe 7) sagen GM und ABR27 richtig die Positivierung des Kohlenstoffs voraus, die daraus folgende Negativierung der Wasserstoffatome durch den induktiven Effekt wird von GM

jedoch nicht erkannt und sogar als deutliche Positivierung bewertet. In Reihe 9 berechnen beide Verfahren für H1 praktisch identische Partiaalladungen, die Regel zur negativen Induktion wird daher nur zufällig erfüllt oder verletzt. Im Toluol reproduziert das GM-Modell nicht die Negativierung in Ortho- und Para-Stellung gegenüber der Meta-Position, da die Ladungen als praktisch identisch ermittelt werden (Reihe 11). Der Vergleich der Carbonylkohlenstoff-Partiaalladungen (Reihe 15) scheitert für das GM-Verfahren an den als identisch ermittelten Werten in Prolin und Essigsäure. Die Partiaalladung des Stickstoffs beim Übergang von Me-L-Ala-Me zu Me-L-Ser-Me in Reihe 16 wird durch die intramolekulare Wasserstoffbrücke praktisch nicht beeinflusst, entsprechend erfolgt bei diesem Modell auch keine Positivierung des Carbonyl-C3. Auch der induktive Effekt durch die Hydroxylgruppe am C1 wird vom GM-Verfahren falsch eingeschätzt. In Reihe 18 klassifiziert ABR27 die Positivierung des Carbonyl-Kohlenstoffs falsch. Anstelle einer Negativierung am C3 im 4-Methyl-1H-imidazol-Kation (Reihe 20) gegenüber der Neutralverbindung ergibt sich bei beiden Ladungsmodellen eine Positivierung.



7 Angewandter Teil



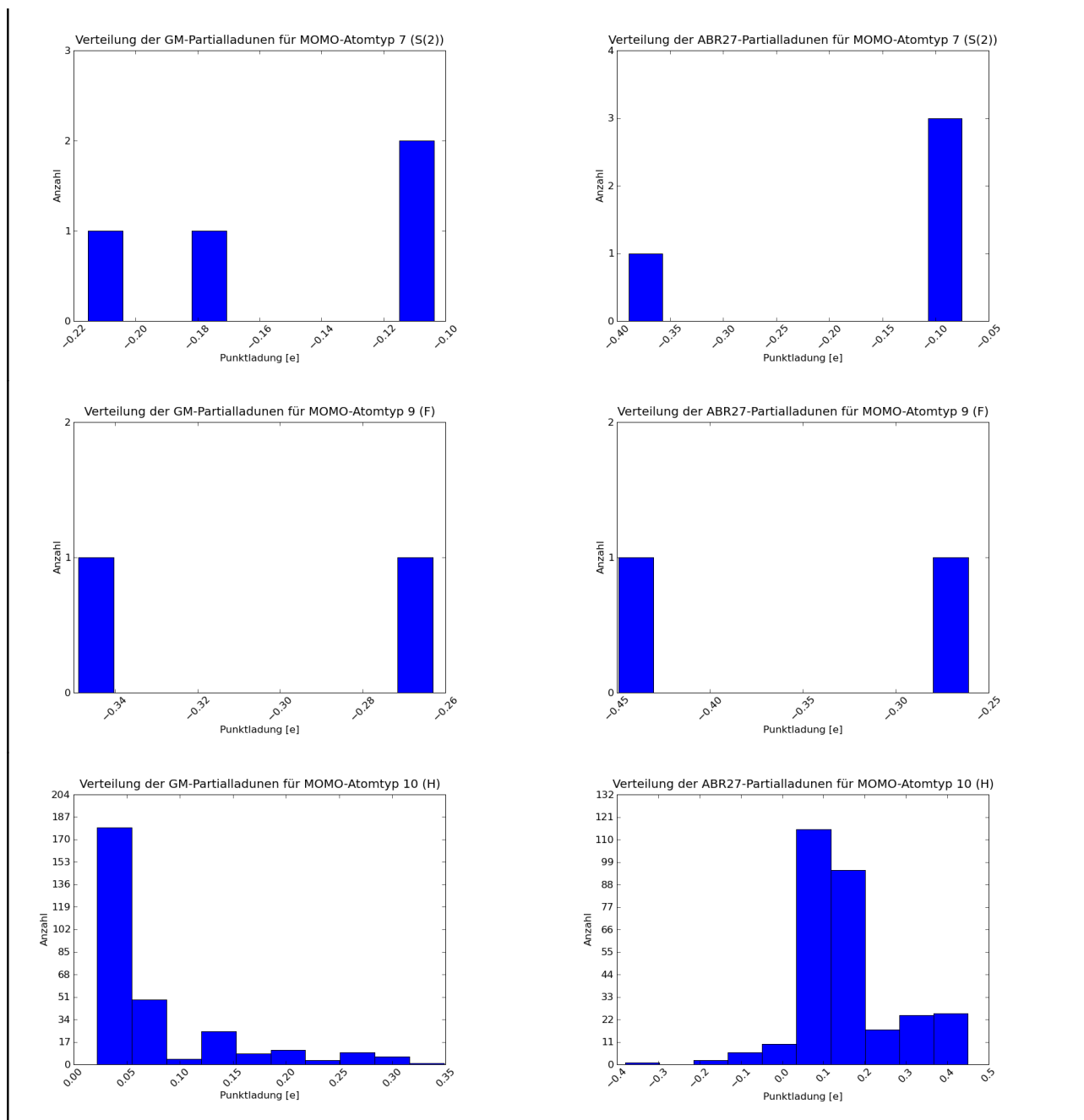


Tabelle 27: Verteilung der Atomladungen für die Modelle GM und ABR27 nach Atomtyp

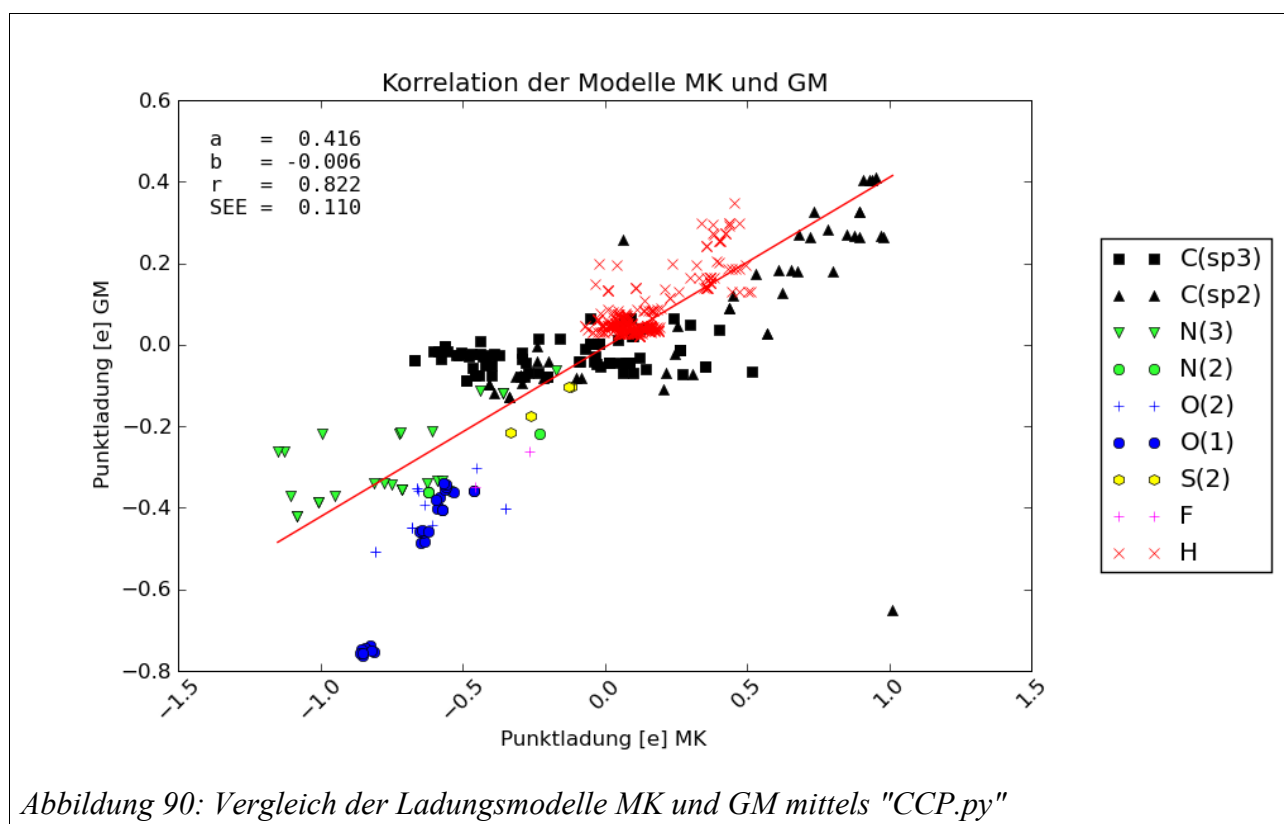
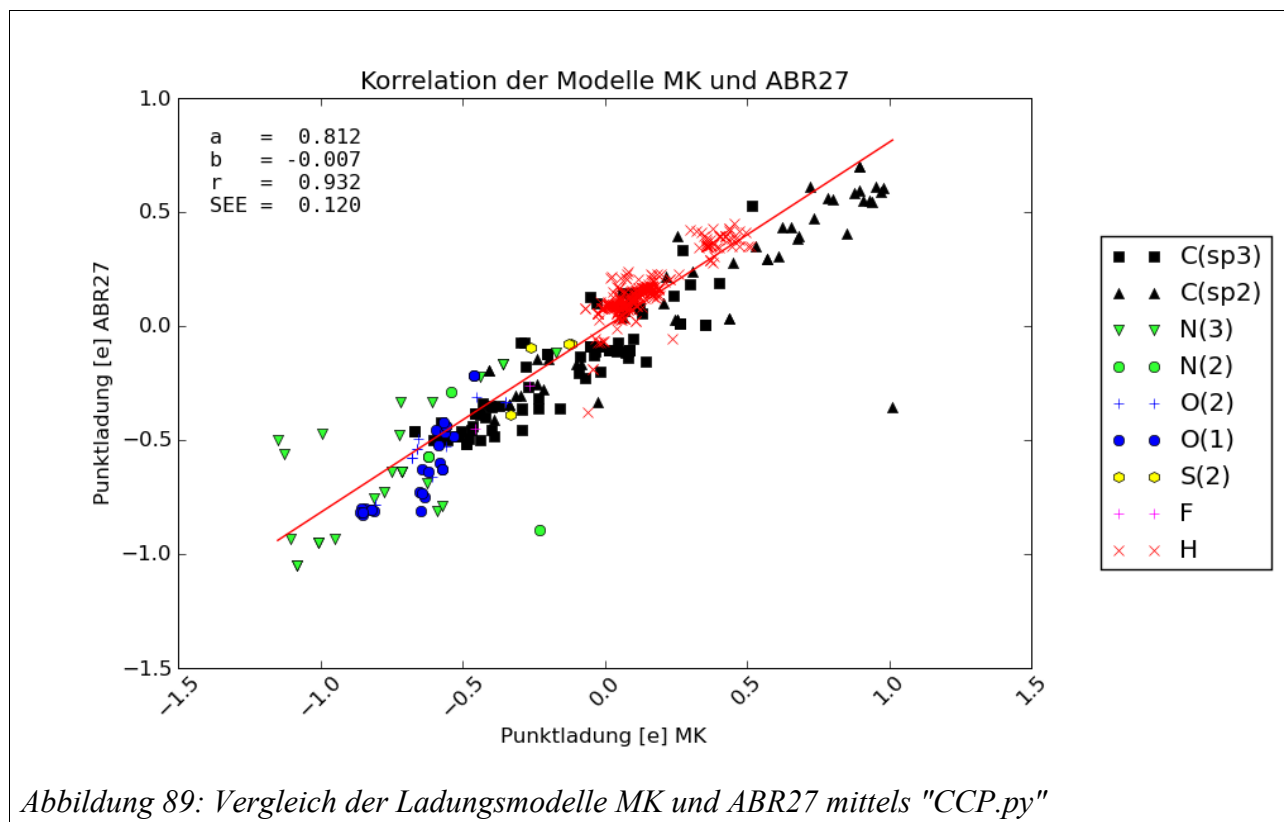
Bei Betrachtung der Ladungsverteilungen in Tabelle 27 fällt sofort der gegenüber den ABR27-Ladungen wesentlich kleinere durch die GM-Ladungen abgedeckte Wertebereich auf. In den Histogrammen zum GM-Modell treten bevorzugt ausgeprägte Peaks auf. Die ABR27-Ladungen decken einen breiten Bereich ab und neigen weniger zu scharf begrenzten Verteilungen. Sowohl ABR27- als auch die GM-Ladungen hängen von der Topologie des Moleküls ab. ABR27 hat allerdings einen von der Molekülgeometrie abhängigen Anteil, sodass die Unterschiede in der

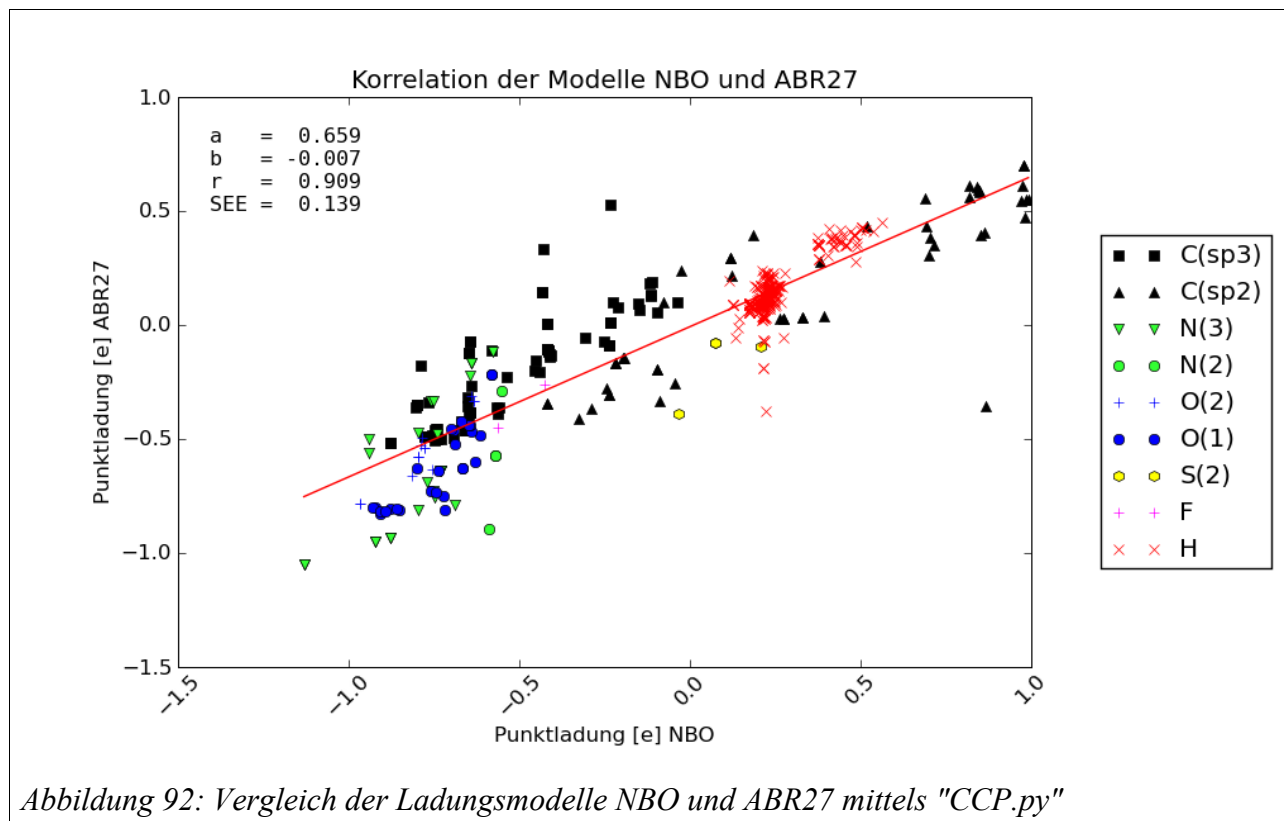
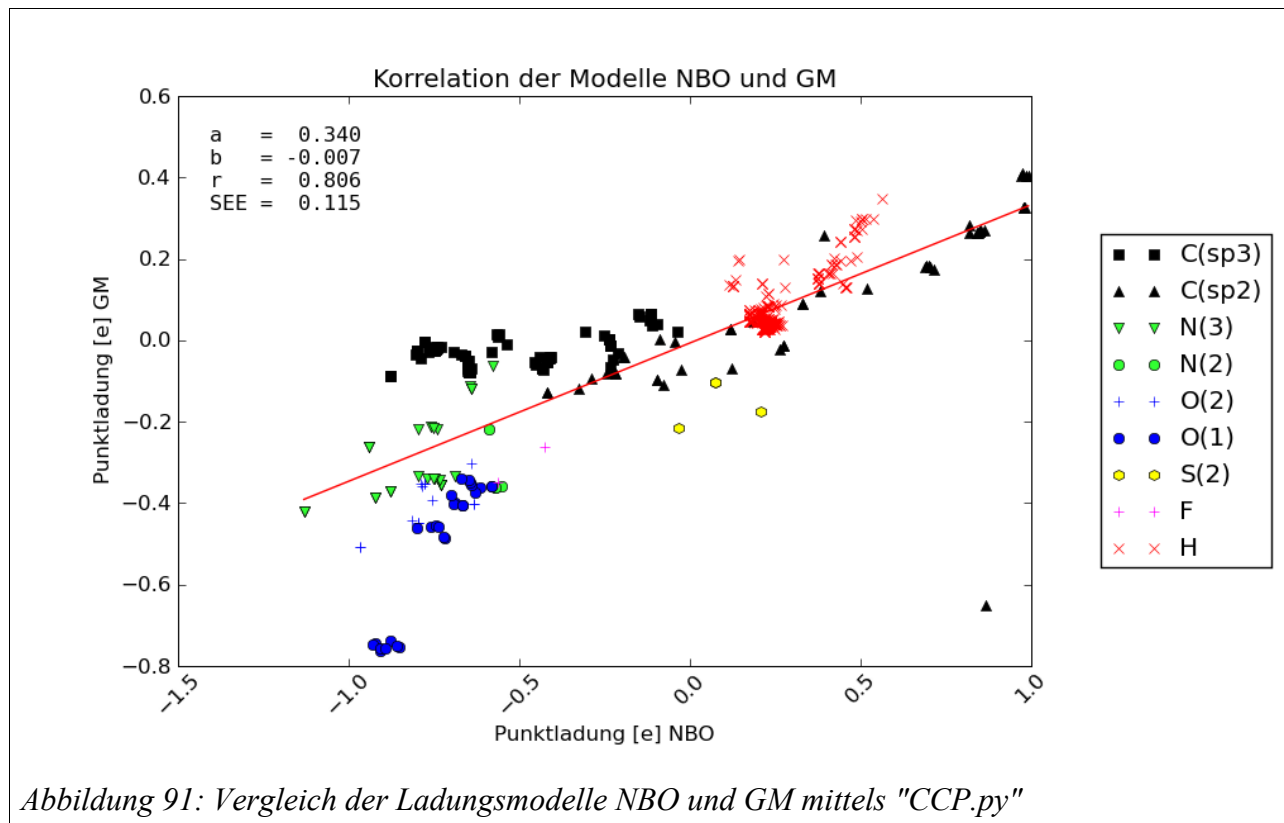
Gestalt der Ladungsverteilung ähnlich wie beim Vergleich zwischen MK und NBO damit erklärt werden können.

Fazit

Das MK-Verfahren eignet sich als Referenz für die nachfolgende Parametrisierung scheinbar am besten, da es alle empirischen Regeln erfüllt und damit vom Betrag als auch von der systematischen Entwicklung der Ladungen her den Erwartungen des Chemikers am besten entspricht. Da MK-Ladungen das elektrostatische Potenzial eines Moleküls modellieren, eignen sie sich gut zur Simulation intermolekularer Wechselwirkungen. Das NBO-Verfahren reproduziert nicht immer die systematische Entwicklung der Partiaalladung nach chemischen Gesichtspunkten, zeigt allerdings nicht die Konformationsabhängigkeit der MK-Ladungen. Vom Ansatz her sollten GM und NBO, sowie MK und ABR27 bei einem Vergleich am besten harmonieren. Abb. 89 zeigt die Korrelation der Modelle MK und ABR27. Tatsächlich harmoniert das ABR27-Modell am besten mit den MK-Ladungen, diese Kombination bietet die mit Abstand beste Übereinstimmung. Ausreißer sind das von den Stickstoffatomen mit dem Atomtyp N(3) benachbarte C4 im N-Propylguanidinium-Kation und das N1 im 4-Methyl-1H-imidazol-Kation.

Bei allen anderen Paarungen der Modelle ist die Korrelation deutlich schlechter. Beim Betrachten der Korrelationsdiagramme fallen sofort vertikale (Abb. 92) oder horizontale (Abb. 90, 91) „Streifen“ auf. Diese kommen durch die Paarung eines Modells mit breiter Ladungsverteilung (MK, ABR27) mit einem solchen mit scharfer Verteilung (NBO, GM) zustande. Die Ladungsverteilung des GM-Verfahrens ist schärfer als jene der NBO-Methode, sodass auch bei dieser Paarung (Abb. 91) „Streifen“ entstehen. Bei den Korrelationen mit den GM-Ladungen hat die Ausgleichsgerade eine besonders niedrige Steigung, hier sollte versucht werden, den Wertebereich durch eine Parametrisierung zu spreizen.





7.2.9 Die Parametrisierung der Punktladungsmodelle

Mit den zur Verfügung stehenden Referenzdaten wurde nun die Parametrisierung des Ladungsmodells von Gasteiger und Marsili und des Modells von Sabine Monz versucht. Dazu wurden verschiedene Parametrisierungsprogramme erstellt, die sich durch das verwendete Ladungsmodell und die benutzten Referenzdaten unterscheiden. Die Parametrisierung geschieht mittels eines evolutionsstrategischen Algorithmus. Ein Atomtyp kann in vielen verschiedenen chemischen Umgebungen auftreten, dennoch sind die verwendeten Parameter immer gleich. Je größer der zur Parametrisierung herangezogene Datensatz ist, desto größer ist auch die Chance, dass eine Vereinbarkeit der Parameter mit verschiedenen chemischen Umgebungen nicht gegeben ist. Ziel der Parametrisierung ist es, auch solche Fälle zu identifizieren.

Da die abgeleiteten empirischen Regeln immer die Verhältnisse in einer Testreihe beschreiben, wurde die Parametrisierung zunächst nur an den isolierten Reihen getestet. Um einen allgemeingültigen Parametersatz zu erhalten, wurden alle Moleküle aus dem Testdatensatz schließlich schrittweise zusammengeführt.

7.2.10 Parametrisierung des Gasteiger-Marsili-Modells

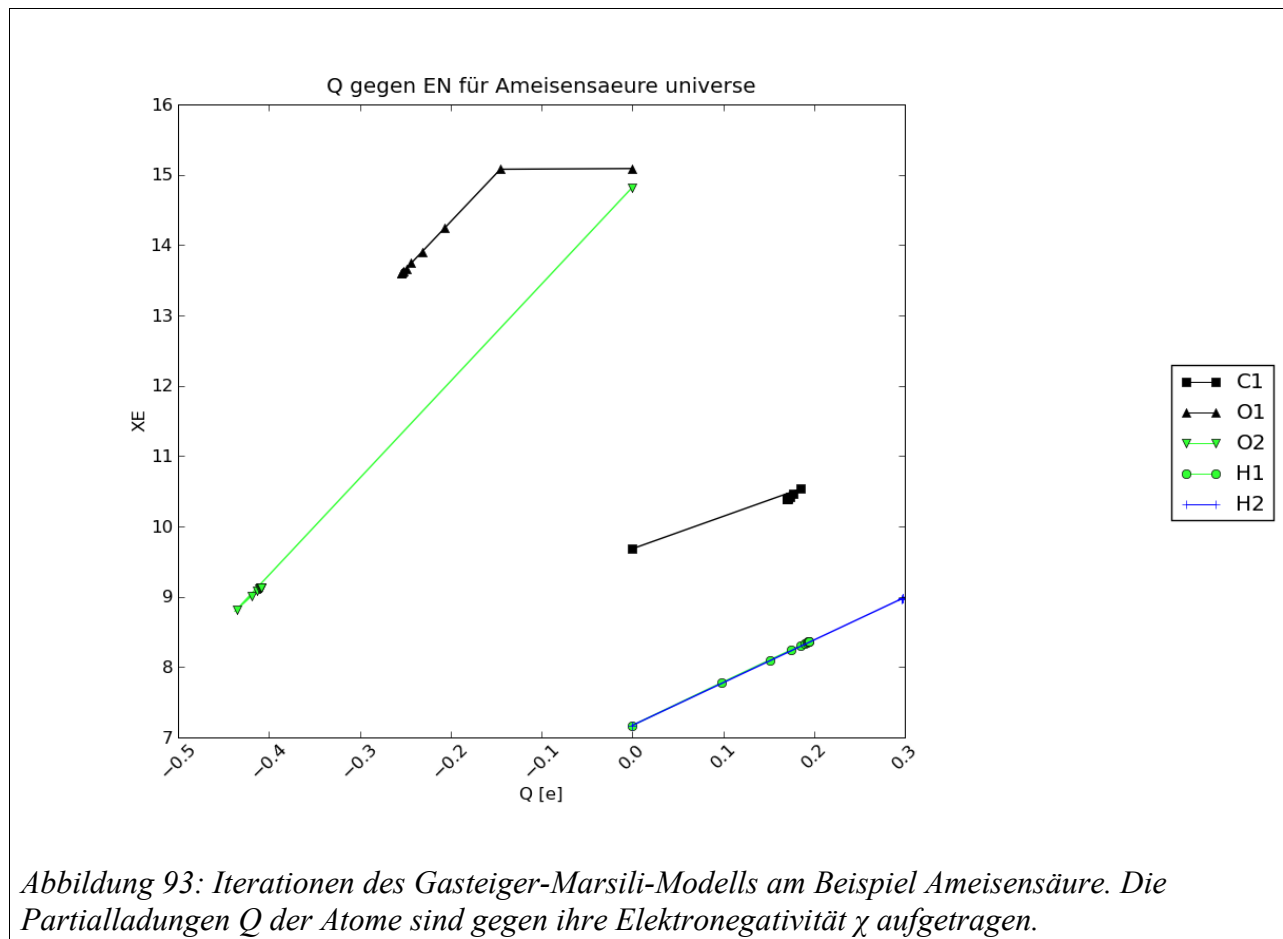
Die Justierung des Gasteiger-Marsili-Modells (Kapitel 3.3.2.1) erfolgt durch die Anpassung der Abhängigkeit der Elektronegativität von der Partialladung eines Atoms beschreibenden Polynom 2. Grades (7.2.1).

$$\chi_{Atom} = a + b \cdot q_{Atom} + c \cdot q_{Atom}^2 \quad (7.2.1)$$

Eine analytische Ableitung der berechneten Ladungen nach den Parametern lässt sich nicht angeben. Dies liegt unter anderem daran, dass das Ergebnis iterativ erhalten wird, π - und σ -Anteil der Partialladungen hängen voneinander ab und werden abwechselnd ermittelt. Zur Bestimmung der optimalen Parameter bleiben deshalb nur stochastische Suchverfahren oder bei wenigen Parametern auch die systematische Suche. Die Verwendung evolutionsstrategischer Algorithmen hat sich hierfür bewährt.

Im verwendeten Testdatensatz sind die acht Atomtypen C(sp³), C(sp²), N(3), O(2), O(1), S(2), F und H enthalten. Bei drei Parametern pro Atomtyp, für die z. B. 10 verschiedene Werte getestet würden, ergäben sich bereits 10²⁴ mögliche Kombinationen. Könnten drei Parametersätze pro Sekunde getestet werden, dauerte die Berechnung bereits über 10¹⁶ Jahre. Hier kommt nur noch der evolutionsstrategische Algorithmus in Frage. Diese Klasse von Algorithmen profitiert, wenn die

Suche in eine bestimmte Richtung gelenkt werden kann und wenn sich die Lösung aus unabhängigen Teilproblemen zusammensetzt. Um dies abzuschätzen, muss der Einfluss der Parameter auf das Endergebnis qualitativ betrachtet werden:



Zu Beginn der Berechnung haben alle Atome (in neutralen Molekülen) die Ladung Null. Der Ladungstransfer wird folglich durch den Achsenabschnitt a (Formel 7.2.1) bestimmt. Ab dem darauf folgenden Schritt werden die π -Ladungen mit in die Iteration einbezogen. Der weitere Ladungstransfer wird dann auch durch die Steigung b beeinflusst. Die Bedeutung des Terms 2. Ordnung (Parameter c) ist eher untergeordnet, dieser wird bei einigen alternativen Implementierungen der PEOE-Methode dementsprechend auch nicht verwendet. Abb. 93 demonstriert die Entwicklung der Partialladungen und Elektronegativitäten am Beispiel der Ameisensäure für alle Atome. Eine Spur repräsentiert immer die Entwicklung für ein Atom, jeder Punkt markiert eine Iteration. Die Ladungen stehen zunächst auf Null, für die π -Atome C1 und O1 ist ab der zweiten Iteration ein abweichender Verlauf, der sich als „Knick“ in der Kurve bemerkbar

macht, zu sehen. Dies hängt mit der erst hier einsetzenden π -SCF-Rechnung zusammen, diese benötigt ein Zwischenergebnis für die σ -Ladungen. Die mit der Zahl der Iterationen zunehmende Dämpfung der Ladungsübertragung bedingt, dass die Punkte, welche die Iterationen markieren, immer enger zusammenliegen. Betrachtet man den Ladungstransfer entlang der Bindung in einer binären Verbindung, so legen die Parameter a der Atomtypen die Richtung des Ladungstransfers fest, und die Werte für b modifizieren dessen Geschwindigkeit. Die Menge der übertragenen Ladung ist eine Kombination von beidem. Um hier auf eine bestimmte Ladung zu kommen, sind verschiedene Kombinationen der Parameter a und b möglich, die allerdings alle bestimmte Randbedingungen einhalten müssen. Bei allen Kombinationen muss die numerische Reihenfolge der Werte für a erhalten bleiben, da sonst der Ladungstransfer in die falsche Richtung lief. Die Funktionen für χ_{Atom} dürfen sich in dem von der Rechnung genutzten Bereich auf keinen Fall schneiden, da sonst ein Atom immer positiver sowie eines immer negativer und die Ladungsübertragung damit praktisch unendlich würde. Auch darf es nicht vorkommen, dass die Elektronegativität mit steigender Partialladung sinkt, hier würde ebenfalls eine unendliche Ladungsübertragung auftreten. Passende Werte lassen sich für den einfachen Fall einer binären Verbindung durch schrittweise Annäherung finden. Für komplexere Moleküle mit einem Netzwerk von Bindungen ist dies nicht möglich, da nicht eindeutig angegeben werden kann, wieviel Ladung von wo nach wo und über welchen Weg fließt, um die endgültige Ladungsverteilung zu erreichen. Wäre dies möglich, ließen sich die Parameter für Gleichung 7.2.1 durch Berechnung eines Ausgleichspolynoms durch die gewünschten Elektronegativitäten in Abhängigkeit von der Partialladung eines Atoms berechnen.

Im Rahmen dieser Arbeit wurden Ansätze getestet, die versuchen, die Parameter der Atomtypen im Testdatensatz schrittweise in der Art zu ändern, dass das Ergebnis der Ladungsberechnung in Richtung der Referenzladung geändert wird. Dazu wird bei der Berechnung der Testmoleküle für jeden Atomtyp Buch geführt, ob die Ladung eines Atoms positiver oder negativer ausfallen soll. Ladung und Elektronegativität wurden dabei für jede Iteration gespeichert und um einen kleinen Betrag in die gewünschte Richtung modifiziert. Die Parameter werden dann durch Berechnung eines Ausgleichspolynoms anhand der modifizierten Werte bestimmt. Dieser Ansatz funktioniert unter anderem deshalb nicht, weil die Ladung eines Atomtyps in verschiedenen chemischen Umgebungen einmal zu hoch und ein anderes Mal zu niedrig im Vergleich zur Referenz ermittelt wird. Das so ermittelte Ausgleichspolynom hat nach einigen Zyklen ein relatives Maximum oder Minimum, was bedeutet, dass die Elektronegativität z. B. bis zu einer bestimmten Ladung steigt und

dann wieder abfällt. Dieses Verhalten kann bei der Ladungsberechnung in einer unendlichen Ladungsübertragung enden. Ein zweiter Ansatz versucht, das Problem der konkurrierenden Anforderungen an den Elektronegativitätsverlauf für einen bestimmten Atomtyp zu lösen, indem versucht wird, die Nachbaratome mit einzubeziehen, um in Konflikt stehende Anforderungen aufzulösen. Eine konfliktfreie Lösung wird sehr schnell unmöglich, es muss also eine Abwägung erfolgen. Wegen der fehlenden Differenzierbarkeit der Ladungen nach den Parametern kann diese so nicht erfolgen. Eine Ermittlung, in welcher Richtung die Elektronegativität eines bestimmten Atomtyps geändert werden sollte, ist auf dieser Grundlage unmöglich. Es ist lediglich möglich, verschiedene Parameter zu testen und anhand der Rechenergebnisse zu beurteilen.

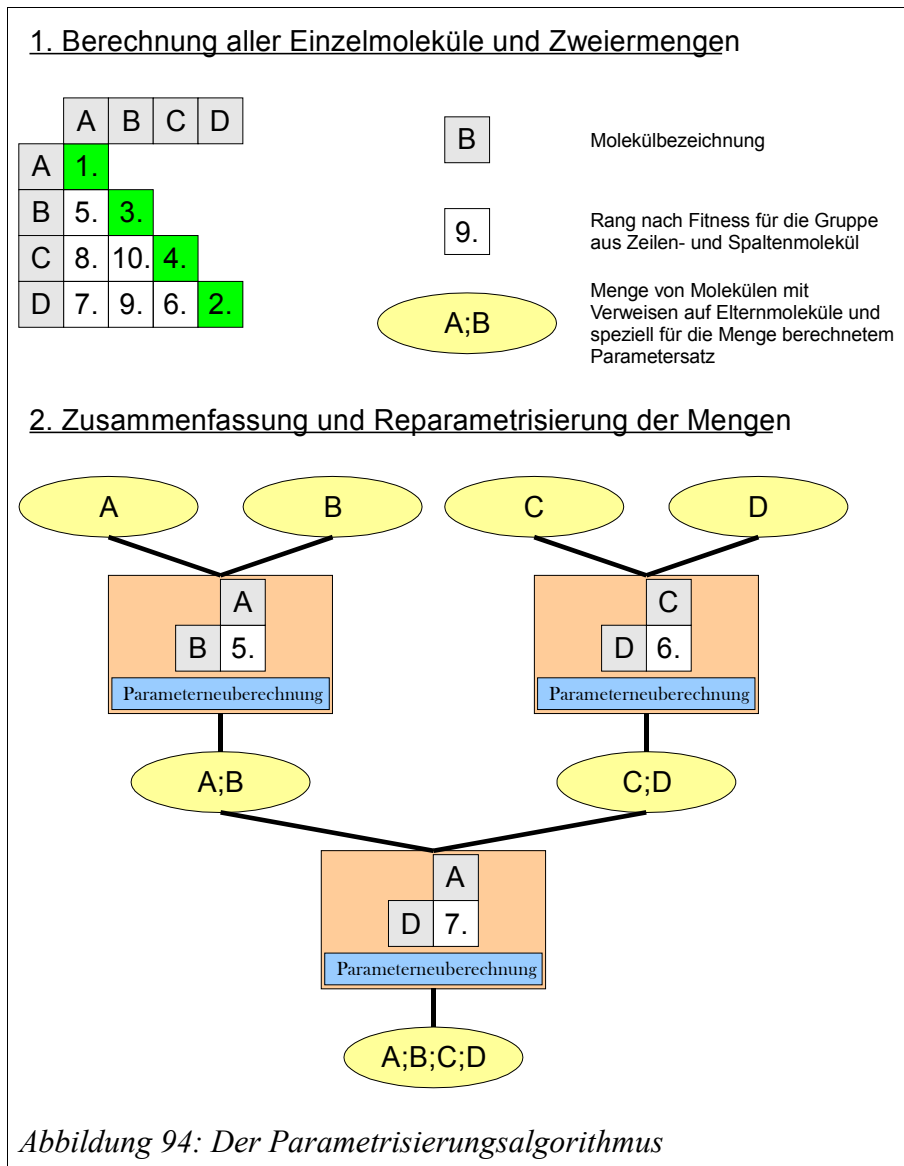
Für einen evolutionsstrategischen Algorithmus erscheint die Problemstellung gut geeignet. Ein Parametersatz für einen Atomtyp kann durch Abstimmung der Parameter für andere Atomtypen darauf optimiert werden. Das Prinzip der schrittweisen Verbesserung während der Evolution ist hier folglich gültig. Die Qualität der Anpassung eines Ladungsmodells an einen Referenzdatensatz lässt sich an

- einem möglichst geringen RMSD-Wert für die Ladungen,
- einer möglichst nahe bei 1.0 liegenden Steigung der Ausgleichsgeraden im Korrelationsdiagramm,
- einem nahe bei 1.0 liegenden R-Wert für die Ausgleichgerade im Korrelationsdiagramm,
- einer kleinen Standardabweichung des Schätzwertes bei der Korrelation mit den Referenzladungen,
- und einer weitgehenden Erfüllung der empirischen Regeln für die systematische Entwicklung der Partialladungen

erkennen.

Als Fitnesskriterium für den evolutionsstrategischen Algorithmus eignen sich besonders die numerischen Kenngrößen wie der RMSD-Wert, die empirischen Regeln können als letaler Faktor oder als „Fitnessstrafe“ bei verletzten Regeln berücksichtigt werden.

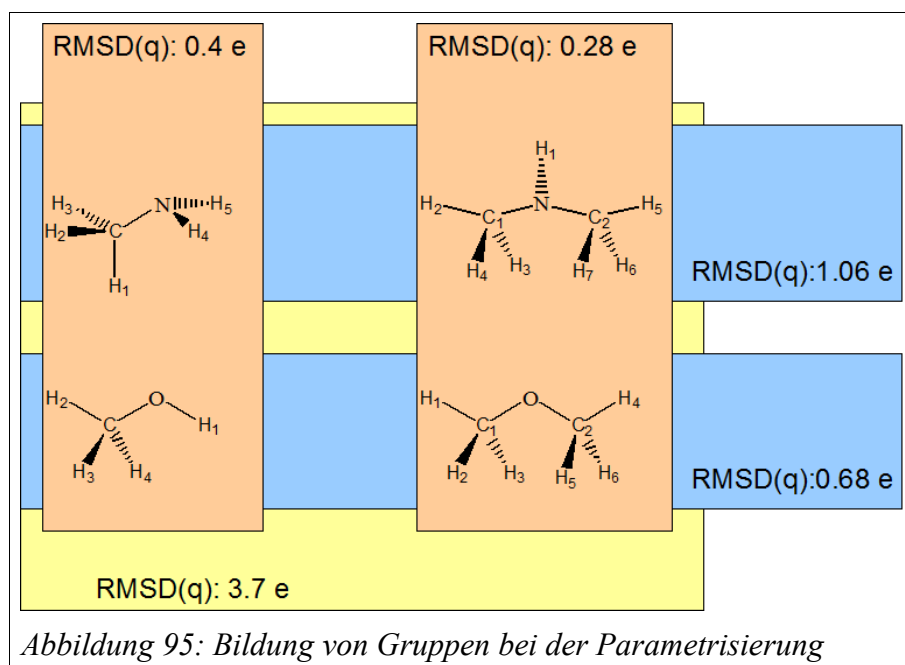
Um die Nutzung der empirischen Regeln zu testen, wurde die Parametrisierung zunächst auf die isolierten in Kapitel 7.2.4 beschriebenen Reihen beschränkt. Die dort definierten Vergleichsanweisungen dienten als letaler Faktor für einen Parametersatz.



Wenn in der Startpopulation keine Individuen vorhanden sind, die alle empirischen Regeln erfüllen, wurde diese Generation komplett gelöscht, denn es ist nicht gelungen, Parameter zu finden, die alle Regeln erfüllen, die in der Startgeneration nicht bereits erfüllt werden. Insgesamt ist kein signifikanter Einfluss der empirischen Regeln auf die ermittelten Parameter festzustellen. Die für die einzelnen Atomtypen in einer Reihe ermittelten Parameter haben sich als schlecht zwischen den Reihen übertragbar herausgestellt. Für die endgültige Parametrisierung der Ladungsmodelle wurde ein Ansatz gewählt, bei dem zunächst ein separater Parametersatz für jedes Molekül individuell berechnet wird. Im weiteren Verlauf werden die Moleküle dann zusammengefasst und für die so gebildeten Gruppen Parameter bestimmt. Abb. 94 illustriert den eingesetzten Algorithmus.

Schließlich sind alle Moleküle des Referenzdatensatzes in einer Gruppe enthalten und parametrisiert. Die Reihenfolge der Gruppierung lässt Rückschlüsse auf Schwierigkeiten bei der Parametrisierung zu. Der Algorithmus wird im Folgenden beschrieben:

1. Zunächst werden alle Moleküle aus den Referenzreihen eingelesen und Doubletten entfernt.
2. Alle Moleküle individuell und alle möglichen Kombinationen zweier Moleküle bilden eine Gruppe; alle Gruppen werden mithilfe des evolutionstrategischen Algorithmus parametrisiert. Dabei werden für jede Gruppe der beste Parametersatz und die damit erreichte Fitness vermerkt.
3. Die Gruppen werden nach ihrer Fitness aufsteigend sortiert.
4. Die nach Fitness sortierte Liste der Gruppen wird beginnend mit der fittesten Gruppe abgearbeitet. Wann immer eine Gruppe Elemente aus verschiedenen Gruppen enthält, werden alle diese Gruppen zu einer neuen Gruppe zusammengefasst und diese erneut mithilfe des evolutionstrategischen Algorithmus parametrisiert. Die Parametersätze der Elterngruppen werden in die Startpopulation für die Parameterberechnung der neuen Gruppe aufgenommen.



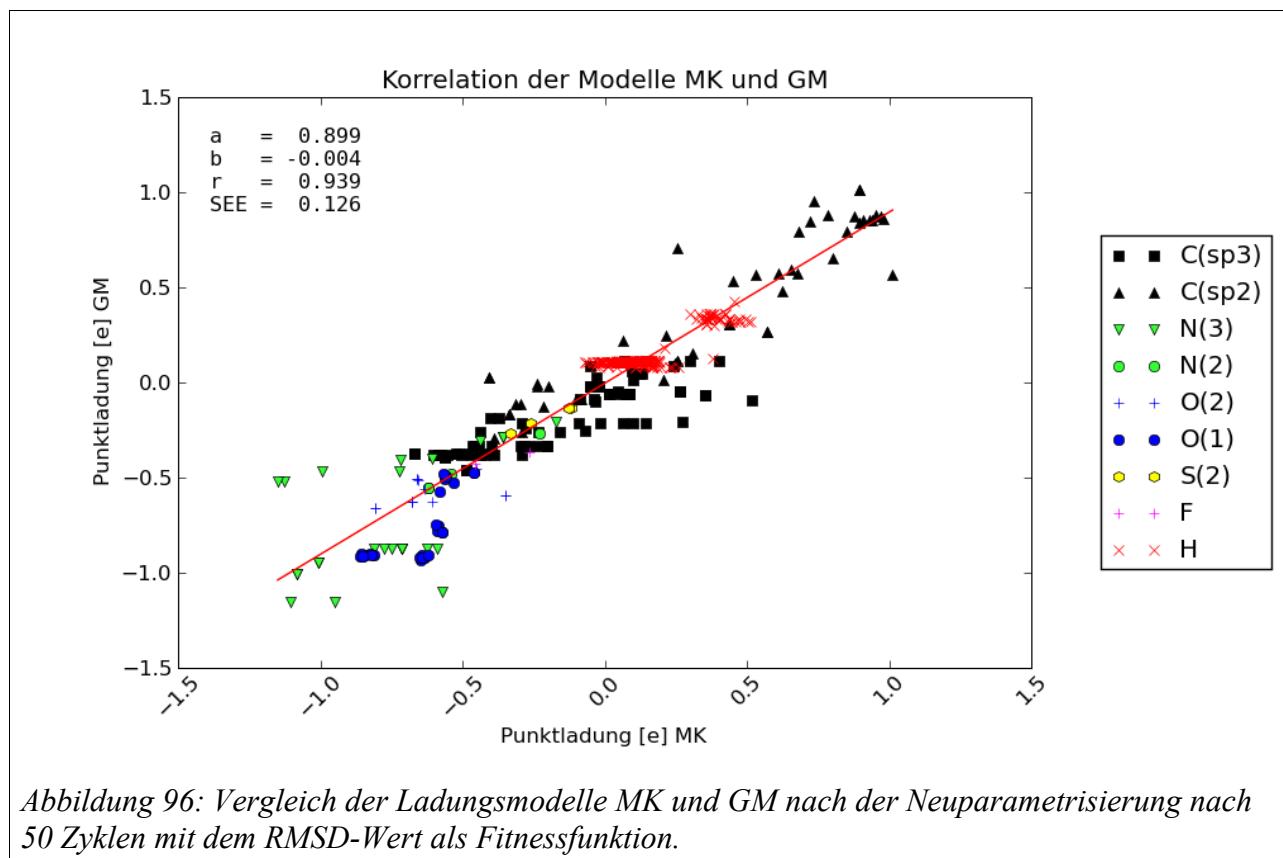
Auf diese Weise entsteht, ausgehend von den einzelnen Molekülen im Referenzdatensatz, eine Art Stammbaum der Parameter, mit dessen Hilfe sich Problemfälle eingrenzen lassen. Dies soll am

Beispiel des Datensatzes der Reihe 10 veranschaulicht werden. In Abb. 95 sind verschiedene Möglichkeiten der Gruppierung der vier Testmoleküle farblich hervorgehoben. Der im so markierten Feld angegebene RMSD-Wert ist der mit dem besten bei der Parametrisierung dieser Gruppe gefundenen Parametersatz erhaltene. Das Beispiel enthält die Verbindungen Methanol, Methylamin, Dimethylether und Dimethylamin. Bedeutsam für die Ladungsberechnung sind die freien Elektronenpaare an Stickstoff und Sauerstoff. Um die daraus resultierende hohe Ladungsdichte in einem Punktladungsmodell wiedergeben zu können, muss den betreffenden Atomtypen eine hohe Elektronegativität zugeordnet werden. Die elektronenziehende Wirkung der Methylgruppen wird dazu passend justiert. Werden Gruppen gebildet, die je eins der beiden Zentralatome enthalten und deren Moleküle ansonsten analog sind, gelingt dies sehr gut (die lachsfarben markierten Gruppen in Abb. 95). Werden nun Gruppen gebildet, bei denen das Zentralatom mit einer oder zwei Methylgruppen substituiert ist (blau markiert), kommt es bei der Einstellung der Elektronegativität der Methylgruppen zu einem Konflikt. Im Fall der einfachen Substitution ist die elektronenziehende Wirkung zu schwach und im Fall der zweifachen Substitution zu stark. Wenn neben dem unterschiedlichen Substitutionsgrad nun noch beide Zentralatome mit einbezogen werden (gelbe Markierung), kommt erschwerend hinzu, dass Sauerstoff zwei freie Elektronenpaare besitzt und Stickstoff nur eines. Die aus diesem Grund verschiedenen Elektronegativitäten müssen dann mit nur einem Parametersatz für die Methylgruppen ausgeglichen werden, was einen Kompromiss und damit einen höheren RMSD-Wert erzwingt.

Das Gasteiger-Marsili-Ladungsmodell wurde mit diesem Verfahren gegen die Referenzladungen jeweils nach dem Merz-Kollman-Verfahren und dem NBO-Verfahren parametrisiert.

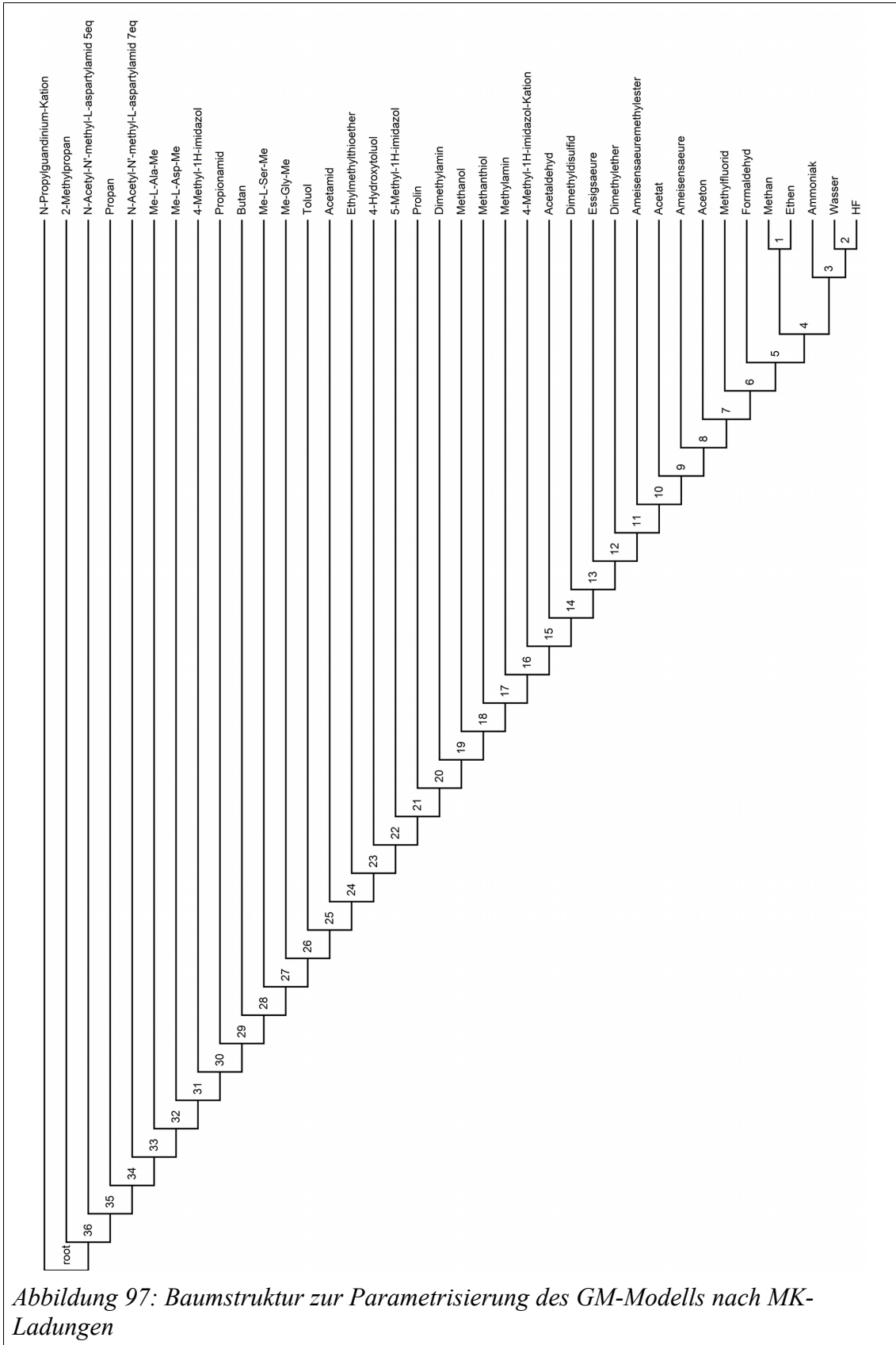
Merz-Kollman-Referenzladungen

Die durch die Neuparametrisierung erreichten Verbesserungen sind durch den Vergleich der beiden Korrelationsdiagramme (Abb. 90 und Abb. 96) sofort zu erkennen. Der R-Wert hat sich von 0.822 auf 0.939 verbessert, und die Steigung der Ausgleichsgeraden liegt mit 0.899 nun wesentlich näher an dem Idealwert 1.0. Die grundsätzlichen Probleme bei der Vereinbarkeit dieser beiden Ladungsmodelle bleiben dennoch bestehen und sind in Abb. 96 weiterhin in Form von horizontalen „Streifen“ zu erkennen. Das GM-Modell kann die breite Verteilung der MK-Ladungen immer noch nicht befriedigend wiedergeben. Besonders gravierend ist dies bei den Ladungen der Atomtypen Wasserstoff und sp^3 -Kohlenstoff. Vor allem jene empirischen Regeln, die auf dem Vergleich von

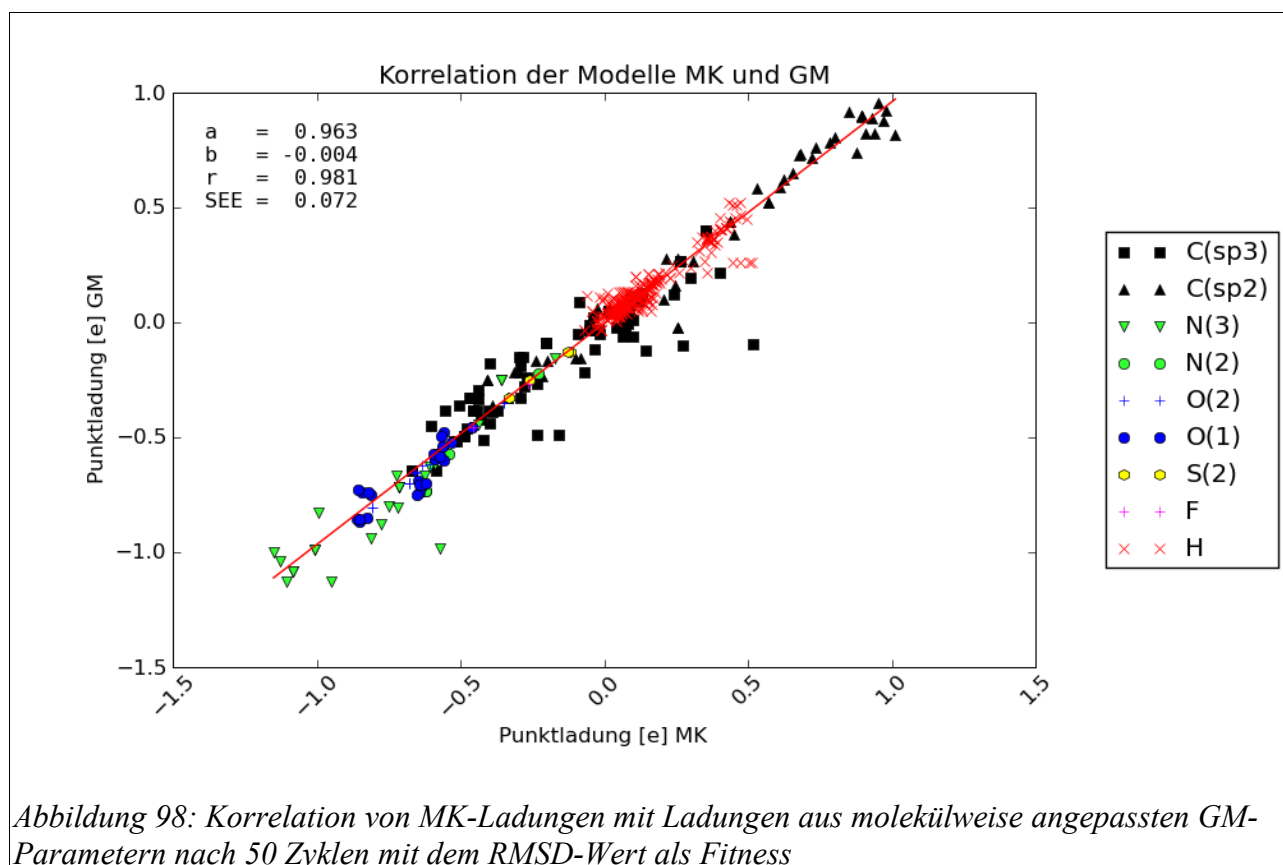


Wasserstoff-Partialladungen beruhen, sind hier oft nicht anwendbar, da die Unterschiede der zu vergleichenden Ladung im Bereich des Fehlers der berechneten Ladungen liegen. Wasserstoff und Kohlenstoff nehmen bei der Parametrisierung Fehler zugunsten anderer Atomtypen auf. Zur Ursachenforschung wurde eine speziell für die Nutzung der Daten aus dem Verlauf der Parametrisierung angepasste Version des Auswertungsprogrammes „CCP.py“ erstellt. Die in Abb. 94 gezeigte Datenstruktur enthält (als gelbe Ovale eingezeichnet) Mengen von gemeinsam parametrisierten Molekülen und den entsprechenden Parametersatz. Die neue Version ermöglicht den Zugriff auf eine solche Teilmenge anhand der in Abb. 97 angegebenen Knotennummern. Zusätzlich ist es möglich, die Ladungen für jedes Molekül mit individuellen Parametern zu bestimmen (mit den Daten aus den „Blättern“ der Baumstruktur), um prinzipielle Schwierigkeiten mit einzelnen Molekülen sichtbar zu machen und eine Art „Grenzwert“ für den möglichen Erfolg der Parametrisierung zu erhalten.

Zunächst wird daher die Korrelation mit den individuellen Parametern getestet. Das Ergebnis überrascht, da die Anpassung der Ladungsmodelle offenbar schlechter, oder je nach betrachteten statistischen Parametern zumindest kaum besser, gelingt, wenn jedes Molekül einzeln parametrisiert



wird. Eigentlich sollten dies die idealen Voraussetzungen für eine Korrelation sein. Abb. 98 zeigt die Grafik zu dieser Anpassung. Das Problem der zusammenschrumpfenden Wertebereiche für einige Atomtypen tritt hier weniger auf, die Korrelation könnte allerdings besser sein. Eine mögliche Erklärung wäre, dass der evolutionsstrategische Algorithmus mehr Zyklen benötigt, um seine Lösung zu verbessern, die zusammenschrumpfenden Wertebereiche weisen auf ein unpassendes Fitnesskriterium für die Optimierung hin.



Die Fitness für einen Parametersatz ergibt sich aus dem RMSD-Wert der berechneten Ladung im Vergleich zu den Referenzladungen. Atomtypen mit vergleichsweise kleinen Partialladungen tragen nur wenig zum RMSD-Wert bei; dass die Abweichungen quadriert werden, verschärft das Problem noch. Dies trifft besonders auf den Wasserstoff-Atomtyp zu.

Um eine Verbesserung der Parametrisierung zu erhalten, wurden eine Erhöhung der Zykenzahl und ein alternativer von Thomas Kleinöder [Kleinöder2005] vorgeschlagener Fitnesswert (Formel 7.2.2) getestet. Bei der Parametrisierung der individuellen Moleküle gibt es bei einigen davon, die nur aus wenigen Atomen bestehen (z. B. HF, Wasser), Probleme, da hier die in die Fitness nach

7.2.2 einfließende lineare Regression nicht sinnvoll berechnet werden kann oder sogar ungültige Werte liefert.

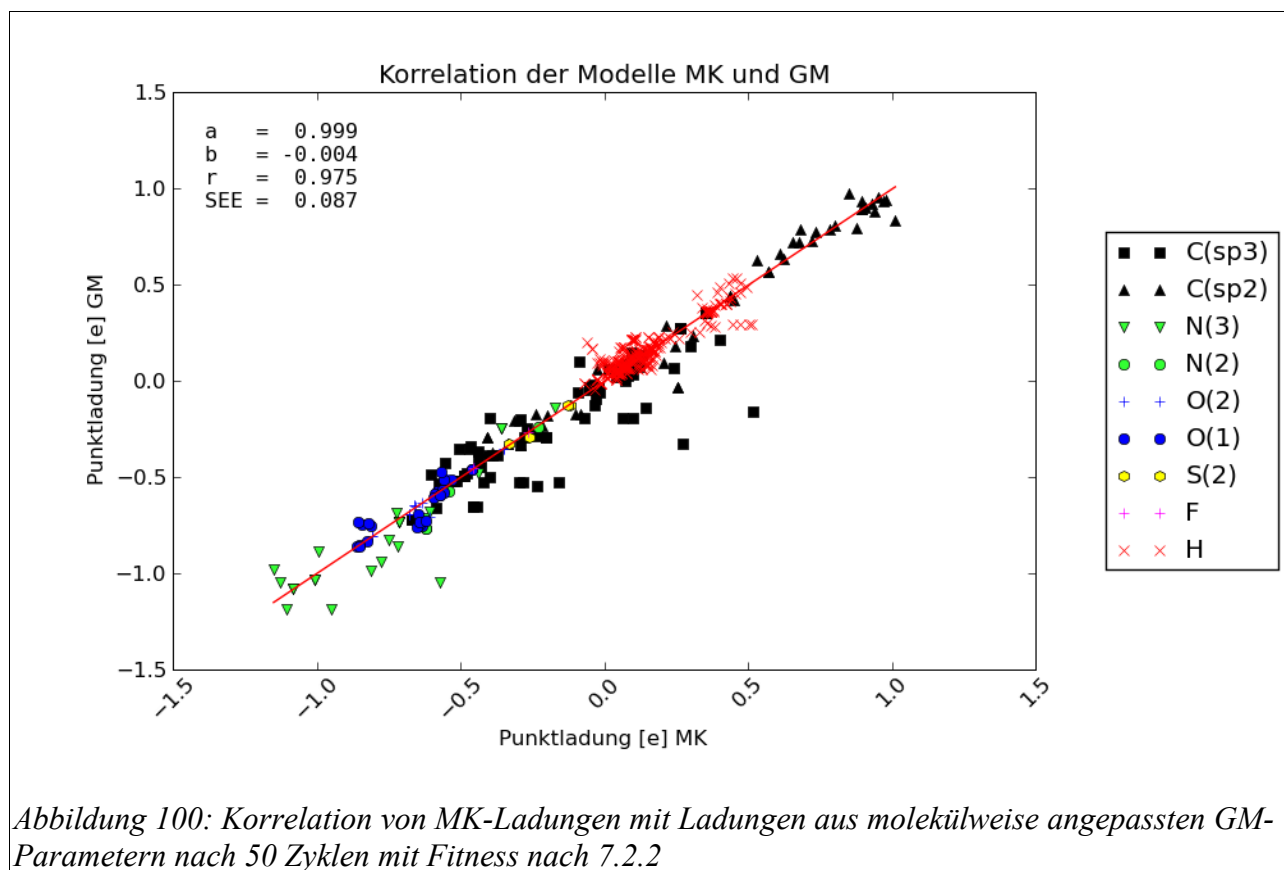
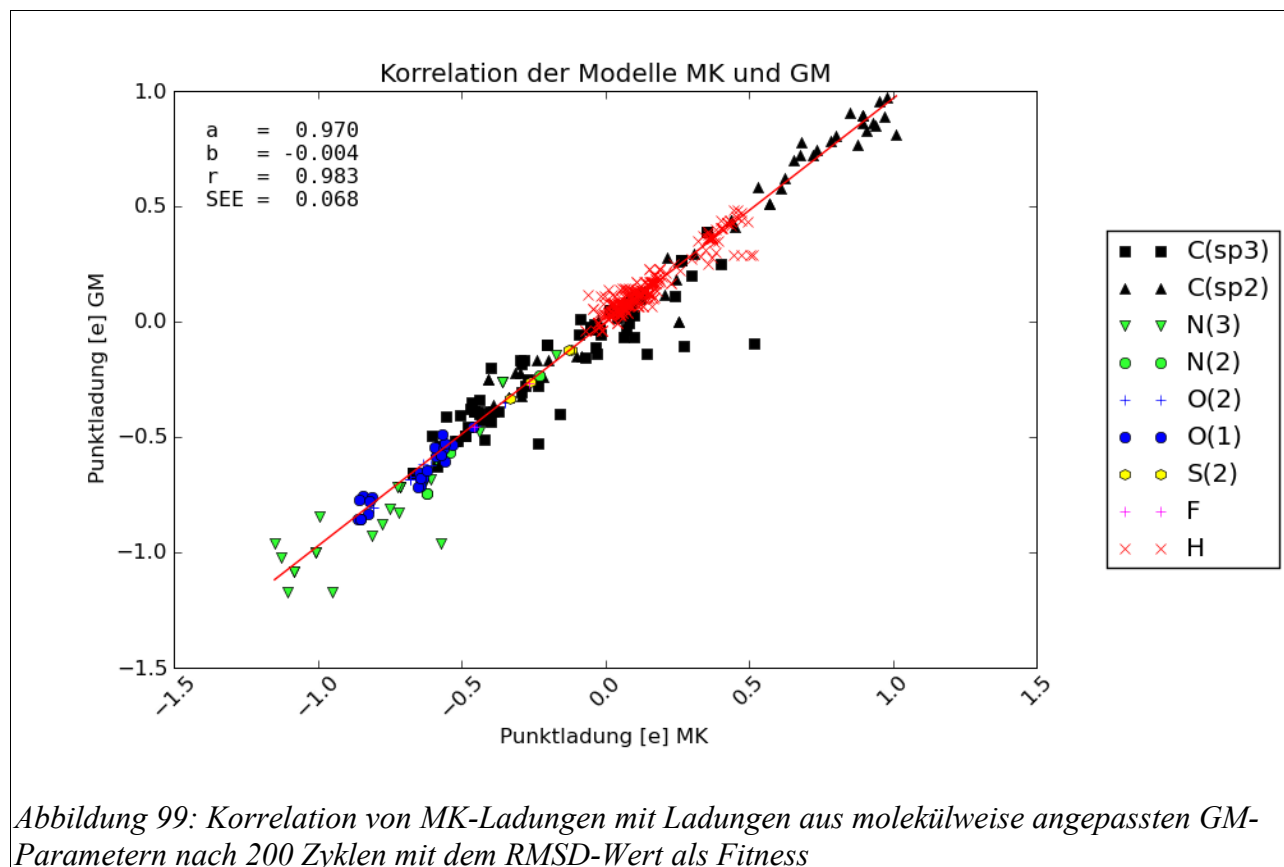
$$Fitness = RMSD_{Ladung} + (1 - r^2) + |1 - a| + \sigma + b$$

$RMSD_{Ladung}$	RMSD Ladungsmodell/Referenz
r	Korrelationskoeffizient
a	Steigung der Ausgleichsgerade
σ	Standardabweichung des Schätzwertes
b	Achsenabschnitt der Ausgleichsgerade

Formel 7.2.2: Fitnessfunktion nach Thomas Kleinöder

Die ungültigen Werte werden lediglich markiert („not a number“) und nicht als Fehler vom Rechner signalisiert. Daher müssen sie gezielt abgefragt werden. Geschieht dies nicht, werden solche Fitnesswerte falsch sortiert, da ihr Verhalten in Verbindung mit Vergleichsoperatoren dafür sorgt, dass ein ungültiger Wert an die Spitzenposition rückt [Goldberg1991]. In diesem Fall wird auf den RMSD-Wert als Fitness zurückgegriffen und eine „Fitnessstrafe“ addiert.

Die Erhöhung der Zyklenzahl von 50 (Abb. 98) auf 200 (Abb. 99) verbessert das Ergebnis nur wenig, sodass eine Berechnung mit 50 Zyklen als ausreichend erachtet wird. Der Wechsel des Fitnesskriteriums gemäß Formel 7.2.2 bringt, wie in Abb. 100 zu erkennen, eine Verbesserung der Steigung auf Kosten der Güteparameter aus der linearen Regression. Da diese Berechnungen quasi den Grenzwert der mit den Ladungsmodellen erzielbaren Übereinstimmung darstellen, interessieren hier Details im Hinblick auf Probleme bei einzelnen Molekülen und den empirischen Regeln. Besonders auffallend sind die offenbar großen Probleme bei der Parametrisierung von Kohlenwasserstoffketten. Bei Molekülen, die nur aus Kohlenstoff und Wasserstoff bestehen (Reihe 5), kann die Verteilung der MK-Ladungen schlicht nicht vom GM-Modell wiedergegeben werden. Es ist prinzipiell nicht möglich, die Konformationsabhängigkeit der MK-Ladungen abzubilden. Schon beim Übergang von Methan zu Propan können bereits keine Parameter mehr gefunden werden, die die Ladungsübertragung zwischen Wasserstoff und Kohlenstoff wiedergeben und gleichzeitig dafür sorgen, dass die Partialladungen der Kohlenstoffatome C1 und C3 so nahe beieinanderliegen, wie das MK-Modell dies angibt.



Gute Übereinstimmungen werden nur dann erzielt, wenn die Kohlenstoff- und Wasserstoffatome nicht Teil von Kohlenwasserstoffketten sind. Auch innerhalb größerer Moleküle ist andernfalls sofort eine Verschlechterung der Übereinstimmung zu beobachten. Beispiele dafür sind beim Vergleich zwischen Acetamid und Propionamid (Reihe 8), Dimethyldisulfid und Ethylmethylthioether (Reihe 14) und beim N-Propylguanidinium-Kation (Reihe 19) zu finden. Interessant ist die Verbesserung der Übereinstimmung in Reihe 11 beim Vergleich zwischen Toluol und 4-Hydroxytoluol. Die zusätzliche Hydroxygruppe mit dem neuen Atomtyp Sauerstoff entlastet den Parametrisierungsalgorithmus offenbar erheblich. An dieser Stelle sei daran erinnert, dass hier jedes Molekül seinen eigenen Parametersatz erhält, um prinzipielle Grenzen der Methode zu testen. Auch die Ladungen anderer funktioneller Gruppen innerhalb von Kohlenwasserstoffketten werden schlechter reproduziert. Empirische Regeln, die nicht mit der beschriebenen Problematik in Konflikt geraten, werden erfüllt.

Bezieht man die Regressionsparameter nach Formel 7.2.2 mit ein, ändert sich die Situation kaum. Es herrscht eine größere Tendenz, Parameter von selten vorhandenen Atomtypen zugunsten der Parameter häufig vorkommender zu opfern, was sich in einer Verschlechterung der Gütekriterien der linearen Regression und einer dafür näher bei eins liegenden Steigung zeigt. So verbessert sich die Wiedergabe der Wasserstoffpartialladungen z. B. etwas auf Kosten des Kohlenstoffs und des Stickstoffs. Diese Änderungen können die grundsätzlichen Probleme mit Kohlenwasserstoffketten jedoch nicht lösen.

Neben statistischen Gütekriterien kann auch die Zahl der verletzten empirischen Regeln in die Fitnessfunktion integriert werden. Aus Formel 7.2.2 wird dann der Ausdruck 7.2.3.

$$Fitness = RMSD_{Ladung} + (1 - r^2) + |1 - a| + \sigma + b + \left(1 - \frac{\text{erfüllte Regeln}}{\text{Regeln gesamt}}\right) \quad 7.2.3$$

Eine Bevorzugung von Lösungen mit möglichst vielen erfüllten empirischen Regeln bringt keine Verbesserungen, da in diesen Beispielen nur wenige der Regeln verletzt werden. Weil die Übertretungen eine Folge der besprochenen Problematik sind, ist es nicht möglich, hier überhaupt akzeptable Lösungen anzugeben.

Bei der Zusammenfassung der Parametersätze der Einzelmoleküle zu einem allgemeinen Parametersatz erhöht sich die Anzahl der einzugehenden Kompromisse zwangsläufig. Es ist zu erwarten, dass dies bei den Atomtypen Wasserstoff und Kohlenstoff am stärksten ausgeprägt ist,

dementsprechend sind die reinen Kohlenwasserstoffe und Moleküle mit entsprechenden Ketten weiter außen im Baumdiagramm (Abb. 97) zu finden.

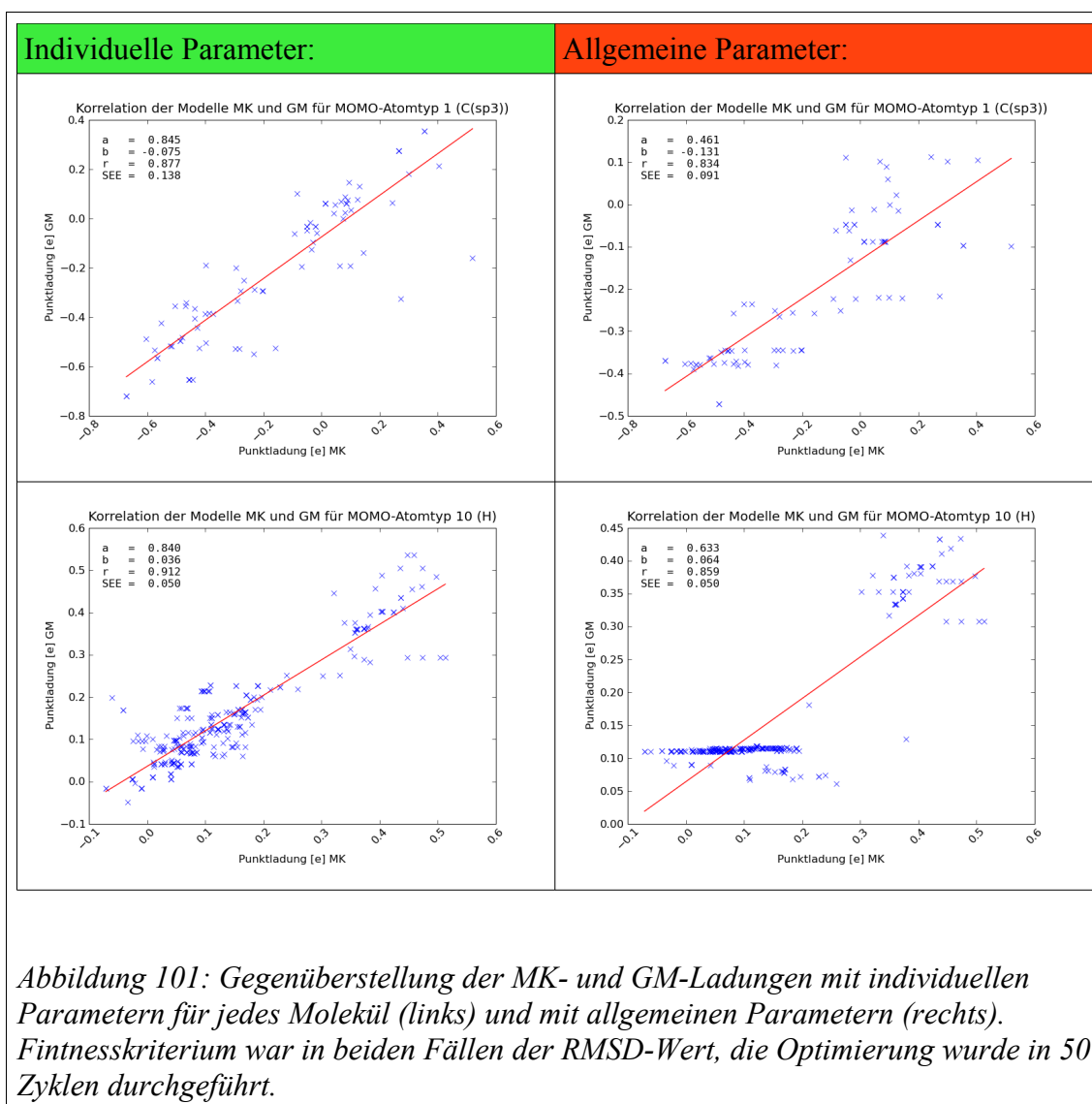


Abb. 101 skizziert den Effekt anschaulich für die Atomtypen Kohlenstoff (sp³) und Wasserstoff. Die Konzentration des GM-Modells auf wenige verschiedene Ladungen ergibt sich daraus, dass die Ladungen für Atome in ähnlichen chemischen Umgebungen in einem auf Elektronegativitäten beruhenden Ladungsmodell wie GM auch ähnliche Werte haben müssen; in Tabelle 28 sind die Wertebereiche für sp³-Kohlenstoff nach Nachbaratomen vergleichend aufgeführt. Hier ist deutlich zu erkennen, dass die Mittelwerte für gleiche Nachbarn für Referenzdaten (MK),

molekülspezifische Parameter und allgemeine Parameter eine akzeptable Übereinstimmung aufweisen, die Verteilung der Ladungen sich jedoch immer mehr auf den Mittelwert konzentriert.

Anzahl	Ladungsmodelle: Nachbarn	MK [me]			GM molekülweise [me]			GM [me]		
		Min.	Max.	Mittelw.	Min.	Max.	Mittelw.	Min.	Max.	Mittelw.
19	C(sp ²); H; H; H;	-671	-290	-507	-720	-334	-500	-389	-350	-374
14	N(3); H; H; H;	-85	354	83	-17	355	93	-96	-61	-85
11	C(sp ³); H; H; H;	-466	-201	-336	-653	-190	-425	-346	-344	-345
8	O(2); H; H; H;	-50	265	88	-31	275	95	-48	-11	-43
6	C(sp ³); C(sp ²); N(3); H;	-51	404	176	-48	212	92	89	113	104
6	C(sp ³); C(sp ³); H; H;	-92	273	78	-324	-60	-161	-224	-217	-221
4	S(2); H; H; H;	-400	-278	-336	-386	-200	-316	-265	-236	-247
4	C(sp ³); C(sp ²); H; H;	-437	-68	-224	-548	-195	-408	-258	-251	-255
2	H; H; H; H;	-487	-487	-487	-495	-495	-495	-472	-472	-472
2	C(sp ³); N(3); H; H;	-30	123	47	-96	77	-10	-12	22	5
1	C(sp ³); O(2); H; H;	94	94	94	147	147	147	61	61	61
1	C(sp ²); N(3); H; H;	101	101	101	35	35	35	-1	-1	-1
1	C(sp ³); C(sp ³); C(sp ³); H;	520	520	520	-161	-161	-161	-98	-98	-98
1	C(sp ³); S(2); H; H;	-35	-35	-35	-125	-125	-125	-131	-131	-131
1	F; H; H; H;	130	130	130	130	130	130	-14	-14	-14

Tabelle 28: Vergleich der Ladungsverteilungen für sp³-Kohlenstoff nach Nachbaratomen, Ladungsmodell und Parametersatz. Fitnesskriterium war in allen Fällen der RMSD-Wert, die Optimierung wurde in 50 Zyklen durchgeführt. Alle Ladungen in Millielementarladungen.

Die Parameter der übrigen Atomtypen lassen diese Tendenz ebenfalls erkennen, allerdings kommen sie in weniger vielen verschiedenen chemischen Umgebungen vor und umfassen wesentlich weniger Datenpunkte. Die Optimierung alleine anhand des RMSD-Wertes verstärkt die Tendenz des Ladungsmodells, sich auf die Mittelwerte zu konzentrieren; nimmt man Kriterien wie in Formel 7.2.2 und 7.2.3 hinzu, verschlechtern sich zwar Kenngrößen wie der RMSD-Wert, der Korrelationskoeffizient und die Standardabweichung des Schätzwertes, aber die als Skalierungsfaktor zu deutende Steigung a der Ausgleichsgerade liegt näher an ihrem Idealwert eins.

Die Wirkung der verschiedenen Fitnessfunktionen kann anschaulich in einem frühen Stadium des Parametrisierungsvorgangs untersucht werden. Je nachdem, mit welcher Formel für die Fitness gerechnet wird, ändert sich die Reihenfolge der bei der Optimierung zusammengefassten Molekülmengen etwas; es lassen sich aber kleine Gruppen bilden, deren Reihenfolge erhalten bleibt, auch wenn die Mengen innerhalb der Gruppen ihre Positionen ändern können. Erwartungsgemäß wählt der Algorithmus zu Beginn kleine Moleküle aus, deren Parameter sich nicht gegenseitig stören. Dies wäre für den verwendeten Datensatz die Menge (Ammoniak, Methan,

Wasser, Fluorwasserstoff). Hier ist jeweils Wasserstoff um ein Zentralatom herum angeordnet. Der Optimierungsalgorithmus braucht hier nur einen Wert für die Elektronegativität des Wasserstoffs festzulegen und dann jene des Zentralatoms daran anzupassen. Die Ladungen können in diesen Fällen perfekt wiedergegeben werden. Fügt man diesem Datensatz nun die Menge (Ethen, Formaldehyd, Methylfluorid) hinzu (entspricht im Baumdiagramm für alle drei Fitnessfunktionen dem Knoten Nr. 6), können Schwächen bei der Vereinbarkeit der beiden Ladungsmodelle sichtbar werden. Was bei der Parametrisierung im Konfliktfall bevorzugt wird, entscheiden die Fitnessfunktionen. Besonders problematisch sind in diesem Beispiel Fluorwasserstoff und Methylfluorid. Die Referenzladung des Fluors unterscheidet sich in beiden Molekülen stark. Legt man eine für Fluorwasserstoff passende Elektronegativität für Fluor fest, ergibt sich im Methylfluorid eine viel zu starke Negativierung. Die elektronenziehende Wirkung der Methylgruppe kann jedoch nicht darauf eingestellt werden, da diese dann nicht mehr mit den Bindungspartnern von Kohlenstoff und Wasserstoff in den übrigen Molekülen harmoniert. Es muss also ein Kompromiss gefunden werden, hier ist die gewählte Fitnessfunktion ausschlaggebend. Zur Wahl stehen der RMSD-Wert sowie Formel 7.2.2 und 7.2.3.

Der RMSD-Wert als Fitnessfunktion löst den geschilderten Konflikt, indem die elektronenziehende Wirkung der Methylgruppe durch Anhebung der Elektronegativitäten von Kohlenstoff und Wasserstoff und Absenken der von Fluor vermindert wird. Diese Änderungen fallen nur gering aus, da sich die daraus resultierenden Wirkungen sofort auf die übrigen Moleküle des Datensatzes übertragen. Kohlenstoff und Wasserstoff sind schließlich jene Atomtypen, die am häufigsten als Bindungspartner auftreten, folglich ergibt sich überall eine Verschlechterung.

Wendet man die Fitnessfunktion nach 7.2.2 an, so ist die Tendenz, die Fehler zu verteilen, geringer ausgeprägt, was noch verstärkt wird, wenn mit Formel 7.2.3 die empirischen Regeln mit einbezogen werden. Dies ist in Tabelle 29 zu erkennen. Fitnessfunktion 7.2.3 erscheint hier am geeignetsten, da der Fehler weitgehend auf das Methylfluorid begrenzt wird. Ein Kompromiss zugunsten des Methylfluorids bringt nicht den gewünschten Erfolg, verursacht jedoch auf breiter Front eine Verschlechterung. Mit zunehmender Zahl der in die Parameterberechnung einbezogenen Moleküle ist zu erwarten, dass sich die Ergebnisse trotz verschiedener Fitnessfunktionen immer mehr angleichen, da die „Ausweichmöglichkeiten“ der Optimierung immer geringer werden. Abb. 102 und 103 zeigen die Korrelation für die vollständige Parametrisierung des Referenzdatensatzes mit Fitnessfunktion 7.2.2 bzw. 7.2.3.

Fitness: Molekül	RMSD-Wert		Formel 7.2.2		Formel 7.2.3	
	a	σ	a	σ	a	σ
Ammoniak	1.006	0.001	1.036	0.001	0.980	0.001
Ethen	0.869	0.000	0.875	0.000	0.930	0.000
Formaldehyd	0.986	0.056	1.003	0.084	1.039	0.052
HF	0.964	n. a.	0.978	n. a.	1.056	n. a.
Methan	0.861	0.000	0.991	0.000	0.942	0.000
Methylfluorid	1.576	0.930	1.377	0.106	1.484	0.081
Wasser	1.006	0.002	1.046	0.002	1.057	0.002

a Steigung der Ausgleichsgerade

σ Standardabweichung des Schätzwertes

n. a. nicht anwendbar

Tabelle 29: Auswirkung verschiedener Fitnessfunktionen auf die Parameterberechnung

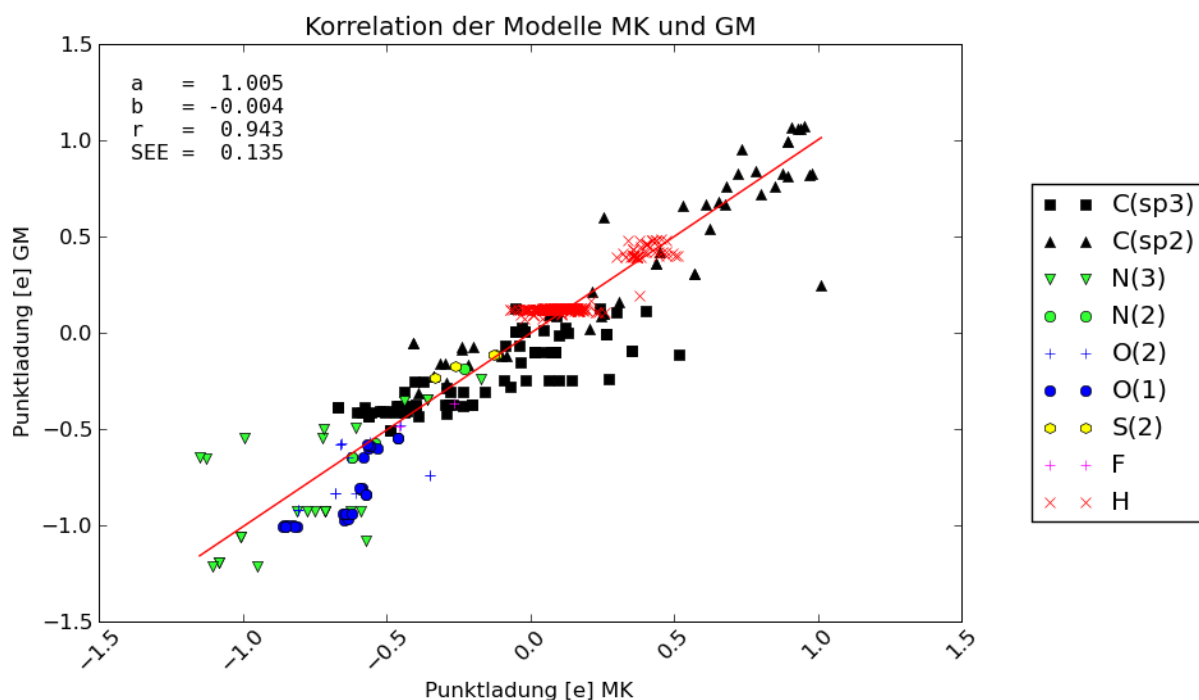
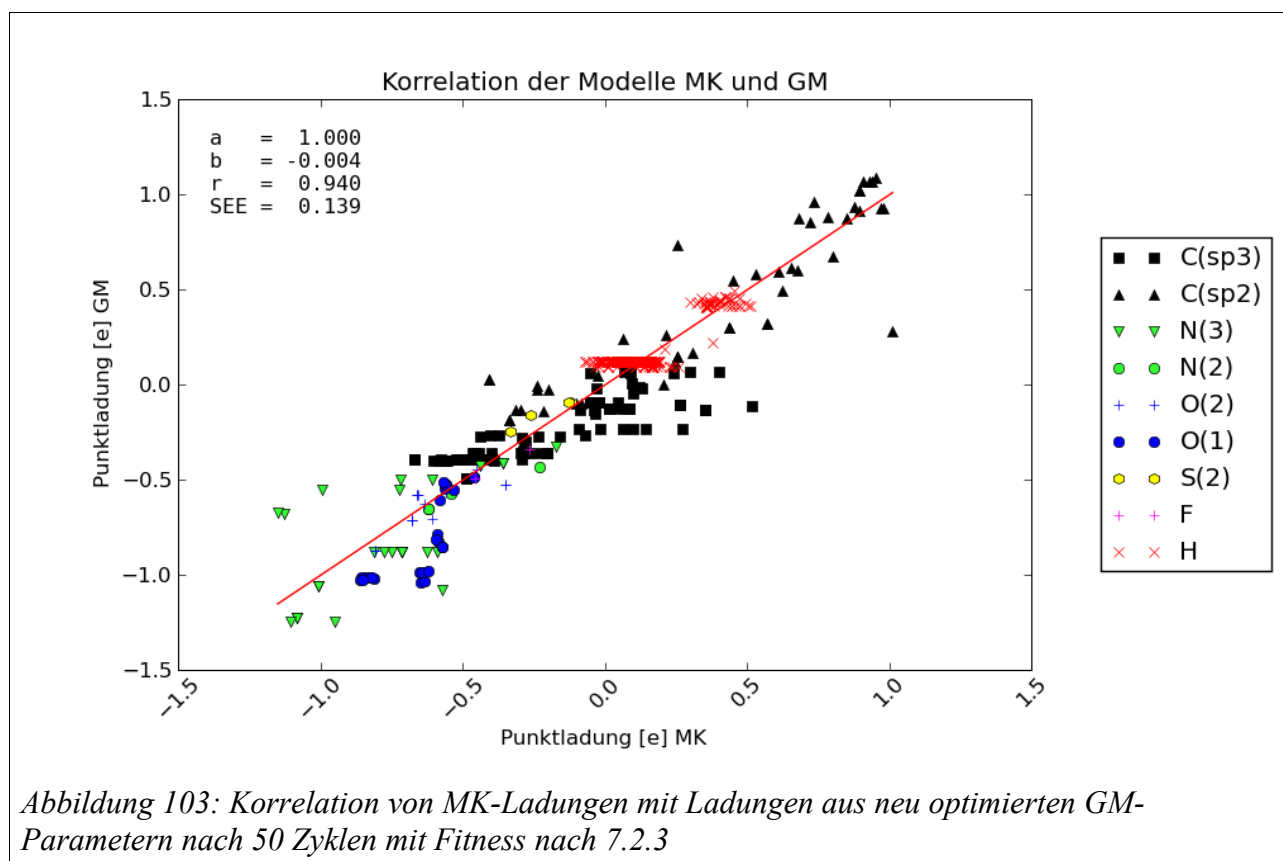
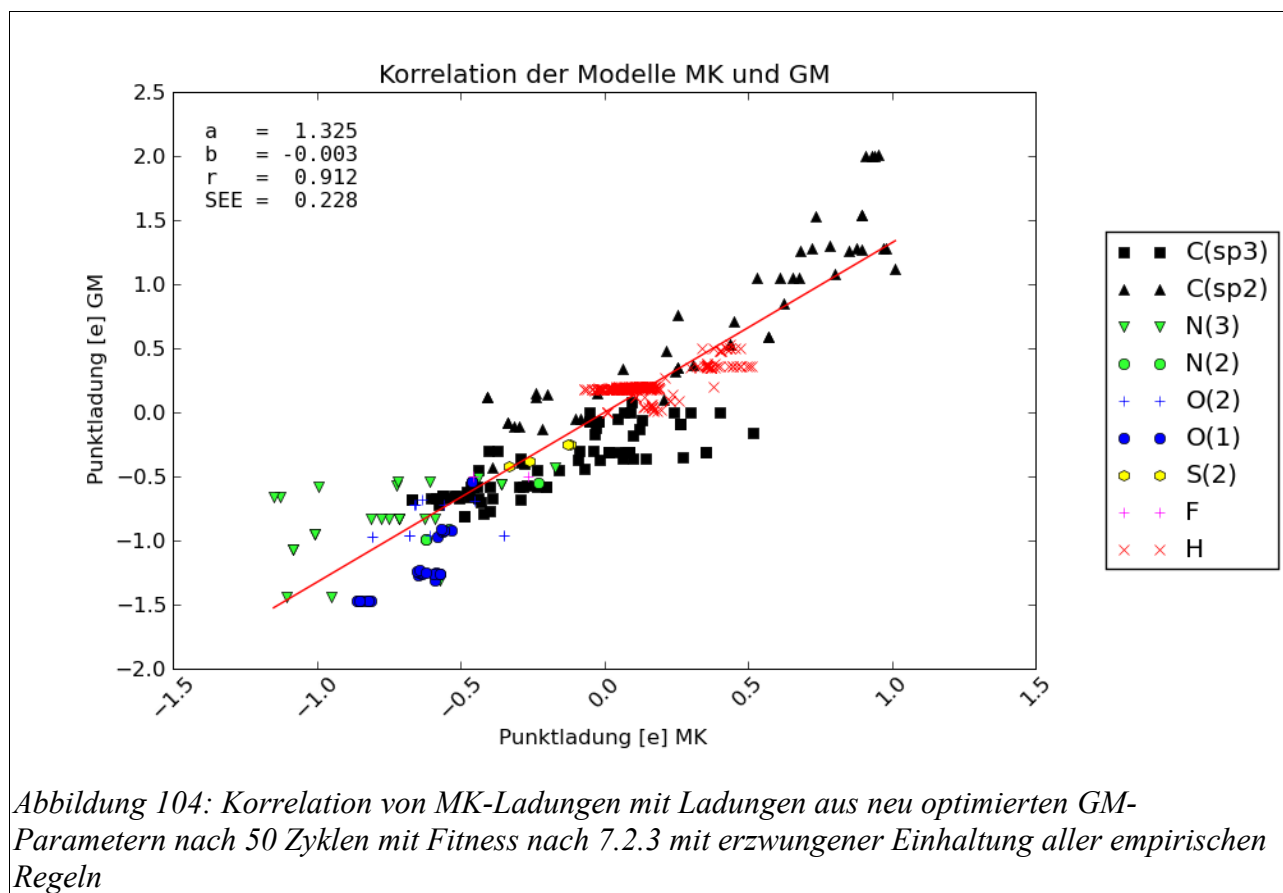


Abbildung 102: Korrelation von MK-Ladungen mit Ladungen aus neu optimierten GM-Parametern nach 50 Zyklen mit Fitness nach 7.2.2

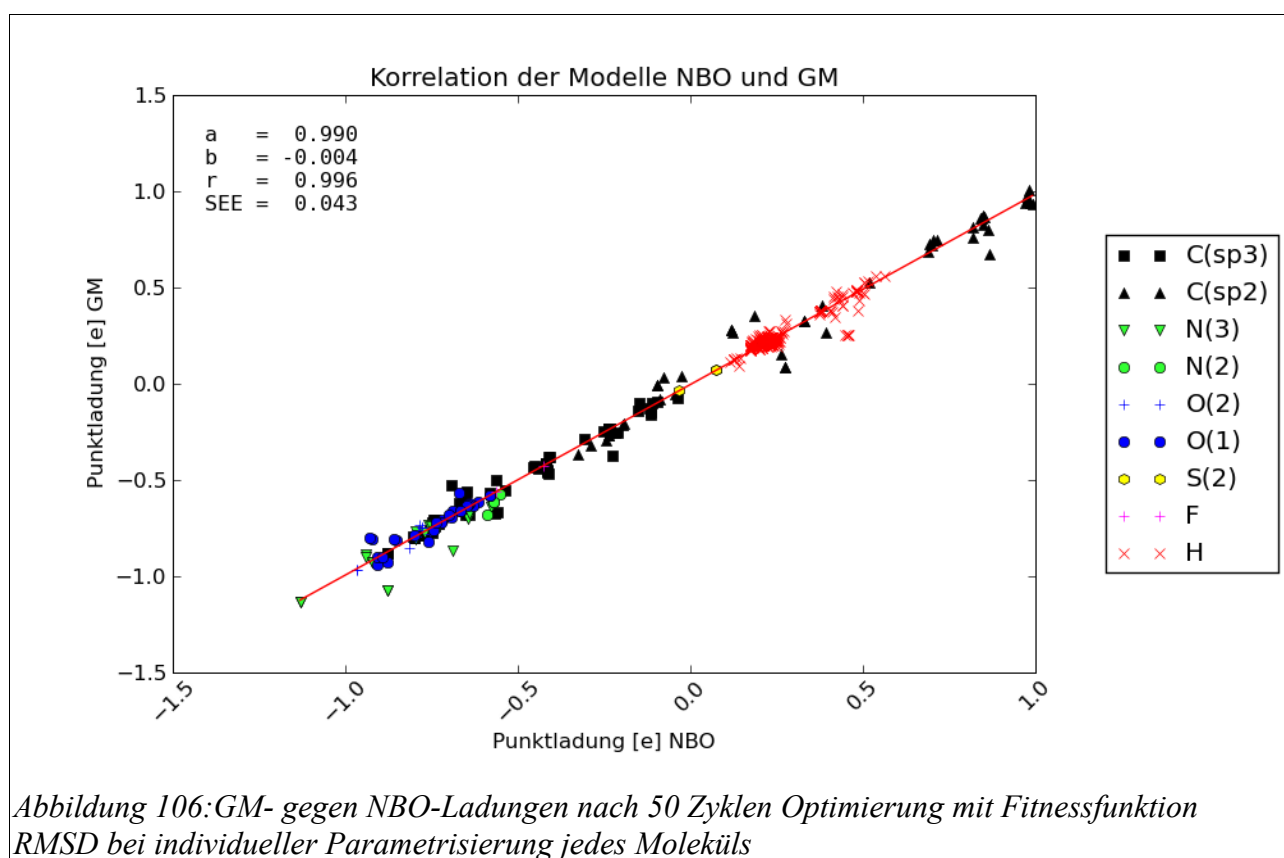
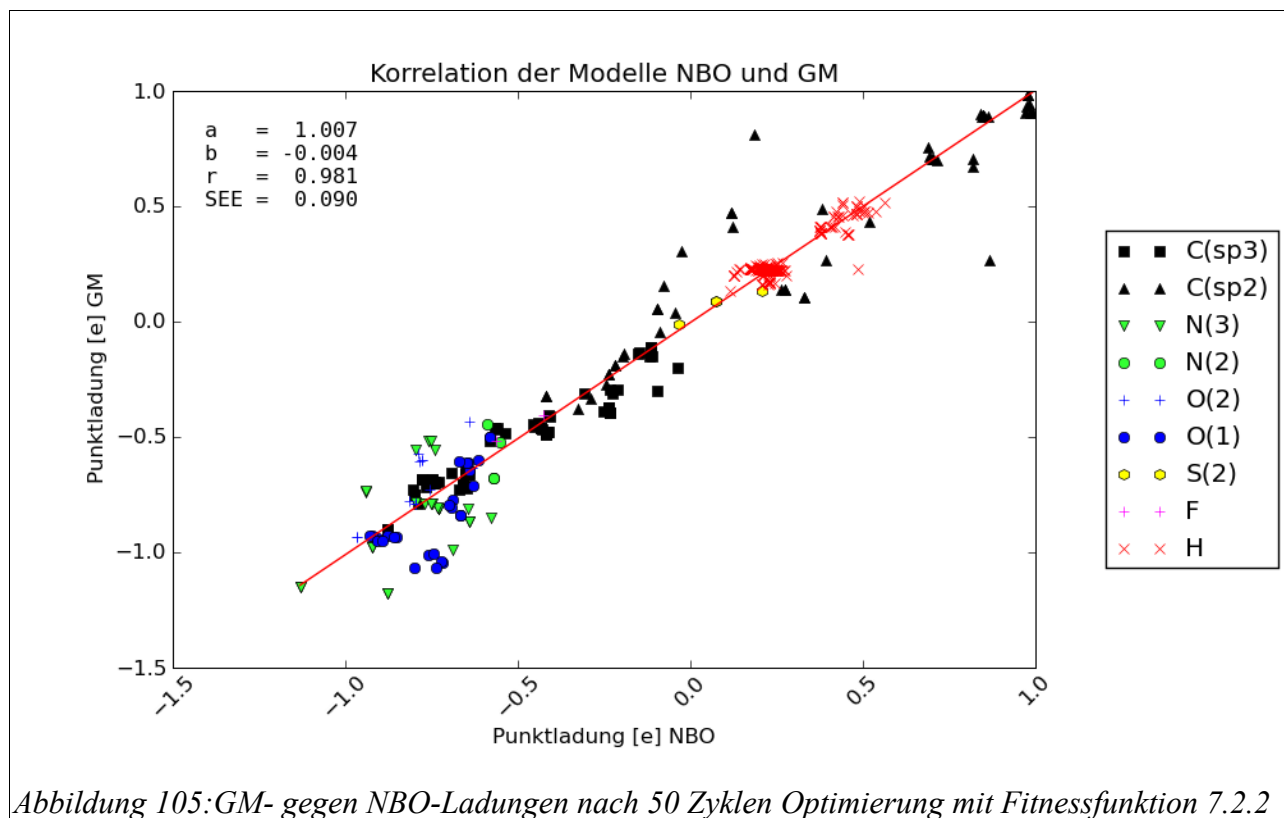


Die Differenz zwischen der Optimierung mit Fitnessfunktion 7.2.2 und 7.2.3 ist sehr gering, im Unterschied zur Fitness nach dem RMSD-Wert wird hier die Konzentration auf den Mittelwert ein wenig abgemildert, und die Steigung der Ausgleichsgerade entspricht annähernd dem Idealwert eins. Dies geht geringfügig auf Kosten der Güteparameter. Die empirischen Regeln in Funktion 7.2.3 beeinflussen das Ergebnis kaum. Die meisten empirischen Regeln werden erfüllt, oft liegen die zu vergleichenden Werte sehr nahe beieinander, sodass gerade eben die Reihenfolge eingehalten wird. Nur eine der empirischen Regeln wird verletzt. Es handelt sich dabei um jene in Reihe 20. Hier wird die Ladung vom C3 im 4-Methyl-1H-imidazol zu negativ und im 4-Methyl-1H-imidazol-Kation zu positiv berechnet. Die Ladungen erscheinen unerreichbar weit von ihren Referenzwerten entfernt. Testweise wurde das Gewicht der empirischen Regeln in der Fitnessfunktion extrem erhöht, um deren vollständige Einhaltung zu erzwingen. Das Ergebnis ist in Abb. 104 zu sehen. Ein entsprechender Parametersatz kann zwar gefunden werden, doch ist der Preis, was die übrigen Qualitätskriterien angeht, zu hoch, zumal viele der Vergleichsanweisungen nur gerade eben, durch fast identische Zahlen, erfüllt werden.



NBO-Referenzladungen

Schon der zuvor durchgeführte Vergleich der vorhandenen Ladungsmodelle verspricht eine bessere Vereinbarkeit des GM-Modells mit der NBO- gegenüber der MK-Referenz (siehe Abb. 90, 91). Dies wird durch die Neuparametrisierung auch bestätigt. Auf den ersten Blick fällt bei der Korrelation von NBO- und mit den neuen Parametern berechneten Ladungen (Abb. 105) auf, dass die mit der MK-Referenz auftretenden horizontalen Streifen (MK-Ladungen hatten in einigen Bereichen eine breitere Verteilung als die GM-Ladungen) erheblich reduziert wurden. Dafür scheint es bei den sp^2 -Kohlenstoff- und den Wasserstoffatomen nun Fälle zu geben, in denen die Ladungen des GM-Modells einen breiteren Bereich abdecken, als dies bei den NBO-Ladungen der Fall ist. Zunächst soll ein Vergleich mit individuell für jedes Molekül berechneten Parametern zeigen, ob es ein prinzipielles Problem gibt. Abb. 106 zeigt die entsprechende Gegenüberstellung. Die Tendenz ist bei den sp^2 -Kohlenstoffatomen bereits auf dieser Stufe zu erkennen, wenn auch nicht so ausgeprägt wie im allgemeinen Parametersatz.



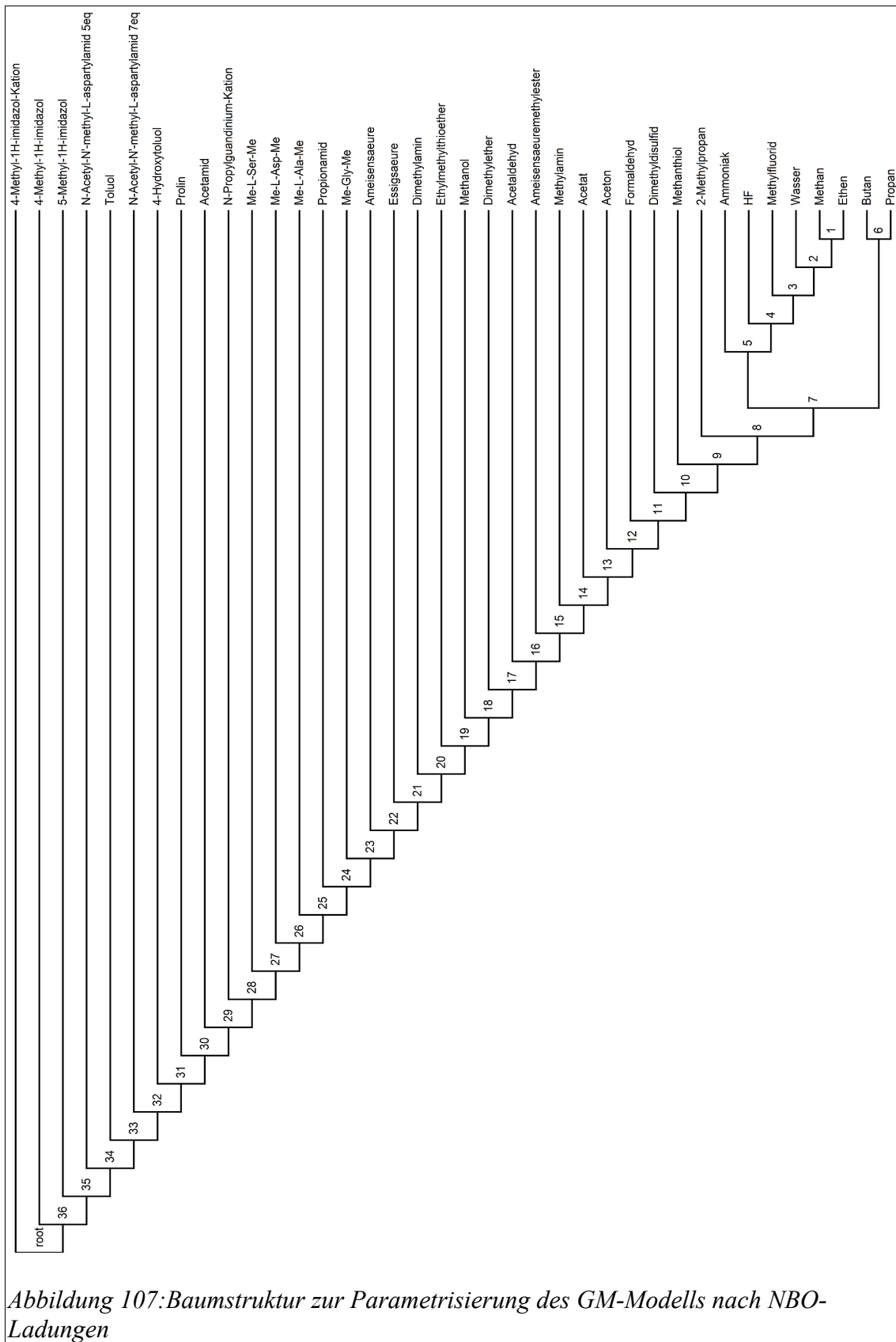


Abbildung 107: Baumstruktur zur Parametrisierung des GM-Modells nach NBO-Ladungen

Um den zugrunde liegenden Konflikt bei der Parametrisierung aufzuspüren, hilft die Baumdarstellung in Abb. 107: Es fällt auf, dass vor allem Moleküle mit einem großen Anteil von π -Systemen weit außen liegen, also schlechter integrierbar sind. Bis zum Knoten Nr. 11 lassen sich weitgehend problemlos Parameter finden. Sobald durch das Hinzufügen von Acetat und Aceton (Knoten Nr. 13) sp^2 -Kohlenstoff als Bindungspartner für Sauerstoff in weiteren chemischen Umgebungen auftritt, treten nicht miteinander vereinbare Parameter auf. Der sp^2 -Kohlenstoff wird dabei zu positiv und der einfach koordinierte Sauerstoff zu negativ. Abhilfe kann hier durch eine verringerte Ladungstrennung im π -System geschaffen werden. Ein Eingriff in die π -SCF-Rechnung erscheint allerdings nicht angebracht: Betrachtet man die Parameter an im Baumdiagramm weiter außen gelegenen Knoten (Nr. 22), werden die entsprechenden Ladungen von sp^2 -Kohlenstoff in Acetaldehyd, Acetat, Ameisensäure, Ameisensäureethylester und Essigsäure gut wiedergegeben, während bei Aceton und Formaldehyd die zu starke Positivierung zu beobachten ist. Die Parameter für Ethen und Carbonylverbindungen lassen sich offenbar schlecht vereinbaren, hier ist die Ladungsübertragung vom Wasserstoff auf den Kohlenstoff zu gering. Das π -System spielt bei Letzterem wegen der Symmetrie des Ethens keine Rolle. Ab Knoten Nr. 28 kommen Verbindungen mit zweifach koordiniertem Stickstoff zum Referenzdatensatz hinzu. In Verbindung mit diesem wird sp^2 -Kohlenstoff zu negativ bewertet.

Abgesehen von diesen Einschränkungen harmonisiert das Gasteiger-Marsili-Modell sehr gut mit den NBO-Referenzladungen.

7.2.11 Parametrisierung des ABR27-Modells

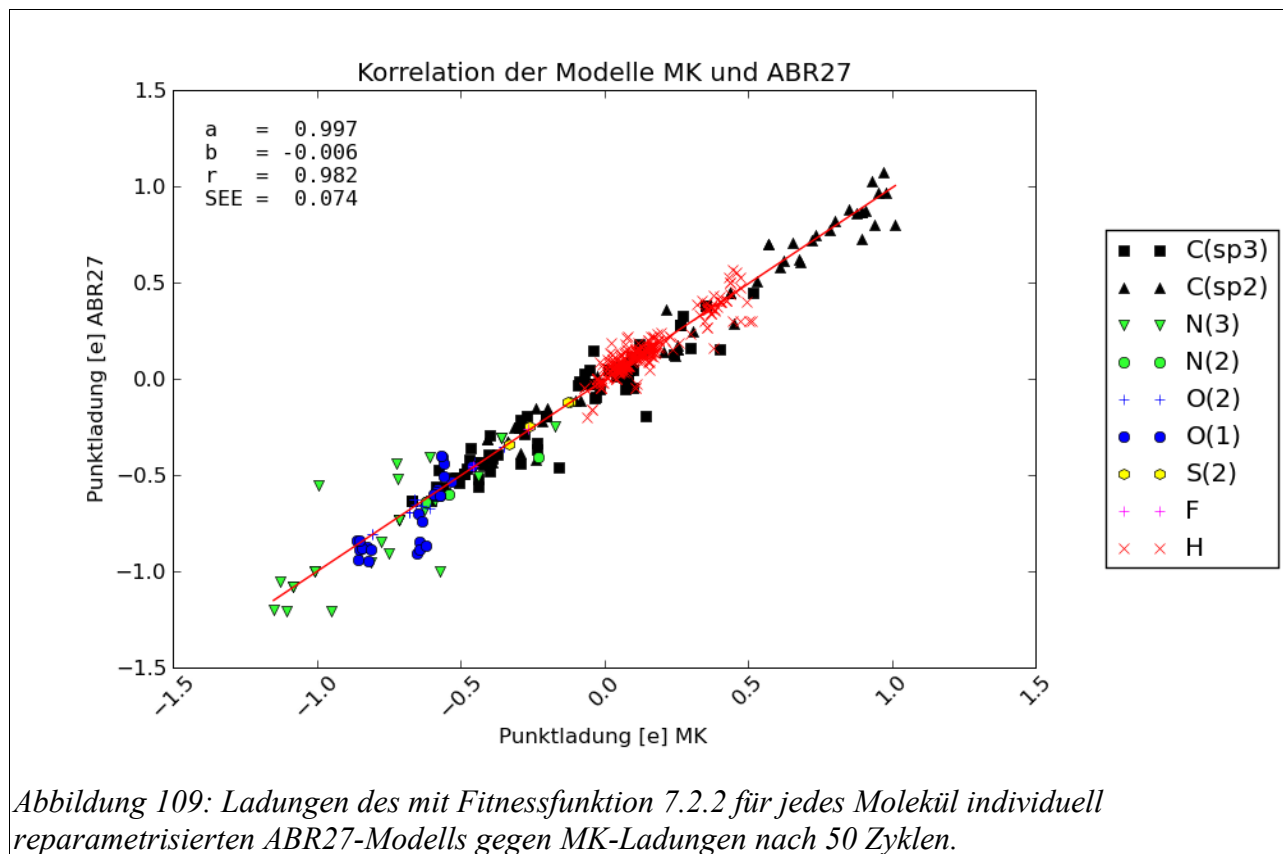
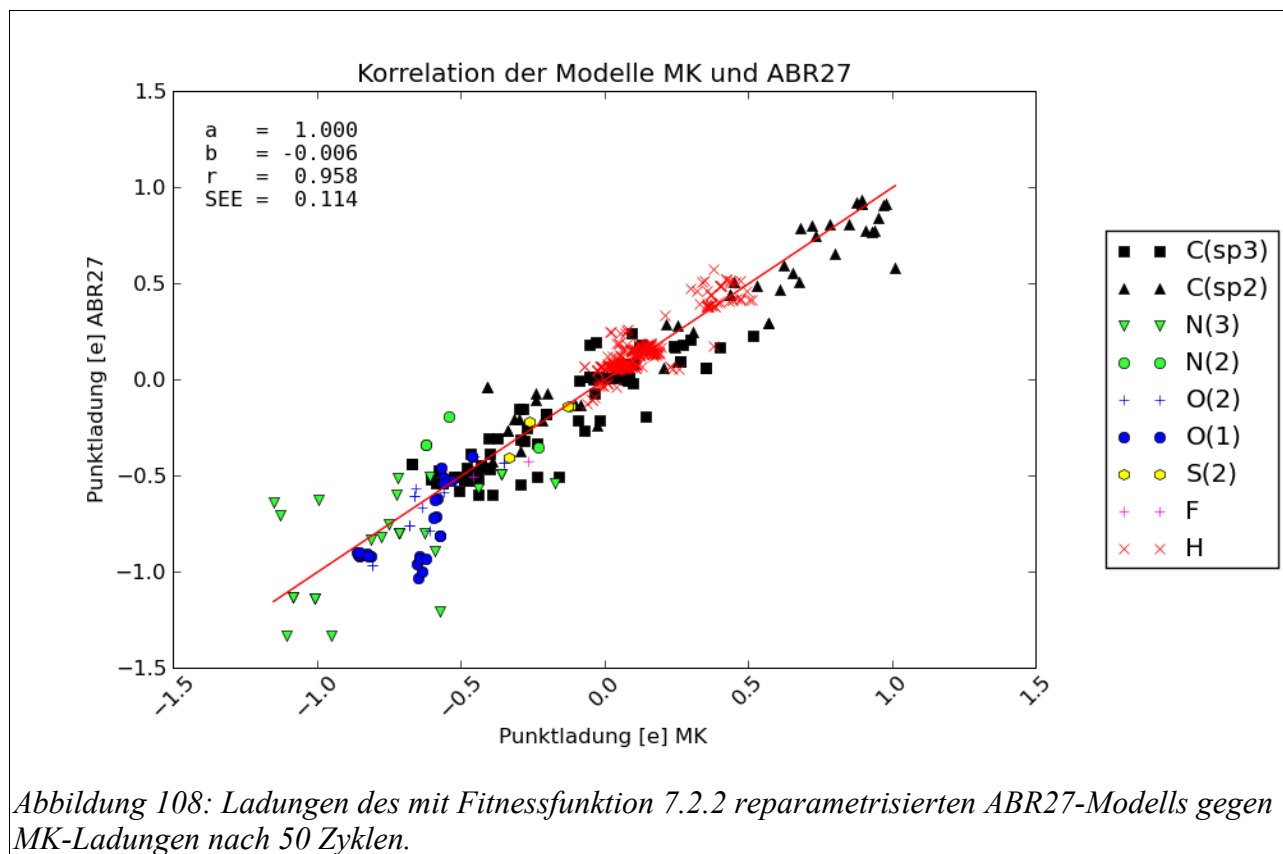
Beim ABR27-Modell (Kapitel 3.3.2.2) handelt es sich um ein Inkrementsystem, bei dem die Inkremente durch schon vorhandene Ladungen modifiziert werden. Dies macht eine iterative Berechnung erforderlich und verhindert, ähnlich wie beim GM-Modell, die Verwendung von auf analytischen Ableitungen basierenden Optimierungsverfahren. Auch hier wurden ein evolutionsstrategischer Algorithmus zur Parameterberechnung eingesetzt und verschiedene Fitnessfunktionen getestet. Der beim GM-Modell verwendete Algorithmus zum Aufbau der Parameter für den gesamten Testdatensatz kam erneut zum Einsatz. Zur Anpassung des ABR27-Modells dienen die Parameter a_{AB} (vgl. Formel 3.3.9), b_{AC} (vgl. Formel 3.3.11) c_{Atom} (vgl. Formel 3.3.13) sowie die atomare Polarisierbarkeit P_{Atom} und die Elektronegativität χ_{Atom} .

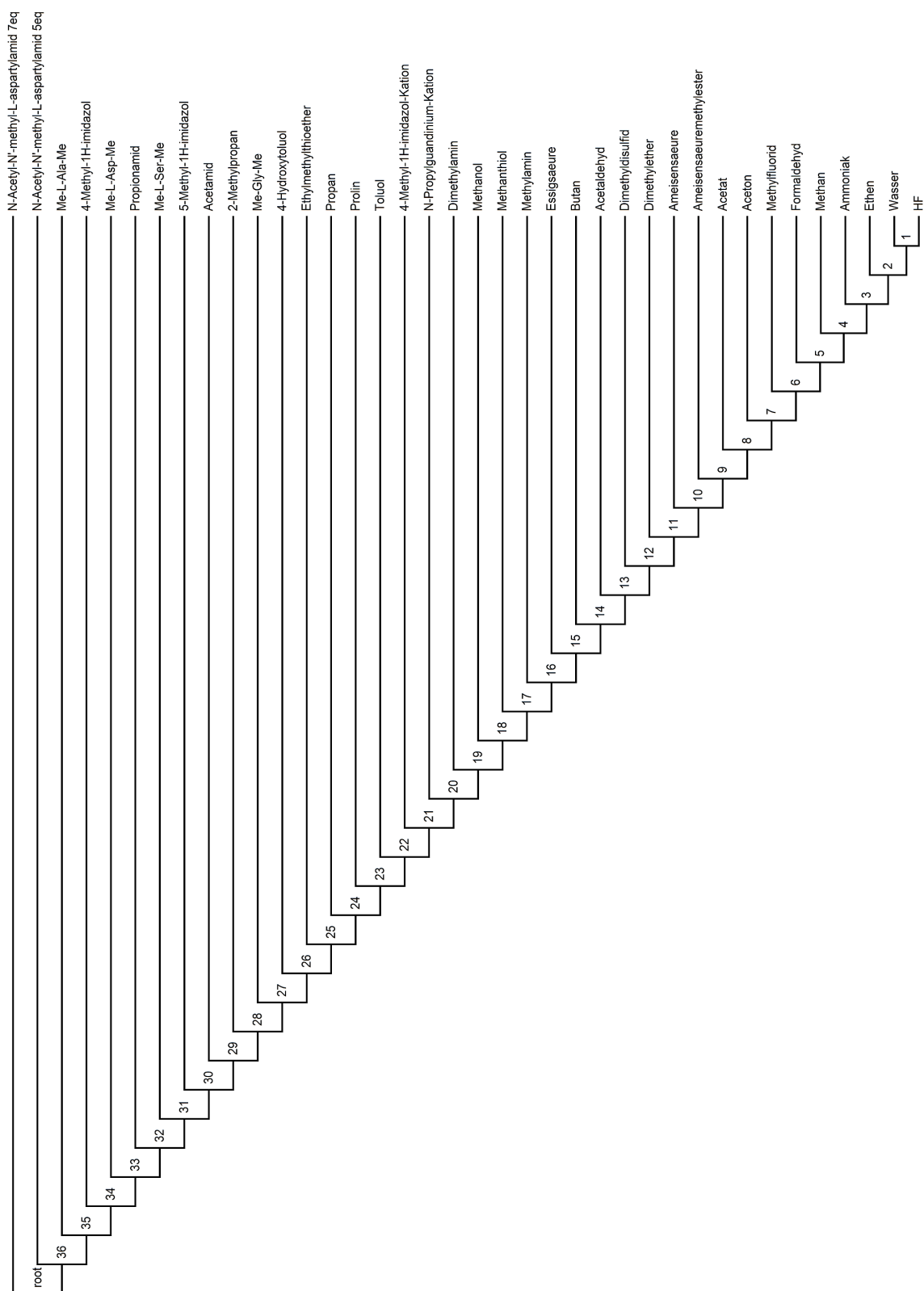
Bei der Entwicklung des ABR27-Modells wurden Ladungen nach dem MK-Modell zugrunde gelegt, bereits mit den Originalparametern konnten deshalb gute Übereinstimmungen erzielt

werden. Auch wenn das Modell in der Lage ist, Konformationsabhängigkeiten zu berücksichtigen, bleibt es dennoch von den Elektronegativitäten der Atome dominiert. Mit den Originalparametern stimmen die Ergebnisse wesentlich besser mit dem MK-Modell als mit dem NBO-Modell überein (vgl. Kapitel 7.2.8, Abb.89 und 92), dies muss aber keine prinzipielle Bedeutung haben.

Merz-Kollman-Referenzladungen

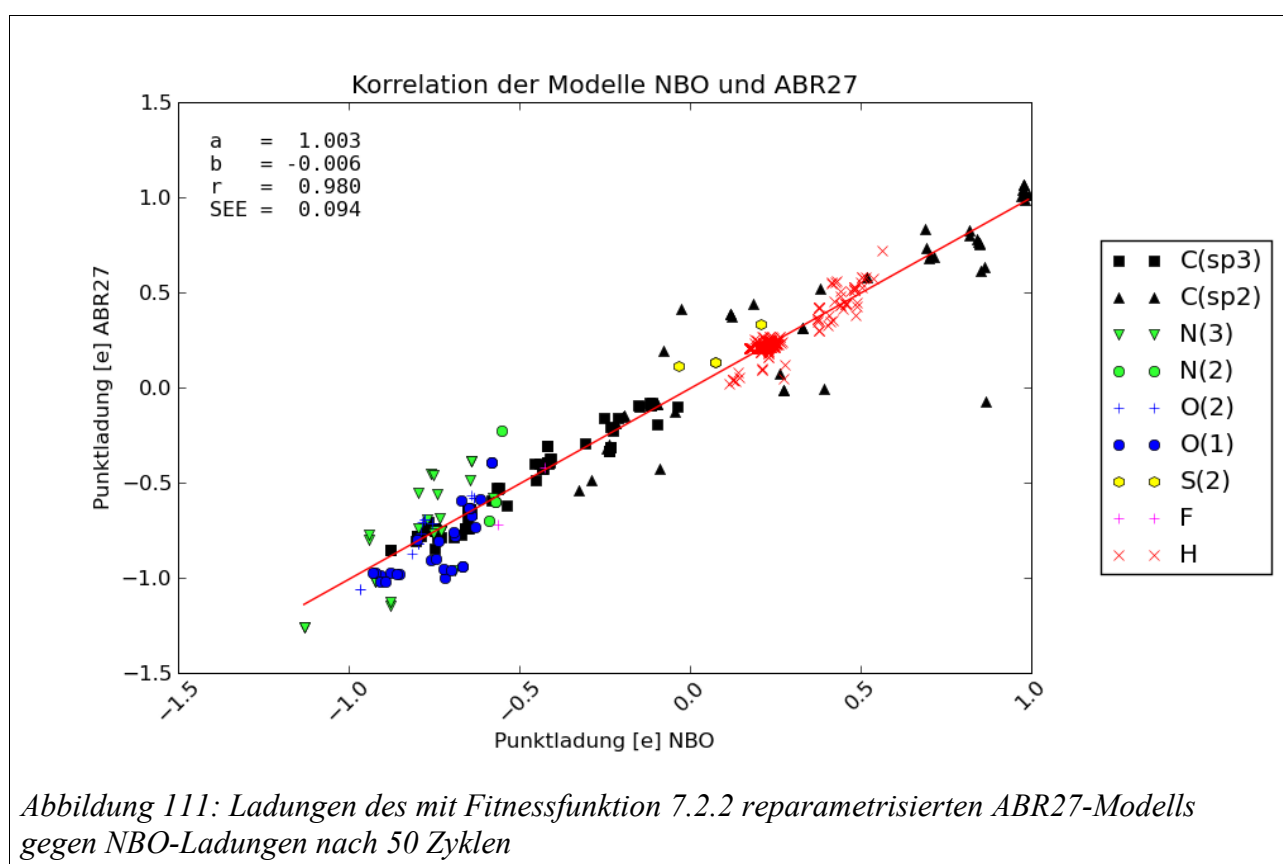
Die statistischen Kenngrößen der Korrelation mit den MK-Ladungen konnten durch die Reparametrisierung noch etwas verbessert werden, aber schon der originale Parameterstanz liefert gute Ergebnisse. In Abb. 108 sind, wie auch schon beim Vergleich der Modelle MK und GM, horizontale Streifen zu erkennen. Eine nähere Analyse der Daten zeigt, dass es sich bei einigen davon eher um nebeneinander liegende vertikale Streifen handelt, also Fälle, in denen sich die vom ABR27-Modell ermittelten Ladungen über einen breiteren Bereich erstrecken als die der MK-Referenz. Beim einfach koordiniertem Sauerstoff ist dies besonders auffällig. Mithilfe der Ergebnisse einer individuellen Parametrisierung (Abb. 109) für jedes Molekül und den vom Parametrisierungsalgorithmus aufgestellten Baum (Abb. 110) soll nun nach den Ursachen gesucht werden. Die beschriebene „Streifenbildung“ findet sich bei der individuellen Parameterberechnung wieder; betrachtet man die Reihenfolge der vom Optimierungsalgorithmus zusammengefassten Moleküle, so scheint es Probleme mit π -Systemen zu geben. Eine nähere Betrachtung beweist, dass ABR27 Schwierigkeiten hat, sich an die zum Teil erheblichen Ladungstrennungen, die durch die π -SCF-Rechnung bestimmt werden, anzupassen, um auf die Referenzwerte zu kommen. Bei Nicht- π -Systemen werden die Eigenheiten des MK-Ladungsmodells gut wiedergegeben, die Probleme des GM-Modells mit Kohlenwasserstoffketten treten hier kaum auf. Die Ladungstrennung durch ABR27 ist oft recht groß, was zu den beschriebenen vertikalen Streifen führt. Abhilfe verspricht hier eine Erweiterung des Bereiches, in dem die Parameter durch das Optimierungsprogramm verstellt werden dürfen. Es ergibt sich zwar eine Verbesserung, das Problem wird aber nicht gelöst. Leider können die zur Beeinflussung der π -Ladungen zu ändernden Parameter nicht ohne Weiteres in die Justierung eines Ladungsmodells mit aufgenommen werden, da damit auch die für andere Kraftfeldkomponenten wichtigen π -Bindungsordnungen beeinflusst werden können. Die Einbeziehung der empirischen Regeln zur Parametrisierung bringt keine Verbesserung, zumal einige schon vom Referenzmodell MK nur schlecht bzw. gar nicht erfüllt werden.





NBO-Referenzladungen

Das ABR27-Modell wurde anhand von Ergebnissen des MK-Modells entwickelt, dennoch werden die berechneten Ladungen zum größten Teil durch Elektronegativitätsunterschiede bestimmt. Daher kann erwartet werden, dass sich durch eine Reparametrisierung eine gute Übereinstimmung mit dem NBO-Modell ergibt. Letzteres hatte, was die Vereinbarkeit mit dem nur auf Elektronegativitätsunterschieden beruhenden GM-Modell angeht, deutliche Vorteile. Wie in Abb. 111 zu sehen, konnte die Übereinstimmung gegenüber den Originalparametern (Abb. 92) deutlich gesteigert werden, die statistischen Gütekriterien übertreffen jede andere getestete Kombination von Ladungsmodellen.



Die von Problemen mit der Anpassung der π -SCF-Rechnung herrührenden vertikalen Streifen beim einfach koordinierten Sauerstoff, dreifach koordiniertem Stickstoff und dem sp^2 -Kohlenstoff sind jedoch auch hier vorhanden. Es ist zu überlegen, die Berechnung der π -Ladungen von den anderen Kraftfeldtermen zu trennen oder hier auf ein rein empirisches Verfahren auszuweichen.

7.3 Parametrisierung der bindenden Kraftfeldterme

Da das MOMO-Kraftfeld zur Berechnung organischer Moleküle verwendet wird und für die dort häufig vorkommenden Atomtypen parametrisiert werden soll, wurde mittels einer Literaturrecherche nach Veröffentlichungen zur Kraftfeldparametrisierung ein Katalog von relativ kleinen organischen Molekülen verschiedener Verbindungsklassen wie Alkane, Alkene, Alkine, Alkohole, Aldehyde, Ketone, Carbonsäuren, Amine und einigen Schwefel- und Phosphorverbindungen zusammengestellt. Die Moleküle sollen repräsentativ für ihre Verbindungsklasse sein. Zu diesen Verbindungen wurden Einträge in der Gasphasen-Molekülspektrendatenbank „MOGADOC 2002/2003“ [Vogt2004] gesucht. Die Verbindungen wurden im Editor aufgebaut und, soweit verfügbar, mit Geometriedaten aus MOGADOC ergänzt. Auch in der Cambridge Structural Database [Allen2002] wurde nach den Verbindungen aus MOGADOC gesucht, was meistens jedoch erfolglos war.

7.3.1 Speicherung der Referenzdaten

Für die Speicherung kommt wieder eine Erweiterung der verwendeten XML-Dateien um spezielle Tags zum Einsatz:

```
<ParameterizationReference>
<bfix atom1="C1" atom2="C2" dist="1.530" molecule="molecule 1"/>
<bfix atom1="C2" atom2="C3" dist="1.534" molecule="molecule 1"/>
<bfix atom1="C1" atom2="H1A" dist="1.094" molecule="molecule 1"/>
<bfix atom1="C1" atom2="H1C" dist="1.093" molecule="molecule 1"/>
<bfix atom1="C1" atom2="H1B" dist="1.095" molecule="molecule 1"/>
<bfix atom1="C2" atom2="H2*" dist="1.095" molecule="molecule 1"/>
<afix atom1="C1" atom2="C2" atom3="C3" ang="114.3" molecule="molecule 1"/>
<afix atom1="H1A" atom2="C1" atom3="C2" ang="110.8" molecule="molecule 1"/>
<afix atom1="H1C" atom2="C1" atom3="C2" ang="111.9" molecule="molecule 1"/>
<afix atom1="H1B" atom2="C1" atom3="C2" ang="110.0" molecule="molecule 1"/>
<afix atom1="H2*" atom2="C2" atom3="C1" ang="109.4" molecule="molecule 1"/>
<tfix atom1="H1A" atom2="C1" atom3="C2" atom4="C3" tor="180.0" molecule="molecule 1"/>
<tfix atom1="H1C" atom2="C1" atom3="C2" atom4="C3" tor="60.0" molecule="molecule 1"/>
<tfix atom1="H2A" atom2="C2" atom3="C1" atom4="H1B" tor="180.0" molecule="molecule 1"/>
<tfix atom1="C1" atom2="C2" atom3="C3" atom4="C4" tor="180.0" molecule="molecule 1"/>
</ParameterizationReference>
```

Abbildung 112: XML Tags zur Aufnahme der Geometrieinformationen aus MOGADOC. Atomnamen können auch sogenannte Wildcards (Muster, die auf mehrere Atomnamen zutreffen) sein (rot umrahmt).

Unterhalb des <ParameterizationReference> Tags werden die Tags <bfix>, <afix> und <tfix> angeordnet.

<bfix> ist eine Referenz für Bindungslängen. Die Attribute „atom1“ und „atom2“ geben die an der Bindung beteiligten Atome an, „dist“ die Referenzbindungslänge in Å und „molecule“ den Namen des Moleküls, in dem die Bindung zu finden ist.

<afix> ist eine Referenz für Bindungswinkel. Die Attribute „atom1“, „atom2“ und „atom3“ sind die Namen der den Winkel aufspannenden Atome, wobei „atom2“ das zentrale Atom des Winkels ist. „ang“ ist der Referenzwinkel in Grad.

<tfix> ist eine Referenz für Torsionswinkel. Die Attribute „atom1“, „atom2“, „atom3“ und „atom4“ sind die Namen der den Winkel aufspannenden Atome, wobei „atom2“ und „atom3“ die zentralen Atome des Winkels sind. „tor“ ist der Referenzwinkel in Grad.

In den Dateien liegen die Moleküle in einer Standardgeometrie vor, die nicht mit denen in den genannten Tags übereinstimmen muss.

7.3.2 Der Referenzdatensatz

Es wurde ein Datensatz von etwa 140 Molekülen zusammengestellt. Nach internen Koordinaten aufgeschlüsselt, ergeben sich die in Abb. 113, 114 und 115 gezeigten Häufigkeiten von Bindungs-, Bindungswinkel- und Torsionswinkeltypen. Wie für eine Sammlung organischer Moleküle zu erwarten, ist der Kohlenstoff am häufigsten vertreten, mit großem Abstand folgen Sauerstoff, Stickstoff und schließlich Schwefel. Gerade für die wenig vertretenen Elemente wäre eine Vergrößerung des Datensatzes wünschenswert, was jedoch nicht das Hauptziel dieser Arbeit war.

Zur Evaluierung des MOMO-Kraftfeldes wurde das Programm „MOMOParamGeo.py“ geschrieben, mit diesem wurden u. a. auch die Abbildungen zur Häufigkeit der verschiedenen Typen interner Koordinaten erstellt. Ausgehend von einem Verzeichnis der Referenzdateien, den zulässigen Toleranzen für Bindungslängen, Bindungswinkel und Torsionswinkel und optional einer Parameterdatei, wird die Qualität des Kraftfeldes bewertet. Das Programm erzeugt einen Bericht, in dem für jede Klasse interner Koordinaten zunächst die Verteilung der Koordinatentypen, dann ein Korrelationsdiagramm der Daten von Kraftfeld gegen Referenz aufgelistet ist. Schließlich wird noch für jeden Untertyp (z. B. C(sp³)-C(sp³)-Bindungslänge) ein kleines Korrelationsdiagramm und ein Tortendiagramm erstellt, das zeigt, wie viele der Referenzdaten innerhalb der vorgegebenen Toleranzen erfüllt oder verletzt werden und wie oft die betreffende Koordinate ohne Referenzdaten im Testdatensatz vorkommt. Jede nicht erfüllte Referenz wird in Textform aufgelistet.

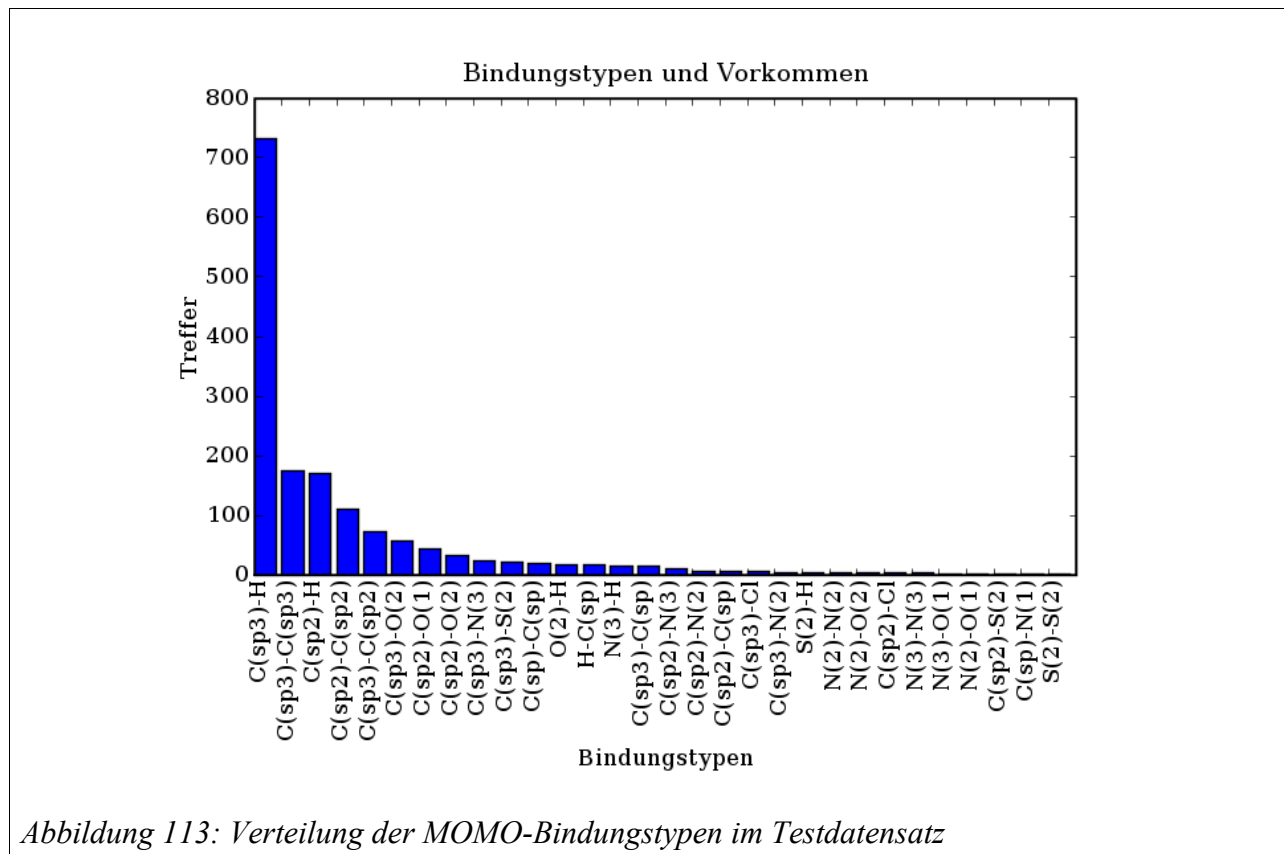


Abbildung 113: Verteilung der MOMO-Bindungstypen im Testdatensatz

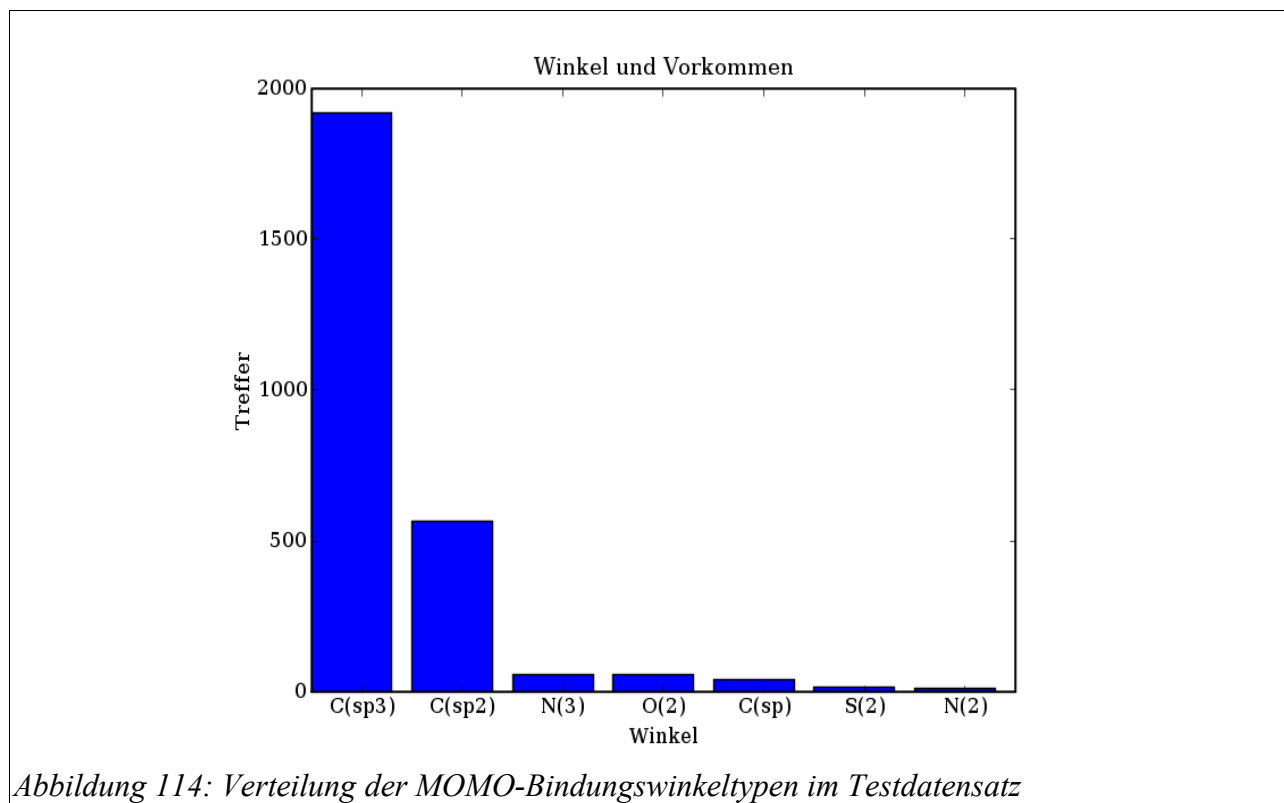


Abbildung 114: Verteilung der MOMO-Bindungswinkeltypen im Testdatensatz

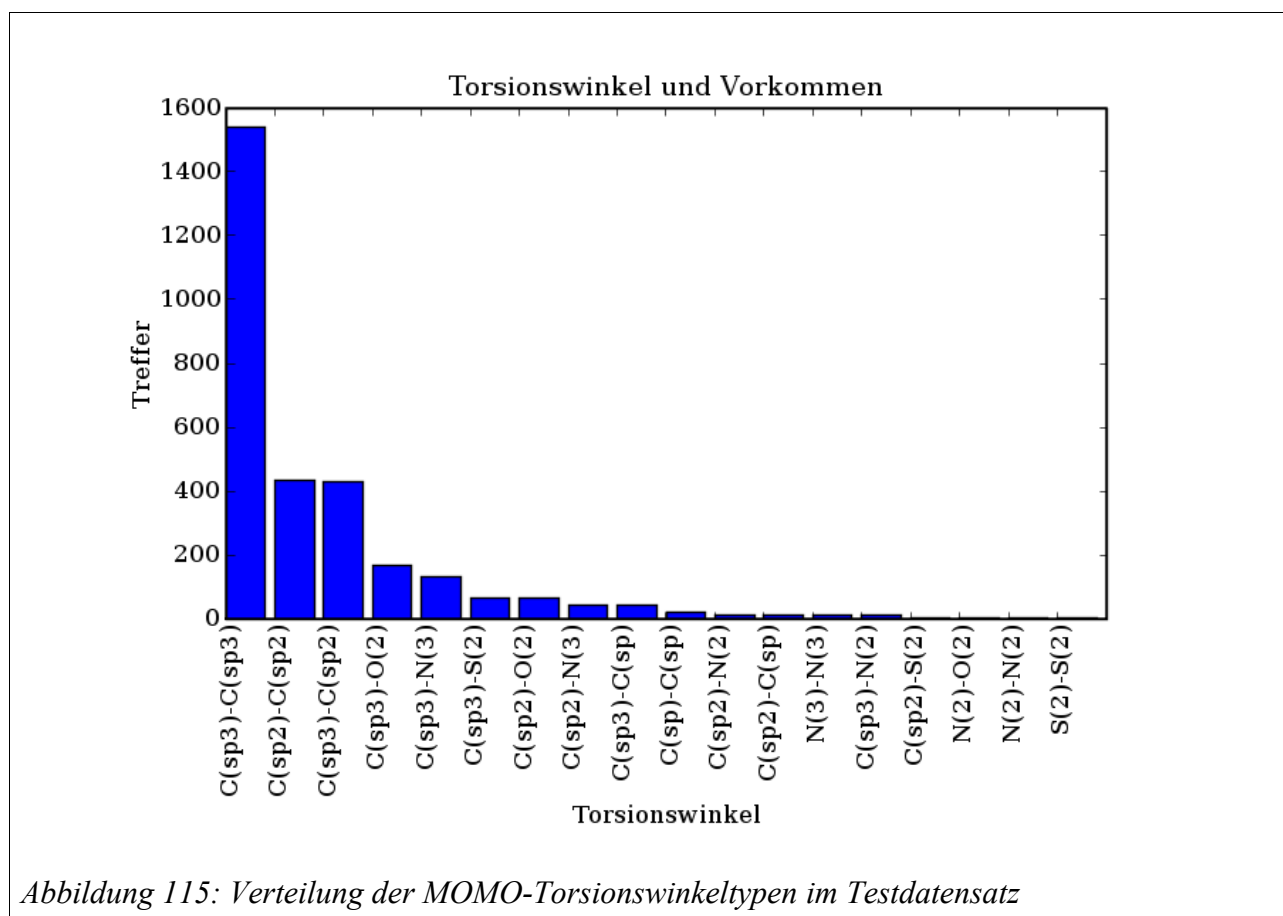
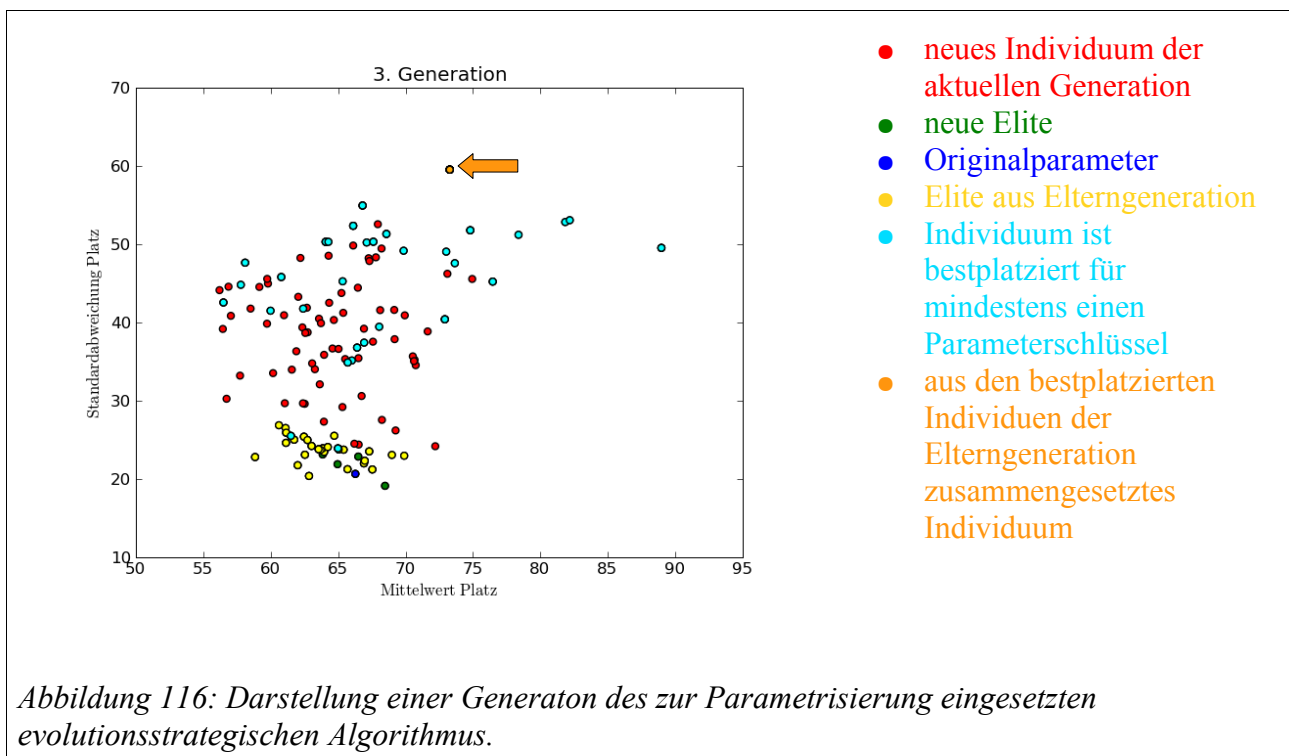


Abbildung 115: Verteilung der MOMO-Torsionswinkeltypen im Testdatensatz

7.3.3 Das Parametrisierungsprogramm

Wie in Kapitel 3 beschrieben, werden die Kraftfeldterme in MOMO oft unter bestimmten Bedingungen modifiziert. Will man von einer internen Koordinate in einem Referenzmolekül auf die dazugehörigen Kraftfeldparameter schließen, so müssen diese Bedingungen mit berücksichtigt werden, damit der Optimierungsalgorithmus auf die richtigen Parameter wirken kann. Als Referenzwerte sind im Testdatensatz ausgewählte Bindungslängen und -winkel sowie Torsionswinkel enthalten. Es ergeben sich so 30 Bindungslängentypen, 7 Bindungswinkeltypen und 18 Torsionswinkeltypen. Aus diesen insgesamt 55 Typen interner Koordinaten werden im vorliegenden Datensatz durch die Modifizierung der Kraftfeldterme insgesamt 69 Schlüssel zu jeweils einer Parameterkombination. Kraftfeldterme sind nie voneinander isoliert, was den Einsatz von Optimierungsalgorithmen erschwert, die Ableitungen nach den Parametern benötigen. Aus diesem Grund wurde erneut auf einen evolutionstrategischen Algorithmus zurückgegriffen. Als Fitnessfunktion wird wie schon bei den Partiaalladungen Formel 7.2.2 eingesetzt. Die Bewertung der Fitness erfolgt allerdings für jeden Parameterschlüssel separat, was auch den Hintergrund hat, dass

die den verschiedenen Koordinatentypen zugeordneten Zahlen, was Größenordnung und Toleranzen angeht, nicht vergleichbar sind. Das Prinzip des Algorithmus wird in Abb. 117 illustriert. Um die in den Individuen des evolutionstrategischen Algorithmus abgelegten Parameterkombinationen in eine Reihenfolge zu bringen, wird für jeden auftretenden Parameterschlüssel eine Rangliste der Individuen aufgestellt. Ein besonders geeigneter Parametersatz ist durchschnittlich auf den vorderen Plätzen eingeordnet, und die Platzierung hat eine geringe Standardabweichung über alle Parameterschlüssel. Die in dieser Hinsicht besten Individuen werden vom Selektionsoperator ausgewählt. Ein Teil von ihnen wird unverändert als „Elite“ in die nächste Generation übernommen, die restlichen Individuen werden durch Mutation und Rekombination innerhalb der 30 % besten Individuen erzeugt. Zusätzlich werden noch einige Individuen unter Einbeziehung schlechter platzierter Parametersätze und ein Individuum, das die Parameter aus den für jeden Schlüssel jeweils bestplatzierten Individuen zusammenfasst, generiert. Das Programm speichert für jede Generation alle Individuen und die Elite für die spätere Auswertung mit „MOMOParamGeo.py“ in Dateien. Ferner erstellt es Grafiken, die die relative Verteilung der Individuen im Hinblick auf deren durchschnittliche Platzierung mit deren Standardabweichung zeigen. Abb. 116 enthält eine solche Darstellung.



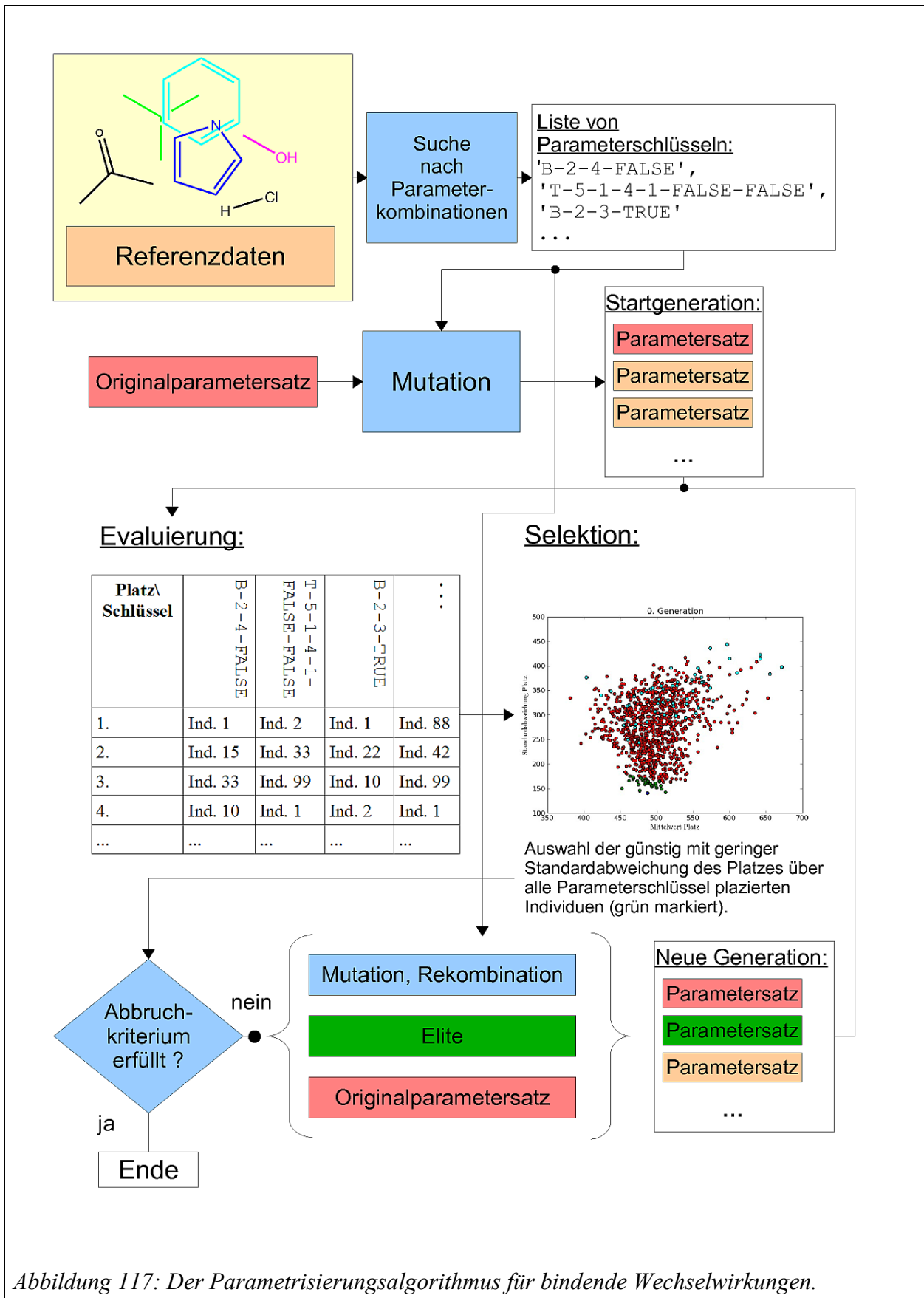


Abbildung 117: Der Parametrisierungsalgorithmus für bindende Wechselwirkungen.

7.3.4 Evaluierung des Kraftfeldes und Ergebnisse der Parametrisierung

Die meisten der verwendeten Referenzdaten wurden mittels Elektronenbeugung und Mikrowellenspektroskopie, zwei Methoden zur Strukturbestimmung freier gasförmiger Moleküle, ermittelt, einige stammen aus Röntgenbeugungsexperimenten. Röntgenbeugung ermittelt die Geometrie von Molekülen im Kristallgitter, die Beugung erfolgt an den die Atomkerne umgebenden Elektronen, bei der Elektronenbeugung werden Elektronen an den Atomkernen und Elektronen gebeugt, Mikrowellenspektroskopie misst Rotationsenergieniveaus, die von den Atomabständen abhängen. Moleküle sind niemals starr, selbst im Schwingungsgrundzustand unterliegen berechnete Abstände einer durch die Nullpunkt-Schwingung bedingten Unschärfe. Ein bedeutender Unterschied zwischen Röntgenbeugung und den Gasphasenmethoden ist, dass Erstere mittlere Atompositionen bestimmt, zwischen denen dann Abstände abgemessen werden, während die Mikrowellenspektroskopie mittlere Atomabstände bestimmt. Trotz der Fehler und prinzipiellen Unterschiede bei der experimentellen Geometriebestimmung geht man von einer Genauigkeit von 0.01\AA für Bindungslängen und 0.5° für Bindungs- und Torsionswinkel aus [Lewars2003]. Für die erzielbare Genauigkeit von Kraftfeldern gelten 0.01\AA für Bindungslängen $1^\circ - 2^\circ$ für Bindungswinkel und bis zu 10° für Torsionswinkel als akzeptabel. Mit diesen Toleranzen wurden die MOMO-Originalparameter anhand des Referenzdatensatzes getestet.

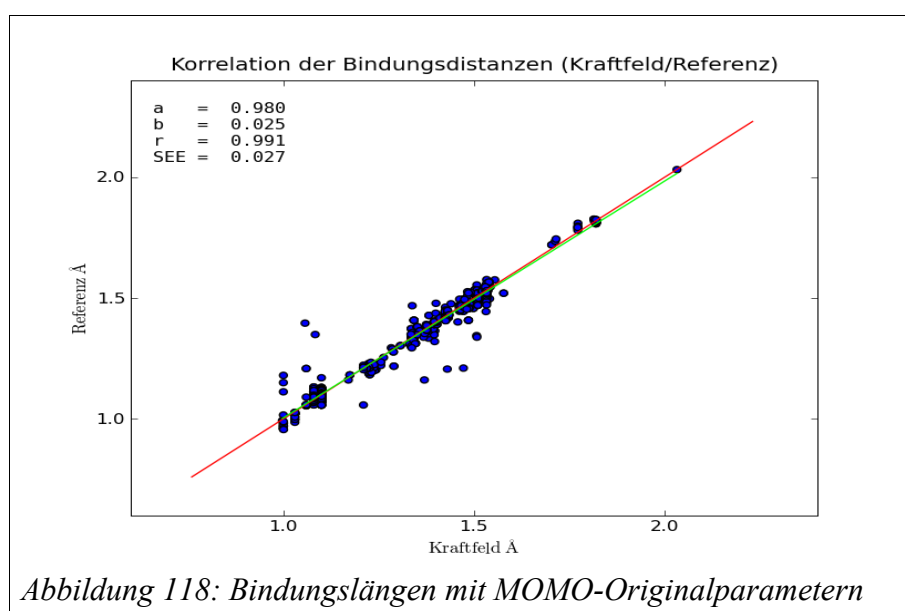


Abb. 118 zeigt eine Gegenüberstellung der mit MOMO ermittelten Bindungslängen und der Referenzdaten. Sofort fällt auf, dass sich die Referenzwerte über einen viel größeren Wertebereich verteilen als die mit dem Kraftfeld bestimmten Werte. Eine detailliertere Auswertung offenbart, dass dies im Wesentlichen durch Bindungen, an denen Wasserstoff beteiligt ist, verursacht wird. Bei den übrigen Bindungstypen treten besonders bei $C(sp^2)-C(sp^2)$; $C(sp^2)-O(1)$; $C(sp)-C(sp)$ und $N(3)-O(1)$ größere Abweichungen auf.

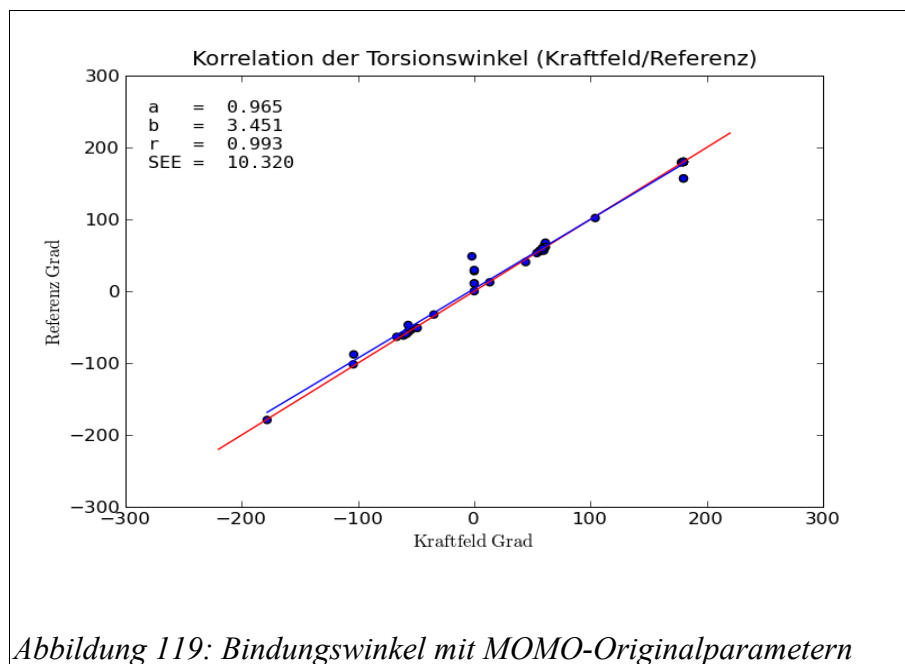


Abb. 119 zeigt eine Gegenüberstellung der mit MOMO ermittelten Bindungswinkel und der Referenzdaten. Die beste Übereinstimmung wird bei Winkeln mit zentralem Kohlenstoff erreicht. Schwierigkeiten treten hier vor allem in kleinen Ringen auf, oder wenn der Bindungswinkel mit anderen schlechter reproduzierten Winkeln in einem Ring konkurriert. Bei Winkeln mit zweifach koordiniertem Sauerstoff werden jene in den OH-Gruppen zu groß gegenüber der Referenz berechnet, C-O-C Fragmente werden jedoch gut wiedergegeben. Die Bindungswinkel mit zweifach koordiniertem Schwefel werden systematisch zu groß berechnet, der Gleichgewichtswinkel ist deutlich größer gewählt als der mittlere in den Referenzdaten auftretende Winkel. In kleinen Ringen ergibt sich teilweise eine Konkurrenz mit anderen Atomtypen, die zu einer Verzerrung führt. Der Gleichgewichtsbindungsinkel für zweifach koordinierten Stickstoff ist ebenfalls zu groß gewählt. Winkel um dreifach koordinierten Stickstoff werden gut reproduziert.

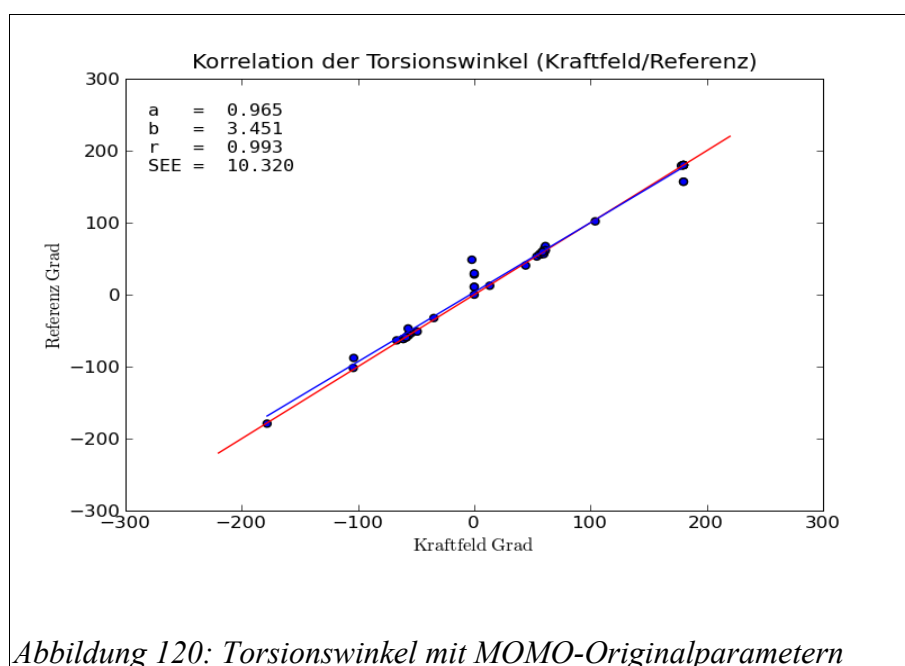


Abbildung 120: Torsionswinkel mit MOMO-Originalparametern

Abb. 120 zeigt eine Gegenüberstellung der mit MOMO ermittelten Torsionswinkel und der Referenzdaten. Hier sind nur sehr wenige Daten vorhanden, denn die meisten der in den Molekülen vorkommenden Torsionswinkel sind ohne Referenzen gespeichert. Eine Verfeinerung von Torsionswinkelparametern scheint auf dieser Grundlage kaum angebracht, der Datensatz sollte unbedingt in diese Richtung erweitert werden. Torsionswinkel mit einer zentralen C-C Bindung werden gut reproduziert, größere Abweichungen gibt es nur bei kleinen Ringen als Folge von dort schlecht reproduzierten Bindungswinkeln. Abgesehen davon enthält der Datensatz noch einige Torsionswinkel für terminale OH Gruppen, deren Konformationen nur geringfügige Energiedifferenzen aufweisen und die deshalb für die Bewertung des Kraftfeldes nicht relevant sind.

Insgesamt liefert das Kraftfeld in der Gegenüberstellung mit den experimentellen Daten gute Übereinstimmungen. Neben einigen wenigen systematischen Abweichungen sind einige interne Koordinaten uneinheitlich im Datensatz vertreten, und besonders bei den Torsionswinkeln fehlt es an Referenzdaten. Einen Ausweg bietet hier nur die Nutzung von ab-initio-Methoden, die die entsprechenden Daten mit guter Genauigkeit berechnen können. Die Neuberechnung der Moleküle aus dem Datensatz mit diesen Methoden verspricht, die Vereinbarkeit von Kraftfeld und Referenzdaten zu verbessern, da hier nicht verschiedene Fehlerarten der zugrundeliegenden experimentellen Methoden vermischt werden.

Trotz der beschriebenen Probleme wurde eine Reparametrisierung des Kraftfeldes anhand der experimentellen Referenzdaten versucht. Die vorhandenen Parameter dienten dem Algorithmus als Startwert. Es konnten keine signifikant besseren Parameter gefunden werden, alle Änderungen spielen sich weit unterhalb des für das Kraftfeld akzeptierten Fehlerbereiches ab; Geometriedaten, die weit außerhalb der Referenz lagen, tun dies praktisch unverändert. Das Ergebnis verwundert nicht, da im Unterschied zu den Wasserstoffbrücken- und Partialladungsmodellen dieser Teil des Kraftfeldes über viele Jahre von Hand optimiert wurde. Trotzdem bleibt die Frage offen, ob der Algorithmus überhaupt funktioniert. Um dies zu klären, kann der Startparametersatz absichtlich verfälscht werden, um zu testen, ob eine den Originalparametern ähnliche Qualität erreicht werden kann. Eine weitere Möglichkeit ist es, die bei den Partialladungen beschriebene Methode zum Auffinden von Störungen, nämlich das schrittweise Vergrößern und Zusammenfassen der Referenzdaten, anzuwenden.

Verfälscht man Parameter zur Bindungslänge, so wird dies bei der Optimierung sehr schnell korrigiert. Dies ist damit zu erklären, dass es sich hierbei um ein einfach aufgebautes und sehr „hartes“ Potenzial handelt, welches nur wenig von anderen Kraftfeldtermen beeinflusst wird. Das Auslenken von Torsionswinkelparametern erscheint aufgrund der dünnen Datenlage nicht angebracht. Für die Bindungswinkel erscheint ein Experiment mit dem dreifach koordinierten Stickstoff geeignet. Dieser kann, je nachdem, ob er Teil eines π -Systems ist, planar (Bindungswinkel 120°) oder pyramidal (109°) vorliegen. Um diese Berechnung zu stören, wurde der Parameter q_N^0 (Formel 3.1.9) so angepasst, dass sich in jedem Fall ein „Sollwinkel“ von 120° ergibt. Der Parametrisierungsalgorithmus war in der Lage, den Parameter sicher wiederzufinden.

Der Ansatz an sich funktioniert folglich, die genannten Möglichkeiten zur Weiterentwicklung wurden bereits in Angriff genommen, sind jedoch nicht mehr Bestandteil dieser Arbeit.

8 Zusammenfassung und Ausblick

Kraftfelder sind ein vielseitiges Werkzeug zur schnellen Berechnung vielfältiger Moleküleigenschaften und spielen eine wichtige Rolle in der chemischen und pharmazeutischen Forschung. Da es sich um empirische Modelle handelt, ist eine Parametrisierung der zugrunde liegenden Gleichungen für das zu bearbeitende Problem erforderlich. Ein solches Modell ist immer eine Vereinfachung der realen Zusammenhänge; die Qualität der damit erhaltenen Vorhersagen ist deshalb auch ein Maß, wie gut die wichtigen Einflussgrößen verstanden und vor allem in das Modell integriert sind. Bei der Parametrisierung müssen viele Effekte gegeneinander ausbalanciert werden, da die Kraftfeldterme leider nicht unabhängig voneinander betrachtet werden können. Umfangreiche Testrechnungen sind erforderlich, um die notwendige Qualität der Parameter sicher zu stellen. Wird ein Bestandteil des Modells geändert oder kommen neue hinzu, so beginnt die Ausbalancierung von Neuem. Eine Automatisierung dieses Prozesses bringt nicht nur eine enorme Zeitersparnis, sie zwingt auch zur sorgfältigen Definition von Vorgaben und Qualitätskriterien, denn Optimierungsalgorithmen nutzen oft überraschende Lücken in der ihnen vorgegebenen Spezifikation. Die Formulierung einer Strategie in einem Programm anstelle von „intelligentem Raten“ fördert zudem ein tieferes Verständnis. Ein weiterer Vorteil ist, dass bei einer Änderung der Strategie nur das entsprechende Programm geändert werden muss und dem Entwickler die mühevollen Handarbeit erspart bleibt. Bei der Datenerfassung von Hand schleichen sich oft, trotz mehrfacher Kontrolle durch verschiedene Personen, Fehler ein. Wann immer Handeingabe notwendig ist, wurden deshalb automatische Methoden zur Plausibilitätsprüfung eingesetzt, erst so waren die Probleme durch Eingabefehler in den Griff zu bekommen. Die programmgesteuerte Erstellung aussagekräftiger Protokolle und Grafiken macht die Fülle der bei der Parametrisierung und Evaluierung eines Kraftfeldes anfallenden Informationen für den Benutzer überschaubar. Probleme und deren Zusammenhang können so leichter erfasst werden. Das Verstehen der Schwächen und Grenzen eines Modelles sowie die sich daraus ergebende Notwendigkeit, Entscheidungen zu treffen, kann kein Computerprogramm leisten; erst der Benutzer kann die gewonnenen Erkenntnisse in eine Änderung des Modells oder eine Anpassung der Parametrisierungsstrategie umsetzen, zumindest kann eine verbesserte Darstellung bei der programmgesteuerten Auswertung erreicht werden, um „versteckte“ Probleme hervorzuheben.

Für das MOMO-Kraftfeld konnten auf diese Weise verbesserte und neue Parameter für Wasserstoffbrücken abgeleitet werden, zwei empirische Punktladungsmodelle und deren Verträglichkeit mit zwei quantenchemischen Modellen verbessert und prinzipielle Probleme bei

deren Vereinbarkeit erkannt werden sowie die automatische Parametrisierung von Bindungslängen, Bindungswinkeln und Torsionswinkeln ermöglicht werden. Bei Letzterem konnte jedoch keine Verbesserung gegenüber den Originalparametern erreicht werden, was nicht weiter verwunderlich ist, da diese seit Jahrzehnten entwickelt worden sind, wohingegen Wasserstoffbrücken und Partialladungen erst später hinzugekommen sind und nicht so umfangreich wie die bindenden Kraftfeldterme getestet wurden.

Voraussetzung für die hier gewählte Vorgehensweise, alle Arbeiten weitgehend zu automatisieren und Strategien immer in Programme umzusetzen, waren sehr umfangreiche Programmierarbeiten, die unter anderem auch eine vollständige Neuimplementierung des MOMO-Kraftfeldes umfassen, nachdem dessen genaue Arbeitsweise aus dem zuvor analysiertem Quellcode der ursprünglichen Version ermittelt werden musste.

Ziel war es, auf einfache Weise die Steuerung des Kraftfeldes aus kleineren Programmen, die spezielle Probleme bearbeiten, zuzulassen. Dazu sollten die Ein- und Ausgabe chemischer und nichtchemischer Daten unterstützt werden, etablierte Standardverfahren aus Mathematik und Molecular Modelling, zur grafischen Darstellung und zur parallelen Berechnung von Problemen in Bibliotheken zur Verfügung stehen. Die kleinen Programme sollten in einer leicht zu erlernenden Sprache geschrieben werden können, die sogar interaktiv genutzt werden kann, da mächtige Bibliotheken alles, was sonst etwa in einer programmeigenen Kommandosprache eingegeben würde, als Eingabe für entsprechende Funktionen annehmen. Die Einfachheit einer eigenen Kommandosprache ist so mit der Flexibilität einer „echten“ Programmiersprache kombiniert. Alles würde dann idealerweise noch Lizenzbedingungen unterliegen, die eine problemlose Weitergabe an andere Nutzer erlauben.

Für einen Einzelnen ist es nicht möglich, dies im zeitlichen Rahmen einer Dissertation zu erreichen, glücklicherweise gibt es aber zahlreiche Open-Source-Projekte, die gemeinsam die gewünschte Funktionalität zur Verfügung stellen. So konnte der Aufwand auf die dazu passende Implementierung des MOMO-Kraftfeldes und das Verbinden mit der von diesen Projekten bereitgestellten Software beschränkt werden. Den gemeinsamen Nenner bildet die Programmiersprache Python. Es handelt sich um eine leicht erlernbare interpretierte Sprache, die mit in Compilersprachen geschriebenen Bibliotheken erweitert werden kann und durch mächtige Funktionen die schnelle Erstellung von kleinen problemorientierten Programmen und sogar eine interaktive Nutzung gestattet. MOMO und fast alle eingesetzte Software macht sich diese

Besonderheiten zunutze. Der Kern des MOMO-Kraftfeldes wurde aus Geschwindigkeitsgründen in der Compilersprache C geschrieben, Datenein- und -ausgabe und die Programme zur Parametrisierung und Auswertung wurden in Python geschrieben.

Die damit erreichte Kombination aus der Flexibilität einer „echten“ Programmiersprache mit einer Kommandosprache erwies sich als so vorteilhaft, dass Forschungspraktikanten ohne Programmierkenntnisse in kurzer Zeit an die Erstellung einfacher Programme für spezielle Problemstellungen herangeführt werden konnten. Berechnung und (grafische) Auswertung konnten so teilweise in wenigen Programmzeilen kombiniert werden. Dies hat mich schließlich dazu bewogen, ein Programmierpraktikum im Rahmen des Masterstudiengangs anzubieten. Die Studenten wurden dabei schrittweise mit der Programmiersprache Python und den für das Aufgabengebiet Molecular Modelling wichtigen Bibliotheken und MOMO vertraut gemacht. In einem dreitägigen Abschlusspraktikum konnte ein Programm zur supramolekularen Konformationsanalyse implementiert werden.

Entwicklungspotenzial besteht beim Parametrisierungsprogramm für die bindenden Wechselwirkungen. Hier könnte, ähnlich wie bei den übrigen Programmen, noch die automatische Berechnung mit ab-initio-Methoden ergänzt werden; auf diese Weise ließen sich auch zusätzliche Stützwerte für die Torsionspotenziale bestimmen. Daneben kann das Programmpaket um eine grafische Benutzeroberfläche ergänzt werden. Eine Erweiterung des Kraftfeldes um ein Solvationsmodell wurde bereits in einer anderen Dissertation in Angriff genommen; eine Erweiterung um die Möglichkeit, auch mit periodischen Randbedingungen in allen kristallografischen Gittertypen rechnen zu können, erscheint vielversprechend. Eine auf ein einziges Problem beschränkte automatische Parametrisierung anhand von quantenchemischen Daten würde die Anwendung von Suchverfahren, für die allgemeine Kraftfelder nicht genau genug sind, ermöglichen. Die Nutzung thermodynamischer Referenzdaten könnte schließlich noch die Parametrisierung des van-der-Waals- und des Coulomb-Potenzials verbessern.

Ich hoffe, dass die im Rahmen dieser Arbeit erstellten und zusammengetragenen Programme und die dazugehörigen Anwendungen als Anregung dienen mögen und die Arbeit bei der Lösung neuer Probleme vereinfachen.

Literaturverzeichnis

- [Abraham1988#45] ABRAHAM, R. J. & GRANT, G.H. (1988): *Charge Calculations in Molecular Mechanics. V. Silicon Compounds and π -Bonding*. *J. Comput. Chem.*, 9, 244.
- [Abraham1992] ABRAHAM, R. J. & GRANT, G.H. (1992): *Charge calculations in molecular mechanics. IX. - A general parameterisation of the scheme for saturated halogen, oxygen and nitrogen compounds*. *J. Comput.-Aided Mol. Des.*, 6, 273.
- [Abraham1984] ABRAHAM, R. J. & HUDSON, B. (1984): *Approaches to Charge Calculations in Molecular Mechanics. 2 - Resonance Effects in Conjugated Systems*. *J. Comput. Chem.*, 5, 562.
- [Abraham1985] ABRAHAM, R. J. & HUDSON, B. (1985): *Charge Calculations in Molecular Mechanics. Amino Acids and Peptides*. *J. Comput. Chem.*, 6, 173.
- [Abraham1988#46] ABRAHAM, R. J. & SMITH, P.E. (1988): *Charge calculations in molecular mechanics 6: the calculation of partial atomic charges in nucleic acid bases and the electrostatic contribution to DNA base pairing*. *Nucleic Acids Res.*, 16, 2639.
- [Abraham1988] ABRAHAM, R. J. & SMITH, P.E. (1988): *Charge Calculations in Molecular Mechanics IV: A General Method for Conjugated Systems*. *J. Comput. Chem.*, 9, 288.
- [Abraham1989] ABRAHAM, R. J. & SMITH, P.E. (1989): *Charge calculations in molecular mechanics 7: Application to polar zc systems incorporating nitro, cyano, amino, C=S and thio substituents*. *J. Comput.-Aided Mol. Des.*, 3, 175.
- [Abraham1991] ABRAHAM, R. J., GRANT, G. H., HAWORTH, I. S. & SMITH, P.E. (1991): *Charge calculations in molecular mechanics. Part 8: Partial atomic charges from classical calculations*. *J. Comput.-Aided Mol. Des.*, 5, 21.
- [Abraham1982] ABRAHAM, R. J., GRIFFITHS, L. & LOFTUS, P. (1982): *Approaches to Charge Calculations in Molecular Mechanics*. *J. Comput. Chem.*, 3 (3), 407.
- [Allen2002] ALLEN, F. H. (2002): *The Cambridge Structural Database: a quarter of a million crystal structures and rising*. *Acta Cryst.*, B58, 380.
- [Allinger1996#53] ALLINGER, N. L., CHEN, K. & LIU, J. (1996): *An Improved Force Field (MM4) for Saturated Hydrocarbons*. *J. Comput. Chem.*, 17, 642.
- [Atkins2001] ATKINS, P. W. (2001): *Physikalische Chemie*. VCH Weinheim.
- [Bader1991] BADER, R. F. W. (1991): *A Quantum Theory of Molecular Structure and Its Applications*. *Chem. Rev.*, 91, 893.
- [Beck1989] BECK, H. (1989): *MOMO - Ein Beitrag zur Weiterentwicklung des Molecular Modelling*, Universität Göttingen.
- [Berman2000] BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P.E. (2000): *The Protein Data Bank*. *Nucleic Acids Res.*, 28, 235.
- [Bray] BRAY, T., PAOLI, J., SPERBERG-McQUEEN, C. M., MALER, E. & YERGEAU, F. ; *Extensible Markup Language (XML) 1.0 (Fourth Edition)*. <http://www.w3.org/TR/2006/REC-xml-20060816/#sec-origin-goals>.
- [CHELPG1989] BRENNEMAN, M. C. & WIBERG, K.B. (1989): *Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis*. *J. Comp. Chem.*, 11, 361.
- [Brooks1983] BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J. & SWAMINATHAN, S.K.M. (1983): *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations*. *J. Comput. Chem.*, 4, 187.
- [CHELP1987] CHIRLIAN, L. E. & FRANCL, M.M. (1987): *Atomic Charges Derived from Electrostatic Potentials: A Detailed Study*. *J. Comp. Chem.*, 8, 894.
- [Coulson1947] COULSON, C. A. & LONGUET-HIGGINS, H.C. (1947): *The Electronic Structure of Conjugated Systems. I. General Theory*. *Proc. R. Soc. London, Ser. A*, 191A, 39.

- [Dakin1965] DAKIN, R. (1965): *A tree-search algorithm for mixed integer programming problems. The Computer Journal*, 8, 250.
- [De_Jong1975] DE JONG, K. A. (1975): *An Analysis of Behavior of a Class of Genetic Adaptive Systems*, University of Michigan.
- [Deb2000] DEB, K. (2000): *An efficient constraint handling method for genetic algorithms. Comput. Methods Appl. Mech. Engrg.*, 186, 311.
- [DeLano] DELANO, W. L. ; *"The PyMOL Molecular Graphics System."* DeLano Scientific LLC, San Carlos, CA, USA. <http://www.pymol.org>.
- [Dinur1995] DINUR, U. & HAGLER, A.T. (1995): *Geometry-Dependent Atomic Charges: Methodology and Application to Alkanes, Aldehydes, Ketones, and Amides. J. Comput. Chem.*, 16, 154.
- [Dudek1995] DUDEK, M. J. & PONDER, J.W. (1995): *Accurate Modeling of the Intramolecular Electrostatic Energy of Proteins. J. Comput. Chem.*, 16, 791.
- [Eads2008] EADS, D. & ROSTEN, E. (2008): *Experiences Using SciPy for Computer Vision Research. Proceedings of the 7th Python in Science Conference*, 0, 22.
- [Fleming1998] FLEMING, P. J. & FONSECA, C.M. (1998): *Multiobjective optimization and multiple constraint handling with evolutionary algorithms. IEEE Trans. Syst. Man Cybern. Part A Syst. Humans*, 28, 26.
- [Fletcher1987] FLETCHER, R. (1987): *Practical Methods of Optimization*. John Wiley & Sons, Chichester.
- [Fletcher1963] FLETCHER, R. & POWELL, M.J.D. (1963): *A Rapidly Convergent Descent Method for Minimization. The Computer Journal*, 6, 163.
- [Fletcher1964] FLETCHER, R. & REEVES, C.M. (1964): *Function minimization by conjugate gradients. The Computer Journal*, 7, 149.
- [Fogel1966] FOGEL, J., OWENS, A. J. & WALSH, M.J. (1966): *Artificial intelligence through simulated evolution*. John Wiley & Sons, New York.
- [Fonseca1995] FONSECA, C. M. & FLEMING, P.J.F. (1995): *An Overview of Evolutionary Algorithms in Multiobjective Optimization. Evolutionary Computation*, 3, 1.
- [Gasteiger1980] GASTEIGER, J. & MARSILI, M. (1980): *Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. Tetrahedron*, 36, 3219.
- [Gammel1999] GEMMEL, E. (1999): *Entwicklung Evolutionärer Algorithmen zur Konformationsanalyse und Parametrisierung für das Kraftfeldprogramm MOMO*, Universität Frankfurt am Main.
- [Goldberg1991] GOLDBERG, D. (1991): *What Every Computer Scientist Should Know About Floating-Point Arithmetic. ACM Comput. Surv.*, 23, 5-48.
- [Grefenstette1987] Goldberg, D. E. & Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. In J. J. Grefenstette (Ed.), *Genetic Algorithms and Their Application: Proceedings of the Second International Conference of Genetic Algorithms* (p. 41). Lawrence Erlbaum, Hillsdale.
- [Halgren1996] HALGREN, T. A. (1996): *Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. J. Comput. Chem.*, 17, 490.
- [Hinsen1997] HINSEN, K. (1997): *The Molecular Modeling Toolkit: a case study of a large scientific application in Python. Proceedings of the 6th International Python Conference*, <http://www.python.org/workshops/1997-10/proceedings/hinsen.html>, 1.
- [Hinsen1998] HINSEN, K. (1998): *Analysis of Domain Motions by Approximate Normal Mode Calculations. Proteins: Struct. Funct. Genet. Bioinf.*, 33, 417.
- [Hinsen2000] HINSEN, K. (2000): *The Molecular Modelling Toolkit: A New Approach to Molecular Simulations. J. Comp. Chem.*, 21, 79.

- [Hinsen1999] HINSEN, K., THOMAS, A. & FIELD, M.J. (1999): *Analysis of Domain Motions in Large Proteins. Proteins: Struct. Funct. Genet. Bioinf.*, 34, 369.
- [Hinze1962] HINZE, J. & JAFFE, H.H. (1962): *Electronegativity. I. Orbital Electronegativity of Neutral Atoms. J. Am. Chem. Soc.*, 84, 540.
- [Hirshfeld1977] HIRSHFELD, F. L. (1977): *Bonded-Atom Fragments for Describing Molecular Charge Densities. Theoret. Chim. Acta*, 44, 129.
- [Holland1995] HOLLAND, J. H. (1995): *Adaption in natural and artificial systems - an introductory analysis with applications to biology, control, and artificial intelligence.* MIT Press, Cambridge USA.
- [Humphrey1996] HUMPHREY, W., DALKE, A. & SCHULTEN, K. (1996): *VMD: Visual Molecular Dynamics. J. Mol. Graphics Modell.*, 14, 33.
- [Hunter2007] HUNTER, J. D. (2007): *Matplotlib: A 2D Graphics Environment. Comp. Sci. Eng.*, 9, 90.
- [Jakob04] JAKOB, W. (2004): *Eine neue Methodik zur Erhöhung der Leistungsfähigkeit Evolutionärer Algorithmen durch die Integration lokaler Suchverfahren*, Universität Karlsruhe.
- [Yongho_Kim2002] KIM, Y., KIM, H. & PARK, K. (2002): *Theoretical Studies for Strong Hydrogen Bonds in Trimethyl Phosphate-(HNO₃)_n Complexes, n=1-3. Bull. Korean Chem.Soc.*, 23(12), 1811.
- [Kirkpatrick1983] KIRKPATRICK, S., GELATT, C. D. J. & VECCHI, M.P. (1983): *Optimization by Simulated Annealing. Science*, 220, 671.
- [Kleinoeder2005] KLEINÖDER, T. (2005): *Prediction of Properties of Organic Compounds Empirical Methods and Management of Property Data*, Universität Erlangen-Nürnberg.
- [Kollman1984] KOLLMAN, P. A. & SINGH, U.C. (1984): *An Approach to Computing Electrostatic Charges for Molecules. J. Comp. Chem.*, 5, 129.
- [Langer2005] LANGER, H. (2005): *Extended Evolutionary Algorithms for Multiobjective and Discrete Design Optimization of Structures*, TU München.
- [Langtangen2004] LANGTANGEN, H. P. (2004): *Python Scripting for Computational Science.* Springer-Verlag Berlin, Heidelberg.
- [Leach2001] LEACH, A. R. (2001): *Molecular Modelling, Principles and Applications.* Pearson Education Limited.
- [Lendvay1998] LENDVAY, G. & MAYER, I. (1998): *Some difficulties in computing BSSE-corrected potential surfaces of chemical reactions. Chem. Phys. Lett.*, 297, 365.
- [Lewars2003] LEWARS, G. E. (2003): *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics.* Springer Netherlands.
- [Lundh2005] LUNDH, F. ; *The cElementTree Module.* <http://effbot.org/zone/celementtree.htm>.
- [Maple1988] MAPLE, J. R., DINUR, U. & HAGLER, T.A. (1988): *Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. Proc. Natl. Acad. Sci. USA*, 85, 5350.
- [Maple1994] MAPLE, J. R., HWANG, M. J., STOCKFISCH, T. P., DINUR, U., WALDMAN, M., EWIG, S. C. & HAGLER, A.T. (1994): *Derivation of Class II Force Fields. I. - Methodology and Quantum Force Field for the Alkyl Functional Group and Alkane Molecules. J. Comput. Chem.*, 15 (2), 162.
- [Metropolis1953] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N. & TELLER, A.H. (1953): *Equation of State Calculations by Fast Computing Machines. J. Chem. Phys.*, 21, 1087.
- [Michalewicz1999] MICHALEWICZ, Z. (1999): *Genetic Algorithms + Data Structures = Evolution Programs.* Springer Verlag, Berlin.
- [Monz1997] MONZ, S. (1997): *Test und Entwicklung von Methoden zur Bestimmung atomarer Punktladungen*, Universität Frankfurt am Main.
- [Mulliken1955] MULLIKEN, R. S. (1955): *Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. J. Chem. Phys.*, 23, 1833.

- [Mulliken1962] MULLIKEN, R. S. (1962): *Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of LCAO-MO Population Analysis*. *J. Chem. Phys.*, 36, 3428.
- [Murray-Rust1999] MURRAY-RUST, P. & RZEPA, H.S. (1999): *Chemical Markup, XML, and the Worldwide Web. 1. Basic Principles*. *J. Chem. Inf. Comput. Sci.*, 39, 928.
- [Nelder1965] NELDER, J. A. & MEAD, R. (1965): *A Simplex Method for Function Minimization*. *The Computer Journal*, 7, 308.
- [Norrby2001] NORRBY, P. & BRANDT, P. (2001): *Deriving force field parameters for coordination complexes*. *Coord. Chem. Rev.*, 212, 79.
- [Norrby1998] NORRBY, P. & LILJEFORS, T. (1998): *Automated Molecular Mechanics Parameterization with Simultaneous Utilization of Experimental and Quantum Mechanical Data*. *J. Comput. Chem.*, 19, 1166.
- [Ortega1996] ORTEGA, J. M. & ROCKOFF, M.L. (1996): *Nonlinear difference equations and Gauss-Seidel type iterative methods*. *SIAM J. Numer. Anal.*, 3, 497.
- [NumRecipes] PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W.T. (1989): *Numerical Recipes*. Cambridge University Press.
- [Quinkert1995] QUINKERT, G., EGERT, E. & GRIESINGER, C. (1995): *Aspekte der Organischen Chemie - Struktur*. Verlag Helvetica Chimica Acta, Basel; VCH Weinheim.
- [Rechenberg1973] RECHENBERG, I. (1973): *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog, Stuttgart-Bad Cannstatt.
- [Reed1985] REED, A. E., WEINSTOCK, R. B. & WEINHOLD, F. (1985): *Natural population analysis*. *J. Chem. Phys.*, 83, 735.
- [Rew1990] REW, R. & DAVIS, G. (1990): *NetCDF: an interface for scientific data access*. *IEEE Comput. Graphics Appl.*, 10, 76.
- [Sigfridsson1998] SIGFRIDSSON, E. & RYDE, U. (1998): *Comparison of Methods for Deriving Atomic Charges from the Electrostatic Potential and Moments*. *J. Comput. Chem.*, 19, 377.
- [PIMM1991] SMITH, A. E. & LINDNER, H.J. (1991): *π -SCF Molecular Mechanics PIMM - Parameters, Formulation, Applications*. *J. Comput.-Aided Mol. Des.*, 5, 235.
- [Smith1951] SMITH, R. P., REE, T., MAGEE, J. L. & EYRING, H. (1951): *The Inductive Effect and Chemical Reactivity. I. General Theory of the Inductive Effect and Application to Electric Dipole Moments of Haloalkanes*. *J. Am. Chem. Soc.*, 73, 2263.
- [Steiner2002] STEINER, T. (2002): *Die Wasserstoffbrücke im Festkörper*. *Angew. Chem.*, 114, 50.
- [Stephens1994] STEPHENS, P. J., DEVLIN, F. J., CHABALOWSKI, C. F. & FRISCH, M.J. (1994): *Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields*. *J. Phys. Chem.*, 98, 11623.
- [Taylor1981] TAYLOR, R. & KENNARD, O. (1981): *Crystallographic Evidence for the Existence of C-H-O, C-H-N, and C-H-Cl Hydrogen Bonds*. *J. Am. Chem. Soc.*, 104, 5063.
- [Thomas1999] THOMAS, A., HINSEN, K., FIELD, M. J. & PERAHIA, D. (1999): *Tertiary and Quaternary Conformational Changes in Aspartate Transcarbamylase: A Normal Mode Study*. *Proteins: Struct. Funct. Bioinf.*, 34, 96.
- [Gaussian03] TRUCKS, G. W., FRISCH, A. & FRISCH, M.J. (2003): *Gaussian 03*. Gaussian, Inc., Pittsburg PA., 2003.
- [van_Rossum1995] VAN ROSSUM, G. ; *Python Reference Manual*. <http://www.python.org>.
- [Vanovschi2005] VANOVSCHI, V. ; *Parallel Python*. <http://www.parallelpython.com>.
- [Vedani1985] VEDANI, A. & DUNITZ, J.D. (1985): *Lone-Pair Directionality in Hydrogen Bond Potential Functions for Molecular Mechanics Calculations: The Inhibition of Human Carbonic Anhydrase II by Sulfonamides*. *J. Am. Chem. Soc.*, 107, 7653.
- [Vogt2004] VOGT, J. & VOGT, N. (2004): *Structure Searching in the MOGADOC Database*. *J. Mol.*

Struct., 695, 237.

[GW2004] WAGNER, G. (2004): *Test und Bewertung atomarer Punktladungsmodelle im Kraftfeldprogramm MOMO*, Universität Frankfurt am Main.

[Wolpert1997] WOLPERT, D. H. & MACREADY, W.G. (1997): *No Free Lunch Theorems for Optimization. IEEE Trans. Evol. Comput.*, 1 (1), 67.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eidesstatt, dass die vorliegende Arbeit von mir eigenhändig und nur unter Verwendung der genannten Hilfsmittel und Literatur angefertigt wurde.

Frankfurt am Main, den

Guido Wagner

Kastanienstraße 8; 61352 Bad Homburg v.d.H.



Lebenslauf

Angaben zur Person

Geburtsdatum/ -ort 17.3.1977, Frankfurt am Main
Familienstand ledig
Staatsangehörigkeit deutsch

Ausbildung

1983 - 1987 Besuch der Grundschule in Bad Homburg/Ober-Erlenbach
1987 - 28.6.1996 Besuch der Kaiserin-Friedrich-Schule in Bad Homburg,
Schulabschluss: Abitur
1.7.1996 - 30.4.1997 Grundwehrdienst
15.4.1997 Beginn des Chemiestudiums an der Goethe-Universität Frankfurt a. M.
24.5.2000 Diplom-Chemiker-Vorprüfung
22.3.2004 Diplom-Chemiker-Hauptprüfung
Thema der Diplomarbeit: „Test und Bewertung empirischer
Ladungsmodelle im Kraftfeldprogramm MOMO“
1.4.2004 Beginn der Dissertation „Entwicklung und Implementierung einer Prozedur
zur automatischen Parametrisierung des MOMO-Kraftfeldes“ unter
Anleitung von Herrn Prof. Dr. Ernst Egert

Akademische Lehrer

Anorganische Chemie Prof. Dr. N. Auner, Prof. Dr. B. O. Kolbesen, Prof. Dr. M. U. Schmidt, Prof.
Dr. Ferdinand Schüth, Prof. Dr. M. Wagner
Organische Chemie Prof. Dr. Ernst Egert, Prof. Dr. Joachim Engels, Prof. Dr. Michael Göbel,
Prof. Dr. Christian Griesinger, Prof. Dr. Dieter Rehm
Physikalische Chemie Prof. Dr. B. Brutschy, Prof. Dr. Karl Hensen, Prof. Dr. T. Prisner
Physik Prof. Dr. Horst Schmidt-Böcking, Prof. Dr. Bruno Lüthi, Prof. Dr. Deitinghoff
Mathematik Prof. Dr. Kohlmaier
Informatik Apl. Prof. Dr. Rüdiger Brause, Prof. Dr. Oswald Drobnik, Prof. Dr. Georg
Schnitger, Prof. Dott. Ing. Roberto Zicari
Eingeschränkte Sachkunde nach § 5 Abs. 2 der Chemikalien-Verbots-Verordnung
Prof. Dr. C. Jochum, Prof. Dr. G. Schmalzing

Frankfurt am Main, den 17.12.09
