

**Quantenchemische Programmentwicklung:  
Automatische Erzeugung von Ableitungen  
und  
Parametrisierung semiempirischer Methoden**

Dissertation

zur Erlangung der naturwissenschaftlichen Doktorwürde  
(Dr. sc. nat.)  
vorgelegt der  
Mathematisch-naturwissenschaftlichen Fakultät  
der  
Universität Zürich

von  
Rainer Steiger Kohler  
aus  
Altstätten SG

Begutachtet von  
Prof. Dr. Jürg Hutter  
Prof. Dr. Walter Thiel

Zürich 2003

Die vorliegende Arbeit wurde von der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich auf Antrag von Prof. Dr. Jürg Hutter und Prof. Dr. John A. Robinson als Dissertation angenommen.

*L'essentiel est invisible pour les yeux.  
Petit Prince, Antoine de Saint-Exupéry*



## Danksagung

Herrn Prof. Dr. Walter Thiel möchte ich an dieser Stelle ganz herzlich danken. Einerseits daher, weil ich eine Gelegenheit bekam, mich in ein relativ unbekanntes Gebiet einzuarbeiten und meiner Begeisterung für das Programmieren nachkommen konnte. Andererseits auch dafür, dass ein sehr grosses Vertrauen mir entgegengebracht wurde, einerlei wo ich bin und arbeite. Danke!

Desweiteren möchte ich mich bei Herrn Prof. Dr. Jürg Hutter bedanken, dass ich meine Arbeit bei ihm an der Universität Zürich einreichen konnte.

Zudem gilt mein Dank Herrn Prof. Dr. Bruno Lang für die schöne Zeit in Aachen und die gute Zusammenarbeit bei dem Projekt ADIFOR & MNDO.

Danken möchte ich auch den Kollegen aus der Zürcher-Zeit, insbesondere Dr. Max Holthausen als auch Nacit Dileme. Aus der Mülheimer-Zeit gilt mein Dank insbesondere Martin Graf und Dr. Mavinahalli N. Jagadeesh. Jagadeesh werde ich als einen sehr herzlichen Roommate in Erinnerung behalten und Martin danke ich für alles, privat wie auch wissenschaftlich. Er hat die Mülheimer-Zeit schön gemacht. Dr. Jürgen Breidung, in Zürich als auch in Mülheim, möchte ich vielmals für die sehr interessanten Gespräche danken. Dr. Axel Koslowski danke ich für den besten Hochzeitsspruch als auch für die unzähligen Hilfsleistungen, sei es in Latex, Fortran oder was auch immer. Axel weiss alles. Horst Lenk gebührt grossen Dank für die hervorragende Systemadministration.

Dank gebührt natürlich auch allen weiteren Kolleginnen und Kollegen, die bisher unerwähnt blieben. Ihr werdet in guter Erinnerung bleiben.

Ohne meine Eltern, Elsbeth und Linus, und meiner Schwester Doris, wären all die bisherigen und nachfolgenden Zeilen nicht möglich gewesen. Für die finanzielle als auch menschliche Unterstützung danke ich ganz fest und möchte an dieser Stelle diese Dissertation Euch widmen.

Ein Riesendank gebührt Evelyne, meiner Geliebten und Frau. Ohne Evelyne wäre alles anders, und das wäre schade. Danke für die Zeit, die war und die kommen wird. Freue mich auf eine Zukunft mit Dir und möchte diese Dissertation auch Dir widmen.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Automatische Erzeugung von Ableitungen</b>	<b>3</b>
2.1	Vorbemerkung . . . . .	3
2.2	Einleitung . . . . .	3
2.3	Numerische Ableitungen . . . . .	5
2.4	Analytische Ableitungen . . . . .	7
2.5	Symbolische Differentiation . . . . .	8
2.6	Automatische Differentiation . . . . .	8
2.6.1	Software zum automatischen Differenzieren . . . . .	10
2.6.2	ADIFOR . . . . .	11
2.7	Analytische Gradienten für MNDO97 . . . . .	14
2.7.1	MNDO97 . . . . .	15
2.7.2	Programmrestrukturierung . . . . .	16
2.7.3	Erzeugung und Implementierung von ADIFOR-Code . . . . .	16
2.8	Resultate . . . . .	19
2.8.1	MNDO97s: Testphase . . . . .	19
2.8.2	Genauigkeit der Gradienten . . . . .	20
2.8.3	Zeitmessungen . . . . .	24
2.8.4	Memoryanforderungen . . . . .	27
2.9	Zusammenfassung . . . . .	29
<b>3</b>	<b>Parametrisierung semiempirischer Methoden</b>	<b>31</b>
3.1	Einleitung . . . . .	31
3.2	Entwicklung semiempirischer Methoden: Übersicht . . . . .	32
3.2.1	Auswahl der Referenzdaten . . . . .	33
3.2.2	Gewichtung der Referenzdaten . . . . .	34
3.2.3	Optimierung der Parameter . . . . .	34
3.3	Optimierungsmethoden . . . . .	37
3.3.1	Klassifizierung . . . . .	37
3.3.2	Evolutionäre Algorithmen . . . . .	38
3.3.3	Genetische Algorithmen . . . . .	38

3.3.3.1	Prinzipielle Funktionsweise . . . . .	38
3.3.3.2	Repräsentation der Parameter . . . . .	40
3.3.3.3	Initialisierung der Populationen . . . . .	40
3.3.3.4	Evaluierung der Fitness . . . . .	40
3.3.3.5	Selektionsschemata . . . . .	41
3.3.3.6	Genetische Operatoren . . . . .	41
3.3.3.7	Genetics at work . . . . .	42
3.4	Das Programm gps1 . . . . .	42
3.4.1	Systematische Suche . . . . .	44
3.4.2	Stochastische Suche mit genetischen Algorithmen . . . . .	45
3.5	Übersicht zu den durchgeführten Parametrisierungen . . . . .	46
3.6	Reparametrisierung von MNDO . . . . .	48
3.6.1	Systematische Suche . . . . .	48
3.6.2	Suche mit genetischen Algorithmen . . . . .	55
3.6.3	Auswahl der Parameter . . . . .	56
3.6.4	Resultate und Diskussion . . . . .	56
3.6.4.1	Standardbildungsenthalpien . . . . .	58
3.6.4.2	Geometrien . . . . .	59
3.6.4.3	Ionisierungspotentiale . . . . .	61
3.6.4.4	Dipolmomente . . . . .	61
3.6.4.5	Schwingungen . . . . .	61
3.6.4.6	Ionen und Radikale . . . . .	62
3.6.5	Wertung . . . . .	62
3.6.6	Tabellen mit detaillierten Ergebnissen . . . . .	63
3.7	Parametrisierung von Fluor für OM2 und OM3 . . . . .	73
3.7.1	Der theoretische Ansatz in OM2 und OM3 . . . . .	73
3.7.2	Parametrisierungsverlauf . . . . .	76
3.7.3	Optimierte Parameterwerte . . . . .	78
3.7.4	Ergebnisse und Diskussion . . . . .	86
3.7.4.1	Referenzdatensatz: Übersicht . . . . .	86
3.7.4.2	Referenzdatensatz: Standardbildungsenthalpien . . . . .	88
3.7.4.3	Referenzdatensatz: Geometrien . . . . .	90
3.7.4.4	Referenzdatensatz: Ionisierungspotentiale . . . . .	92
3.7.4.5	Referenzdatensatz: Dipolmomente . . . . .	92
3.7.4.6	Referenzdatensatz: Schwingungen . . . . .	94
3.7.4.7	Validierungssatz A . . . . .	95
3.7.4.8	Validierungssatz B . . . . .	96
3.7.4.9	Validierungssatz C . . . . .	99
3.7.4.10	Zusammenfassung . . . . .	101
3.7.5	Tabellen mit Einzelergebnissen . . . . .	101



---

<b>A Programm-Listing</b>	<b>135</b>
<b>B Beispiel für den Vorwärtsmodus</b>	<b>137</b>
<b>C Programmbeschreibung von Parma</b>	<b>139</b>
<b>D Technische Anmerkungen zum Programm gps1</b>	<b>141</b>
D.1 Kompilation . . . . .	141
D.2 Programmaufruf . . . . .	141
D.3 Programmkonzept . . . . .	142
D.3.1 Programmübersicht . . . . .	143
D.3.2 Masterfile . . . . .	143
D.3.3 Erzeugung von Inputfiles . . . . .	144
D.3.4 SSQ-Berechnung . . . . .	144
D.4 Jobberechnung . . . . .	145
D.5 Diversa . . . . .	146
D.6 Technische Probleme . . . . .	148
D.7 Schlussbemerkungen . . . . .	149
<b>Literaturverzeichnis</b>	<b>149</b>
<b>Lebenslauf</b>	<b>159</b>



# Kapitel 1

## Einleitung

In der Theoretischen Chemie gibt es eine Reihe allgemeiner Techniken, die in vielen Teilbereichen eingesetzt werden können und die daher in verschiedenen Rechenprogrammen oft zu unterschiedlichen Zwecken implementiert worden sind. Hierzu gehören die Bildung von Ableitungen durch Differenzieren einer vorgegebenen Funktion und die Optimierung von Parametern bei Modellierungsaufgaben. Die vorliegende Arbeit befasst sich mit diesen beiden Themenkreisen im Zusammenhang mit der Weiterentwicklung semiempirischer Verfahren in der Quantenchemie.

Im ersten Teil werden die Methoden des automatischen Differenzierens angewendet, um aus einem vorhandenen semiempirischen Code einen neuen Code zu erzeugen, der die analytische erste Ableitung der Energie bezüglich der Kernkoordinaten berechnen kann. Dieser automatisch generierte Code wird hinsichtlich der Präzision der Ergebnisse und hinsichtlich des Rechenaufwandes mit den verfügbaren Alternativen verglichen (analytischer Gradient bei etablierten semiempirischen Verfahren, numerischer Gradient bei neueren Verfahren), um die Vorteile und Nachteile der Technik des automatischen Differenzierens bei quantenchemischen Programmen auszuloten.

Im zweiten Teil wird ein neues Programmsystem vorgestellt, mit dem die Parametrisierung semiempirischer Verfahren systematisiert und automatisiert werden kann. Es erlaubt zum einen eine systematische und iterative Durchsuchung des Parameterraums, jeweils kombiniert mit lokaler gradientenbasierter Optimierung günstiger Parametersätze, und zum anderen den Einsatz genetischer Algorithmen, um eine möglichst globale Optimierung vorzunehmen. Die Leistungsfähigkeit dieses neuen Programmsystems wird anhand von zwei Anwendungsbeispielen aufgezeigt, welche die testweise Reparametrisierung des etablierten MNDO Verfahrens und die Herleitung der bislang nicht verfügbaren Fluorparameter für die neueren OM2 und OM3 Verfahren betreffen.



# Kapitel 2

## Automatische Erzeugung von Ableitungen

### 2.1 Vorbemerkung

Dieses Kapitel beschreibt die Erstellung analytischer Gradienten in dem semiempirischen MNDO99 Programm mittels automatischem Differenzieren. Die folgenden Abschnitte schildern zunächst die zugrunde liegende Theorie und die programmtechnische Implementierung. Danach werden die erzielten Resultate besprochen, insbesondere hinsichtlich Präzision, Zeit- und Hauptspeicherbedarf. Dieses Projekt wurde in Zusammenarbeit mit PD B. Lang, RWTH Aachen, durchgeführt.

### 2.2 Einleitung

Ausgehend von der Schrödinger-Gleichung sind heutzutage viele Eigenschaften von Molekülen mit dem Computer berechenbar. Oft sind diese Eigenschaften als Ableitungen definiert [1]. Dies gilt insbesondere dann, wenn der Einfluss eines Systemparameters  $x$  auf die Observable  $f$  durch eine Taylor-Entwicklung dargestellt werden kann:

$$f(x_0 + \Delta x) = f(x_0) + \left. \frac{\partial f}{\partial x} \right|_{x_0} \Delta x + \frac{1}{2} \Delta x \left. \frac{\partial^2 f}{\partial x^2} \right|_{x_0} \Delta x + O^3 \quad (2.1)$$

So kann beispielsweise die Energie  $E$  beim Anlegen eines homogenen elektrischen Feldes  $\mathbf{F}$  nach den Feldkomponenten  $F_i (i = x, y, z)$  entwickelt werden. In Matrix-

schreibweise:

$$E(\mathbf{F}) = E_0 - \mathbf{F}^T \boldsymbol{\mu} - \frac{1}{2} \mathbf{F}^T \boldsymbol{\alpha} \mathbf{F} + O(\mathbf{F}^3) \quad (2.2a)$$

$$\mu_i = - \left. \frac{\partial E}{\partial F_i} \right|_0$$

$$\alpha_{ij} = - \left. \frac{\partial^2 E}{\partial F_i \partial F_j} \right|_0 \quad (2.2b)$$

Die angegebenen Ableitungen sind die Elemente des Dipolmomentvektors  $\boldsymbol{\mu}$  und des Polarisierbarkeitstensors  $\boldsymbol{\alpha}$ . Gemäss 2.2a beschreiben sie die Wechselwirkung des Systems mit einem externen elektrischen Feld in erster bzw. zweiter Ordnung.

Wird die Gleichung 2.1 auf die Born-Oppenheimer Potentialenergie-Hyperfläche am Punkt  $\mathbf{x} = \mathbf{x}_0$  (Koordinationspunkt) angewandt, so erhält man:

$$E(\mathbf{x}) = E(\mathbf{x}_0) + \Delta \mathbf{x}^T \mathbf{g} + \frac{1}{2} \Delta \mathbf{x}^T \mathbf{H} \Delta \mathbf{x} + O(\Delta \mathbf{x}^3) \quad (2.3a)$$

$$g_i = \left. \frac{\partial E}{\partial x_i} \right|_{\mathbf{x}_0} \quad (2.3b)$$

$$H_{ij} = \left. \frac{\partial^2 E}{\partial x_i \partial x_j} \right|_{\mathbf{x}_0} \quad (2.3c)$$

Hier bezeichnet  $\mathbf{x}$  den Vektor der Kernkoordinaten, der bei Verwendung kartesischer Koordinaten für  $N$  Atome  $3N$  Komponenten aufweist. Die ersten und zweiten Ableitungen der Energie nach den Kernkoordinaten definieren die Gradientenkomponenten  $g_i$  (2.3b, Kräfte) und die harmonischen Kraftkonstanten  $H_{ij}$  (2.3c, 'Hessian'). Diese geometrischen Ableitungen sind von zentraler Bedeutung für die Lokalisierung und Charakterisierung stationärer Punkte auf der Born-Oppenheimer Potentialfläche. Jede effiziente Geometrieoptimierung bedingt die Kenntnis des Gradienten  $\mathbf{g}$ , und die Diagonalisierung der Kraftkonstantenmatrix  $\mathbf{H}$  erlaubt die Klassifizierung der optimierten stationären Punkte als Minima, Übergangszustände oder höhere Sattelpunkte. Die zweiten (und höheren) Ableitungen dienen darüber hinaus zur Berechnung der spektroskopischen Konstanten in der Rotations-Vibrations-Spektroskopie.

Diese Beispiele sind nur ein kleiner Bruchteil dessen, wofür Ableitungen verschiedenster Art und Ordnung innerhalb der Computerchemie gebraucht werden können. Für eine vollständigere Übersicht sei auf die Literatur verwiesen [1]. Ableitungen können im Prinzip auf zwei verschiedene Arten erhalten werden: **numerisch** oder **analytisch**. Im Nachfolgenden wird auf die Unterschiede eingegangen.

## 2.3 Numerische Ableitungen

Wenn man die Taylorreihe 2.1 um den Punkt  $x$  anwendet und die höheren Terme vernachlässigt, so erhält man folgende Näherung :

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} \quad (2.4a)$$

$$\frac{f(x+h) - f(x)}{h} = f' + \frac{1}{2}hf'' + \dots \quad (2.4b)$$

Unter Verwendung der Abkürzungen  $f_0 = f(x_0)$ ,  $f_{+1} = f(x_0 + h)$ ,  $f_{+2} = f(x_0 + 2h)$ ,  $f_{-1} = f(x_0 - h)$ ,  $f_{-2} = f(x_0 - 2h)$  ergeben sich bei analogen Taylor-Entwicklungen um geeignete Punkte folgende weitere Näherungen:

$$f'_0 \approx \frac{1}{h}(-f_0 + f_{+1}) - \frac{1}{2}hf''_0 \quad (2.5a)$$

$$f'_0 \approx \frac{1}{2h}(-f_{-1} + f_{+1}) - \frac{1}{6}h^2f'''_0 \quad (2.5b)$$

$$f'_0 \approx \frac{1}{2h}(-3f_0 + 4f_{+1} - f_{+2}) + \frac{1}{3}h^2f'''_0 \quad (2.5c)$$

$$f'_0 \approx \frac{1}{12h}(f_{-2} - 8f_{-1} + 8f_{+1} - f_{+2}) + \frac{1}{30}h^4f^{(5)}_0 \quad (2.5d)$$

Differentiationsformeln des Typs 2.5b und 2.5d nennt man zentrale Formeln, da die Gitterstellen  $x_0 + nh$  symmetrisch um den zu differenzierenden Punkt  $x_0$  liegen; wohingegen die Formeln 2.5a und 2.5c fortschreitende Differenzenquotienten genannt werden, die Gitterstellen liegen hier asymmetrisch hinsichtlich des Punktes  $x_0$ .

Ableitungen, welche mittels eines Differenzenquotienten erhalten werden, sind notwendigerweise ungenau. Hauptsächlich kann zwischen zwei Arten von Fehlern unterschieden werden: Rundungsfehler ('roundoff error') und Abbruchfehler ('truncation error'). Auf beide möglichen Ursachen der Rechenungenauigkeit soll im folgenden genauer eingegangen werden [2, p. 180]. Im Allgemeinen können reelle Zahlen nicht exakt innerhalb des binären Zahlenraums dargestellt werden, da der Computer immer eine endliche maschinelle Präzision von  $\epsilon_m$  aufweist, die nicht umgangen werden kann. Zur Minimierung numerischer Fehler bei Anwendung von 2.4a wird man daher zunächst fordern, dass  $h$  hinreichend genau darstellbar sein muss. Selbst unter der Annahme, dass  $h$  exakt darstellbar ist, beträgt der Rundungsfehler in 2.4a  $\epsilon_r \sim \epsilon_f |f(x)/h|$ , wobei  $\epsilon_f$  der Fehler ist, womit  $f$  berechnet wird. Für einfache Funktionen gilt im besten Fall  $\epsilon_f \approx \epsilon_m$ , allerdings wird  $\epsilon_f$  bei komplizierteren Funktionen rasch grösser. Die Grössenordnung des Abbruchfehlers  $\epsilon_t$  liegt gemäss Gleichung 2.4b im Bereich von  $\epsilon_t \sim |hf''(x)|$ . Die optimale Grösse von  $h$  kann durch Minimierung der Summe von  $\epsilon_r + \epsilon_t$  erhalten werden:

$$h \sim \sqrt{\frac{\epsilon_f f}{f''}} \quad (2.6)$$

Für diese optimale Schrittweite beträgt die maximale Rechengenauigkeit in 2.4a:

$$(\epsilon_r + \epsilon_t)/|f'| \sim \sqrt{\epsilon_f}(f f''/f'^2)^{1/2} \sim \sqrt{\epsilon_f} \quad (2.7)$$

Bei Anwendung der einfachen Differenzenquotientenmethode nach der Gleichung 2.4a ist die Rechengenauigkeit somit proportional zur Quadratwurzel der Präzision der Funktionsberechnung (d.h. bestenfalls zur Quadratwurzel der Maschinenpräzision  $\epsilon_m$ ).

Wird nun anstelle der Gleichung 2.4a oder 2.5a die symmetrische Variante 2.5b verwendet, so liegt der Abbruchfehler in der Größenordnung  $\epsilon_t \sim h^2 f'''$ . Die durch die maschinelle Genauigkeit gegebene maximale Rechengenauigkeit liegt nun bei:

$$(\epsilon_r + \epsilon_t)/|f'| \sim (\epsilon_f)^{2/3} \quad (2.8)$$

Generell kann gesagt werden, dass die Rechengenauigkeit von 2.5a nach 2.5b zunimmt; die mögliche Maschinenpräzision wird jedoch nicht erreicht. Die Genauigkeit kann weiter gesteigert werden, indem man weitere Terme innerhalb der Taylorentwicklung berücksichtigt oder andere Methoden als die der Differenzenquotienten wählt [2]. Diese höhere Rechengenauigkeit muss aber in allen Fällen mit einem erhöhten Aufwand an Rechenschritten gewonnen werden.

Trotz der skizzierten Nachteile – schwierig einzugrenzende Rechengenauigkeit und oft hoher Rechenaufwand – sind numerische Ableitungen weit verbreitet. Der Hauptvorteil liegt in der einfachen Implementation: bei Verfügbarkeit von Code zur Berechnung der Funktionswerte  $f(x)$  lassen sich die numerischen Ableitungen in einer einfachen Schleife über geeignete Funktionsaufrufe leicht bestimmen. Ein guter Kompromiss zwischen Rechenaufwand und Genauigkeit ist die zentrale Differenzenquotientenmethode nach Gleichung 2.5b.

In nachfolgender Tabelle 2.1 (entsprechender Source-Code im Anhang A, Seite 135 beigelegt) wird die Präzision der einseitigen sowie zentralen Differenzenquotientenmethode anhand eines Beispiels dokumentiert. Die Rechnungen wurden auf einer Compaq-Maschine (tru64) durchgeführt. Es ist klar ersichtlich, dass in beiden Fällen – single sowie double precision – die zentrale Differenzenquotientenmethode besser das exakte Resultat reproduziert. Ebenso erkennt man, dass bei sehr kleiner Schrittweite  $h$  die Resultate unvorhersehbar und unbrauchbar werden. Demzufolge gibt es eine ideale Schrittweite. Die Rechnungen mit doppelter Genauigkeit weisen zudem den Vorteil auf, dass zuverlässige Zahlenwerte noch für kleine Werte von  $h$  berechnet werden können.



**Tabelle 2.1:** Numerische Gradienten für die Funktion  $f(x) = -\frac{1}{x}$  am Punkt  $x = \sqrt{1/42}$ . Vergleich von einseitiger sowie zweiseitiger Differenzenquotientenmethode.

n <sup>a</sup>	Exakt <sup>b</sup>	single precision		double precision	
		mono <sup>c</sup>	two <sup>d</sup>	mono <sup>c</sup>	two <sup>d</sup>
1	42.0	25.4842930	72.4137802	25.48429149428791	72.41379310344828
2	42.0	39.4437790	42.1771545	39.44375266787379	42.17714400482024
3	42.0	41.7299232	42.0019608	41.72956153225194	42.00176407409150
4	42.0	41.9759750	42.0022011	41.97279851763547	42.00001764000039
5	42.0	42.0093536	41.9855118	41.99727826534883	42.00000017644534
6	42.0	41.9616699	41.9616699	41.99972781115946	42.00000000187387
7	42.0	42.9153442	42.9153442	41.99997278853118	42.00000000231796
8	42.0	95.3674316	71.5255737	41.99999725784664	41.99999996679082
9	42.0	0.0000000	0.0000000	41.99999992238190	41.99999992238190
10	42.0	0.0000000	0.0000000	42.00000347509558	42.00000347509558
11	42.0	0.0000000	0.0000000	42.00000347509558	42.00000347509558
12	42.0	0.0000000	0.0000000	42.00018111077952	42.00018111077952
13	42.0	0.0000000	0.0000000	42.00195746761892	42.00195746761892
14	42.0	0.0000000	0.0000000	42.01083925181592	41.96643033083092
15	42.0	0.0000000	0.0000000	42.63256414560601	42.18847493575595

<sup>a</sup> Schrittweite  $h = 10^{-n}$

<sup>b</sup> exakte Lösung

<sup>c</sup> einseitige Differenzenquotientenmethode, gemäss Gleichung 2.5a

<sup>d</sup> zentrale Differenzenquotientenmethode, gemäss Gleichung 2.5b

## 2.4 Analytische Ableitungen

Ist eine Funktion  $f(x)$  gegeben, kann sie differenziert werden, sofern folgende Kriterien erfüllt sind.

$$y' = f'(x) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \quad (2.9a)$$

$$\lim_{x \rightarrow x_0-0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \rightarrow x_0+0} \frac{f(x) - f(x_0)}{x - x_0} \quad (2.9b)$$

Die Gleichung 2.9a bedingt, dass der Grenzwert existiert und endlich ist, Gleichung 2.9b hingegen verlangt, dass die linksseitige und die rechtsseitige Ableitung an der Stelle  $x_0$  übereinstimmen müssen.

Sind diese Bedingungen erfüllt, so können Ableitungen analytisch berechnet werden. Verglichen mit numerischen Ableitungen sind die analytischen Ableitungen prinzipiell genauer und in der Regel auch effizienter zu berechnen. In der Quantenchemie sind die zu differenzierenden Gleichungen jedoch oft alles andere als einfach, und die Herleitung der entsprechenden Formeln sowie ihre Implementation

und Übersetzung in eine Programmiersprache sind schwierig, langwierig und fehlerträchtig.

Für viele quantenchemische Standardverfahren sind analytische Ableitungen bereits verfügbar (vgl. hierzu die einschlägigen Übersichten [1, 3, 4, 5] und Originalarbeiten [6, 7, 8, 9]), aber bei neu entwickelten Methoden ergibt sich immer wieder der Bedarf, analytische Ableitungen neu zu implementieren. Hierzu stehen seitens der Informatik mächtige Hilfsmittel bereit, auf die im folgenden eingegangen wird.

## 2.5 Symbolische Differentiation

Programme wie Maple, Reduce oder Mathematica sind in der Lage, Funktionen fast beliebiger Art abzuleiten und die resultierenden Ausdrücke zu optimieren; teilweise können die Ergebnisse sogar in Form von Fortran77 Code ausgegeben werden. Jedoch muss gesagt werden, dass diese Fähigkeiten ihre Grenzen haben. Häufig muss der bestehende Computercode (z. B. Fortran77) in seine zu differenzierenden Teile aufgegliedert werden, oder er muss in für die betreffende Software verständliche Befehle übersetzt werden, bevor mit der eigentlichen Differentiation begonnen werden kann. Die generierten symbolischen Ausdrücke müssen danach wieder in den Computercode zurückübersetzt und schlussendlich in das Programmpaket integriert werden. Einer der wesentlichen Nachteile der symbolischen Differentiation liegt jedoch nicht in diesem 'administrativen' Aufwand, sondern viel eher in der Tatsache, dass solche Programme (noch) nicht in der Lage sind, grössere Programmkonstrukte wie Schleifen oder Unterprogramme sinnvoll zu handhaben. Zudem stellt die Übersetzung in z. B. für Mathematica verständliche Befehle und die Integration des neuen Codes eine erhebliche Quelle von Fehlern dar. Dennoch sind diese Programme eine grosse Hilfe bei der Implementierung analytischer Ableitungen und bereits mehrmals erfolgreich eingesetzt worden [6, 7, 8, 9].

## 2.6 Automatische Differentiation

Die automatische Differentiation (AD) bezeichnet einen Satz von Techniken, welche ein vorgegebenes Computerprogramm so erweitern, dass der erzeugte neue Code nicht nur die ursprüngliche Funktion  $\mathbf{f}$  an einem Punkt  $\mathbf{x}$  auswertet, sondern auch die zugehörige Jacobi-Matrix  $J = \partial\mathbf{f}/\partial\mathbf{x} \in \mathbb{R}^{m \times n}$ . In unserem Fall ist die abzuleitende Funktion zwar kein Vektor der Länge  $m$ , sondern ein skalare Grösse (Energie), aber im folgenden werden trotzdem die Prinzipien der AD zunächst allgemein formuliert. Die automatische Differentiation basiert auf der Einsicht, dass jede Berechnung im Prinzip eine (möglicherweise sehr lange) Sequenz von elementaren Rechenschritten wie Addition, Multiplikation etc. darstellt, für welche die Ableitungen bekannt sind. Durch Kombination dieser 'elementaren' Ableitungen

gemäss der Kettenregel kann man prinzipiell immer die gewünschte Ableitung erhalten. Die Kettenregel ist assoziativ und kann daher in verschiedener Weise angewendet werden. Im folgenden werden die zwei bekanntesten Strategien kurz umrissen.

Bei dem **Vorwärtsmodus** (forward mode) wird jeder Variablen  $u$  in dem Originalcode ein Ableitungsobjekt  $\mathbf{u}'$  zugeordnet. Sobald das Programm der Variablen  $u$  einen neuen Wert zuweist, wird  $\mathbf{u}'$  derart modifiziert, dass es immer die Ableitung  $\partial u / \partial \mathbf{x}$  enthält. Als Beispiel sei die Zuweisung  $u = u + \exp(v)$  im Originalcode betrachtet. In Anwendung der Ableitungsregeln für die Addition und die Exponentialfunktion wird der erweiterte Code dann zusätzlich zur Funktionsberechnung ein Statement der Form  $\mathbf{u}' = \mathbf{u}' + \exp(v) \cdot \mathbf{v}'$  ausführen, um den Wert der Ableitung entsprechend an die Änderung des Funktionswertes anzupassen. Um die Jacobi-Matrix für die komplette Berechnung zu erhalten, werden die Ableitungsobjekte der *unabhängigen* Variablen  $x_j$  (d.h. der  $n$  Inputvariablen, bezüglich derer abgeleitet wird) gemäss  $\mathbf{x}'_j = \mathbf{e}_j^T \in \mathbb{R}^{1 \times n}$  initialisiert, wobei  $\mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)^T$  den  $j$ -ten Einheitsvektor bezeichnet; danach werden die Ableitungsobjekte dann jeweils parallel zu der Funktionsberechnung modifiziert (wie oben beschrieben). Auf diese Weise erreicht man, dass nach Ablauf des Programms die Ableitungsobjekte der *abhängigen* Variablen (d.h. der  $m$  Outputvariablen, deren Ableitungen gesucht sind) die Zeilen der Jacobi-Matrix  $J$  enthalten. Da die Ableitungsobjekte Vektoren der Länge  $n$  sind, kann der Vorwärtsmodus zu einem  $O(n)$  Anstieg in der Operationszahl und in den Memoryanforderungen führen.

Bei dem **Rückwärtsmodus** (reverse mode) wird jedem Zwischenergebnis  $u$  bei der Berechnung von  $\mathbf{f}$  ein adjungiertes Objekt  $\bar{\mathbf{u}}$  derart zugeordnet, dass dieses die Ableitung  $\partial \mathbf{f} / \partial u$  darstellt. Hier beginnt die Berechnung der Ableitungen mit den bekannten Werten  $\bar{\mathbf{f}}_i = \mathbf{e}_i \in \mathbb{R}^{m \times 1}$  und erfolgt dann rückwärts, bis die gewünschten Ableitungen  $\bar{\mathbf{x}}_j$  (d.h. die Spalten von  $J$ ) erhalten werden. Betrachtet man wieder die Zuweisung  $u = u + \exp(v)$  in dem Originalcode, so muss der erweiterte Code jetzt ein zusätzliches Statement der Form  $\bar{\mathbf{v}} = \bar{\mathbf{v}} + \exp(v) \cdot \bar{\mathbf{u}}$  ausführen, wobei  $\bar{\mathbf{u}}$  der Wert des adjungierten Objekts von  $u$  nach der erfolgten Änderung von  $u$  ist. Bei dieser Vorgehensweise ist die Funktionsberechnung beendet, bevor die Bestimmung der Ableitungen beginnt, und es ist daher erforderlich, die Zwischenergebnisse und die relevante Information über Programmverzweigungen während der Funktionsberechnung zu speichern (oder jeweils wieder neu zu erzeugen). Die adjungierten Objekte sind Vektoren der Länge  $m$ , und daher darf man bei dem Rückwärtsmodus einen  $O(m)$  Anstieg der Rechenzeit und der Memoryanforderungen erwarten.

Formal sollte der Rückwärtsmodus in unserem Fall besonders attraktiv sein, weil die abzuleitende Funktion  $\mathbf{f}$  (d.h. die Energie  $E$ ) eine skalare Grösse ist ( $m=1$ ), so dass die Berechnung des Gradienten hinsichtlich Operationszahl und Memory genauso skalieren sollte wie die Funktionsberechnung selbst. Allerdings

ist der Kontrollfluss bei dem Vorwärtsmodus einfacher und robuster als bei dem Rückwärtsmodus, und mit der von uns verwendeten AD-Software konnte nur mit dem Vorwärtsmodus ein korrekter Gradientencode erstellt werden (siehe unten). Daher wird der Rückwärtsmodus hier nicht weiter besprochen. Für den Vorwärtsmodus ist im Anhang B, Seite 137, ein einfaches Beispiel [10] vorgerechnet, um nach der obigen abstrakten Darstellung die konkreten Rechenschritte bei der automatischen Differentiation zu illustrieren.

Abschliessend sollen noch die wesentlichen Unterschiede zu anderen Ansätzen festgehalten werden: Im Gegensatz zur symbolischen Differentiation, bei welcher die mathematischen Gleichungen für die Ableitungen bestimmt werden, arbeitet die automatische Differentiation durchgängig nur mit den *Werten* der Ableitungen (siehe Anhang B). Im Vergleich zur numerischen Differentiation werden numerische Fehler (z.B. Abbruchfehler) vermieden, ausserdem entfällt die Bestimmung der optimalen Schrittweiten (vgl. Kapitel 2.3).

### 2.6.1 Software zum automatischen Differenzieren

Die Änderungen am Code, die bei dem automatischen Differenzieren anfallen, sind so regelmässig, dass sie automatisch erzeugt werden können. Für diesen Zweck stehen mittlerweile eine Reihe von AD-Werkzeugen zur Verfügung. Erwähnenswert sind insbesondere die Programme ADIFOR [11, 12, 13, 14], ADOL-F [15], Odyssee/Tapenade [16] und TAF [17], welche zur entsprechenden Bearbeitung von Fortran-Code dienen. Alle diese Programme (ausser ADOL-F) führen eine sogenannte Quellentransformation ('source transformation') durch, d.h. die zusätzlichen Anweisungen zur Berechnung der Ableitungen werden direkt in den bestehenden Code eingefügt. Im Gegensatz dazu beruht ADOL-F auf dem Prinzip des *operator overloading*, d.h. die zusätzlichen Berechnungen werden in vordefinierten Modulen durchgeführt. Alle genannten Programme unterstützen den Vorwärtsmodus und - mit Ausnahme von ADIFOR2.0 - auch den Rückwärtsmodus; letzterer ist bei ADIFOR bislang nur in der Entwicklungsversion (3.0) verfügbar.

Die AD-Programme bedienen sich einer globalen Analyse des gesamten Programms ('dependence analysis'), um die aktiven Variablen zu identifizieren, deren Werte von den unabhängigen Variablen beeinflusst werden *und* die ihrerseits zu den abhängigen Variablen beitragen. Für die restlichen inaktiven Variablen werden keine Ableitungsobjekte bzw. adjungierten Objekte benötigt, so dass aufgrund einer solchen Aufteilung in aktive und inaktive Variable die Operationszahl und der Memorybedarf stark gesenkt werden können.

Das AD-Konzept wurde ursprünglich im Jahr 1959 vorgeschlagen [18]. Seither sind viele Veröffentlichungen zu diesem Thema erschienen, als Standardreferenz gelten die Bücher von Rall [19] und Griewank [20]. Einen guten Überblick gibt auch

ein kürzlich erschienener Übersichtsartikel [13].

### 2.6.2 ADIFOR

In der vorliegenden Arbeit wurde die ADIFOR-Software eingesetzt. ADIFOR [11] ist ein Akronym und steht für **A**utomatic **D**ifferentiation of **F**ORtran. Dieses Programm kann via WEB unter den URL's <http://www.mcs.anl.gov/adifor> oder <http://www.cs.rice.edu/adifor> bezogen werden und unterstützt zahlreiche Plattformen. Bisher ist ADIFOR nur in der Lage, analytische Gradienten für Programme zu erzeugen, die in Fortran77 geschrieben sind. Fortran90 soll in naher Zukunft unterstützt werden [21].

Das Programm ADIFOR wurde in seiner aktuellen Version (2.0) und in einer Entwicklungsversion (3.0) benutzt. Beide Versionen unterstützen den Vorwärtsmodus, ADIFOR3.0 auch den Rückwärtsmodus. ADIFOR3.0 liefert generell einen ähnlichen Code wie ADIFOR2.0, es bietet jedoch zahlreiche zusätzliche Optionen in den Steuerfiles, und der erzeugte Code kann meist von den Compilern besser optimiert werden [22].

Um einen Code zu generieren, der aus einem Programm die Ableitungen von abhängigen Variablen bezüglich der unabhängigen Variablen berechnet, geht ADIFOR in drei Schritten vor:

- 1. Schritt:** Der Fortran77 Code wird bereinigt und in eine Standard-Form übersetzt. Dabei werden unter anderem lange Funktionsberechnungen innerhalb von Subroutinen derart umgeschrieben, dass sie kürzer und kompakter werden. Dies erleichtert die Integration weiterer Libraries (z. B. SparsLinC) in den Code.
- 2. Schritt:** Die aktiven und inaktiven Variablen werden identifiziert. Eine Variable wird als aktiv definiert, wenn sie folgende Eigenschaft aufweist<sup>1</sup>: The preprocessor associates a gradient object with every variable whose value may depend on the value of a variable considered 'independent' with respect to differentiation, and whose value impacts a variable considered 'dependent' with respect to differentiation. Alle restlichen Variablen sind inaktiv.
- 3. Schritt:** Für jede Anweisung, die eine aktive Variable deklariert, wird ein Ableitungscode generiert, unter Allozierung der entsprechenden Gradientenobjekte. Anweisungen, die intrinsische Funktionen (Sinus, Cosinus etc.) enthalten, werden gemäss einer Bibliothek in ihre Ableitungen übersetzt.

ADIFOR benötigt zwei Steuerfiles, welche die Namen der beteiligten Unterprogramme und die Optionen für das Vorgehen bei der automatischen Differentiation spezifizieren.

---

<sup>1</sup>aus ADIFOR2.0, Users's Guide, Originalkommentar

Im folgenden wird ein kurzes Beispiel vorgeführt, komplett mit allen erforderlichen Schritten, um analytische Gradienten eines Programmes zu erhalten. Die angegebenen Listings und Steuerfiles beziehen sich auf ADIFOR2.0. Ausgangspunkt ist ein Programm (Listing 1), das eine Subroutine aufruft (Listing 2). Das

---

```

1  program main
2  real x, y
3  read *, x
4  call squareroot(x,y)
5  print *, y
6  end

```

---

### Programmlisting 1: Das Hauptprogramm

---

```

1  subroutine squareroot(x,y)
2  real x, y
3  y = sqrt(x)
4  end

```

---

### Programmlisting 2: Die Subroutine

ADIFOR-Programm erzeugt aus der vorgegebenen Subroutine 'squareroot(x,y)' eine neue Subroutine 'g\_squareroot(g\_p,x,g\_x,ldg\_x,y,g\_y,ldg\_y)', welche Gradientenobjekte für alle aktiven Variablen enthält, hier insbesondere g\_x für x und g\_y für y (siehe Listing 3). Die Variablen ldg\_x und ldg\_y geben die Dimension der Felder g\_x und g\_y an, sie sind in dem gewählten trivialen Beispiel beide gleich 1. Die Variable g\_p\_ entspricht der aktuellen Zahl  $n$  der unabhängigen Variablen (hier  $n=1$ ). Der Parameter g\_pmax\_ spezifiziert die Maximalzahl der unabhängigen Variablen, welche von ADIFOR aus einer Inputoption übernommen wird (hier ebenfalls gleich 1). Der erzeugte ausführbare Code testet zunächst in einer Abfrage (Zeile 9-12), ob der aktuelle Wert von g\_p\_ zulässig ist. Nach der Funktionsberechnung (Zeile 13) wird dann die Ableitung gemäss den Differentiationsregeln für die Wurzelfunktion bestimmt, mit entsprechender Fehlerbehandlung (Zeile 14-18). Wie in Kapitel 2.6 erläutert, werden bei dem AD-Vorwärtsmodus zunächst die Ableitungsobjekte als Einheitsvektoren initialisiert und dann parallel zur Funktionsberechnung modifiziert: dies geschieht hier in den Zeilen 19-21, wobei die Initialisierung von g\_x im aufrufenden Programm erfolgt; im vorliegenden Fall ist diese Operation eine triviale Multiplikation mit 1. Im allgemeinen Fall von  $n$  unabhängigen Variablen ( $n > 1$ ) erhält man auf diese Weise die Werte aller  $n$  Ableitungen, die dann gemeinsam mit dem Funktionswert (Zeile 22) an das aufrufende Programm übergeben werden.

Um die neu erhaltene Subroutine 'g\_squareroot' einzubinden, muss die Routine 'main.f' gemäss Programmlisting 4 umgeschrieben werden: In der neuen

---

```

1  subroutine g_squareroot(g_p,x,g_x,ldg_x,y,g_y,ldg_y)
2  real x, y
3  integer g_pmax_
4  parameter (g_pmax_ = 1)
5  integer g_i,g_p,ldg_y,ldg_x
6  real r1_p,r2_v,g_y(ldg_y),g_x(ldg_x)
7  integer g_ehfid
8  data g_ehfid /0/
9  if (g_p_ .gt. g_pmax_) then
10     print *, 'parameter g_p_ is greater than g_pmax_'
11     stop
12 endif
13 r2_v = sqrt(x)
14 if (x .gt. 0.0e0) then
15     r1_p = 1.0e0 / (2.0e0 * r2_v)
16 else
17     call ehufSV (9,x,r2_v,r1_p,'g_squareroot.f',37)
18 endif
19 do g_i_ = 1,g_p
20     g_y(g_i_) = r1_p * g_x(g_i_)
21 enddo
22 y = r2_v
23 end

```

---

**Programmlisting 3:** Die abgeleitete Subroutine

---

```

1  program driver
2  real x, y
3  real g_x(1), g_y(1)
4  read *, x
5  g_x(1) = 1.0
6  call g_squareroot(1,x,g_x,1,y,g_y,1)
7  call ehrpt
8  print *, y
9  print *, g_y(1)
10 end

```

---

**Programmlisting 4:** Das neue Hauptprogramm

Routine ('driver') werden zunächst die nötigen Felder definiert (Zeile 3). Danach erfolgen die Initialisierung der Ableitungsobjekte (Zeile 5) und der Aufruf der von ADIFOR erzeugten Subroutine mit geeigneten Argumenten (Zeile 6). Der Aufruf der Library-Routine 'ehrpt' (exception handler **report**) in Zeile 7 kann deaktiviert werden, wenn die Testläufe erfolgreich verlaufen sind. Ein Executable, welches nun schlussendlich die Berechnung der gewünschten Ableitung erlaubt, kann gemäss Tabelle 2.2 erhalten werden: Die hinzugelinkten Libraries sind wichtig für allfällige

**Tabelle 2.2:** Makefile

```
f77 -c driver.f
f77 -c g_squareroot.f
f77 -o driver driver.o g_squareroot.o
      AD_LIB/lib/ReqADIntrinsics-$AD_OS.o
      AD_LIB/lib/libADIntrinsics-$AD_OS.a
```

Fehlerhandhabungen (z. B. Divisionen durch Null). Es ist aber möglich, Executables ohne die entsprechenden Libraries zu erstellen.

## 2.7 Analytische Gradienten für MNDO97

An der Winterschool (Februar 2000) im Forschungszentrum Jülich, Deutschland, wurde von Prof. Dr. C. H. Bischof automatisches Differenzieren vorgestellt [13]. Im Anschluss daran begann ein Projekt mit Dr. B. Lang, RWTH Aachen, welches der Frage nachgehen sollte, ob analytische Gradienten der Energie hinsichtlich der Kernkoordinaten für das Programm MNDO97 mittels automatischen Differenzierens erhalten werden können. Folgende Fragen standen zu Beginn des Projektes im Vordergrund:

- Können einfache Energieausdrücke wie die Molekülmechanik (MM)-Korrekturen zur Peptidbindung in PM3 mit ADIFOR differenziert werden?
- Ist es möglich, mittels ADIFOR analytische Gradienten für Standardmethoden (MNDO, MNDO/d, AM1 und PM3) zu erzeugen und die Ergebnisse des hierfür bereits vorhandenen analytischen Codes zu reproduzieren?
- Ist es möglich, analytische Gradienten für OM1 und OM2 zu erstellen, wo bisher nur numerische Ableitungen zur Verfügung stehen?

In der Testphase wurde eine modifizierte Version von MNDO97 [23] benutzt, MNDO97s ('s' steht für simple). Diese vereinfachte Version erlaubt nur Single-Point-Energie Rechnungen und numerische Gradientenberechnungen über finite Differenzen. MNDO97s enthält ca. 25000 Zeilen, im Gegensatz zu MNDO97, welches ca. 155000 Zeilen aufweist.



### 2.7.1 MNDO97

Analytische erste und zweite Ableitungen der Energie hinsichtlich der Kernkoordinaten stehen bei den etablierten semiempirischen Methoden wie MNDO [24], AM1 [25], PM3 [26, 27] sowie MNDO/d [28, 29] zur Verfügung, jeweils für closed-shell self-consistent-field (SCF) Wellenfunktionen [9]. Weiterhin sind analytische erste Ableitungen für open-shell half-electron SCF Wellenfunktionen sowie für CI-Wellenfunktionen erfolgreich implementiert [7, 8, 30].

Analytische Ableitungen fehlen jedoch in MNDO97 für die neueren semiempirischen Methoden OM1 [31] und OM2 [32], welche Orthogonalisierungskorrekturen beinhalten.

Das ADIFOR-Projekt hatte das Ziel, für alle verfügbaren semiempirischen Methoden (einschliesslich OM1 und OM2) automatisch den Code zur analytischen Berechnung der Gradienten zu erzeugen (bei Verwendung von closed-shell SCF Wellenfunktionen). Die Gesamtenergie  $E$  ist in diesem Fall generell definiert als

$$E = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}) + V_{NN} \quad , \quad (2.10a)$$

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda} \sum_{\sigma} P_{\lambda\sigma} \left[ (\mu\nu, \lambda\sigma) - \frac{1}{2} (\mu\lambda, \nu\sigma) \right] \quad , \quad (2.10b)$$

wobei  $P_{\mu\nu}$ ,  $H_{\mu\nu}$ , und  $F_{\mu\nu}$  die Elemente der Dichtematrix, der Einelektronen-Matrix sowie der Fock-Matrix (Ein- und Zweielektronenterme) bezeichnen. Desweiteren entspricht  $V_{NN}$  der Kern-Kern-Abstossung und  $(\mu\nu, \lambda\sigma)$  beschreibt ein Zwei-Elektronen-Integral. Wird nun nach der Ortskoordinate  $x_a$  differenziert, so kann man die explizite Ableitung der variationsmässig optimierten Dichtematrix vermeiden [3]. In der *ab initio* MO Theorie können die entsprechenden Beiträge zum Gradienten durch die Ableitungen der Überlappungsintegrale ausgedrückt werden, welche in semiempirischen MO Verfahren vom MNDO-Typ wegen der vorgenommenen Integralapproximationen wegfallen. Daher ergibt sich in solchen semiempirischen Verfahren für die analytische Ableitung bezüglich der Kernkoordinate  $x_a$  (Superskript a) folgender Ausdruck:

$$E^a = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu}^a + F_{\mu\nu}^a) + V_{NN}^a \quad , \quad (2.11a)$$

$$F_{\mu\nu}^a = H_{\mu\nu}^a + \sum_{\lambda} \sum_{\sigma} P_{\lambda\sigma} \left[ (\mu\nu, \lambda\sigma)^a - \frac{1}{2} (\mu\lambda, \nu\sigma)^a \right] \quad , \quad (2.11b)$$

Zu differenzieren sind somit lediglich die Ein-Elektronen-Integrale  $H_{\mu\nu}^a$ , die Zwei-Elektronen-Integrale  $(\mu\lambda, \nu\sigma)^a$  und der Term für die Kern-Kern-Abstossung ( $V_{NN}^a$ ).

## 2.7.2 Programmrestrukturierung

Die Berechnung der Energie wird in MNDO97 in einer Subroutine namens SCFCAL bewerkstelligt. Eine direkte Anwendung von ADIFOR auf SCFCAL und die nachgeordneten Unterprogramme wäre jedoch ineffizient, weil der Rechenaufwand in SCFCAL durch die iterative Optimierung der Molekülorbitale und der zugehörigen Dichtematrix dominiert wird. Diese Rechenoperationen sind für die Berechnung des Gradienten irrelevant, weil dort nur die konvergierten Dichtematrixelemente als konstante Vorfaktoren auftreten, vgl. 2.11a und 2.11b.

Somit wurde die Subroutine SCFCAL derart ergänzt, dass nach SCF-Konvergenz eine zusätzliche Subroutine GRDCAL aufgerufen wird. Diese neue Subroutine berechnet die Energie nochmals gemäss den Formeln 2.10a und 2.10b, unter Verwendung der konvergierten Dichtematrix. Dieses Wiederberechnen aller benötigten Integrale benötigt im Schnitt weniger als 10 % der CPU-Zeit einer vollständigen SCF-Berechnung.

Das ADIFOR-Programm wurde auf zwei verschiedene Versionen von GRDCAL angewendet: a) eine konventionelle Implementierung, welche alle notwendigen Integrale vor der Energieberechnung berechnet und speichert, sowie b) eine direkte Implementierung, welche die einzelnen Beiträge aufsummiert ohne die Speicherung der Matrizen, die in den Formeln 2.10a und 2.10b auftauchen.

## 2.7.3 Erzeugung und Implementierung von ADIFOR-Code

Die Implementierung der konventionellen und der direkten GRDCAL-Version erfolgen völlig analog. Sie beinhalten jeweils die folgenden drei Schritte:

- Definieren der Files `adifor.adf` und `adifor.cmp`
- Aufruf des Programms
- Einbau einer Schnittstelle zwischen ADIFOR- und MNDO-Code

Die beiden **Steuerfiles** sind in Tabelle 2.3 angegeben.

Das File 2.3(a) definiert das Verhalten von ADIFOR2.0. Die Variable `AD_TOP` kennzeichnet diejenige Subroutine, ab welcher abgeleitet werden soll, wobei mittels `AD_EXCLUDE_PROCS` gewisse Subroutinen vom Differenzieren ausgeschlossen werden können. `AD_IVARS` resp. `AD_DVARS` kennzeichnen die inaktiven resp. aktiven (dependent) Variablen. In unserem Falle sind wir interessiert an den Ableitungen der Energie (`DENER`) hinsichtlich den Ortskoordinaten (`COORD`). `AD_PMAX` kennzeichnet die maximale Zahl der unabhängigen Variablen. Diese Variable ist nötig, da Fortran77 keine Standardmechanismen für ein dynamisches Speicherallocieren zur Verfügung stellt. Die Variable `AD_PMAX` ist jeweils die erste Dimension von jedem 'gradient-object'. Die Wahl von `AD_EXCEPTION_FLAVOR = performance` hat zur Folge, dass der erzeugte ADIFOR-Code keine Tests hinsichtlich Divi-

AD_TOP	=	grdcal	block.f
AD_PMAX	=	456	cpusec.f
AD_IVARS	=	COORD	dcart.f
AD_DVARS	=	DENER	fock.f
AD_PROG	=	adifor.cmp	gmetry.f
AD_EXCLUDE_PROCS	=	etime	guessp.f
AD_EXCEPTION_FLAVOR	=	performance	hcore.f
			input1.f
			input2.f
			iter.f
			linalg.f
			mndo97.f
			print.f
			scf.f

(a) adifor.adf

(b) adifor.cmp

**Tabelle 2.3:** Steuerfiles für ADIFOR2.0

sion durch 0 und dergleichen durchführt. Der Flag 'performance' bewirkt ausserdem, dass der erzeugte Fortran77 Code frei von Library-Routinen bleibt.

Das File 2.3(b) kennzeichnet alle Files des Quellcodes, die abgeleitet werden sollen. ADIFOR2.0 erkennt selbständig, welche Subroutinen in diesen Files abgeleitet werden müssen und welche nicht. Mit Ausnahme der Einführung von GRDCAL (siehe oben) wurden keinerlei Modifikationen am Quellcode vorgenommen.

Bei der **Programmcode-Erzeugung** wurden die Programme ADIFOR2.0 und ADIFOR3.0 eingesetzt. Die meisten Tests und Rechnungen wurden mit 'ADIFOR2.0-Code' durchgeführt, aber es wurde auch versucht, die neuen Funktionalitäten von ADIFOR3.0 zu nutzen. Hierbei waren die berechneten analytischen Gradienten via 'forward-mode' (im Vergleich jeweils zu einer selbst programmierten numerischen finiten Differenzenmethode) korrekt, jedoch lieferten die im 'reverse-mode' mit ADIFOR3.0 erzeugten Gradienten falsche Werte. Der Fehler in dem von ADIFOR3.0 erzeugtem Gradientencode konnte nicht genau lokalisiert werden, aber vermutlich behandelt ADIFOR3.0 im Rückwärtsmodus einige Fortran77-Konstrukte nicht korrekt. Im folgenden werden daher nur die Resultate diskutiert, welche mittels ADIFOR2.0, Revision D, im Vorwärtsmodus erhalten wurden.

Das Erzeugen der Gradienten mit ADIFOR beansprucht ca. 200 Sekunden auf einem Laptop mit einem PII, 300 MHz, 128 MB RAM unter dem OS Linux. Total werden ca. 29000 neue Zeilen erzeugt.

Die **Schnittstelle** zwischen dem analytischen Gradientencode und MNDO97s ist in dem Programmlisting 5 dokumentiert. Anzumerken gilt, dass `g_pmax_ = 456`

---

```

1  SUBROUTINE x_grdcal(a,lm5)
2  C
3  IMPLICIT DOUBLE PRECISION (A-H,O-Z)
4  C
5  INTEGER g_pmax_
6  PARAMETER (g_pmax_ = 456)
7  PARAMETER (LM1 = 600)
8  PARAMETER (LMV=3*LM1)
9  PARAMETER (LM1M=20000)
10 DOUBLE PRECISION g_coord(g_pmax_,3,LM1)
11 DOUBLE PRECISION g_dener(g_pmax_)
12 DOUBLE PRECISION a(lm5)
13 C
14 COMMON /ATOMS / NUMAT, NAT(LMV)
15 COMMON /g_atomc / g_coord
16 COMMON /g_denerg / g_dener
17 C
18 C - extract memory, machdies =
19 C ps -a -o "cmd rss vsz" | grep mndo99mem | grep -v grep > rainer
20 C
21 CALL SYSTEM ("rm -f rainer ; sh ./machdies")
22 CALL SYSTEM ("echo 'x_grdcal enter' 'cat rainer' >> cmemory")
23 C
24 C — construct the unit-matrix —
25 C
26 DO 10 ir=1,3
27   DO 20 ia=1,NUMAT
28     DO 30 j=1,3*NUMAT
29       g_coord(j,ir,ia) = 0.0d0
30   30 CONTINUE
31     g_coord(ir+(ia-1)*3,ir,ia) = 1.0d0
32   20 CONTINUE
33   10 CONTINUE
34 C
35 CALL g_grdcal(3*NUMAT,a,lm5)
36 C
37 CALL SYSTEM ("rm -f rainer ; sh ./machdies")
38 CALL SYSTEM ("echo 'x_grdcal exit' 'cat rainer' >> cmemory")
39 C
40 END

```

---

### Programmlisting 5: Der Driver

eine der Variablen ist, welche im Steuerfile 2.3(a) angegeben wurde. Diese Grösse ist fest vergeben und limitiert die Zahl der möglichen Koordinaten. Die Zeilen 21–23 sowie 38–40 dienen dazu, den Memorybedarf direkt aus dem laufenden Programm zu messen; siehe dazu auch das Kapitel 2.8.4.

## 2.8 Resultate

Die beiden Gradienten-Implementierungen von ADIFOR – konventionell sowie direkt – sollen im folgenden mit bestehenden Gradientenmethoden (numerischer sowie analytischer Art) verglichen werden; und zwar hinsichtlich Präzision, Zeitverbrauch sowie Memoryanforderungen. Um endlose Wiederholungen zu verhindern, sollen folgende Abkürzungen gelten:

P steht für Präzision, M für Memoryanforderungen sowie T für Zeitverbrauch. G steht für die Art und Weise, wie der Gradientencode erstellt wurde. Wird auf Resultate von ADIFOR2.0 verwiesen, so werden diese mit dem Index 'A<sub>di2</sub>' bezeichnet. Soll genauer zwischen den beiden Implementierungen unterschieden werden, so geschieht dies mit dem Index 'k' für konventionell respektiv 'd' für direkt (z. B. M<sub>A<sub>di2</sub>.d</sub>). Die gleiche Logik gilt für die analytischen Gradienten von MNDO99 (G<sub>Analyt</sub>, nicht vorhanden in MNDO97s) sowie für die numerischen Gradienten via finite Differenzenmethode (G<sub>Diff.m</sub>), welche in MNDO97s implementiert sind.

Um faire Benchmarks zu gewährleisten, wurden auf der Grundlage von (2.10a) weitere numerische Gradienten über finite Differenzen implementiert: G<sub>Diff.d</sub> entspricht der direkten Integralversion, G<sub>Diff.k</sub> der konventionellen Version. Alle Rechnungen hinsichtlich der Genauigkeit sowie für Benchmarks wurden auf einer ES40 EV68 (7.0 GB RAM, 4\*833 MHz), Betriebssystem tru64 durchgeführt.

Ein Programm namens 'Parma' (Akronym für 'Programm für automatisches Erzeugen von z-Matrizen') wurde geschrieben, um automatisch verschiedene beliebig lange Molekülketten zu generieren. Diese Molekülketten können aus einfachen Einheiten wie z.B. -CH<sub>2</sub>- (Alkane), -C- (Alkine) oder Glycin (Peptide) bestehen. Desweiteren ist es möglich, die Bindungswinkel und Bindungslängen innerhalb gewisser Schranken zufällig zu variieren. Eine genauere Programmbeschreibung kann aus dem Anhang C ab Seite 139 oder direkt aus dem Source-Code von 'Parma' entnommen werden.

### 2.8.1 MNDO97s: Testphase

In der Test- resp. Validierungsphase wurden die mittels ADIFOR2.0 erhaltenen Gradienten (G<sub>A<sub>di2</sub></sub>) mit denjenigen der finiten Differenzenmethode (G<sub>Diff.m</sub>) hinsichtlich der Genauigkeit der einzelnen Gradientenkomponenten und der kartesischen Gradientennorm verglichen. Hierdurch wurde geprüft, ob ADIFOR2.0 einen korrekten (analytischen) Gradientencode für MNDO97s liefert. Der Grund, dass nicht mit G<sub>Diff.d</sub> oder G<sub>Diff.k</sub> verglichen wurde, liegt darin, dass zwei unabhängige Methoden für die Berechnung der Gradienten verwendet werden sollen. So beruht z. B. G<sub>Diff.k</sub> auch auf GRDCAL, derjenigen Subroutine also, welche von ADIFOR2.0 differenziert wurde.

Zu Beginn der Testphase traten bei einem von 50 Standard-Testmolekülen Differenzen zwischen den beiden Methoden  $G_{\text{Adi2}}$  und  $G_{\text{Diff.m}}$  auf, wobei die Unterschiede weit ausserhalb der zu akzeptierenden Toleranz lagen. In der Analyse zeigte sich, dass der Fehler bei einem Peptidmolekül auftrat, bei dem die selten gebrauchte Option `immok = 1` gesetzt war. Diese Option ruft eine Subroutine 'MMOKG' innerhalb MNDO97s auf, welche MM-Korrekturen für Peptidbindungen berechnet [26, 33]. Ein fehlender Faktor von 2 in dem Standardcode war schlussendlich verantwortlich für die gefundenen Differenzen (genauer: für den falschen Gradienten von  $G_{\text{Diff.m}}$ ). Nach der Behebung dieses Bugs lieferten beide Methoden ( $G_{\text{Adi2}}$  und  $G_{\text{Diff.m}}$ ) die gleichen Gradienten innerhalb der erwarteten Fehlerschranken.

## 2.8.2 Genauigkeit der Gradienten

Während der Testphase sowie auch später wurden die Gradienten von Adifor2.0 (sowohl  $G_{\text{Adi2.k}}$  als auch  $G_{\text{Adi2.d}}$ ) jeweils auf Genauigkeit geprüft. Das Prüfkriterium, woran sich die neuen analytischen Gradienten messen müssen, sind die in MNDO99 vorhandenen analytischen Gradienten für MNDO, AM1, PM3 und MNDO/d; analytische Gradienten für neuere semiempirische Methoden wie OM1 und OM2 standen ja bislang noch nicht zur Verfügung. Verwendet wurde ein Testsatz aus 50 unverzweigten Alkanen ( $C_kH_{2k+2}$  ( $k = 1, \dots, 50$ )). Für  $G_{\text{Adi2.k}}$  konnte nur bis  $k=46$  getestet werden, da oberhalb dieses Wertes die Memoryanforderungen (Stacksize) zu gross sind und die Berechnungen abgebrochen werden mussten (siehe dazu auch das Kapitel 2.8.4). Alle Moleküle wurden in einem ersten Schritt optimiert, und zwar auf demselben Level, auf dem danach die verschiedenen Genauigkeits- sowie Benchmarktests durchgeführt wurden. Alle folgenden Vergleiche beziehen sich auf Single-Point-Rechnungen an identischen optimierten Geometrien. Die Gradientenkomponenten an den optimierten Geometrien liegen typischerweise im Bereich von  $10^{-3}$  a.u.

Um zu gewährleisten, dass beide analytischen Codes von der exakt gleichen Dichtematrix ausgehen, wurden sowohl  $G_{\text{Adi2.k}}$  und  $G_{\text{Adi2.d}}$  als auch die numerischen Pendanten  $G_{\text{Diff.d}}$  und  $G_{\text{Diff.k}}$  in den bestehenden Code von MNDO99 integriert. Dazu wurden zwei zusätzliche Directories (ADIFOR2 und ADIFOR3) zu dem bestehenden MNDO99-Paket hinzugefügt. Zusätzlich waren einige Erweiterungen der entsprechenden Makefiles nötig. Erweitert wurde auch die Subroutine HFOCKP.f, welche für die direkte Integralberechnung zuständig ist. Diese Erweiterung war für die analytischen Gradienten in OM1 nötig. Als Schnittstelle für alle 'Gradienten-Codes' fungiert SCF.f, die generelle Kontrollroutine für die Energie und Gradientenberechnung. In dieser Subroutine sind alle Aufrufe zu den entsprechenden Gradientencodes implementiert, analytischer sowie numerischer Art. Diese Schnittstelle wurde auch benutzt, um die Zeit- und Memorymessungen

durchzuführen.

Zum Vergleich der mit  $G_{\text{Analyt}}$  und  $G_{\text{Adi}}$  berechneten Ergebnisse diene ein Hilfsprogramm, welches alle kartesischen Gradientenkomponenten im gesamten Alkan-Testsatz ( $C_kH_{2k+2}$  ( $k = 1, \dots, 50$ )) auf Übereinstimmung innerhalb eines vorgegebenen Toleranzkriteriums überprüfte ( $10^{-m}$  a.u.;  $m=1, \dots, 15$ ).

Anzumerken gilt, dass in MNDO99 die kartesischen Gradientenkomponenten in (kcal/(mol\*Angstrom)) ausgegeben werden. Der Umrechnungsfaktor in atomaren Einheiten (a. u.) beträgt 1 a.u. = 1185.82 kcal/(mol\*Angstrom). Im nachfolgenden ist, wenn nichts anderes angegeben, immer die Rede von atomaren Einheiten.

#### Analytische Gradienten von AM1 versus $G_{\text{Adi2}}$ .

Analytische Gradienten von AM1, implementiert in MNDO99, wurden mit den neuen analytischen Gradienten  $G_{\text{Adi2},k}$  sowie  $G_{\text{Adi2},d}$  verglichen. Desweiteren wurden zur internen Konsistenzprüfung die mit  $G_{\text{Adi2},k}$  und  $G_{\text{Adi2},d}$  berechneten ADIFOR-Gradienten verglichen. Die Resultate sind in der Tabelle 2.4 dargestellt.

**Tabelle 2.4:** Genauigkeiten analytischer Gradienten (AM1) versus  $G_{\text{Adi2}}$ . Totale Anzahl aller Gradientenkomponenten, welche ein gewisses Toleranzkriterium  $m$  erfüllen.

$m^a$	$G_{\text{Analyt}}$ vs. $G_{\text{Adi2},d}^b$		$G_{\text{Analyt}}$ vs. $G_{\text{Adi2},k}^c$		$G_{\text{Adi2},d}$ vs. $G_{\text{Adi2},k}^c$	
	ok	not ok	ok	not ok	ok	not ok
1	11775	0	10005	0	10005	0
2	11775	0	10005	0	10005	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	11775	0	10005	0	10005	0
11	8778	2997	7460	2545	10005	0
12	8044	3731	6834	3171	10005	0
13	4670	7105	3985	6020	10005	0
14	1399	10376	1203	8802	10003	2
15	271	11504	237	9768	7773	2232

<sup>a</sup> Toleranzkriterium  $10^{-m}$  (a. u.)

<sup>b</sup> Alkylketten ( $C_kH_{2k+2}$  ( $k = 1, \dots, 50$ ))

<sup>c</sup> Alkylketten ( $C_kH_{2k+2}$  ( $k = 1, \dots, 46$ ))

Ausgehend von der Annahme, dass die MNDO99 Referenzwerte ( $G_{\text{Analyt}}$ ) numerisch korrekt sind, gelten folgende Beobachtungen:

- Die analytischen Gradienten von  $G_{\text{Adi2},d}$  weisen eine Genauigkeit von mindestens  $10^{-10}$  a.u. auf, wobei jedoch oft auch höhere Genauigkeiten (bis zu  $10^{-15}$  a.u.) gemessen werden.

- Gleiches gilt für  $G_{\text{Adi2.k}}$ , die Genauigkeit ist wiederum mindestens bei  $10^{-10}$  a.u.
- Die Gradienten nach  $G_{\text{Adi2.d}}$  sowie  $G_{\text{Adi2.k}}$  sind bis zu einer Präzision von mindestens  $10^{-13}$  a.u. identisch.

Die gemessenen Genauigkeiten lassen den Schluss zu, dass die neuen analytischen Gradienten für AM1 korrekt sind. Analoge Testrechnungen für PM3 sowie MNDO/d liefern ähnliche Ergebnisse, wenn auch die Genauigkeiten insgesamt geringfügig niedriger sind als für AM1. Den Gründen für diese Unterschiede wurde nicht nachgegangen. Insgesamt darf man festhalten, dass die neuen analytischen Gradienten ( $G_{\text{Adi2.d}}$  sowie  $G_{\text{Adi2.k}}$ ) numerisch korrekt sind.

Um eine mögliche Korrelation der Präzision mit der Molekülgrösse zu prüfen, wurde  $G_{\text{Adi2.d}}$  im Vergleich zu  $G_{\text{Analyt}}$  genauer analysiert, berechnet jeweils auf dem AM1 Niveau. Dazu wurden wiederum die 'beiden' Gradienten eines Moleküls miteinander verglichen und die Verteilung der Genauigkeiten beobachtet. Einige wenige repräsentative Beispiele sind in der Tabelle 2.5 wiedergegeben. Das Tole-

**Tabelle 2.5:** Genauigkeiten der Gradienten für spezielle Moleküle, AM1-Niveau. Totale Anzahl aller Gradientenkomponenten, welche ein gewisses Toleranzkriterium  $m$  erfüllen.

$C_k H_{2k+2}$	$l^b$	Präzision $m^a$					
		1.. 10	11	12	13	14	15
$k$							
1	15	15	15	15	15	15	14
5	51	51	40	36	31	20	14
10	96	96	71	66	48	20	5
20	186	186	141	131	86	28	3
30	276	276	211	187	105	34	6
40	366	366	271	248	136	44	8
50	456	456	327	309	192	50	6

<sup>a</sup> Toleranzkriterium  $10^{-m}$  (a. u.)

<sup>b</sup> Totale Anzahl der Gradienten,  $l=9k+6$

ranzkriterium ist für alle Moleküle bis zu  $10^{-10}$  a.u. erfüllt, im Einklang natürlich mit Tabelle 2.4. Für das einfachste Molekül,  $C_1H_4$ , findet man praktisch perfekte Übereinstimmung zwischen  $G_{\text{Adi2.d}}$  und  $G_{\text{Analyt}}$ . Für grössere Moleküle nimmt die Zahl der Übereinstimmungen mit der Schärfe des Toleranzkriteriums ( $m = 11, \dots, 15$ ) ab, und zwar in ähnlicher Weise für verschieden grosse Moleküle: so findet man z.B. für  $m = 12$  noch 68 - 71% Übereinstimmung ( $k = 5, \dots, 50$ ). In diesem Sinne ist die Genauigkeit der ADIFOR-Gradienten nicht von der Molekülgrösse



abhängig. Ergänzend kann angefügt werden, dass für  $G_{\text{Adi2}\cdot k}$  das gleiche Verhalten beobachtet wird.

### Numerische Gradienten von OM1 versus $G_{\text{Adi2}}$ .

Analytische Gradienten für OM1 sowie OM2 sind bisher nicht in MNDO99 implementiert, nur numerische Gradienten sind bis anhin verfügbar. Ein Impetus für diese Arbeit bestand ja unter anderem auch in dieser Sachlage. Mittels ADIFOR sind nun analytische Gradienten für OM1 zugänglich. Die Genauigkeit derselben ist aus Tabelle 2.6 ersichtlich:

**Tabelle 2.6:** Genauigkeiten analytischer sowie numerischer Gradienten für OM1. Totale Anzahl aller Gradientenkomponenten welche ein gewisses Toleranzkriterium  $m$  erfüllen

$m$ <sup>a</sup>	$G_{\text{Diff}\cdot m}$ vs. $G_{\text{Adi2}\cdot d}$ <sup>b</sup>		$G_{\text{Diff}\cdot m}$ vs. $G_{\text{Adi2}\cdot k}$ <sup>c</sup>		$G_{\text{Adi2}\cdot d}$ vs. $G_{\text{Adi2}\cdot k}$ <sup>c</sup>	
	ok	not ok	ok	not ok	ok	not ok
1	11775	0	10005	0	10005	0
2	11775	0	10005	0	10005	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
7	11775	0	10005	0	10005	0
8	11759	16	10002	3	10005	0
9	9108	2667	7755	2250	10005	0
10	1413	10362	1212	8793	10005	0
11	138	11637	120	9885	10005	0
12	15	11760	14	9991	10005	0
13	4	11771	4	10001	10005	0
14	1	11774	1	10004	10004	1
15	0	11775	0	10005	7995	2010

<sup>a</sup> Toleranzkriterium  $10^{-m}$ , (a. u.)

<sup>b</sup> Alkylketten ( $C_kH_{2k+2}$  ( $k = 1, \dots, 50$ ))

<sup>c</sup> Alkylketten ( $C_kH_{2k+2}$  ( $k = 1, \dots, 46$ ))

- Die Unterschiede zwischen den analytischen Gradienten nach  $G_{\text{Adi2}\cdot d}$  sowie  $G_{\text{Adi2}\cdot k}$  und den numerischen Gradienten  $G_{\text{Diff}\cdot m}$  sind stets geringer als  $10^{-7}$  a.u.
- Die beiden analytischen Gradienten unterscheiden sich frühestens bei einem Toleranzkriterium von  $10^{-14}$  a.u.

Vor dem Hintergrund der AM1 Ergebnisse (siehe oben) legen diese Befunde den Schluss nahe, dass die analytischen ADIFOR-Gradienten für OM1 eine höhere Präzision aufweisen als die entsprechenden numerischen Gradienten. Letztere wurden aus (2.11a) ermittelt, wobei die Integralableitungen über finite Differenzen mit einer Schrittweite von  $10^{-6}$  Å berechnet werden.

Wie im vorangehenden Kapitel wurde weiterhin eine Korrelation zur Molekülgrösse untersucht. Die Befunde für OM1 sind analog zu denen bei AM1 und werden daher nicht im einzelnen diskutiert.

Zusammenfassend kann man festhalten, dass der mit ADIFOR erzeugte Code für die analytischen Gradienten korrekte und numerisch präzise Ergebnisse liefert.

### 2.8.3 Zeitmessungen

Die benötigten Rechenzeiten sind in der Praxis ein wichtiges Kriterium für die Güte der verschiedenen Gradientencodes. Der Zeitbedarf ist eine Funktion der Zahl an Basisorbitalen ( $N$ ) und an Kernkoordinaten ( $n$ ). Für die Alkane  $C_kH_{2k+2}$ ,  $k = 1, 2, \dots$  gilt  $N = 6k + 2$  (4 Orbitale pro C, 1 Orbital pro H) und  $n = 9k + 6$  (3 Freiheitsgrade pro Atom).

Folgende Rechenzeiten werden miteinander verglichen:

$T_{\text{Adi2-d}}$ , analytische Gradienten, ADIFOR-Code, erzeugt durch automatisches Differenzieren von GRDCAL, direkte Integralversion;

$T_{\text{Adi2-k}}$  analog zu  $T_{\text{Adi2-d}}$ , konventionelle Integralversion;

$T_{\text{Analyt}}$ , handkodierte analytische Gradienten von MNDO99, wobei bekannt ist, dass diese mit  $O(N^2)$  skalieren [9];

$T_{\text{Diff-d}}$ , numerische Gradienten, zentrale finite Differenzen, direkte Integralversion, die Subroutine GRDCAL wird für jede Gradientenkomponente zweimal aufgerufen;

$T_{\text{Diff-k}}$ , numerische Gradienten, konventionelle Integralversion, wiederum zweimaliger Aufruf der Subroutine GRDCAL;

$T_{\text{Diff-full}}$ , numerische Gradienten, zentrale finite Differenzen auf der Grundlage vollständiger SCF-Rechnungen für die Energien (igrad=1), skaliert mit  $O(nN^3)$ ;

$T_{\text{SCF}}$ , Zeit für die SCF-Energie-Berechnung, die in jedem Fall gemacht werden muss und daher als ideale Referenzgrösse dient, skaliert mit  $O(N^3)$ .

Für die analytischen sowie numerischen Gradienten, welche auf der Subroutine GRDCAL basieren, wird ein Skalierungsverhalten von  $O(nN^2)$  erwartet, da die zugrunde liegenden Formeln der Energieberechnung via 2.10a und 2.10b mit  $O(N^2)$  skalieren. Um systembedingte Zeitschwankungen der Tests zu verhindern, wurden die Zeittests für alle Gradientenmethoden insgesamt 10 Mal durchgeführt. Die Ausnahme davon bildet  $T_{\text{Diff-full}}$ , da die entsprechenden Zeiten sehr gross werden, so dass hier die Zeit jeweils nur einmal gemessen wurde.

Die Gradientencodes  $G_{\text{Adi2-d}}$  und  $G_{\text{Adi2-k}}$  berechnen sowohl die Gradienten als auch die Terme der abzuleitenden Funktion. Gezwungenermassen! Dies ist unabdingbar, da der von ADIFOR erzeugte Code die einzelnen Operationen bei der Berechnung der abzuleitenden Funktion ja jeweils durch die entsprechenden Operationen für die Gradientenberechnung ergänzt. Um die Zeiten für die effektive Gradientenberechnung zu erhalten, wurde von den gemessenen Gesamtzeiten jeweils die Zeit für eine Funktionsberechnung (GRDCAL) subtrahiert.

Die Zeittests wurden wiederum für AM1 und OM1 mit dem Alkantestsatz durchgeführt. Als Hardware diente wiederum eine Compaq ES40 EV8 (4\*833 MHz, 7 GB RAM), wobei jeweils nur eine CPU verwendet wurde.

### Zeittests für AM1.

Die Tabelle 2.7 gibt das Zeitverhalten verschiedener Gradienten wieder. Die Zeit

**Tabelle 2.7:** CPU-Zeiten [s]: Energie und Gradienten, AM1

$k^b$	$n^c$	$N^d$	mndo99		ADIFOR <sup>a</sup>		numerisch <sup>a</sup>		
			$T_{\text{SCF}}$	$T_{\text{Analyt}}$	$T_{\text{Adi2-d}}$	$T_{\text{Adi2-k}}^e$	$T_{\text{Diff-d}}$	$T_{\text{Diff-k}}$	$T_{\text{Diff-full}}^f$
5	51	32	0.01	0.00	0.02	0.02	0.07	0.04	0.35
10	96	62	0.04	0.01	0.14	0.13	0.53	0.27	4.07
15	141	92	0.12	0.02	0.42	0.40	1.68	0.88	17.72
20	186	122	0.27	0.03	0.94	0.90	3.88	2.06	54.03
25	231	152	0.60	0.04	1.85	1.65	7.26	3.71	125.25
30	276	182	1.01	0.06	3.19	2.82	12.16	6.41	254.79
35	321	212	1.54	0.08	5.05	4.36	18.59	10.44	459.29
40	366	242	2.27	0.11	7.39	6.25	29.64	16.25	767.93
45	411	272	3.19	0.14	10.71	8.84	39.87	24.33	1198.85
50	456	302	4.36	0.16	14.18		51.61	30.13	1833.77

<sup>a</sup> Zeit nur für die Gradientenberechnung

<sup>b</sup>  $C_k H_{2k+2}$

<sup>c</sup>  $n=9k+6$ , Zahl der Basisorbitale

<sup>d</sup>  $N=6k+2$ , Zahl der Kernkoordinaten

<sup>e</sup> Stacksizeprobleme für  $k > 46$

<sup>f</sup> eine Messung, alle anderen Werte gemittelt aus 10 Messungen

für die (iterative) Energieberechnung ( $T_{\text{SCF}}$ ) dient als Referenzgrösse, welche bekanntermassen mit  $O(N^3)$  skaliert. Die handkodierte analytischen Gradienten [9] ( $T_{\text{Analyt}}$ ) skalieren mit  $O(N^2)$  und sind somit sehr effizient. Für die betrachteten Alkane beträgt die benötigte Zeit 10–3% von  $T_{\text{SCF}}$ , Tendenz fallend für grössere Moleküle. Am anderen Ende der Skala liegt  $T_{\text{Diff-full}}$ , ein numerischer Ansatz, bei dem die Gradienten aus vollständigen SCF-Rechnungen an ausgelenkten Geometrien ermittelt werden. Die Skalierung hierfür beträgt  $O(nN^3)$ . Wie aus

der Tabelle 2.7 ersichtlich, ergibt sich für  $k=50$  aus den Zeiten für  $T_{\text{Diff.full}}$  resp.  $T_{\text{SCF}}$  ein Verhältnis von 421, in ziemlich guter Übereinstimmung zum theoretisch erwarteten Wert von  $n=456$ . Bei der direkten Integral-Version sind die Zeiten für den numerischen ( $T_{\text{Diff.d}}$ ) sowie den analytischen ( $T_{\text{Adi2.d}}$ ) Gradienten jeweils um die Faktoren 11–14 sowie 3–3.5 grösser als  $T_{\text{SCF}}$ . Beachtenswert ist, dass der analytische Gradient  $G_{\text{Adi2.d}}$  um einen Faktor von 3–4 schneller ist als sein numerisches Pendant  $G_{\text{Diff.d}}$ . Die beobachteten Zeiten entsprechen dem erwarteten Skalierungsverhalten von  $O(nN^2)$ . Die konventionelle Implementation ist in allen Zeittests der direkten Implementation überlegen, sie weist jedoch erhebliche Probleme mit dem Memorymanagement auf (siehe Kapitel 2.8.4). Jedenfalls betragen die Faktoren im Vergleich zu  $T_{\text{SCF}}$  für  $T_{\text{Diff.k}}$  6–7.5 und 2.75–3.5 für  $T_{\text{Adi2.k}}$ . Wiederum ist der analytische Gradient  $G_{\text{Adi2.k}}$  um einen Faktor von 2–2.75 schneller als  $G_{\text{Diff.k}}$ . Gesamthaft sind die Gradienten  $G_{\text{Adi2.k}}$  um 8–18% schneller als  $G_{\text{Adi2.d}}$ , je grösser also das Molekül ist, umso grösser ist der Zeitvorteil von  $T_{\text{Adi2.k}}$ .

Die Zeiten für  $T_{\text{Diff.d}}$  und  $T_{\text{Diff.k}}$  stehen in einem Verhältnis von knapp 2:1. Dieses Verhältnis entspricht in etwa der unterschiedlichen Zahl an Rechenoperationen in der direkten und konventionellen Implementierung. Das entsprechende Verhältnis für  $T_{\text{Adi2.d}}$  und  $T_{\text{Adi2.k}}$  ist niedriger (nur wenig höher als 1:1). Dies mag daran liegen, dass die konventionelle Version durch das Verarbeiten grosser Felder mit Integralableitungen verlangsamt werden könnte, während die direkte Version vermutlich von effizienten Cache-Operationen profitiert. Eine genauere Analyse wurde jedoch nicht durchgeführt.

### Zeittests für OM1.

Die gleichen Zeittests wie für AM1 wurden auch für OM1 gemacht. Zu beachten ist, dass für OM1 keine analytischen Gradienten in MNDO99 vorhanden sind. Statt dessen wurde mit handkodierten numerischen Gradienten für OM1 verglichen, die bei der Bestimmung der Integralableitungen in (2.11a) für eine gegebene Koordinate nur jeweils diejenigen Terme neu berechnen, welche sich bei der betreffenden Auslenkung ändern können (Skalierungsverhalten  $O(N^2)$ ). Die Tabelle 2.8 gibt das Zeitverhalten verschiedener Gradienten wieder. Als Referenzgrösse dient wiederum  $T_{\text{SCF}}$ . Die handkodierten numerischen Gradienten  $G_{\text{Diff.m}}$  sind sehr effizient, die benötigte Zeit liegt bei 45–17% von  $T_{\text{SCF}}$ , Tendenz fallend für grössere Moleküle.  $T_{\text{Diff.full}}$  beansprucht am meisten Zeit, das Verhältnis zu  $T_{\text{SCF}}$  beträgt 493, in guter Übereinstimmung zu dem erwarteten Wert von 456. Die Zeitvergleiche bei der direkten Implementation liefern Faktoren von 53–63 für  $T_{\text{Diff.d}}$  und 17–18.5 für  $T_{\text{Adi2.d}}$  im Vergleich zu  $T_{\text{SCF}}$ . Der analytische Gradient  $G_{\text{Adi2.d}}$  ist gegenüber der numerischen Variante  $G_{\text{Diff.d}}$  um einen Faktor von 3–3.5 schneller. Die konventionelle Implementation verhält sich ähnlich wie bei der AM1-Untersuchung:  $T_{\text{Diff.k}}$  ist im Vergleich zu  $T_{\text{SCF}}$  einen Faktor von 27–31 langsamer,  $T_{\text{Adi2.k}}$  um 9.5–10.5. Der analytische Gradient  $G_{\text{Adi2.k}}$  ist wiederum um einen Faktor von 2.5–3 schneller als der numerische Gradient  $G_{\text{Diff.k}}$ . Analog zu den Betrachtungen zu AM1 ist  $G_{\text{Adi2.k}}$  um einen Faktor von 1.7 schneller als  $G_{\text{Adi2.d}}$ .

**Tabelle 2.8:** CPU-Zeiten: Energie und Gradienten, OM1

$k^b$	$n^c$	$N^d$	mndo99		ADIFOR <sup>a</sup>		numerisch <sup>a</sup>		
			$T_{\text{SCF}}$	$T_{\text{Diff}\cdot\text{m}}$	$T_{\text{Adi2}\cdot\text{d}}$	$T_{\text{Adi2}\cdot\text{k}}^e$	$T_{\text{Diff}\cdot\text{d}}$	$T_{\text{Diff}\cdot\text{k}}$	$T_{\text{Diff}\cdot\text{full}}^f$
5	51	32	0.01	0.02	0.17	0.10	0.56	0.28	0.59
10	96	62	0.05	0.05	0.93	0.53	3.18	1.58	5.58
15	141	92	0.15	0.09	2.74	1.56	9.05	4.51	22.49
20	186	122	0.33	0.15	5.85	3.42	19.26	9.69	65.37
25	231	152	0.62	0.23	11.00	6.30	36.89	17.61	146.22
30	276	182	1.06	0.31	18.05	10.28	57.18	28.80	300.50
35	321	212	1.58	0.40	27.99	15.86	87.98	44.52	535.96
40	366	242	2.30	0.52	42.61	23.37	127.12	65.52	903.48
45	411	272	3.47	0.64	59.38	33.57	178.09	92.39	1420.53
50	456	302	4.52	0.77	78.60		240.42	123.24	2228.91

<sup>a</sup> Zeit nur für die Gradientenberechnung

<sup>b</sup>  $C_k H_{2k+2}$

<sup>c</sup>  $n=9k+6$ , Zahl der Basisorbitale

<sup>d</sup>  $N=6k+2$ , Zahl der Kernkoordinaten

<sup>e</sup> Stacksizeprobleme für  $k > 46$

<sup>f</sup> eine Messung, alle anderen Werte gemittelt aus 10 Messungen

In Analogie zu AM1 liegt das Verhältnis von  $T_{\text{Diff}\cdot\text{d}}$  und  $T_{\text{Diff}\cdot\text{k}}$  bei knapp 2, während das Verhältnis von  $T_{\text{Adi2}\cdot\text{d}}$  und  $T_{\text{Adi2}\cdot\text{k}}$  bei OM1 auf etwa 1.7–1.8 ansteigt (AM1 1.0–1.2) und damit den Erwartungen besser entspricht (siehe oben). Dies hängt damit zusammen, dass der Rechenaufwand für die Integrale in OM1 insgesamt deutlich höher ist als in AM1 (vgl. z.B.  $T_{\text{Diff}\cdot\text{d}}$  in Tab. 2.7 und 2.8) und dass die relative Zahl der Rechenoperationen dann den Vergleich von  $T_{\text{Adi2}\cdot\text{d}}$  und  $T_{\text{Adi2}\cdot\text{k}}$  dominiert.

Zusammenfassend gilt für OM1, dass die analytischen Gradienten  $G_{\text{Adi2}}$  erheblich langsamer sind als die handkodierte numerischen Gradienten in MNDO99. Sie sind jedoch schneller (Faktoren von ca. 2–4) als ihre entsprechenden numerischen Implementationen und zudem numerisch präziser als alle verfügbaren Alternativen.

### 2.8.4 Memoryanforderungen

Der Memorybedarf ist ein weiteres wichtiges und oft auch limitierendes Kriterium der verschiedenen Gradienten-Codes. Fortran77 als auch Fortran90 weisen keine Konstrukte auf, die es erlauben würden, aus einem laufenden Programm die Memoryanforderungen für spezifische Subroutinen auszulesen. Dieses Problem kann dadurch umgangen werden, dass vor und nach dem Aufruf der entsprechenden Subroutinen Systemaufrufe eingeführt werden. Diese Systemaufrufe können zwischen

den verschiedenen Betriebssystemen variieren, für tru64 (COMPAQ) hat sich folgendes Konstrukt bewährt:

```
CALL SYSTEM ("sh skript")
CALL g_GRDCAL
CALL SYSTEM ("sh skript")
wobei skript =
      ps -a -o "cmd rss vsz" | grep mndoadi >> cfile
```

Hierbei ist g\_GRDCAL die Subroutine, an deren Memorybedarf man interessiert ist; mndoadi ist der Name des Executables, und ps resp. grep sind typische UNIX-Kommandos. Der obige Befehl liefert die unmittelbare Belegung des wirklich gebrauchten Speicherplatzes (resident size, rss) und den reservierten Memoryplatz (virtual size, vsz).

Ein weiterer limitierender Faktor bei einem Programmlauf kann die 'Stacksize' sein. Diese systeminterne Grösse definiert die maximale Grösse (in MByte) der Argumentliste beim Aufruf von Subroutinen. Wenn über die Argumentliste grosse und viele Arrays an Subroutinen weitergegeben werden und diese Subroutinen eventuell auch noch weitere Subroutinen aufrufen (etc.), so kann es sein, dass das laufende Programm an die gegebenen Grenzen der Stacksize stösst und die Berechnung abbricht. Der einzige Ausweg besteht dann darin, die Stacksize innerhalb des Kernels neu zu definieren oder aber das Programm neu zu strukturieren. Für alle Zeit- und Memorytests lag die Grösse der Stacksize bei 2 GB.

Im folgenden wird nur auf das Memoryverhalten von AM1 und OM1 bei dem Alkan-Testsatz eingegangen. Dazu wurde vor und nach Aufruf der zu beobachtenden Subroutine ein CALL SYSTEM in den Source-Code eingebaut (siehe oben) und die Differenz der 'resident size' berechnet.

Für die numerischen Gradienten  $G_{\text{Diff},d}$  und  $G_{\text{Diff},k}$  wurden keine Memorytests durchgeführt, weil hierbei über die Energieberechnung hinaus kein zusätzliches Memory benötigt wird. Desweiteren wurden auch keine Memorytests für die in MNDO99 implementierten analytischen Gradienten durchgeführt, da diese sehr wenig zusätzliches Memory brauchen. Der durch ADIFOR erzeugte analytische Gradientencode verwendet für alle aktiven Variablen (d.h. abhängig von den Kernkoordinaten) ein zusätzliches Feld für die entsprechenden Ableitungen. Für diese Felder muss eine fixe obere Grenze (maximale Zahl der unabhängigen Variablen) angegeben werden, weil Fortran77 keine dynamische Memoryallozierung zulässt. Dies hat zur Folge, dass der ADIFOR-Gradienten-Code den Memoryverbrauch aufbläht. Die gemessenen Werte sind in der Tabelle 2.9 wiedergegeben.

Offensichtlich verhalten sich die beiden Implementationen unterschiedlich.

**Tabelle 2.9:** Memorybedarf [MB] der konventionellen sowie direkten Implementation für AM1 und OM1

$k^a$	AM1		OM1	
	$M_{\text{Adi2-d}}$	$M_{\text{Adi2-k}}$	$M_{\text{Adi2-d}}$	$M_{\text{Adi2-k}}$
1	1.2	1.7	3	3.6
5	3.0	18.7	6.1	21.6
10	4.1	67.6	7.1	70.5
20	6.0	258.0	8.9	261.0
30	7.9	575.0	10.8	578.0
40	10.3	985.0	13.2	982.0
50	12.3	1610.0	15.2	1604.0

<sup>a</sup>  $C_k H_{2k+2}$

Die direkte Implementation  $G_{\text{Adi2-d}}$  weist nur wenige kleine zusätzliche Arrays auf, demzufolge ist der zusätzlich beanspruchte Memoryverbrauch gering und skaliert linear. Für das grösste Testmolekül ( $k=50$ ) werden nur 12.3 MB für AM1 resp. 15.2 MB für OM1 beansprucht. Völlig anders sieht die Situation für die konventionelle Implementation  $G_{\text{Adi2-k}}$  aus, welche grosse Felder verwendet, die mit  $O(nN^2)$  skalieren. Demzufolge steigen die entsprechenden Memoryanforderungen mindestens quadratisch an; für  $k=50$  beträgt der Memoryverbrauch ca. 1.6 GB für AM1 sowie OM1. Dieser steile Anstieg begrenzt die Anwendbarkeit der konventionellen Implementierung und spricht klar zugunsten der direkten Implementierung.

## 2.9 Zusammenfassung

Unter Verwendung von ADIFOR2.0 im Vorwärtsmodus wurde analytischer Gradienten-Code für alle in MNDO97 implementierten semiempirischen SCF Methoden erzeugt. Seine Richtigkeit und numerische Präzision wurden durch Vergleiche mit unabhängig berechneten Gradienten bestätigt (analytische Vergleichswerte bei MNDO, AM1, PM3 und MNDO/d; numerische Vergleichswerte bei OM1 und OM2). Ein Fehler in dem existierenden Code wurde hierbei entdeckt und korrigiert. Die ADIFOR-erzeugten Routinen für den analytischen Gradient sind relativ langsam verglichen mit existierenden handkodierten Versionen, aber sie sind schneller als ihre numerischen Pendanten. Die integral-direkte Version des ADIFOR-basierten Codes ist der konventionellen Version klar vorzuziehen, weil sie viel weniger Memory beansprucht und nur wenig höhere Rechenzeiten erfordert.

Ein wesentlicher Fortschritt der in diesem Kapitel geschilderten Arbeiten besteht darin, dass der ADIFOR-erzeugte Code die bisher nicht verfügbaren

analytischen SCF-Gradienten für OM1 und OM2 bereitstellt. Für die Zukunft ist geplant, mittels ADIFOR auch die analytischen CI-Gradienten für OM1 und OM2 zu erzeugen. Weiterhin erscheint es wünschenswert, mit einer konsolidierten Version (3.0) von ADIFOR (sobald verfügbar) zu untersuchen, ob man im Rückwärtsmodus automatisch einen effizienteren Gradientencode erzeugen kann, wie dies theoretisch zu erwarten ist (siehe Kapitel 2.6).



# Kapitel 3

## Parametrisierung semiempirischer Methoden

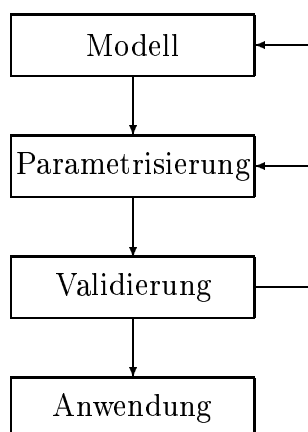
### 3.1 Einleitung

Mit den semiempirischen Methoden der Quantenchemie können die Eigenschaften grosser Moleküle schnell und dennoch oft mit brauchbarer Genauigkeit berechnet werden. Dieser Vorteil muss in der Entwicklungsphase mit einem beträchtlichen Aufwand an Zeit für die Parametrisierung erkaufte werden, da die Güte semiempirischer Methoden von derselben in einem hohen Masse abhängt [34].

Bislang erfolgte die Parametrisierung üblicherweise durch viele einzelne lokale Optimierungen, bei denen jeweils für einen Satz von Startparametern die Fehlerquadratsumme bezüglich gewisser Referenzdaten minimiert wurde und danach die Qualität der besten erhaltenen Parametersätze anhand verschiedener Validierungen bewertet wurde. In dem hier geschilderten Projekt wurde nun versucht, diesen langwierigen Parametrisierungsprozess zu systematisieren und zu automatisieren. Dazu wurde ein Programm geschrieben, welches selbstständig die zu optimierenden Parameter variiert, die Ergebnisse der einzelnen Optimierungsläufe hinsichtlich gewisser Kriterien auswertet, die besten erzeugten Parametersätze automatisch auswählt und wieder in den nächsten Parametrisierungszyklus einbindet. Neben der systematischen Durchsuchung des Parameterraums sind in dem Programm auch genetische Algorithmen implementiert, welche durch stochastische Variation der Parameter innerhalb eines vorgegebenen Bereichs eine globale Optimierung anstreben. Im folgenden werden zunächst die Methoden und Algorithmen der Parametrisierung allgemein besprochen, bevor das neu entwickelte Programmsystem vorgestellt wird. Danach wird dessen Anwendung für die testweise Reparametrisierung der Elemente C, H, N und O in dem etablierten MNDO Verfahren sowie für die Parametrisierung des Elementes Fluor in den neuen OM2 und OM3 Verfahren beschrieben.

## 3.2 Entwicklung semiempirischer Methoden: Übersicht

Die Entwicklung semiempirischer Methoden erfolgt gemäss Abbildung 3.1 in einem mehrstufigen Prozess. Zunächst muss das zugrunde liegende **Modell** spezifiziert wer-



**Abbildung 3.1:** Übersicht der semiempirischen Methodenentwicklung

den, das den theoretischen Ansatz und alle vorgenommenen Näherungen definiert (ausgehend von der Schrödingergleichung). Innerhalb eines Entwicklungsprojektes ist dieses Modell meist fest vorgegeben, aber generell gilt, dass das Modell sich ändern kann, weil man langfristig eine immer bessere Annäherung an die Wirklichkeit anstrebt und deshalb fortlaufend versucht, das gewählte Modell zu verfeinern. Die **Parametrisierung** ist ein zentraler Schritt bei der Entwicklung semiempirischer Methoden. Da das verwendete Modell die physikalischen Gegebenheiten immer nur unzulänglich wiedergeben kann, enthalten einzelne Terme in dem Modell justierbare Parameter, die anhand geeigneter Referenzdaten kalibriert werden, um die modellbedingten Fehler zumindest teilweise zu kompensieren. Dabei werden die Parameter so angepasst, dass der Unterschied zwischen den berechneten Moleküleigenschaften und den entsprechenden experimentellen Werten möglichst klein wird. Dies geschieht durch Minimierung der Fehlerquadratsumme (SSQ, sum of squares):

$$SSQ = \sum_i [(c_i - e_i)w_i]^2$$

$c_i$  = Berechnete Moleküleigenschaft  $i$

$e_i$  = Experimentelle Moleküleigenschaft  $i$

$w_i$  = Gewichtungsfaktor

Ein optimaler Parametersatz weist also einen möglichst kleinen SSQ-Wert für die gewählten Referenzmoleküle und Referenzdaten auf. Wegen der Existenz vieler lo-

kaler Minima für SSQ ist ein solcher optimaler Satz jedoch nur schwer zu finden, so dass normalerweise viele separate Parametrisierungsläufe mit unterschiedlichen Startparametern gemacht werden müssen.

Die **Validierung** ist der dritte Schritt bei der Methodenentwicklung. Dabei werden mit dem besten erhaltenen Parametersatz P' weitere Moleküle und weitere Eigenschaften berechnet, welche während der Parametrisierung nicht verwendet worden sind. Es kann sich herausstellen, dass der Satz P' zwar geeignet ist für die 'Parametrisierungsmoleküle', jedoch nicht für die 'Validierungsmoleküle' bzw. die entsprechenden Eigenschaften. In solch einem Fall muss man sich wieder der Parametrisierung zuwenden und die Referenzdaten durch Hinzunahme neuer Moleküle und Eigenschaften ergänzen, ggf. auch unter veränderter Gewichtung einzelner Referenzdaten. Es kann sein, dass auch nach solchen Reparametrisierungsversuchen gewisse Moleküleigenschaften noch unzureichend berechnet werden. Dies deutet dann darauf hin, dass diese Eigenschaften von dem gewählten Modell nicht gut wiedergegeben werden können und dass man deshalb eine Änderung des zugrunde liegenden Modells in Erwägung ziehen muss.

Die **Anwendung** kommt zum Zuge, wenn alle obigen Schritte erfolgreich waren und die Methode deshalb für den alltäglichen Gebrauch empfohlen werden kann.

### 3.2.1 Auswahl der Referenzdaten

Die gewählten Referenzdaten sollten repräsentativ sein für die Moleküle und Eigenschaften, auf welche die betreffende semiempirische Methoden später in der Praxis angewendet werden soll [34]. Es macht keinen Sinn, Eigenschaften als Referenzdaten zu verwenden, die im Rahmen des zugrunde liegenden Modells nicht adäquat beschrieben werden können. Für eine allgemeine Parametrisierung auf Grundzustandseigenschaften haben sich folgende Arten von Referenzdaten bewährt:

**Molekülgeometrien**, insbesondere Bindungslängen, Bindungswinkel und Torsionswinkel für die charakteristischen Strukturelemente in einem Molekül. Die Referenzwerte werden meist aus spektroskopischen Messungen oder aus Beugungsexperimenten entnommen. Gleichgewichtsdaten ( $r_e$ ) sind vorzuziehen, oft hat man aber nur Grundzustandsdaten ( $r_0$ ) oder andere Werte ( $r_g$ ) zur Verfügung.

**Bildungsenthalpien** als Mass für relative Stabilitäten. Referenzdaten müssen von gasförmigen Systemen stammen, da die semiempirischen Rechnungen sich auf isolierte Moleküle beziehen. Man verwendet Standardbildungsenthalpien in der Gasphase bei 298K ( $\Delta H_f$ , heat of formation). Zusätzlich werden auch relative Energien von Isomeren und Konformeren verwendet.

**Dipolmomente** als Indikator für die Elektronendichte-Verteilung innerhalb eines Moleküls.

**Ionisierungspotentiale** als Mass für die relativen Energien von Neutralelektronenmolekül und zugehörigem Kation.

**Weitere Moleküleigenschaften** werden manchmal als Referenzdaten hinzugezogen, beispielsweise Elektronenaffinitäten, Polarisierbarkeiten und Schwingungsfrequenzen.

Traditionell werden als Referenzdaten in semiempirischen Parametrisierungen ausschliesslich experimentelle Daten verwendet, weil semiempirische Rechnungen konzeptionell an der experimentellen Wirklichkeit kalibriert werden sollen (und nicht an den Ergebnissen eines anderen theoretischen Modells). Seit einigen Jahren haben hochkorrelierte *ab initio* Rechnungen aber eine derart hohe Genauigkeit erreicht, dass sie anstelle von experimentellen Daten als Referenzdaten bei Parametrisierungen und Validierungen eingesetzt werden können. Dies geschieht auch im folgenden, sofern experimentelle Referenzdaten überhaupt nicht oder nur in zweifelhafter Qualität zur Verfügung stehen.

### 3.2.2 Gewichtung der Referenzdaten

Die verschiedenen Referenzdaten müssen einerseits gewichtet werden, um sie in dimensionslose Grössen zu überführen. Andererseits dient die Gewichtung natürlich auch dazu, den Beitrag der einzelnen Referenzdaten zu der Fehlerquadratsumme festzulegen. Die typischerweise verwendeten Gewichtungsfaktoren sind in der Tabelle 3.1 aufgelistet. Bei Bedarf können diese Gewichtungsfaktoren abgeändert werden,

**Tabelle 3.1:** Gewichtungsfaktoren für Referenzdatensätze

Referenzwert	Faktor
$\Delta H_f$	1.0 mol kcal <sup>-1</sup>
Bindungslänge	100.0 Å <sup>-1</sup>
Bindungswinkel	0.66 deg <sup>-1</sup>
IP	10.0 eV <sup>-1</sup>
Dipolmoment	20.0 D <sup>-1</sup>

IP = Ionisierungspotential

und zwar generell für jede gewünschte Eigenschaft als auch individuell für einzelne Referenzdaten. Auf diese Weise kann die Parametrisierung bis zu einem gewissen Grad gesteuert werden.

### 3.2.3 Optimierung der Parameter

In diesem Abschnitt wird der Algorithmus geschildert, der in dem vorhandenen Parametrisierungsprogramm (pp99) implementiert und in früheren Parametrisierungen

eingesetzt worden ist. Hierbei werden die Parameter durch eine lokale nichtlineare Minimierung der gemäss Formel 3.1 berechneten Fehlerquadratsumme (SSQ) optimiert. Der Vektor  $x$  enthält die zu optimierenden Parameter.

$$SSQ = \sum_{i=1}^m [(c_i(x) - e_i)w_i]^2 = \sum_{i=1}^m f_i^2(x) = f(x)^T f(x) \quad (3.1)$$

$c_i$  = Berechnete Moleküleigenschaft  $i$

$e_i$  = Experimentelle Moleküleigenschaft  $i$

$w_i$  = Gewichtungsfaktor

Die gewichteten Abweichungen zwischen den berechneten und gemessenen Daten sind in dem Vektor  $f(x)$  zusammengefasst. Die Jacobi-Matrix  $J_f$  ist folgendermassen definiert:

$$J_f = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

Die erste Ableitung, der Gradient  $g$ , ergibt sich somit zu:

$$g = \begin{bmatrix} \frac{\partial SSQ}{\partial x_1} \\ \frac{\partial SSQ}{\partial x_2} \\ \vdots \\ \frac{\partial SSQ}{\partial x_n} \end{bmatrix} = 2 \sum_{i=1}^m f_i(x) \nabla f_i(x) = 2J_f^T f$$

Die zweite Ableitung, auch Hesse-Matrix  $H$  genannt, berechnet sich zu:

$$\begin{aligned} H &= 2 \sum_{i=1}^m \nabla f_i(x) \nabla f_i(x) + 2 \sum_{i=1}^m f_i(x) \nabla^2 f_i(x) \\ &= 2J_f^T(x) J_f(x) + 2 \sum_{i=1}^m f_i(x) \nabla^2 f_i(x) \approx 2J_f^T J_f \end{aligned} \quad (3.2)$$

Der zweite Summand in Gleichung 3.2 kann häufig vernachlässigt werden, da er verglichen mit dem ersten Summanden meist klein ist.

Ausgehend von diesen Definitionen können nun die Parameter optimiert werden. Da mehrere Variablen vorhanden sind, benutzt man ein iteratives Verfahren. Dazu werden Startwerte gewählt, die sukzessiv modifiziert werden, bis ein Konvergenzkriterium erfüllt ist. Die neuen Parameter erhält man jeweils gemäss Gleichung 3.3

$$x^{k+1} = x^k + \alpha p \quad (3.3)$$

wobei  $k$  den Zyklus,  $p$  den Suchvektor und  $\alpha$  die zugehörige Schrittweite bezeichnet (letztere wird in jeder Iteration neu bestimmt). Der Suchvektor  $p$  kann auf verschiedene Arten definiert werden. Die grundlegende Idee besteht darin, das Taylor-Theorem auf eine Funktion  $f$  und ihre Umgebung  $\alpha p$  anzuwenden:

$$f(x + \alpha p) = f(x) + \alpha p^T \nabla f(x) + \frac{1}{2} \alpha^2 p^T \nabla^2 f(x) p + \text{weitere Terme}$$

Je nach Verfügbarkeit der benötigten Ableitungen können verschiedene Suchvektoren benutzt werden:

$$p = -\nabla f(x) = J_f^T f \quad (3.4)$$

$$p = -\nabla^2 f^{-1}(x) \nabla f(x) = H^{-1} g = -[J_f^T J_f]^{-1} J_f^T f \quad (3.5)$$

$$p = [J_f^T J_f + \lambda^2 1]^{-1} J_f^T f \quad (3.6)$$

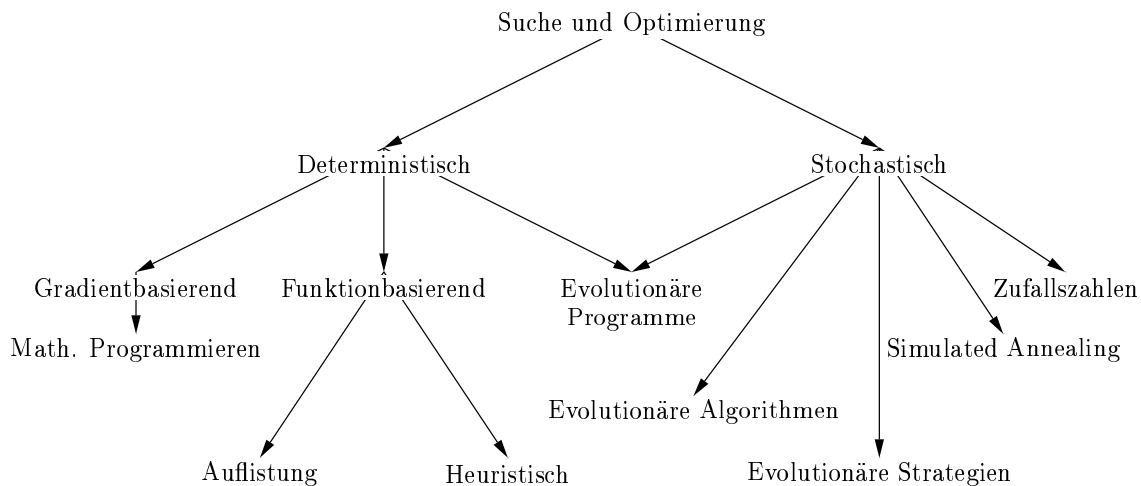
Ein Suchvektor laut (3.4) wird in den 'Steepest-Descent' Methoden verwendet, die zwar robust hinsichtlich der Suchrichtung sind, aber gegen Ende schlecht konvergieren. Ein Suchvektor gemäss (3.5) führt zur Gauss-Newton Methode, die zu Beginn Probleme mit der groben Suchrichtung haben kann, in der Nähe eines Minimums jedoch schnell konvergiert. Einen guten Kompromiss zwischen (3.4) und (3.5) bietet die Levenberg-Marquardt Methode gemäss (3.6) mit einem Parameter  $\lambda$ , welcher das Verhalten des Suchvektors steuert. So führt ein grosses  $\lambda$  zur 'Steepest-Descent' Methode, ein kleines hingegen zur Gauss-Newton Methode.  $\lambda$  selbst ist definiert durch  $\lambda = c_\lambda \sqrt{SSQ}$ , wobei  $c_\lambda$  durch eine quadratische Interpolation bestimmt wird. Das vorhandene Parametrisierungsprogramm verwendet die Levenberg-Marquardt Methode, die benötigten Ableitungen werden über numerische finite Differenzen erhalten.

Der skizzierte Algorithmus ist in der Lage, ausgehend von einem geratenen Startparametersatz für einen gegebenen Referenzdatensatz das nächste lokale Minimum auf der Parameter-Hyperfläche zu finden. Die Erfahrung hat jedoch gezeigt, dass die Parameter-Hyperfläche meist viele solche lokale Minima aufweist. Dies gilt auch dann, wenn man sicherstellt, dass die Zahl der Referenzdaten wesentlich grösser ist als die Zahl der zu optimierenden Parameter und dass die Parameter nur physikalisch sinnvolle Werte annehmen können. Es ist daher eine nicht-triviale Aufgabe, das beste lokale Minimum (d.h. das globale Minimum) auf der Parameter-Hyperfläche innerhalb des Bereichs der physikalisch sinnvollen Parameterwerte zu finden. In der Literatur existieren eine Reihe von Algorithmen, welche genau dieses Problem behandeln und zu lösen versuchen. Hierzu gehören genetische Algorithmen, aber auch systematische Suchalgorithmen kombiniert mit lokalen Optimierungen. Das folgende Kapitel gibt hierzu eine kurze Übersicht.

## 3.3 Optimierungsmethoden

### 3.3.1 Klassifizierung

Die Suche nach Minima, Maxima und weiteren stationären Punkte einer gegebenen Funktion kann auf vielerlei Arten geschehen. Prinzipiell lässt sich die Vorgehensweise gemäss Abbildung 3.2 klassifizieren [35, 36].



**Abbildung 3.2:** Klassifikation von Such- und Optimierungsmethoden

**Deterministische Methoden** versuchen, das Optimum dadurch zu erreichen, dass sie die Umgebung des jeweiligen Funktionswertes untersuchen und gegebenenfalls in Richtung eines (analytischen oder numerischen) Gradienten optimieren. Zu diesen Methoden gehören alle Algorithmen, welche in irgendeiner Art die Ableitungen der Funktion benutzen. Des weiteren zählen hierzu heuristische Methoden wie z.B. Tabu-Search [37, 38], welche auch eine kurzfristige Verschlechterung des Funktionswertes zulassen, zudem können gewisse Suchregionen ausgeschlossen werden. Enumerative Methoden beinhalten ein komplettes (systematisches) Absuchen eines wohldefinierten Parameterraumes.

**Stochastische Methoden** hingegen versuchen eine Funktion zu optimieren, indem sie eine 'vernünftige' Zufallskomponente einführen, welche einerseits den Funktionsraum ausgiebig absucht und andererseits geeignete Suchregionen lokalisiert, die dann eventuell mit weiteren Optimierungsmethoden untersucht werden. Zu den stochastischen Methoden werden unter anderem die genetischen Algorithmen, Simulated Annealing, Zufallssuche und evolutionäre Strategien gezählt.

Unter dem Begriff evolutionäre Programme sind diejenigen Ansätze zusammengefasst, welche sich sowohl deterministischer als auch stochastischer Methoden

bedienen. Ein typisches Beispiel sind genetische Algorithmen in Kombination mit einer zielgerichteten gradientenbasierten Optimierung der stochastisch erzeugten Ergebnisse.

### 3.3.2 Evolutionäre Algorithmen

Die evolutionären Algorithmen wurden in den 70'er Jahren entwickelt, wobei drei verschiedene Implementationen unterschieden werden können: Genetische Algorithmen (GA), welche erstmals von Holland [39] benutzt und im folgenden durch Goldberg [40] populär wurden; evolutionäre Strategien (ES), die in Deutschland von Rechenberg [41] und Schwefel [42] entwickelt wurden; und evolutionäres Programmieren (EP), welches durch L. J. Fogel [43] und D. B. Fogel [44] bekannt wurde.

Die Grundidee ist bei allen Implementationen die gleiche: Man geht von einer Population aus, die aus  $\mu$  Individuen besteht. Ein Individuum repräsentiert einen bestimmten Punkt des Funktionsraumes, welcher durch die zu optimierenden unbekannt Grössen (z.B. Parameter) definiert wird. In einem ersten Schritt wird nun der zu diesem Punkt zugehörige Funktionswert berechnet. Diesem Funktionswert wird dann eine Fitness zugewiesen, welche eine Aussage macht über die Qualität des Wertes hinsichtlich anderer Kriterien. Prinzipiell gilt, dass ein hoher Fitnesswert einer guten Lösung innerhalb des Funktionsraumes entspricht. Indem an der Anzahl  $\lambda$  der besten Individuen die Originalwerte (z.B. Parameter) durch verschiedenste Operatoren (z.B. Mutation, Crossover) verändert werden, können  $\lambda$  neue Individuen und  $\mu - \lambda$  alte Individuen an einem neuen Zyklus mitmachen. Dieser iterative Prozess wird solange fortgesetzt, bis ein Abbruchkriterium erfüllt ist. Ein Abbruchkriterium kann z.B. die Zahl der Zyklen sein.

Die Analogie zu den Begriffen bei Darwin [45] ist gewollt, da viele Ideen konkrete Vorbilder in der Evolution haben. Im folgenden werden die Konzepte der genetischen Algorithmen genauer dargelegt, da diese in unserem Programm implementiert worden sind.

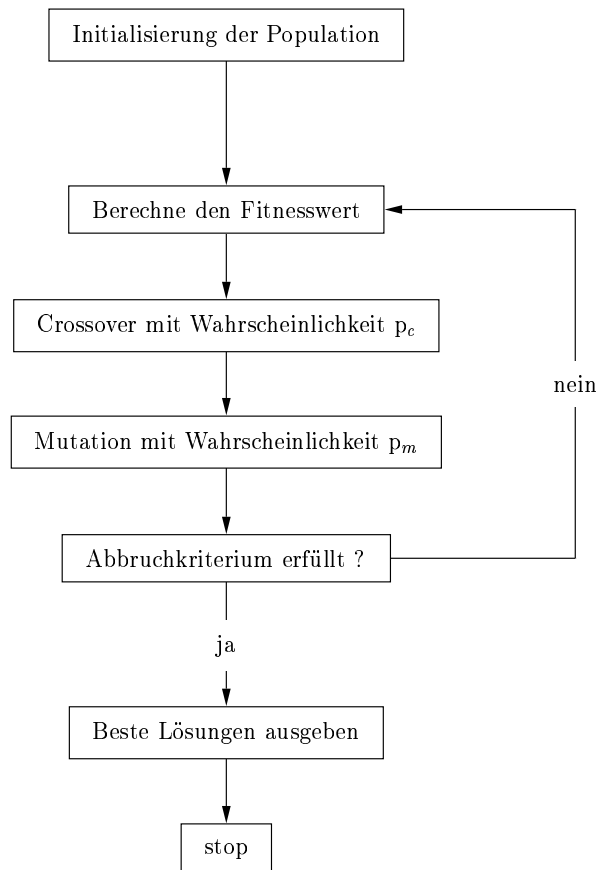
### 3.3.3 Genetische Algorithmen

Die Anwendungen von genetischen Algorithmen (GA) sind vielseitig, beispielsweise in der Chemie [46, 47, 48], im Ingenieurwesen [49, 50, 51], beim Molecular Modelling [52, 53, 54, 55] und in vielen weiteren Gebieten. Daher existiert eine umfassende Literatur zum Thema GA, erwähnenswert sind sicherlich die Bücher von Goldberg [40], Michalewicz [56] sowie Bäck [57] und Schwefel [42].

#### 3.3.3.1 Prinzipielle Funktionsweise

Genetische Algorithmen können in der Regel mit dem Flussdiagramm 3.3 beschrieben werden. Sie arbeiten folgendermassen: Die Menge aller Kombinationen der zu





**Abbildung 3.3:** Übersicht eines einfachen genetischen Algorithmus

optimierenden Parameter bildet den Suchraum. Eine Menge von Punkten daraus stellt eine **Population** für den GA dar. Die **Individuen** in der Population sind in Strings kodiert. Für alle Individuen der aktuellen Population wird nun ein Funktionswert – im nachfolgenden häufig auch als **Fitness** bezeichnet – berechnet, der den jeweiligen Parameterwerten entspricht. Bei einer Parametrisierung ist dies typischerweise der SSQ-Wert. Nach gewissen **Selektionskriterien** (z.B. tiefer SSQ-Wert) werden einige Individuen für **genetische Operationen** ausgewählt. Diese entsprechen Vorgängen aus der Genetik, zu erwähnen sind als Beispiele die **Mutation** sowie das **Crossover**. So entstehen neue Individuen aus Individuen der Vorgängergeneration. Diese neue **Generation** wird für die nächste Reproduktion wieder dekodiert und bewertet. Das Kreieren neuer Generationen wiederholt sich, bis ein **Abbruchkriterium** erfüllt ist. Ein solches Abbruchkriterium kann z.B. das Erreichen eines bestimmten Fitnesswertes oder einer bestimmten Generationszahl sein. Im nachfolgenden werden einige Schritte in einem genetischen Algorithmus etwas genauer beschrieben.

### 3.3.3.2 Repräsentation der Parameter

Die meisten Implementationen übersetzen die zu optimierenden Parameter in ihre binäre Repräsentation und schreiben die erhaltene Folge von 0 und 1 in einen Vektor, welcher häufig auch Parent-Vektor genannt wird [35, 58]. Als 'Chromosom' wird nun die Gesamtheit aller zu optimierenden Parameter bezeichnet. Die Teilabschnitte, welche jeweils einen Parameter definieren, werden Gene genannt.

### 3.3.3.3 Initialisierung der Populationen

Ein Charakteristikum von genetischen Algorithmen besteht darin, dass während der Optimierung viele Parametersätze – Individuen – gleichzeitig verwendet werden. Bei der Initialisierung stellt sich somit zunächst die Frage, wie gross die Population gewählt werden soll. Grosse Populationen erlauben eine effektive Selektion, erfordern aber auch immer viel Zeit für die Funktionsevaluierung. Laut Untersuchungen von Syswerda [59] wird die optimale Grösse der Population von der Art der Aufgabenstellung, den verwendeten genetischen Operationen und anderen Faktoren dermassen stark beeinflusst, dass sie am besten problemspezifisch festgelegt wird, weil man keine zuverlässigen allgemeinen Regeln angeben kann.

Zu Beginn besteht die Population bei unseren Anwendungen meist nur aus einem einzigen Individuum, nämlich aus dem Satz der Startparameter. Um mit dem eigentlichen GA beginnen zu können, muss die Population nun mit weiteren Parametersätzen aufgefüllt werden. Dies kann auf vielerlei Arten erfolgen, man könnte z.B. Zufallswerte nehmen. Dies wäre keine gute Wahl, wenn die vorhandenen Startparameter schon eine relativ hohe Fitness aufweisen, weil dann die nach dem Zufallsprinzip erzeugten neue Parametersätze tendenziell weit entfernt von der angestrebten Lösung sein dürften und der GA dann viel Zeit damit verbringen würde, irrelevante Regionen im Parameterraum zu durchsuchen. In einem solchen Fall bietet sich die Alternative, die neuen Parametersätze aus den Startparametern zu erzeugen, indem Zufallsvariationen innerhalb eines sinnvoll gewählten Bereiches um diese Startwerte herum durchgeführt werden. Eine weitere Möglichkeit besteht darin, den Parameterraum systematisch abzudecken; allerdings kann dies zu sehr vielen neuen Individuen führen, so dass man hier sicher die Suchtiefe vernünftig wählen muss.

### 3.3.3.4 Evaluierung der Fitness

Das Berechnen des Funktionswertes (Fitness) aller Individuen mit den jeweiligen Parametern ist im allgemeinen der zeitaufwendigste Schritt einer GA-Optimierung. Dieser Schritt kann beschleunigt werden, indem die Rechnungen für die einzelnen Individuen auf verschiedene Computer verteilt werden. Natürlich brauchen nur diejenigen Individuen neu berechnet werden, die sich aufgrund einer genetischen Operation verändert haben; die unveränderten Individuen haben denselben Funktionswert wie in der vorangehenden Generation. Nach Durchführung aller erforderlichen Rechnungen wird jedem Individuum eine Fitness zugeteilt, wobei die erhaltenen Funkti-

onswerte oft in geeigneter Form skaliert werden, um eine für den GA-Algorithmus günstige Relation zwischen maximaler und mittlerer Fitness zu erreichen [56, p. 65].

### 3.3.3.5 Selektionsschemata

Nachdem den Individuen ihre (skalierten) Fitnesswerte zugewiesen wurden, gilt es, im Sinne der Evolution fortzufahren: die besten Individuen sollen überleben. Hierzu wird ein gewisser Anteil von Individuen mit höchster Fitness in die nächste Generation übernommen, während andere durch genetische Operationen verändert werden, die sowohl einzelne Individuen als auch Paare von Individuen betreffen können (siehe unten). Es gibt viele mögliche Verfahren zur Bildung solcher Paare, am bekanntesten und am häufigsten verwendet wird die Roulette-Wheel-Selection [60].

### 3.3.3.6 Genetische Operatoren

Die genetischen Operatoren erstellen aus den Individuen der aktuellen Population neue Individuen. Man unterscheidet dabei hauptsächlich zwischen Mutation und Crossover. Der Einfachheit halber beschränken sich die folgenden Erläuterungen auf eine binäre Repräsentation der Individuen. Ein **Crossover** ist dadurch gekennzeichnet, dass zwei Individuen komplette Teilabschnitte ihres Genoms untereinander austauschen. Der einfachste Typ, der 'one-point crossover', wird in Figur 3.4 dargestellt. Weitere Crossovertypen können der Literatur entnommen werden [60].

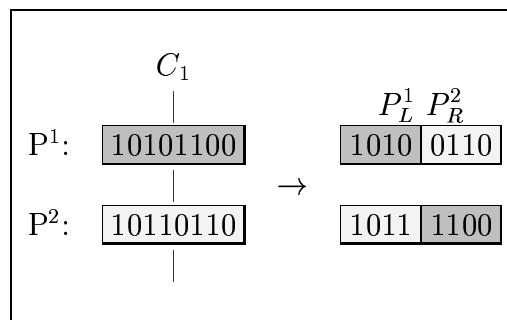
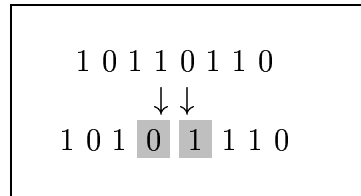


Abbildung 3.4: One-point Crossover

Der zufällige Schnittpunkt  $C_1$  unterteilt die beiden Chromosomen  $P^1$  und  $P^2$  in jeweils zwei Teile:  $P_L^1$  und  $P_R^1$  respektive  $P_L^2$  und  $P_R^2$ , wobei  $P$  für Parent,  $R$  für Rechts und  $L$  für Links stehen. Die Chromosomen bilden nun neue Individuen, indem die neuen Chromosomen sich jeweils aus  $P_L^1$  und  $P_R^2$  respektive  $P_L^2$  und  $P_R^1$  zusammensetzen. Stellt das Chromosom mehrere zu optimierende Parameter dar, so wird pro Gen (d.h. pro Parameter) jeweils ein Schnittpunkt gewählt. Offensichtlich können durch den Crossover-Operator gewisse Sequenzen nie erhalten werden: So ist es unmöglich, das erste Bit beider Chromosomen  $P^1$  und  $P^2$  zu ändern; nie wird es seinen Wert von 1 zu 0 wechseln. Somit gibt es gewisse Suchräume, die durch Crossover nicht

erschlossen werden können, sondern nur durch andere Operatoren wie die Mutation. Eine Mutation kann gemäss Figur 3.5 neue Information kreieren. Als Mutation



**Abbildung 3.5:** Mutation

wird der Vorgang bezeichnet, welcher eine Informationseinheit in eine andere Einheit umwandelt (Bit-Flip in binärer Darstellung). Da die Auswirkung einer solchen Umwandlung je nach der Position des Bits innerhalb des Chromosoms drastisch verschieden sein kann, gibt es nicht-uniforme Mutationen mit entsprechender Gewichtung innerhalb des Chromosoms. Für weitere problemangepasste Möglichkeiten der Mutation sei auf die Literatur verwiesen [56].

Normalerweise verursacht ein Crossover grössere Änderungen als eine Mutation. Bei der Anwendung von GA-Algorithmen auf die Parametrisierung semiempirischer Methoden kann man deshalb davon ausgehen, dass Crossover-Operationen ein weiträumiges Absuchen der Parameter-Hyperfläche bewirken, während Mutationen eher die Optimierung innerhalb der näheren Umgebung vorantreiben.

### 3.3.3.7 Genetics at work

Nachdem jetzt alle wichtigen Komponenten eines genetischen Algorithmus diskutiert sind, wird in Abbildung 3.6 anhand eines einfachen Beispiels gezeigt, wie der GA funktioniert [61]. Nachdem die erste Population initialisiert ist, werden deren zugehörigen Funktionswerte, d.h. die Fitnesswerte  $f(x)$ , berechnet. Dies beansprucht gewöhnlich die meiste Zeit. Danach werden gemäss einem vorgegebenen Kriterium (hier die binäre Tournamenauswahl) die entsprechenden Populationen ausgewählt und zufällig für den Mating-Prozess bereitgestellt. An gewissen Stellen wird ein Crossover sowie eine Mutation durchgeführt, so dass eine neue Population entsteht. Deren Fitnesswerte  $f(x)$  werden neu berechnet, und der Prozess wird wiederholt, bis ein Abbruchkriterium erfüllt ist.

## 3.4 Das Programm gps1

Der Name gps1 ist ein Akronym und steht für **g**lobal **p**arameter **s**earcher, version **1**. Dieses Programm wurde entwickelt, um verschiedene Parametrisierungsstrategien einheitlich zu implementieren und den Parametrisierungsvorgang zu automatisieren. Es besteht im wesentlichen aus zwei grossen Modulen, welche voneinander unabhängig sind:

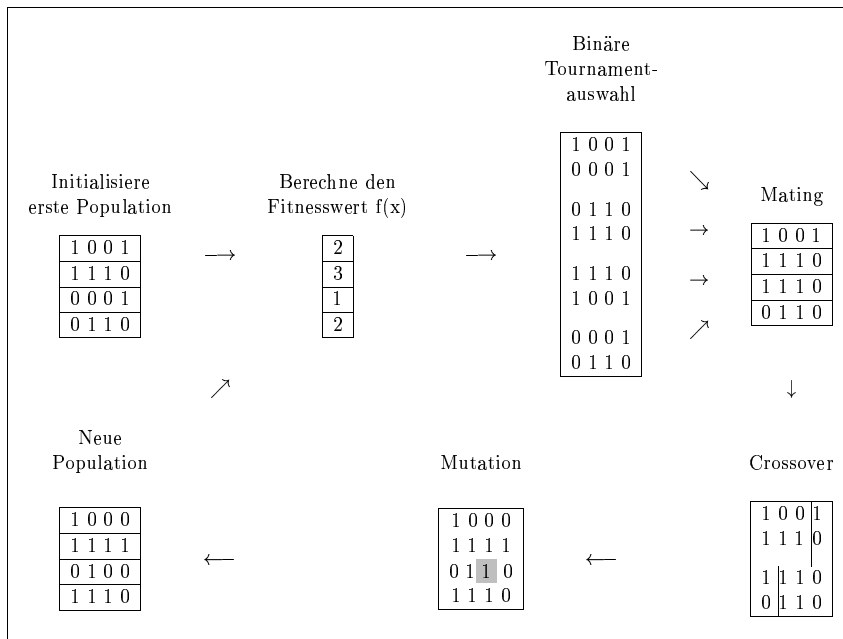


Abbildung 3.6: GA at work

- einer systematischen Suche und Optimierung der Parameter mittels gradientenbasierter Verfahren, im folgenden grid-search genannt;
- einer stochastischen Suche und Optimierung der Parameter mittels eines genetischen Algorithmus, im folgenden ga-search genannt.

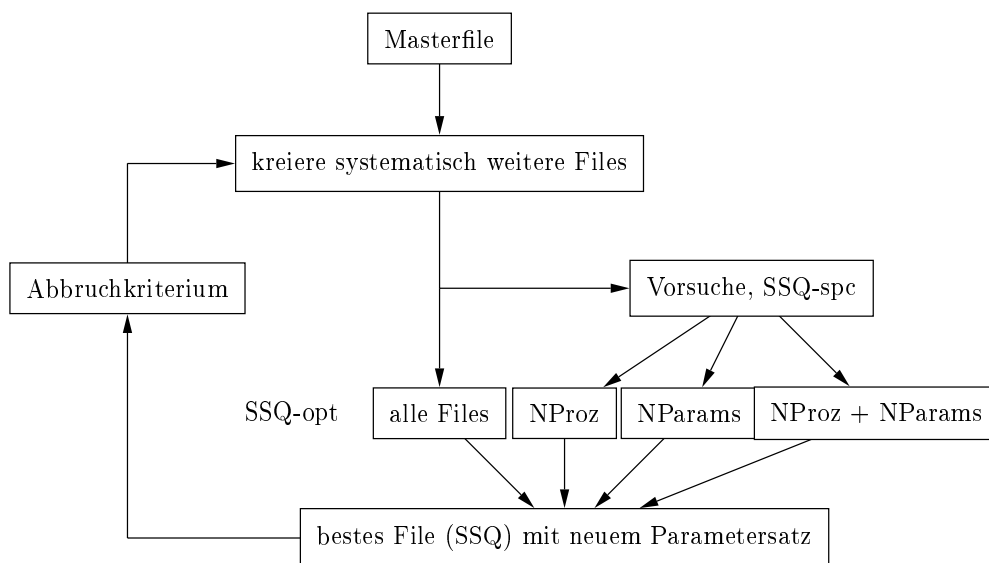
gps1 nutzt das vorhandene Parametrisierungsprogramm (pp99), um für gegebene Parameter den Funktionswert zu berechnen oder um gegebene Startparameter durch nichtlineare Minimierung der Fehlerquadratsumme für eine gewisse Zahl von Iterationen zu optimieren (vgl. Kap. 2.2.3). Diese beiden Aufrufe von pp99 werden im folgenden mit SSQ-spc (SSQ single point calculation) und SSQ-optn (SSQ optimization for  $n$  iterations) abgekürzt. Prinzipiell möglich ist auch eine vollständige Optimierung (SSQ-optfull) mittels pp99, allerdings ist dies im Kontext von gps1 wegen der meist langsamen Konvergenz gegen Ende der Optimierung nicht effizient. Technisch ist gps1 in der Lage, selbständig mehrere verschiedene Inputfiles für pp99 zu generieren, die Berechnungen mittels pp99 lokal oder parallel auf der Batchqueue durchführen zu lassen, die Ergebnisse automatisch auszuwerten, die besten Parametersätze zu identifizieren, diese gemäss geeigneter Kriterien und Algorithmen zu modifizieren, neue Inputfiles für pp99 zu generieren und die SSQ-Berechnung erneut zu starten. Dieser Zyklus kann bis zur Konvergenz durchlaufen werden.

Das Programm ist in Fortran f90 geschrieben und umfasst ca. 36000 Zeilen. Während der Entwicklungsarbeit (ca. 3 Jahre) wurde mittels forcheck [62] gps1 jeweils auf syntaktisch korrekten Code geprüft und gegebenenfalls korrigiert. Programmaufruf,

Programmkonzept, technische Probleme und weitere Details sind genauer im Anhang D, beginnend ab Seite 141, beschrieben. An dieser Stelle werden nur diejenigen Eigenschaften von `gps1` erläutert, die für das Verständnis der nachfolgenden Kapitel wichtig sind.

### 3.4.1 Systematische Suche

Die Optimierung der Parameter auf der Grundlage einer systematischen Suche erfolgt gemäss Figur 3.7. Ausgangspunkt ist ein Masterfile, der den Inputkonventionen



**Abbildung 3.7:** Ablauf einer systematischen Suche, grid-search

von `pp99` entspricht und im wesentlichen die zu optimierenden Parameter sowie die Referenzmoleküle und Referenzdaten enthält. Durch systematische Variation der Parameter kreiert `gps1` weitere Inputfiles. Dabei kann jeder der insgesamt  $NParams$  Parameter innerhalb eines physikalisch sinnvoll gewählten Bereiches  $NValues$  verschiedene Werte annehmen. Zwei Vorgehensweisen sind implementiert:

- eine komplette Variation aller Parameter, die zu insgesamt  $NParams^{NValues}$  Kombinationen von Parametern führt;
- eine eingeschränkte Variation, bei der jeweils ein Parameter den vorgegebenen Wertebereich durchläuft, während alle anderen Parameter ihre Startwerte beibehalten; dies ergibt  $NParams * NValues$  Parametersätze.

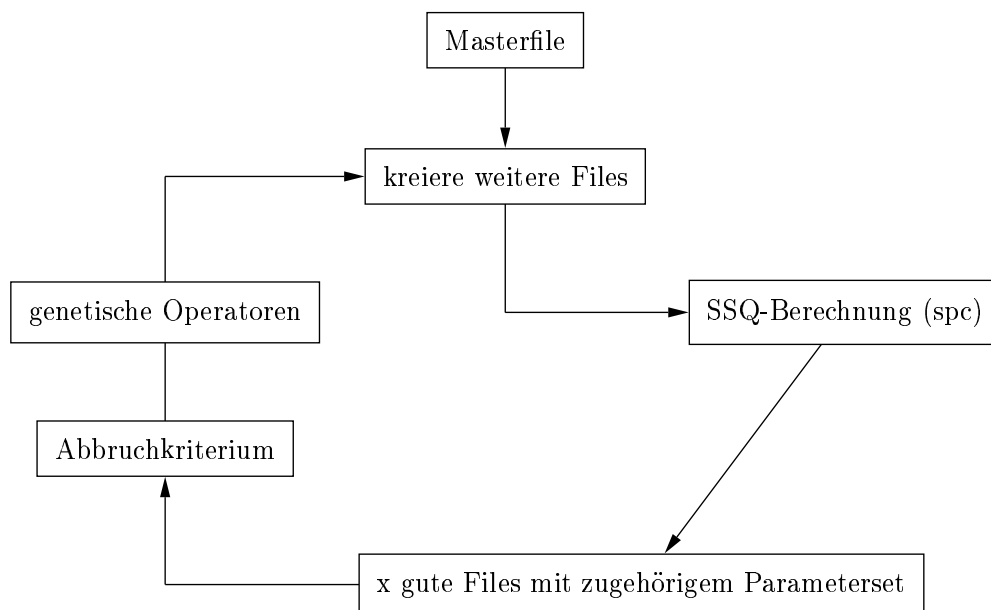
Wünschenswert wäre natürlich die komplette Variation aller Parameter, die jedoch wegen der exponentiell steigenden Zahl an Kombinationen nur bei einer sehr kleinen Zahl von Parametern möglich ist und bei realistischen Anwendungen nicht in Frage kommt.

In der Praxis kann man daher meist nur die eingeschränkte Variation verwenden. Es ist denkbar, dann alle dabei generierten  $NParams * NValues$  Parametersätze zu optimieren (SSQ-optn), allerdings führen solche Optimierungen bei 'schlechten' Startparametern erfahrungsgemäss selten zum Erfolg. Um diese Fälle abzufangen und damit den Rechenaufwand zu senken, ist eine Vorsuche implementiert, bei welcher der SSQ-Wert für jeden der erzeugten  $NParams * NValues$  Parametersätze berechnet wird (SSQ-spc) und anhand dessen entschieden wird, welche Sätze der Optimierung unterworfen werden. Selektiert wird normalerweise ein bestimmter Prozentsatz ( $NProz$ ) aller Parametersätze, wobei niedrige SSQ-Werte als Selektionskriterium dienen (also insgesamt  $NParams * NValues * NProz/100$  Sätze). Alternativ kann auch für jede einzelne Parametervariation der jeweils beste Satz mit dem niedrigsten SSQ-Wert für die weitere Optimierung ausgewählt werden (also insgesamt  $NParams$  Sätze), und es ist zudem möglich, beide Selektionskriterien gemeinsam anzuwenden. Die so selektierten Parametersätze werden dann jeweils optimiert (SSQ-optn), und der beste erhaltene Satz mit dem niedrigsten SSQ-Wert dient wiederum als Ausgangspunkt für die erneute Variation der Parameter (siehe Figur 3.7). Dieser Zyklus wird solange durchlaufen, bis ein vorgegebenes Abbruchkriterium erfüllt ist (z.B. bis der SSQ-Wert sich nicht mehr ändert).

Bei dem geschilderten Vorgehen hat sich die eingeschränkte Variation der Parameter mit Vorsuche in den meisten Anwendungen als effizient erwiesen. Dabei ist  $NParams$  durch die Aufgabenstellung vorgegeben, während  $NValues$  und  $NProz$  problemspezifische Inputgrössen sind.

### 3.4.2 Stochastische Suche mit genetischen Algorithmen

Die Optimierung der Parameter unter Verwendung genetischer Algorithmen ist in Figur 3.8 dargestellt. Ausgangspunkt ist wiederum ein Masterfile gemäss den Inputkonventionen von pp99. Zur Initialisierung wird durch Variation der Startparameter eine Population von Parametersätzen erzeugt (mit den zugehörigen Inputfiles). In Analogie zum grid-search können die neuen Parameter innerhalb eines gewissen Bereiches systematisch um die Startwerte verteilt sein, alternativ ist auch eine Zufallsverteilung innerhalb des zulässigen Wertebereiches implementiert. In der Praxis wird meist der zweite Ansatz bevorzugt, wobei insgesamt  $NPop$  Parametersätze durch simultane Zufallsvariation aller Parameter, jeweils in einem Bereich von  $NVar$  Prozent um den Startwert, generiert werden. Danach wird für jeden dieser Parametersätze der SSQ-Wert berechnet (SSQ-spc). Die  $Ntop$  besten Sätze mit den niedrigsten SSQ-Werten bleiben unverändert in der Population erhalten. Die  $Ngen$  besten Sätze mit den niedrigsten SSQ-Werten ( $Ngen > Ntop$ ) werden genetischen Operationen (z.B. Mutationen und Crossover) unterworfen, um die restlichen  $Npop - Ntop$  Sätze zu erzeugen, mit denen die Population wieder auf ihren Sollwert aufgefüllt wird. Mit dieser Population wird dann der nächste



**Abbildung 3.8:** Ablauf einer genetischen Suche, grid-search

Zyklus begonnen (siehe Figur 3.8). Es wird solange iteriert, bis ein vorgegebenes Abbruchkriterium erfüllt ist.

In der skizzierten Implementierung erfolgt an keiner Stelle eine gradientenbasierte Optimierung, die Parameter werden ausschliesslich über die genetischen Operationen verändert. Die genannten Kontrollvariablen für den genetischen Algorithmus ( $NPop$ ,  $NVar$ ,  $Ntop$ ,  $Ngen$ ) sind ebenso wie andere relevante Optionen (siehe Anhang D und die dort erwähnte Inputbeschreibung) jeweils problemspezifisch zu wählen.

### 3.5 Übersicht zu den durchgeführten Parametrisierungen

Das neu entwickelte Programmpaket ist nach einer umfangreichen Testphase mit einfach konstruierten Beispielen auf zwei grössere Aufgaben aus dem Bereich der semiempirischen Quantenchemie angewendet worden:

- eine Reparametrisierung der Elemente H, C, N und O für MNDO;



- eine Neuparametrisierung des Elementes F für OM2 und OM3.

Im ersten Fall bestand das Ziel nicht darin, die etablierte MNDO Methode durch die Einführung neuer Parameter grundlegend zu verbessern, sondern es sollte untersucht werden, wie stark die mit den neuen Techniken optimierten Parameter von den etablierten MNDO Parametern abweichen und inwieweit sich dadurch die entsprechenden Ergebnisse ändern. Im zweiten Fall sollten optimale Werte für die bisher nicht vorhandenen Fluorparameter in zwei neuartigen semiempirischen Methoden gefunden werden. Die Methoden (OM2 und OM3) verwenden Orthogonalisierungskorrekturen zu den Einelektronen und haben sich für Verbindungen der Elemente H, C, N und O gegenüber den etablierten Verfahren (MNDO, AM1 und PM3) als deutlich überlegen erwiesen, so dass eine Erweiterung der Parametrisierung auf andere Elemente wie Fluor sehr wünschenswert ist.

Bei der Parametrisierung wurden die beiden in `gps1` implementierten Strategien (`grid-search` und `ga-search`) eingesetzt, um deren Leistungsfähigkeit anhand konkreter und realer Beispiele zu überprüfen. Für die Reparametrisierung von MNDO wurde mit beiden Ansätzen eine grössere Zahl von Optionen getestet, von denen dann in der nachfolgenden Parametrisierung bei OM2 und OM3 nur die erfolgreichereren Kombinationen genutzt wurden. Die in diesen beiden Fällen gewonnenen Erfahrungen sollten für künftige Parametrisierungen mit `gps1` hilfreich sein.

Die **statistische Auswertung** der erhaltenen Ergebnisse wurde mittels eines vorhandenen Programmes (`eva99`) durchgeführt. Dieses Programm erlaubt die Berechnung der mittleren absoluten Fehler (mean unsigned error, MUE), der mittleren Fehler (mean signed error, MSE) und der Standardabweichungen (SD) gemäss folgenden Definitionen:

$$MUE = \frac{1}{N} \sum_{i=1}^n (ABS(x_i - y_i)) \quad (3.7a)$$

$$MSE = \frac{1}{N} \sum_{i=1}^n (x_i - y_i) \quad (3.7b)$$

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^n (x_i - y_i)^2} \quad (3.7c)$$

$x_i$  und  $y_i$  entsprechen dabei den berechneten resp. experimentellen Werten. In der Diskussion wird meist vom MUE die Rede sein, bei systematischen Fehlern wird auch der MSE angegeben.

## 3.6 Reparametrisierung von MNDO

Wie bereits erwähnt, zielte die Reparametrisierung von MNDO für C, H, N und O hauptsächlich darauf ab, das neue Programmpaket `gps1` zu testen und herauszufinden, ob und inwieweit eine etablierte Parametrisierung aus dem Jahr 1977 unter Verwendung neuer Optimierungsstrategien, schnellerer Computer und grösserer Referenzdatensätze verfeinert werden kann.

Als Startparameter dienen die bekannten Standardwerte von MNDO [24], die in Tabelle 3.7 angegeben sind. Bei den Resonanzintegralen sind die  $\beta_s$  und  $\beta_p$  Parameter für N und O in MNDO jeweils identisch, diese Einschränkung gilt auch in den Reparametrisierungen.

Die ursprüngliche MNDO Parametrisierung [24] wurde mit 35 Referenzmolekülen durchgeführt, während die jetzigen Reparametrisierungen einen grösseren Satz von 227 Referenzmolekülen benutzen, der im wesentlichen dem ursprünglichen MNDO Validierungssatz [24] entspricht (mit einigen Ergänzungen). Ursprünglich umfassten die Referenzdaten vornehmlich Standardbildungsenthalpien und Molekülgeometrien (mit vergleichbarem Gewicht) und daneben noch Dipolmomente und Ionisierungspotentiale (mit geringerem Gewicht), neu hinzukommen sind jetzt einige andere Eigenschaften wie relative Energien und Schwingungsfrequenzen.

Die Standardparameter von MNDO ergeben mit dem jetzigen Referenzdatensatz einen SSQ-Wert von 41280 (SSQ-spc), der durch Minimierung der Fehlerquadratsumme mit `pp99` leicht auf 40797 absinkt (Konvergenz nach zwei Iterationen). In der Reparametrisierung gilt es, durch Optimierung der 20 justierbaren Parameter diese Werte zu unterschreiten.

### 3.6.1 Systematische Suche

Jeder Parameter wurde innerhalb eines Bereiches von  $\pm 10\%$  um den MNDO-Standardwert variiert. Die Schrittweite war jeweils konstant, die Anzahl der Gridpunkte betrug entweder `NValues=11` (Standard) oder `NValues=101` (feines Gitter). Bei 20 zu optimierenden Parametern (`NParams=20`) wurden demnach 220 resp. 2020 Parametersätze bei der systematischen Suche mit eingeschränkter Variation erzeugt. Nach der Versuche wurden maximal die besten 40% dieser Parametersätze (`NProz=40`) einer Optimierung unterworfen (SSQ-optn mit  $n = 1, 2$  oder 10), daneben wurden auch schärfere Selektionskriterien getestet (`NProz=30,20,10`).

Es ist instruktiv, den Rechenaufwand für solche Parametrisierungen abzuschätzen. Wir gehen dabei von folgenden Optionen aus: `NParams = 20`, `NValues=11`, `NProz = 10`. Die Versuche erfordert `NParams*NValues = Nspc = 220` Funktionsberechnungen. Eine Optimierung erfolgt für `Nspc*NProz/100 = NSet = 22` Parametersätze. In jeder Iteration werden dabei zur SSQ-Minimierung typischerweise 5 Funktionsberechnungen benötigt (minimal 3 und maximal 10 gemäss den Defaultoptionen). In der ersten und zweiten sowie jeder zehnten folgenden Iteration der SSQ-Minimierung wird zusätzlich defaultmässig die Jacobi-Matrix über finite Differenzen berechnet

(insgesamt  $N_{\text{Jacobi}}\text{-mal}$ ), was jeweils  $N_{\text{Params}}*2 = 40$  Funktionsberechnungen erfordert. Für einen vollständigen Zyklus der systematischen Suche braucht man demnach die  $N_{\text{Funct}} = N_{\text{spc}} + N_{\text{Set}}*n*5 + N_{\text{Set}}*N_{\text{Jacobi}}*40$  Funktionsberechnungen (Defaultoptionen). Bei  $n = 1$  ergibt dies eine Zahl von  $220 + 110 + 880 = 1210$ , und bei  $n = 2$  sowie  $n = 10$  entsprechend 2200 bzw. 3080.

Der Rechenaufwand skaliert linear mit der Zahl der Gridpunkte ( $N_{\text{Values}}$ ), da  $N_{\text{spc}}$  und  $N_{\text{Set}}$  proportional zu  $N_{\text{Values}}$  ansteigen. Er skaliert ebenfalls nahezu linear mit dem Anteil der selektierten Parametersätze ( $N_{\text{Proz}}$ ), da  $N_{\text{Set}}$  proportional zu  $N_{\text{Proz}}$  ist und die  $N_{\text{Set}}$ -abhängigen Terme in der Gleichung für  $N_{\text{Funct}}$  dominieren. Schliesslich steigt der Rechenaufwand trivialerweise mit der Gesamtzahl an Zyklen, die in der systematische Suche bis zur Konvergenz benötigt werden.

Bei dem verwendeten Referenzdatensatz mit 227 Molekülen dauert eine einzelne Funktionsberechnung auf der verfügbaren Hardware (Compaq ES40, 667 MHz CPUs) bei erstmaliger Ausführung ca. 20 Sekunden, danach bei Verwendung der vorher konvergierten Dichtematrizen etwa die Hälfte dieser Zeit. Für einen kompletten Zyklus mit den oben diskutierten Optionen ( $N_{\text{Params}}=20$ ,  $N_{\text{Values}}=11$ ,  $N_{\text{Proz}}=10$ ,  $n=1$ ) sind etwa 3 Stunden an Rechenzeit erforderlich. Die einzelnen Funktionsberechnungen werden von `gps1` auf die in der Abteilung vorhandenen CPUs verteilt, wobei im Normalbetrieb auf einen Nutzer durchschnittlich drei CPUs entfallen, so dass ein solcher Zyklus dann in etwa 1 Stunde Realzeit bearbeitet werden kann. Eine komplette derartige Parametrisierung mittels `gps1` kann somit oft innerhalb eines Tages abgeschlossen werden.

Ein typischer Verlauf einer Parametrisierung ist in der Tabelle 3.2 dokumentiert (Ausgangspunkt MNDO-Standardparameter,  $N_{\text{Values}}=11$ ,  $N_{\text{Proz}}=40$ ,  $n=1$ ). In jedem Zyklus sind der beste SSQ-Wert aus der Vorsuche (Spalte SSQ-`spc`) und der beste SSQ-Wert aus der folgenden Optimierung angegeben (Spalte SSQ-`opt1`), welcher seinerseits den Startpunkt für den nächsten Zyklus definiert. Die Konvergenz wird nach 13 Zyklen erreicht. Technische Daten zu Parametrisierungsläufen mit verschiedenen Optionen sind in Tabelle 3.3 zusammengestellt, die zugehörigen optimierten Parameter finden sich in den Tabellen 3.4 und 3.5. Zu Benchmark-Zwecken wurde weiterhin eine Parametrisierung mit einem feinen Gitter durchgeführt (Ausgangspunkt MNDO-Standardparameter,  $N_{\text{Values}}=101$ ,  $N_{\text{Proz}}=40$ ,  $n=1$ ), welche bei einem SSQ-Wert von 30794 konvergierte. Die zugehörigen Parameter werden mit  $\text{MNDO}_{\text{neu}}$  gekennzeichnet und in Tabelle 3.7 gemeinsam mit den ursprünglichen MNDO-Parametern und mit einem weiteren unten diskutierten Satz ( $\text{MNDO}_{\text{arom}}$ ) wiedergegeben.

Für alle neuen Parametersätzen in den Tabellen 3.4-3.7 gilt, dass die erhaltenen Parameterwerte den Originalwerten insgesamt recht ähnlich sind. Oft streuen die neuen um die ursprünglichen Werte und weichen dabei um weniger als 1% ab. Allerdings gibt es auch einige systematische Unterschiede mit Abweichungen von mehreren Prozent. So liegen für Wasserstoff die neuen  $U_{ss}$  Parameter meist zwischen -11.4 und -11.5 eV (statt -11.9 eV), die neuen Orbitalexponenten  $\zeta$  bei

**Tabelle 3.2:** Grid-Optimierung der MNDO-Parameter:  
SSQ-Verlauf bei Parametersatz A

Zyklus	SSQ-spc	SSQ-opt1
1	41270.19	35725.52
2	35278.06	34707.81
3	34449.79	34132.21
4	34181.80	33963.31
5	33931.49	33617.62
6	33558.04	33356.94
7	33368.91	32961.02
8	33071.85	32922.16
9	32459.71	32459.71
10	32459.71	32298.25
11	32255.32	32063.71
12	31834.66	31834.66
13	31834.66	31834.66

MNDO-Referenzwerte:

SSQ-spc 41280.44

SSQ-optfull 40797.18

Optionen NValues=11, NProz=40,  $n=1$ **Tabelle 3.3:** Grid-Optimierung der MNDO-Parameter: Technische Daten

Satz	NValues	NProz	$n^a$	SSQ-Start	SSQ-Ende	Zyklen	Rechnungen <sup>b</sup>	Files <sup>c</sup>
A	11	40	1	41280	31834.66	12	50160	2640/1056
B	11	40	2	41280	29405.79	51	415140	11220/4488
C	11	40	10	41280	30128.08	9	104940	1980/792
D	11	30	1	41280	33814.09	12	38280	2640/792
E	11	30	9999	41280	29512.87	28		6160/1848
F	11	20	1	41280	31698.47	12	26400	2640/528
G	11	20	9999	41280	29530.57	10		2200/440
H	11	10	1	41280	35285.23	16	19360	3520/352
I	11	10	2	41280	29475.59	47	103400	10340/1034
J	11	10	10	41280	30163.35	9	27720	1980/198
K	11	10	20	41280	29579.30	17	86020	3740/374
L	11	10	30	41280	29302.86	12	84480	2640/264
M	11	10	9999	41280	29194.97	14		3080/308

<sup>a</sup>  $n=9999$  fordert eine komplette SSQ-Minimierung bis zur Konvergenz.<sup>b</sup> Zahl der Funktionsberechnungen (SSQ); für komplette SSQ-Minimierungen mit  $n=9999$  nicht erfasst, da unterschiedlich viele SSQ-Berechnungen pro SSQ-Minimierung erfolgen.<sup>c</sup> Totale Anzahl an Jobfiles, SSQ-spc / SSQ-optn.

Tabelle 3.4: Typische optimierte MNDO-Parametersätze: A - F

	Original	Satz A	Satz B	Satz C	Satz D	Satz E	Satz F	
H	$U_{ss}^a$	-11.90627600	-11.43097918	-11.47713339	-11.55756534	-11.37285026	-11.80218133	-11.51125596
	$\zeta^b$	1.33196700	1.24692942	1.22634507	1.23990650	1.27842055	1.22349005	1.25112062
	$\beta_s^c$	-6.98906400	-6.58383005	-6.37975684	-6.49714267	-6.91070768	-6.32620423	-6.70230605
	$\alpha^d$	2.54413410	2.54119997	2.56673696	2.55624529	2.51006852	2.57048947	2.52867457
C	$U_{ss}$	-52.27974500	-52.40345541	-53.06025024	-53.08471262	-52.34206627	-53.13695464	-52.35277429
	$U_{pp}^a$	-39.20555800	-39.13245811	-39.26457807	-39.28938587	-39.19654251	-39.29555595	-39.22816911
	$\zeta$	1.78753700	1.80091204	1.80808169	1.81399632	1.78451714	1.80402573	1.78675309
	$\beta_s$	-18.98504400	-18.89172327	-18.83033785	-18.78179001	-19.00568901	-18.82160452	-19.00995228
	$\beta_p^c$	-7.93412200	-8.49219932	-8.91083884	-9.00000000	-8.09447029	-9.00000000	-8.08635908
	$\alpha$	2.54638000	2.52512871	2.52845820	2.53175837	2.52303698	2.51878687	2.52677353
N	$U_{ss}$	-71.93212200	-70.34013358	-70.15715793	-70.89172866	-71.71439266	-70.61540082	-70.44259115
	$U_{pp}$	-57.17231900	-57.23266848	-57.25133355	-57.46115024	-57.46549688	-57.44149072	-57.32607518
	$\zeta$	2.25561400	2.25568246	2.24157987	2.22643108	2.24592038	2.22480509	2.25677928
	$\beta_s$	-20.49575800	-20.44091110	-20.27669624	-20.10449909	-20.45992462	-20.17138939	-20.40250763
	$\alpha$	2.86134200	2.82444724	2.81289862	2.81709864	2.83637668	2.80589279	2.83208118
O	$U_{ss}$	-99.64430900	-107.65089236	-107.81462411	-107.83505509	-107.80000000	-107.80000000	-107.79914493
	$U_{pp}$	-77.79747200	-77.89606365	-77.99949635	-78.09690748	-77.93881429	-78.12054331	-77.99535940
	$\zeta$	2.69990500	2.61468827	2.59304500	2.61023178	2.59919315	2.58624849	2.55919664
	$\beta_s$	-32.68808200	-32.92039795	-32.62375624	-32.77373324	-32.65974780	-32.63904646	-31.91284626
	$\alpha$	3.16060400	3.09573553	3.07810213	3.09522721	3.08720397	3.06814152	3.06564321
SSQ	41280	31835	29406	30128	33814	29513	31698	

<sup>a</sup> Einzentrenterme [eV]<sup>b</sup> Orbitalexponent [a.u.]<sup>c</sup> Resonanzintegrale [eV]<sup>d</sup> Rumpf-Rumpf-Abstossungen [ $\text{\AA}^{-1}$ ]

Tabelle 3.5: Typische optimierte MNDO-Parametersätze: G - M

		Satz G	Satz H	Satz I	Satz J	Satz K	Satz L	Satz M
H	$U_{ss}^a$	-11.29950714	-11.39792175	-11.52970194	-11.48081363	-11.51587089	-11.41509036	-11.50206030
	$\zeta^b$	1.24665771	1.25232723	1.24206394	1.23521836	1.23597524	1.24310208	1.26191635
	$\beta_s^c$	-6.54243597	-6.83062068	-6.52875788	-6.55699125	-6.52343275	-6.55665775	-6.61778505
	$\alpha^d$	2.54629459	2.50275961	2.54686846	2.53394344	2.54942408	2.53921698	2.54303943
C	$U_{ss}$	-52.97887688	-52.37675929	-52.89483880	-52.95437776	-52.93594314	-52.94965956	-52.90780980
	$U_{pp}^a$	-39.20759506	-39.23535222	-39.20639028	-39.18602757	-39.27150264	-39.21739737	-39.25376671
	$\zeta$	1.81050165	1.78125268	1.80909983	1.80462401	1.81197529	1.80595051	1.80812738
	$\beta_s$	-18.74956351	-18.97151025	-18.75880162	-18.71902243	-18.60507979	-18.71882290	-18.64486096
	$\beta_p^c$	-8.99690624	-8.09740461	-8.99188272	-8.99943018	-8.99950703	-9.00000000	-8.98454662
	$\alpha$	2.52720402	2.51878772	2.52177932	2.52024899	2.53070440	2.51998104	2.52501740
N	$U_{ss}$	-70.05006660	-71.58107678	-70.12353395	-70.07732274	-70.07199278	-70.09168708	-70.02679893
	$U_{pp}$	-57.20473512	-57.47082117	-57.25759596	-57.21774028	-57.31531799	-57.26446703	-57.25388090
	$\zeta$	2.22905926	2.24415984	2.23077344	2.22779559	2.23441389	2.23192939	2.22146186
	$\beta_s$	-20.06707031	-20.43322910	-20.17062077	-20.13031818	-20.08114615	-20.19078396	-19.99263243
	$\alpha$	2.81126444	2.83785061	2.80537334	2.80812917	2.81571690	2.80479592	2.80380248
O	$U_{ss}$	-105.69000000	-99.72088089	-107.80000000	-101.47400142	-105.69000000	-107.80000000	-107.80172901
	$U_{pp}$	-77.79583561	-77.78439721	-77.94678028	-77.77008733	-77.96404177	-77.95190796	-77.96611606
	$\zeta$	2.60375082	2.65794207	2.58743293	2.62127728	2.59939298	2.59116897	2.59199299
	$\beta_s$	-32.62705114	-32.59502889	-32.62844106	-32.42490777	-32.40929915	-32.76714850	-32.62766081
	$\alpha$	3.08356697	3.12004206	3.07337097	3.09057202	3.08596000	3.06742825	3.07836193
SSQ	29531	35267	29476	30163	29579	29302	29195	

<sup>a</sup> Einzentrenterme [eV]<sup>b</sup> Orbitalexponent [a.u.]<sup>c</sup> Resonanzintegrale [eV]<sup>d</sup> Rumpf-Rumpf-Abstossungen [ $\text{\AA}^{-1}$ ]

etwa 1.25 a.u. (statt 1.33 a.u.) und die neuen Resonanzparameter  $\beta_s$  meist zwischen -6.4 und -6.6 eV (statt -7.0 eV), während für Sauerstoff die neuen  $U_{ss}$  Parameter oft Werte von etwa -108 eV annehmen (statt -99.6 eV). Wegen des Auftretens solcher systematischer Unterschiede weisen die neuen Parametersätze untereinander eine grössere Ähnlichkeit auf als zu den MNDO-Originalparametern. Trotz dieser ausgeprägten Ähnlichkeit bleibt festzuhalten, dass sich die neu optimierten Parametersätze durchaus unterscheiden und dass verschiedene systematische Suchen mittels `gps1` demnach nicht zu einem identischen globalen Minimum konvergieren.

Die Fehlerquadratsummen für die neuen Parametersätze schwanken zwischen 29195 und 35267, verglichen mit einem SSQ-Wert von 41280 für die ursprünglichen MNDO-Parameter. In 11 von 13 dokumentierten Fällen (Tabelle 3.3) werden SSQ-Werte unter 32000 erreicht. Vergleicht man die Optimierungen H-M (Tabelle 3.3), so stellt man fest, dass für ansonsten vorgegebene Optionen bei der systematischen Suche der SSQ-Wert tendenziell mit steigender Anzahl  $n$  der erlaubten Iterationen in den zwischengeschalteten SSQ-Minimierungen absinkt, wie dies auch zu erwarten ist. Der niedrigste SSQ-Wert wird für jeweils komplette SSQ-Minimierungen ( $n=9999$ ) erzielt, allerdings ist dies wegen der ungleichmässigen und oft langsamen Konvergenz der Einzelminimierungen mit hohem Rechenaufwand verbunden. Blickt man bei den Sätzen H-M auf die Zahl der nötigen SSQ-Rechnungen und der erzeugten Files (Tabelle 3.3), so scheint die Wahl  $n=10$  ein guter Kompromiss zwischen Rechenaufwand und Optimierungsergebnis zu sein. Eine umfangreichere Selektion nach der Versuche (NProz=40,30,20 statt NProz=10) führt bei ansonsten identischen Optionen meist ebenfalls zu niedrigeren SSQ-Werten (vgl. die Optimierungen A, D, F und H mit  $n=1$ ), allerdings um den Preis eines stark erhöhten Rechenaufwandes; auch hier findet man verbesserte SSQ-Werte für  $n > 1$ . Weiterhin fällt auf (Tabelle 3.3), dass die systematische Suche bei A-M meist in 9-17 äusseren Zyklen konvergiert (in 10 von 13 Fällen) und dass die langsamste Konvergenz für  $n=2$  beobachtet wird (51 resp. 47 Zyklen bei B und I). Die vorgestellten Beispiele erlauben zwar noch keine allgemeingültige Empfehlung für die besten Optionen bei `gps1`-Läufen, aber sie deuten doch darauf hin, dass gewisse Kombinationen von Optionen vorteilhaft sind (z.B.  $n=10$ , NProz=10 oder  $n=1$ , NProz=40).

Zur Dokumentation sind in der Tabelle 3.6 einige statistische Kenndaten für den Referenzdatensatz zusammengestellt, die aus Rechnungen mit den Parametersätzen A-M resultieren. Man erkennt, dass die neuen Parameter für die betrachteten Eigenschaften jeweils zu sehr ähnlichen mittleren absoluten Fehlern führen und dass sie aus statistischer Sicht den MNDO-Originalparametern überlegen sind.

**Tabelle 3.6:** Mittlere absolute Fehler von Grundzustandseigenschaften: Parametersätze A-M

	Anzahl	Original	A	B	C	D	E	F
Enthalpie [kcal/mol] <sup>a</sup>	140	6.35	6.30	5.74	5.87	5.99	5.73	5.82
Bindungslänge [Å]	242	0.014	0.020	0.020	0.019	0.022	0.021	0.023
Bindungswinkel [°]	101	2.56	2.37	2.32	2.17	2.12	2.25	2.37
IP [eV] <sup>b</sup>	52	0.46	0.31	0.28	0.30	0.31	0.34	0.32
Dipolmoment [D]	53	0.35	0.22	0.21	0.25	0.26	0.24	0.22
Rel. Energie [kcal/mol] <sup>c</sup>	16	9.37	8.01	7.52	8.04	8.83	7.80	7.93
Barrieren [kcal/mol]	60	3.44	3.17	3.09	3.02	3.16	3.01	3.24
Wellenzahlen [cm <sup>-1</sup> ]	112	246.0	206.0	201.6	205.3	210.1	198.7	199.3
SSQ		41280	31835	29406	30128	33814	29513	31698
	Anzahl	G	H	I	J	K	L	M
Enthalpie [kcal/mol] <sup>a</sup>	140	5.85	6.11	5.72	5.73	5.76	5.78	5.79
Bindungslänge [Å]	242	0.019	0.023	0.020	0.020	0.020	0.021	0.019
Bindungswinkel [°]	101	2.33	2.44	2.34	2.50	2.39	2.35	2.32
IP [eV] <sup>b</sup>	52	0.29	0.30	0.30	0.29	0.29	0.28	0.30
Dipolmoment [D]	53	0.21	0.26	0.22	0.22	0.22	0.21	0.22
Rel. Energie [kcal/mol] <sup>c</sup>	16	7.69	8.60	7.81	7.75	7.62	7.74	7.80
Barrieren [kcal/mol]	60	3.11	3.16	3.12	3.11	3.12	3.12	3.11
Wellenzahlen [cm <sup>-1</sup> ]	112	207.9	206.2	205.1	205.2	204.8	203.9	209.4
SSQ		29531	35267	29476	30163	29579	29302	29195

<sup>a</sup> Standardbildungsenthalpie bei 298 K

<sup>b</sup> Ionisierungspotential (Koopmans' Näherung)

<sup>c</sup> Relative Energie



### 3.6.2 Suche mit genetischen Algorithmen

Die implementierten genetischen Algorithmen wurden während der Testphase unter anderem auch auf 'kleine' Parametrisierungsprobleme angewendet, so z.B. auf eine Reparametrisierung von MNDO nur für Kohlenwasserstoffe mit einer geringen Zahl an justierbaren Parametern. Hierbei bewährten sich diese Algorithmen in dem Sinne, dass die Fehlerquadratsumme nahezu genauso weit abgesenkt werden konnte wie mit einer entsprechenden systematischen Suche, wobei die erhaltenen Parameter jeweils zu ähnlichen Werten konvergierten.

Bei der in diesem Kapitel diskutierten Reparametrisierung von MNDO für die Elemente H, C, N und O mit insgesamt 20 freien Parametern ergab sich bei der stochastischen Suche mit genetischen Algorithmen ein weniger günstiges Bild. Die Startpopulationen wurden hierbei jeweils durch simultane Zufallsvariation aller Parameter um die MNDO-Originalwerte erzeugt (mit verschiedenen grossen Variationsbreiten), wobei der MNDO-Parametersatz selbst nicht in der Startpopulation beibehalten wurde. In den anschliessenden Optimierungen wurden viele verschiedene Kombinationen von Optionen für den genetischen Algorithmus getestet (siehe Kapitel 3.3.3), ohne dass ein durchschlagender Erfolg zu verzeichnen war. Es gelang zwar, ausgehend von hohen Startwerten mit der Fehlerquadratsumme den MNDO-Standardwert von 41280 zu unterschreiten, aber in keinem Fall wurden SSQ-Werte von weniger als 32000 erzielt, die bei den systematischen Suchen regelmässig anfallen (siehe Kapitel 3.6.1). Gemessen an den Rechenzeiten für typische systematische Suchen wurde den genetischen Algorithmen bei diesen Optimierungen deutlich mehr Rechenzeit eingeräumt (typischerweise mehr als das Dreifache), bevor die entsprechenden Jobs abgebrochen wurden.

Aufgrund dieser Erfahrungen muss man schliessen, dass genetische Algorithmen für derartige semiempirische Parametrisierungen weniger gut geeignet sind als systematische Suchen in Kombination mit lokalen gradientenbasierten Minimierungen. Dies ist rückblickend nicht allzu überraschend, weil man bei der Reparametrisierung von MNDO bereits von einem sehr guten Parametersatz ausgehen kann, der lediglich etwas nachjustiert wird. Daher ist in diesem Fall eher eine lokale als eine globale Optimierung gefragt, so dass die Überlegenheit eines systematischen Vorgehens unter Verwendung der relevanten Ableitungen plausibel ist.

Genetische Algorithmen haben den Vorteil, dass sie weitgehend automatisch arbeiten und dabei grosse Bereiche der Parameterfläche abtasten können. Die rein lokalen Minimierungen der Fehlerquadratsumme (pp99) haben den Nachteil, dass sie nur zum nächsten lokalen Minimum führen können. Dieser Nachteil wird in der jetzt implementierten systematischen Suche weitgehend überwunden und kann daher nicht mehr als Argument zugunsten der genetischen Algorithmen dienen. Für semiempirische Parametrisierungen mit guten Startparametern deutet demnach

alles darauf hin, dass eine systematische Suche in Verbindung mit partiellen gradientenbasierten SSQ-Minimierungen den effizientesten Weg der Optimierung darstellt.

### 3.6.3 Auswahl der Parameter

Für einen detaillierten Vergleich mit den Ergebnissen aus MNDO-Standardrechnungen erscheint es empfehlenswert, sich auf einen repräsentativen neuen Parametersatz zu konzentrieren. Die Wahl fiel auf einen Satz, der aus einer systematischen Suche (siehe Kapitel 3.6.1) mit einem feinen Gitter stammt (Ausgangspunkt MNDO-Standardparameter, NValues=101, NProz=40,  $n=1$ ). Diese Optimierung führte in der ersten Phase unserer Tests mit  $n=1$  zu dem niedrigsten SSQ-Wert, und der zugehörige Satz MNDO<sub>neu</sub> wurde daher ausgedehnten Validierungen unterworfen. In der zweiten Phase (Tests mit  $n > 1$ ) wurden andere Parametersätze mit geringfügig niedrigeren SSQ-Werten gefunden (siehe Tabelle 3.6), die sich jedoch in ihrem qualitativen Verhalten und in ihren statistischen Kenndaten nur sehr wenig von dem Satz MNDO<sub>neu</sub> unterscheiden. Daher sollten die vorhandenen Validierungsrechnungen für MNDO<sub>neu</sub> genügen, um die möglichen Vorteile und Grenzen einer Reparametrisierung von MNDO auszuloten. Ein Vergleich unter Verwendung eines analogen anderen Satzes würde zu ähnlichen Schlussfolgerungen führen.

Tabelle 3.7 enthält die Originalwerte der MNDO-Parameter [24], die ausgewählten neuen Parameterwerte MNDO<sub>neu</sub> sowie einen weiteren Satz MNDO<sub>arom</sub>, der in einer speziellen Parametrisierung mit Ausrichtung auf Aromaten erhalten wurde (siehe unten in Kapitel 3.6.4.1). Offensichtlich sind die beiden neuen Sätze einander recht ähnlich. Die Unterschiede zu den Originalwerten sind ebenfalls klein und liegen typischerweise im Bereich von 1% (vgl. auch die Diskussion in Kapitel 3.6.1).

### 3.6.4 Resultate und Diskussion

Im folgenden werden die Ergebnisse für verschiedene Eigenschaften und Molekülklassen miteinander verglichen, welche mit den MNDO-Standardparametern und mit dem ausgewählten neuen Parametersatz (MNDO<sub>neu</sub>, siehe Tabelle 3.7) erhalten worden sind. Für die Moleküle in dem Referenzdatensatz gibt die Tabelle 3.8 eine Übersicht über die mittleren absoluten Fehler der Grundzustandseigenschaften. Die Spalte MNDO<sub>neu</sub> enthält die Ergebnisse mit dem neuen Parametersatz, die Spalte MNDO analog jene mit den Standardparametern. Zum Vergleich sind die Daten für die etablierten Verfahren AM1 und PM3 sowie für die neuere OM2 Methode [63] hinzugefügt. Die einzelnen Referenzmoleküle sind in der Tabelle 3.12 für die Standardbildungsenthalpien aufgelistet (mit Molekülnamen in Englisch, automatisch aus dem Inputfile übernommen).

Man sieht auf den ersten Blick, dass die meisten Moleküleigenschaften mit den neuen Parametern besser berechnet werden als mit den MNDO-Standardparametern. Die mittleren absoluten Fehler konnten dabei um den in Klammern angegebenen Fak-

**Tabelle 3.7:** MNDO-Parameter: Vergleich der Originalwerte mit zwei ausgewählten neuen Parametersätzen (in Klammern das Verhältnis zwischen neuen und ursprünglichen Werten).

	Originalwert	MNDO <sub>neu</sub>	MNDO <sub>arom</sub>	
H	$U_{ss}$ <sup>a</sup>	-11.90627600	-11.43810895 (0.960)	-11.59020513 (0.973)
	$\zeta$ <sup>b</sup>	1.33196700	1.24572657 (0.935)	1.28809218 (0.967)
	$\beta_s$ <sup>c</sup>	-6.98906400	-6.61898133 (0.947)	-6.72747278 (0.962)
	$\alpha$ <sup>d</sup>	2.54413410	2.54048522 (0.998)	2.54833640 (1.001)
C	$U_{ss}$	-52.27974500	-52.37911057 (1.001)	-52.38538773 (1.002)
	$U_{pp}$ <sup>a</sup>	-39.20555800	-39.13220599 (0.998)	-39.09037133 (0.997)
	$\zeta$	1.78753700	1.80175311 (1.007)	1.77493726 (0.992)
	$\beta_s$	-18.98504400	-18.91708705 (0.996)	-18.91323393 (0.996)
	$\beta_p$	-7.93412200	-8.49897545 (1.071)	-7.98723678 (1.006)
	$\alpha$	2.54638000	2.52287253 (0.990)	2.53708388 (0.996)
N	$U_{ss}$	-71.93212200	-70.33844354 (0.977)	-70.42730053 (0.979)
	$U_{pp}$	-57.17231900	-57.23629966 (1.001)	-57.28629840 (1.001)
	$\zeta$	2.25561400	2.25672404 (1.000)	2.20638336 (0.978)
	$\beta_s$	-20.49575800	-20.45125596 (0.997)	-19.62492660 (0.957)
	$\alpha$	2.86134200	2.82558464 (0.987)	2.81755611 (0.984)
O	$U_{ss}$	-99.64430900	-107.27336636 (1.076)	-109.60741528 (1.099)
	$U_{pp}$	-77.79747200	-77.88398569 (1.001)	-77.97951148 (1.002)
	$\zeta$	2.69990500	2.61349766 (0.967)	2.61700833 (0.969)
	$\beta_s$	-32.68808200	-32.92337253 (1.007)	-32.68440278 (0.999)
	$\alpha$	3.16060400	3.08925868 (0.977)	3.11446149 (0.985)
SSQ	41280.44	30793.75	32920.29	

<sup>a</sup> Einzentrenterme [eV]

<sup>b</sup> Orbitalexponent [a.u.]

<sup>c</sup> Resonanzintegrale [eV]

<sup>d</sup> Rumpf-Rumpf-Abstossungen [ $\text{\AA}^{-1}$ ]

**Tabelle 3.8:** Mittlere absolute Fehler von Grundzustandseigenschaften bei neutralen geschlossenschaligen Molekülen

	Anzahl	MNDO <sub>neu</sub>	MNDO	AM1	PM3	OM2
Enthalpie [kcal/mol] <sup>a</sup>	140	5.90	6.34	5.51	4.19	3.10
Bindungslänge [ $\text{\AA}$ ]	242	0.020	0.014	0.017	0.011	0.016
Bindungswinkel [ $^\circ$ ]	101	2.38	2.56	1.89	2.08	2.24
IP [eV] <sup>b</sup>	52	0.31	0.46	0.36	0.42	0.26
Dipolmoment [D]	53	0.22	0.35	0.26	0.27	0.28
Rel. Energie [kcal/mol] <sup>c</sup>	15	7.62	8.09	5.20	3.36	1.50
Barrieren [kcal/mol]	61	3.14	3.44	2.02	2.95	1.62
Wellenzahlen [ $\text{cm}^{-1}$ ]	112	206.8	246.0	172.1	150.6	155.2

<sup>a</sup> Standardbildungsenthalpie bei 298 K

<sup>b</sup> Ionisierungspotential (Koopmans' Näherung)

<sup>c</sup> Relative Energie

tor verbessert werden: Standardbildungsenthalpien (1.08), relative Energien (1.06), Barrieren (1.09), Bindungswinkel (1.07), Ionisierungspotentiale (1.49), Wellenzahlen (1.19) und Dipolmomente (1.56). Einzig der Fehler für die Bindungslängen ist angestiegen (Faktor 1.37).

Trotz dieser Fortschritte bleibt festzuhalten, dass die Fehler mit den neuen MNDO Parametern bei den meisten Eigenschaften grösser bleiben als jene für die anderen Verfahren, welche entweder aufgrund einer höheren Zahl von justierbaren Parametern flexibler sind (AM1, PM3) oder ein verbessertes theoretisches Modell benutzen (OM2). Die Reparametrisierung kann somit die statistischen Fehler generell leicht verringern, aber die modellbedingten Grenzen des MNDO Ansatzes nicht umgehen. In den folgenden Abschnitten werden einige Ergebnisse für die Moleküle im Referenzdatensatz näher besprochen.

#### 3.6.4.1 Standardbildungsenthalpien

Detailliertere statische Auswertungen für die Standardbildungsenthalpien sind in Tabelle 3.10 zusammengestellt, die zugehörigen Daten für die einzelnen Moleküle finden sich in Tabelle 3.12. Der mittlere absolute Fehler ist von 6.34 kcal/mol (MNDO) auf 5.89 kcal/mol (MNDO<sub>neu</sub>) gesunken, und es sind merkliche Verbesserungen vor allem bei den stickstoffhaltigen Verbindungen zu beobachten (siehe CHN, CHNO und HNO Verbindungen), aber auch bei anderen Molekülklassen wie bei den Dreiringen und Vierringen. Für manche Verbindungsklassen findet man aber auch erhöhte Fehler (z.B. für Aromaten und bicyclische Systeme, siehe unten). Die Zahl der Moleküle mit einem Fehler von mehr als 10 kcal/mol ist von 28 auf 24 gesunken, und der maximale Fehler beträgt jetzt -35.7 statt -49.6 kcal/mol (Cuban). Diese Verbesserungen gegenüber dem MNDO-Standard dürften im wesentlichen auf die verbesserte Parametrisierungstechnik zurückzuführen sein, welche unter anderem dafür sorgt, dass der mittlere Fehler (MSE) jetzt näher an Null liegt (-0.41 statt -1.22 kcal/mol). Wenn man die Fehler im einzelnen betrachtet (Tabelle 3.12), so stellt man meistens analoge Muster bei den neuen MNDO-Parametern und den Standardparametern fest. Dies betrifft sowohl die Vorzeichen als auch die Grösse der Fehler. So zeigen beispielsweise von den 13 Molekülen, die in MNDO einen Fehler von mehr als 15 kcal/mol aufweisen, neun auch mit den neuen Parametern einen entsprechenden Fehler von mehr als 15 kcal/mol (die restlichen vier zwischen 9 und 15 kcal/mol). In beiden Fällen findet man oft die gleichen qualitativen Trends, so sind z.B. sterisch gehinderte Moleküle zu instabil (vgl. Neopentan) und Vierringe zu stabil (vgl. Cyclobutan).

Deutliche Verschlechterungen in den berechneten Standardbildungsenthalpien erhält man mit den neuen MNDO-Parametern bei Aromaten (z.B. Naphthalin, Fehler von 12.7 statt 2.4 kcal/mol) und auch bei bi- und polycyclischen Systemen (z.B. Adamantan, Fehler von 20.6 statt 5.8 kcal/mol). Diese Molekülklassen waren in dem ursprünglichen MNDO-Referenzdatensatz [24] stärker repräsentiert als in dem jetzigen, und es liegt daher nahe, hierin die Ursache für die Verschlechterungen zu

sehen.

Um dies zu prüfen, wurde eine weitere Reparametrisierung durchgeführt, bei der die Standardbildungsenthalpien von Benzol, Naphthalin und Adamantan in den Referenzdaten höher gewichtet wurden (Faktor 5.0 statt 1.0), während ansonsten Standardoptionen verwendet wurden (Ausgangspunkt MNDO-Parameter, Variation  $\pm 10\%$ ,  $NValues = 11$ ,  $NProz = 40$ ,  $n = 1$ ). Der resultierende optimierte Parametersatz (MNDO<sub>arom</sub>, siehe Tabelle 3.7) landet mit den modifizierten Wichtungsfaktoren bei einem SSQ-Wert von 33855, dies entspricht bei Verwendung der Standardfaktoren einem SSQ-Wert von 32920 und liegt damit nur leicht oberhalb des Vergleichswerts von 31835 für den analog mit Standardfaktoren optimierten Satz A (Tabelle 3.4). Eine statische Auswertung (Tabelle 3.11) zeigt, dass dieser neue Satz MNDO<sub>arom</sub> etwa von gleicher Qualität ist wie MNDO<sub>neu</sub>, wobei die Fehler allerdings anders verteilt sind: man findet kleinere Fehler bei Aromaten (3.1 statt 7.5 kcal/mol), Alkinen (3.8 statt 6.9 kcal/mol) und bei bi- und polycyclischen Molekülen (5.8 statt 9.3 kcal/mol), hingegen grössere Fehler bei CHN Verbindungen (6.6 statt 5.2 kcal/mol) und bei Vierringen (21.7 statt 16.6 kcal/mol), wobei der statistische Fehler insgesamt etwas ansteigt (6.0 statt 5.9 kcal/mol). Diese Daten belegen, dass durch die vorgenommene Gewichtung die Ergebnisse für Aromaten sowie bi- und polycyclische Systeme in der Tat spürbar verbessert werden können. Dies wird untermauert durch Rechnungen an einem Validierungssatz [64], der aus 20 grösseren (meist aromatischen) Molekülen besteht (Tabelle 3.13). Hier sinkt der mittlere absolute Fehler der Standardbildungsenthalpien von 17.8 kcal/mol (MNDO<sub>neu</sub>) auf 9.4 kcal/mol (MNDO<sub>arom</sub>), wenn man zu der auf Aromaten ausgerichteten Reparametrisierung übergeht, und man erreicht damit einen ähnlichen Wert wie mit den MNDO-Standardparametern (9.2 kcal/mol).

Dieses Beispiel ist relativ ausführlich besprochen worden um hervorzuheben, wie die Fehler durch eine leichte Veränderung der Referenzdaten (z.B. bei der Gewichtung) umverteilt werden können und wie man auf diese Weise die Ergebnisse qualitativ beeinflussen kann. Bei der Parametrisierung einer neuen semiempirischen Methode muss man oft entscheiden, wo modellbedingte Fehler am ehesten zu tolerieren sind und ob man Verbesserungen in bestimmten Bereichen anstreben soll. Für solche Zwecke sind effiziente Werkzeuge sehr wichtig, um die entsprechenden Nachjustierungen der Parameter schnell zu bewerkstelligen. Im vorliegenden Fall benötigte die Optimierung des modifizierten Parametersatzes (MNDO<sub>arom</sub>) mittels `gps1` etwa zwei Tage an Rechenzeit und verlief praktisch automatisch. Dies zeigt, dass `gps1` den obigen Anforderungen genügt.

#### 3.6.4.2 Geometrien

Die räumliche Struktur eines Moleküls wird üblicherweise durch interne Koordinaten beschrieben, z.B. Bindungslängen, Bindungswinkel und Diederwinkel. Letztere bestimmen im wesentlichen die Konformation und damit das dreidimensionale Aussehen des Moleküls. Sowohl die MNDO-Standardparameter als auch die neu-

en MNDO-Parameter geben in der Regel die relevanten Konformationen qualitativ korrekt wieder, und sie versagen beide in den gleichen Problemfällen (z.B. planare statt abgewinkelte Vierringe und Fünfringe; orthogonale statt koplanare Anordnung gewisser Substituenten wie -COOH oder -CHO am Benzolring). Bei den Diederwinkeln findet man daher jeweils eine grosse Zahl von Ergebnissen mit kleinen Fehlern und einige wenige qualitative Ausreisser mit grossen Fehlern, so dass eine statistische Auswertung daher wenig Sinn macht. Im folgenden werden daher nur Bindungslängen und Bindungswinkel betrachtet.

Bei den **Bindungslängen** ergeben die neuen MNDO-Parameter einen mittleren absoluten Fehler von 0.0195 Å für den Referenzdatensatz (Tabelle 3.14). Dieser Wert ist akzeptabel, auch wenn er höher liegt als bei den MNDO-Standardparametern (0.0142 Å), welche auch bei den meisten individuellen Bindungslängen genauer sind. Dies gilt insbesondere für einige Bindungen mit Heteroatomen (X=N,O), so überschätzen die neuen MNDO-Parameter beispielsweise die O-H und C=O Bindungsabstände systematisch (Tabelle 3.14). Andererseits werden viele Trends in den Bindungslängen von beiden Parametersätzen richtig wiedergegeben, z.B. für CH und CC Abstände in Abhängigkeit von der Hybridisierung des C Atoms. Dies ist in Tabelle 3.9 für durchschnittliche CC Bindungslängen illustriert, die in allen Fällen mit guter Genauigkeit berechnet werden.

**Tabelle 3.9:** Durchschnittliche CC-Bindungslängen [Å] im Referenzdatensatz. Die Werte in Klammern geben die maximale absolute Abweichung vom Experiment an.

	Anzahl	exp <sup>a</sup>	MNDO	MNDO <sub>neu</sub>
C-C	69	1.487	1.493 (0.047)	1.494 (0.046)
C=C <sup>b</sup>	9	1.402	1.416 (0.023)	1.417 (0.025)
C=C	19	1.343	1.350 (0.032)	1.349 (0.030)
C≡C	7	1.210	1.198 (0.019)	1.195 (0.022)

a J.H.Callomon, E.Hirota, K.Kuchitsu, W.J.Lafferty, A.G.Maki, and C.S.Pote, "Structure Data of Free Polyatomic Molecules", Landolt-Boernstein, New Series, Vol. II/7, Springer-Verlag, Berlin, 1976. J.H.Callomon, E.Hirota, T.Iijima, K.Kuchitsu, and W.J.Lafferty, "Structure Data of Free Polyatomic Molecules", Landolt-Boernstein, New Series, Vol. II/15, Springer-Verlag, Berlin, 1987. E.Hirota, T.Iijima, K.Kuchitsu, W.J.Lafferty, and D.A.Ramsay, "Structure Data of Free Polyatomic Molecules", Landolt-Boernstein, New Series, Vol. II/21, Springer-Verlag, Berlin, 1992.

b bei Aromaten

Bei den **Bindungswinkeln** sind die mittleren absoluten Fehler von 2.55° mit den MNDO-Standardparametern leicht auf 2.38° mit den neuen MNDO-Parametern ab-

gesunken (Tabelle 3.15). Dabei sind die Fehler für die Winkel an C und besonders an O Atomen generell etwas kleiner geworden, während sie für die Winkel an N Atomen etwas angestiegen sind. Bei beiden Parametersätzen sind die maximalen Fehler für einen gegebenen Winkeltyp meist ähnlich gross und oft bei dem gleichen Molekül anzutreffen (z.B. HCH Bicyclobutan, CCC Acrolein). Generell liegen die Fehler in beiden Fällen in einer tolerierbaren Grössenordnung.

### 3.6.4.3 Ionisierungspotentiale

Die vertikalen Ionisierungspotentiale werden als negative Orbitalenergien berechnet (Koopmans' Theorem). Dabei vernachlässigt man Relaxationseffekte, welche das Kation stabilisieren, und differentielle Korrelationseffekte, die das Neutalmolekül begünstigen. Die damit verbundenen Fehler kompensieren sich tendenziell, wenn auch nicht vollständig. Für die Moleküle im Referenzdatensatz darf man erwarten, dass die ersten Ionisierungspotentiale bei Anwendung von Koopmans' Theorem eher etwas überschätzt werden sollten (Relaxationseffekte dominant, insbesondere bei einsamen Elektronenpaaren). Dieser erwartete Trend wird bei den MNDO-Standardparametern in der Tat beobachtet (MSE=0.25 eV, siehe Tabelle 3.16), während bei den neuen MNDO-Parametern ein deutlich kleinerer systematischer Fehler (MSE = -0.10 eV) in entgegengesetzter Richtung auftritt. Dies hat zur Folge, dass der mittlere absolute Fehler von 0.46 eV (MNDO) auf 0.31 eV (MNDO<sub>neu</sub>) absinkt, mit entsprechenden Verbesserungen bei nahezu allen Molekülklassen (siehe Tabelle 3.16). Wie bereits betont, bezieht sich dieser Erfolg der systematischen Reparametrisierung auf Ionisierungspotentiale, die mittels Koopmans' Theorem berechnet werden; eine explizite Berücksichtigung der Relaxationseffekte durch separate SCF-Rechnungen von Neutalmolekül und Kation mag ein anderes Bild ergeben.

### 3.6.4.4 Dipolmomente

Wie aus der Tabelle 3.17 ersichtlich ist, werden die Dipolmomente der Referenzmoleküle durch die neuen MNDO-Parameter deutlich besser reproduziert als durch die MNDO-Standardparameter: der mittlere absolute Fehler sinkt von 0.35 auf 0.22 D, mit deutlichen Verbesserungen bei den meisten Molekülklassen. Mit beiden Parametersätzen findet man den grössten Fehler bei Methylocyanid (1.68 resp. 1.39 D).

### 3.6.4.5 Schwingungen

Ein nicht unwesentlicher Beitrag zum totalen SSQ-Wert stammt bei dem jetzigen Referenzdatensatz von den Schwingungswellenzahlen, die bei der ursprünglichen MNDO Parametrisierung nicht zu den Referenzdaten gehörten. Von daher ist es nicht überraschend, dass die Reparametrisierung auch hier zu merklichen Fortschritten führt: bei den 112 untersuchten Schwingungen sinkt der mittlere absolute Fehler von 246 auf 207  $\text{cm}^{-1}$  (siehe Tabelle 3.18). Mit beiden Parametersätzen werden die Schwingungsfrequenzen generell überschätzt, wie dies aufgrund der Verwendung

der Hartree-Fock Näherung und der Vernachlässigung von Anharmonizitätskorrekturen zu erwarten ist (MSE-Werte von 200 und  $154\text{ cm}^{-1}$ , Tabelle 3.18). Die durch die Reparametrisierung bedingten Verbesserungen bei den Streckschwingungen sind wohl zumindest teilweise durch Verschlechterungen bei den Bindungslängen erkauft: wie bereits diskutiert, liefern die neuen MNDO-Parameter generell etwas grössere Bindungsabstände als die MNDO-Standardparameter, was dann erfahrungsgemäss niedrigere Schwingungsfrequenzen impliziert. Die grössten absoluten Fehler treten für beide Parametersätze bei den Streckschwingungen des Ozons auf, welche allerdings selbst auf korreliertem *ab initio* Niveau aufgrund des Multireferenzcharakters von Ozon nur schwierig zu behandeln sind.

#### 3.6.4.6 Ionen und Radikale

Die bisherigen Auswertungen haben sich im wesentlichen auf die neutralen geschlossenschaligen Moleküle im Referenzdatensatz beschränkt. Bei der semiempirischen Methodenentwicklung ist es gängige Praxis, darüber hinaus viele weitere Validierungsrechnungen zur Bewertung der erhaltenen Parameter durchzuführen. Hier werden nur Ergebnisse für die Standardbildungsenthalpien von Ionen (Tabelle 3.19) und Radikalen (Tabelle 3.20) vorgestellt, welche nicht im Referenzdatensatz vorkommen. Die ursprünglichen und die neuen MNDO-Parameter ergeben nahezu gleiche mittlere absolute Fehler bei den untersuchten Kationen (11.6 vs. 11.5 kcal/mol) und Anionen (14.4 vs. 14.7 kcal/mol), während erstere bei den Radikalen leicht überlegen sind (11.0 vs. 13.0 kcal/mol). Die Qualität der Ergebnisse ist insgesamt ähnlich, und auch die systematischen Fehler sind teilweise analog (siehe MSE-Werte): Mit beiden Parametersätzen werden sowohl die Anionen als auch die Radikale im Schnitt zu instabil berechnet (jeweils um etwa 10 kcal/mol), während bei den Kationen kein einheitlicher Trend zu sehen ist (MNDO ohne grossen systematischen Fehler, MNDO<sub>neu</sub> zu stabil).

#### 3.6.5 Wertung

Die testweise Reparametrisierung der etablierten MNDO Methode mittels `gps1` hat gezeigt, dass die Fehlerquadratsumme für den Referenzdatensatz durchaus merklich gesenkt werden kann (von 41280 auf 30794 im oben betrachteten Fall, wobei Optimierungen mit anderen Optionen sogar zu noch etwas niedrigeren SSQ-Werten im Bereich von 29000-30000 führen, siehe Tabelle 3.6). Dabei erweist sich die in `gps1` implementierte systematische Suche (ausgehend von den ursprünglichen MNDO Parametern) in Kombination mit gradientenbasierten Optimierungen als die effizienteste Strategie.

Ein genauerer Vergleich der Ergebnisse für die ursprünglichen und die ausgewählten neuen MNDO-Parameter verdeutlicht, welche Fortschritte bei den Referenzmolekülen im einzelnen erzielt werden konnten. Dabei stellt sich heraus, dass die Fehler im statistischen Sinne zwar für die meisten der untersuchten Eigenschaften



kleiner werden, die beiden Parametersätze sich in qualitativer Hinsicht aber sehr ähnlich verhalten. So findet man oft analoge Muster hinsichtlich der Vorzeichen und Grösse der individuellen Abweichungen, und die maximalen Fehler treten oft bei denselben Molekülen auf. Bei Validierungsrechnungen an Systemen, die nicht im Referenzdatensatz enthalten sind, findet man ebenfalls mit beiden Parametersätzen vergleichbare Fehler.

Festzuhalten bleibt, dass typische qualitative Probleme des MNDO Ansatzes durch die Reparametrisierung nicht behoben werden, sondern bei allen erzeugten Parametersätzen auftreten. Dies gilt auch für verschiedene andere Anwendungsfelder, die hier der Kürze halber nicht besprochen werden (z.B. Rotationsbarrieren, Wasserstoffbrücken und elektronisch angeregte Zustände). Diese Befunde sind nicht überraschend, da die mit `gps1` optimierten Parametersätze sich generell nur wenig von den ursprünglichen MNDO-Parametern unterscheiden (vgl. die Tabellen 3.4-3.7). Man kann daher zusammenfassend sagen, dass das Programm `gps1` sich bei der testweisen Reparametrisierung von MNDO technisch bewährt hat. Wie von vornherein erwartet, führen die neuen MNDO-Parameter zwar zu statistischen Verbesserungen, aber nicht zu wirklichen qualitativen Fortschritten, so dass sie nicht zur allgemeinen Verwendung empfohlen werden. Es erscheint daher erfolgversprechender, sich bei der Anwendung von `gps1` auf Methoden zu konzentrieren, die vom Modell her über den MNDO Ansatz hinausgehen und daher qualitative Fortschritte erlauben sollten.

### 3.6.6 Tabellen mit detaillierten Ergebnissen

Die Tabellen 3.10 - 3.20 dokumentieren die Daten, die in den vorangehenden Abschnitten diskutiert worden sind. Sie umfassen sowohl detaillierte Statistiken für verschiedene Eigenschaften und Molekülklassen als auch Einzelergebnisse für Standardbildungsenthalpien. Sie basieren auf automatisch erzeugten Dateien und enthalten daher eine Reihe von Bezeichnungen in englischer Sprache.

**Tabelle 3.10:** Nach Molekülklassen getrennte Fehler der Standardbildungsenthalpien [kcal/mol] in dem Referenzdatensatz

	Standardbildungsenthalpien [kcal/mol]						
	Anzahl	MNDO <sub>neu</sub>			MNDO		
		MSE	MUE	SD	MSE	MUE	SD
All molecules	140	-0.41	5.89	8.02	-1.22	6.34	9.34
CH compounds	57	-1.36	6.18	8.89	-3.14	5.90	9.62
CHN compounds	32	0.53	5.20	6.98	-1.79	6.24	8.46
CHO compounds	39	0.12	4.98	6.41	0.63	4.80	6.68
CHNO compounds	4	-1.98	9.15	12.31	3.25	13.63	17.98
HNO compounds	8	0.70	9.43	11.94	3.52	13.73	17.19
Aromatic Rings	5	7.52	7.52	8.97	1.92	1.92	2.18
Triple bonds	5	-6.89	6.89	8.28	-4.02	5.38	6.78
Three-membered ring	6	-3.34	3.42	4.99	-4.08	4.78	6.16
Four-membered rings	5	-16.56	16.56	21.90	-21.26	21.26	28.99
Bicyclic Systems	8	6.18	9.27	11.65	0.20	5.99	7.45
Amines	10	1.32	4.47	6.19	1.28	3.83	5.43
Cyanides	9	2.03	2.95	4.01	-3.06	4.56	6.75
With NN bonds	10	0.88	8.47	10.50	-0.54	10.59	13.93
Alcohols, Ethers, Water	12	-0.06	4.80	6.37	-2.71	4.47	5.82
Carbonyl compounds	12	-3.42	4.49	6.56	-1.08	3.84	5.18
Acids, Esters, Anhydrides	11	2.00	4.07	4.86	2.85	3.70	4.55
With OO bonds	7	0.77	8.11	10.15	4.35	10.38	13.09
With NO bonds	3	-9.75	13.26	17.66	-9.52	19.23	23.90

**Tabelle 3.11:** Mittlere absolute Fehler der Standardbildungsenthalpien [kcal/mol] in dem Referenzdatensatz (mittlere Fehler in Klammern angegeben)

	Anzahl	MNDO	MNDO <sub>neu</sub>	MNDO <sub>arom</sub>
All molecules	140	6.34	5.89	6.04
CH compounds	57	5.90	6.18	6.01
CHN compounds	32	6.24	5.20	6.61
CHO compounds	39	4.80	4.98	4.61
CHNO compounds	4	13.63	9.15	7.93
HNO compounds	8	13.73	9.43	10.00
Aromatic Rings	5	1.92	7.52	3.11
Triple bonds	5	5.38	6.89	3.83
Three-membered rings	6	4.78	3.42	3.99
Four-membered rings	5	21.26	16.56	21.71
Bicyclic Systems	8	5.99	9.27	5.75
Amines	10	3.83	4.47	3.82
Cyanides	9	4.56	2.95	6.22
With NN bonds	10	10.59	8.47	10.25
Alcohols, Ethers, Water	12	4.47	4.80	4.23
Carbonyl compounds	12	3.84	4.49	4.13
Acids, Esters, Anhydrides	11	3.70	4.07	3.40
With OO bonds	7	10.38	8.11	7.40
With NO bonds	3	19.23	13.26	12.33

**Tabelle 3.12:** Standardbildungsenthalpien [kcal/mol] der Referenzmoleküle

Standardbildungsenthalpien [kcal/mol]							
Molekül	Exp.	Berechnet		Fehler		$\delta_{exp}$	Lit.
		MNDO <sub>neu</sub>	MNDO	MNDO <sub>neu</sub>	MNDO		
Hydrogen	0.0	2.8	0.7	2.8	0.7		a
Methane	-17.9	-13.4	-11.9	4.5	6.0	0.1	b
Ethane	-20.0	-20.9	-19.7	-0.9	0.3	0.1	b
Ethylene	12.5	12.1	15.4	-0.4	2.9	0.1	b
Acetylene	54.5	51.4	57.9	-3.1	3.4	0.2	b
Propane	-25.0	-25.3	-24.9	-0.3	0.1	0.1	b
Propene ecl	4.8	3.2	5.0	-1.6	0.2	0.2	b
Propyne	44.2	38.3	41.4	-5.9	-2.8	0.2	b
Allene	45.5	41.6	43.9	-3.9	-1.6	0.3	b
n-Butane	-30.0	-29.4	-29.7	0.6	0.3	0.2	b
Isobutane	-32.1	-25.9	-26.8	6.2	5.3	0.2	b
But-1-ene H-ecl	0.0	-1.1	-0.2	-1.1	-0.2	0.2	b
trans-2-Butene	-2.7	-5.5	-5.1	-2.8	-2.4	0.2	b
cis-2-Butene	-1.7	-4.1	-3.9	-2.4	-2.2	0.2	b
Isobutene	-4.0	-1.8	-2.0	2.2	2.0	0.2	b
1,2-Butadiene	38.8	32.9	33.5	-5.9	-5.3	0.1	b
1,3-tr-Butadiene	26.3	27.0	29.0	0.7	2.7	0.3	b
1-Butyne	39.5	34.0	36.1	-5.5	-3.4	0.2	b
2-Butyne	34.8	25.2	24.9	-9.6	-9.9	0.3	b
Vinylacetylene	73.0	62.6	65.6	-10.4	-7.4		c
Butatriene	83.0	70.9	71.3	-12.1	-11.7		c
n-Pentane	-35.1	-33.3	-34.4	1.8	0.7	0.2	b
Neopentane	-40.2	-22.1	-24.6	18.1	15.6	0.2	b
n-Hexane	-39.9	-37.3	-39.1	2.6	0.8	0.2	b
tr-1,3-Pentadiene	18.2	18.3	18.8	0.1	0.6	0.2	b
cis-1,3-Pentadiene	19.5	19.7	20.1	0.2	0.6	0.3	b
1,4-Pentadiene	25.2	24.5	26.0	-0.7	0.8	0.3	b
Cyclopropane	12.7	10.0	11.2	-2.7	-1.5	0.1	b
cis-Dimethylcyclopropane	1.3	0.1	-2.1	-1.2	-3.4		d
Cyclopropene	66.2	66.4	68.3	0.2	2.1	0.6	b
1-Methylcyclopropene	58.2	54.7	53.7	-3.5	-4.5	0.3	b
1,2-Dimethylcyclopropene	46.4	43.2	39.2	-3.2	-7.2		d
Methylenecyclopropan	47.9	38.3	37.9	-9.6	-10.0	0.4	b
Cyclobutane	6.8	-10.3	-11.9	-17.1	-18.7	0.1	b
Cyclobutene plan	37.5	31.5	31.0	-6.0	-6.5	0.4	b
1,2-Dimethylcyclobutene plan	19.8	11.6	6.4	-8.2	-13.4		d
Methylenecyclobutane	29.0	13.2	10.9	-15.8	-18.1	0.2	b
Cyclopentane envelope	-18.3	-27.1	-30.5	-8.8	-12.2	0.2	b
Cyclopentene	8.1	1.7	-0.3	-6.4	-8.4	0.3	b
Cyclopentadiene	32.1	32.9	32.1	0.8	0.0	0.4	b
Cyclohexane chair	-29.5	-30.1	-34.8	-0.6	-5.3	0.2	b
Cyclohexene half-chair	-1.2	-6.1	-9.9	-4.9	-8.7	0.2	b
1,3-Cyclohexadiene	25.4	17.8	14.5	-7.6	-10.9	0.2	b
Benzene	19.7	24.0	21.3	4.3	1.6	0.2	b
Toluene ecl	12.0	18.4	13.6	6.4	1.6	0.1	b

... Continued ...

Tabelle 3.12: (continued)

Molekül	Standardbildungsenthalpien [kcal/mol]						$\delta_{exp}$	Lit.
	Exp.	Berechnet		Fehler				
		MNDO <sub>neu</sub>	MNDO	MNDO <sub>neu</sub>	MNDO			
Ethylbenzene stag	7.1	14.4	8.8	7.3	1.7	0.3	b	
Styrene	35.3	42.1	37.6	6.8	2.3	0.4	b	
Cycloheptatriene	43.2	38.5	33.8	-4.7	-9.4	0.5	b	
Bicyclobutane	51.9	65.1	64.1	13.2	12.2	0.2	b	
Spirocyclopentane	44.3	36.3	33.7	-8.0	-10.6	0.2	b	
trans-Bicyclopropyl	30.9	30.8	28.6	-0.1	-2.3	0.9	b	
Bicyclo[2.1.0]pentan	37.8	33.5	30.2	-4.3	-7.6		c	
Norbornane	-13.1	-2.5	-10.4	10.6	2.7	1.1	b	
Norbornadiene	58.8	69.1	62.9	10.3	4.1	0.7	b	
Bicyclo[2.2.2]octane	-23.7	-16.6	-26.3	7.1	-2.6	0.3	b	
Naphthalene	35.9	48.6	38.3	12.7	2.4	0.4	b	
Adamantane	-32.2	-11.6	-26.4	20.6	5.8	0.5	b	
Cubane	148.7	113.0	99.1	-35.7	-49.6	0.9	b	
Nitrogen	0.0	17.0	8.3	17.0	8.3		a	
Ammonia	-11.0	-5.7	-6.4	5.3	4.6	0.1	b	
Methylamine	-5.5	-8.8	-7.5	-3.3	-2.0	0.1	b	
Dimethylamine	-4.4	-8.8	-6.6	-4.4	-2.2	0.2	b	
Trimethylamine	-5.7	-4.7	-2.8	1.0	2.9	0.2	b	
Ethylamine	-11.3	-14.4	-13.2	-3.1	-1.9	0.2	b	
n-Propylamine	-16.8	-18.5	-18.2	-1.7	-1.4	0.1	b	
Isopropylamine	-20.0	-15.4	-16.4	4.6	3.6	0.1	b	
tert-Butylamine	-28.9	-13.6	-15.5	15.3	13.4	0.2	b	
Aziridine	30.2	26.9	25.1	-3.3	-5.1	0.2	b	
Pyrrole	25.9	28.2	32.5	2.3	6.6	0.1	b	
Pyridine	33.6	31.5	28.8	-2.1	-4.8	0.2	b	
Pyridazine (1,2)	66.5	50.2	43.6	-16.3	-22.9	0.2	b	
Pyrimidine (1,3)	46.8	37.8	35.0	-9.0	-11.8	0.3	b	
Pyrazine (1,4)	46.9	41.1	37.8	-5.8	-9.1	0.3	b	
Aniline	20.8	23.7	21.7	2.9	0.9	0.2	b	
Hydrogen cyanide	32.3	33.5	35.3	1.2	3.0	2.0	e	
Acetonitrile	18.0	20.7	19.2	2.7	1.2	0.2	c	
Methylisocyanide	39.1	52.4	60.3	13.3	21.2	1.7	b	
Propionitrile	12.3	16.1	13.8	3.8	1.5	0.1	b	
Acrylonitrile	43.2	45.6	43.8	2.4	0.6	0.4	b	
Fumaronitrile	81.0	82.1	74.7	1.1	-6.3		b	
Maleonitrile	81.3	82.6	75.0	1.3	-6.3	0.5	b	
Dicyanoacetylene	126.5	122.4	111.4	-4.1	-15.1	0.3	b	
Cyanogen	73.3	74.4	66.6	1.1	-6.7	0.2	b	
Benzonitrile	51.6	60.5	52.0	8.9	0.4	0.5	b	
Hydrazine	22.8	17.0	14.2	-5.8	-8.6	0.2	e	
Methylhydrazine	22.6	15.8	14.4	-6.8	-8.2	0.1	b	
1,1-Dimethylhydrazine	20.1	18.1	18.3	-2.0	-1.8	0.8	b	
1,2-Dimethylhydrazine	22.0	14.9	15.0	-7.1	-7.0	1.0	b	
trans-Azodiisopropane	8.6	9.6	1.5	1.0	-7.1	0.2	b	
Hydrogen azide	70.3	77.1	73.1	6.8	2.8		c	
Oxygen Triplet	0.0	-17.1	-15.3	-17.1	-15.3		a	

... Continued ...

Tabelle 3.12: (continued)

Molekül	Standardbildungsenthalpien [kcal/mol]						Lit.
	Exp.	Berechnet		Fehler		$\delta_{exp}$	
		MNDO <sub>neu</sub>	MNDO	MNDO <sub>neu</sub>	MNDO		
Ozone	34.0	43.6	48.5	9.6	14.5		c
Water	-57.8	-57.0	-60.9	0.8	-3.1		b
Methanol	-48.2	-55.5	-57.4	-7.3	-9.2	0.1	b
Ethanol	-56.2	-60.4	-63.0	-4.2	-6.8	0.1	b
Propanol	-61.0	-64.5	-67.7	-3.5	-6.7	0.1	b
Isopropanol	-65.2	-62.9	-65.1	2.3	0.1	0.1	b
tert-Butanol	-74.7	-60.4	-64.3	14.3	10.4	0.2	b
Dimethylether	-44.0	-49.8	-51.2	-5.8	-7.2	0.2	b
Diethylether	-60.3	-59.2	-62.0	1.1	-1.7	0.2	b
Oxirane	-12.6	-20.8	-15.5	-8.2	-2.9	0.1	b
Furan	-8.3	-8.6	-8.6	-0.3	-0.3	0.2	b
Phenol	-23.0	-19.3	-26.7	3.7	-3.7	0.2	b
Anisole plan	-16.2	-10.0	-17.7	6.2	-1.5	0.2	b
Hydrogen peroxide	-32.5	-39.2	-38.2	-6.7	-5.7		e
Dimethylperoxide	-30.0	-31.9	-28.3	-1.9	1.7	0.3	b
Diethylperoxide	-46.1	-41.5	-39.1	4.6	7.0	0.6	b
Diethylperoxide	-46.1	-41.0	-38.9	5.1	7.2	0.6	b
Carbon monoxide	-26.4	-10.8	-5.9	15.6	20.5		b
Carbon dioxide	-94.1	-84.8	-75.1	9.3	19.0		b
Carbon suboxide	-22.0	-25.8	-23.5	-3.8	-1.5	0.5	c
Formaldehyde	-26.0	-36.0	-32.9	-10.0	-6.9	0.1	b
Acetaldehyde free	-39.7	-45.6	-42.3	-5.9	-2.6	0.1	b
Propionaldehyde ecl	-44.4	-49.9	-47.5	-5.5	-3.1	0.2	b
Acetone ecl	-51.9	-52.2	-49.4	-0.3	2.5	0.2	b
Ketene	-11.4	-12.0	-6.8	-0.6	4.6	0.4	b
Acrolein free	-18.0	-21.1	-17.9	-3.1	0.1		c
Glyoxal free	-50.7	-66.8	-62.5	-16.1	-11.8	0.2	b
Biacetyl (C <sub>2</sub> )	-78.2	-81.9	-78.8	-3.7	-0.6	0.3	b
Biacetyl (C <sub>2h</sub> )	-78.2	-78.4	-75.6	-0.2	2.6	0.3	b
Acetylacetone	-91.0	-87.1	-84.3	3.9	6.7	0.4	b
p-Quinone	-29.4	-31.6	-32.9	-2.2	-3.5	0.8	b
Benzaldehyde	-8.8	-6.3	-9.9	2.5	-1.1	0.7	b
Formic acid (Z)	-90.5	-94.3	-92.6	-3.8	-2.1	0.1	b
Acetic acid stag (Z)	-103.4	-101.3	-101.1	2.1	2.3	0.4	b
Propionic acid C-ecl	-108.4	-104.9	-105.6	3.5	2.8	0.1	b
Oxalic acid 4-ring	-173.0	-171.9	-175.1	1.1	-2.1	1.2	b
Benzoic acid	-70.3	-60.7	-67.7	9.6	2.6	0.5	b
Methyl formiate (Z)	-85.0	-87.0	-85.5	-2.0	-0.5	0.2	b
Methyl acetate ecl	-98.5	-93.5	-93.7	5.0	4.8	0.4	b
Acetic anhydride ecl	-136.8	-132.8	-132.6	4.0	4.2	0.4	b
Maleic anhydride	-95.2	-89.8	-88.5	5.4	6.7	1.2	b
Formamide	-44.0	-49.5	-40.2	-5.5	3.8		c
Dimethylformamide	-45.8	-43.1	-37.0	2.7	8.8	0.5	b
Nitrous oxide	19.6	27.7	31.0	8.1	11.4	0.1	e
Dinitrotetroxide free	2.2	16.1	29.9	13.9	27.7		b
Nitrous acid trans	-18.3	-36.1	-40.7	-17.8	-22.4	0.3	e

... Continued ...

Tabelle 3.12: (continued)

Molekül	Standardbildungsenthalpien [kcal/mol]						$\delta_{exp}$	Lit.
	Exp.	Berechnet			Fehler			
		MNDO <sub>neu</sub>	MNDO	MNDO <sub>neu</sub>	MNDO			
Nitric acid	-32.1	-26.8	-17.5	5.3	14.6	0.1	e	
Methylnitrite	-15.9	-32.6	-36.7	-16.7	-20.8	0.2	b	
Nitromethane	-17.8	-6.1	3.3	11.7	21.1	0.1	b	
Mean error				-0.41	-1.22			
Mean absolute error				5.89	6.34			
Number of comparison				140	140			

a G.Pilcher and H.A.Skinner, in "The Chemistry of the Metal-Carbon Bond", edited by F.R.Hartley and S.Patai, Wiley, New York, 1982, pp. 43-90. b J.B.Pedley, R.D.Naylor, and S.P.Kirby, "Thermochemical Data of Organic Compounds", 2nd ed., Chapman and Hall, London, 1986. c S.G.Lias, J.E.Bartmess, J.F.Liebman, J.L.Holmes, R.D.Levin, and W.G.Mallard, "Gas Phase Ion and Neutral Thermochemistry", J.Phys.Chem.Ref.Data 17, Suppl. 1 (1988). d R.B. Turner, P. Goebel, B.J. Mallon, W. von E. Doering, J.F. Coburn Jr., and M. Pomerantz, J.Am.Chem.Soc. 90, 4315 (1968). See c in Table II, DT paper. e M.W.Chase, C.A.Davies, J.R.Downey, D.R.Frurip, R.A.McDonald, and A.N.Syverud, "JANAF Thermochemical Tables", 3rd edition, J.Phys.Chem.Ref.Data 14, Suppl.1 (1985).

Tabelle 3.13: Standardbildungsenthalpien [kcal/mol] ausgewählter grosser Moleküle [65]

Molekül	Standardbildungsenthalpien [kcal/mol]							$\delta_{exp}$	Lit.
	Exp.	Berechnet			Fehler				
		MNDO <sub>neu</sub>	MNDO <sub>arom</sub>	MNDO	MNDO <sub>neu</sub>	MNDO <sub>arom</sub>	MNDO		
Tetrahydro-2H-pyran	-53.4	-57.3	-60.6	-62.0	-3.9	-7.2	-8.6	0.2	a
Maleic anhydride	-95.2	-89.8	-91.8	-88.5	5.4	3.4	6.7	1.2	a
Cyclohexanone	-54.0	-57.7	-61.6	-60.1	-3.7	-7.6	-6.1	0.5	a
Benzoquinone	-29.4	-31.6	-35.3	-32.9	-2.2	-5.9	-3.5	0.8	a
Hydroquinone	-63.4	-61.9	-64.7	-74.0	1.5	-1.3	-10.6	0.5	a
2-Methoxytetrahydropyran	-95.5	-94.5	-99.2	-103.0	1.0	-3.7	-7.5	0.4	a
Naphthalene	35.9	48.6	39.4	38.3	12.7	3.5	2.4	0.4	a
2-Aminobenzoic acid	-70.7	-60.5	-64.9	-66.3	10.2	5.8	4.4	0.3	a
1-Naphthol	-7.1	8.3	-1.5	-6.8	15.4	5.6	0.3	0.2	a
2-Naphthol	-7.2	5.3	-4.2	-9.7	12.5	3.0	-2.5	0.3	a
1,4-Naphthoquinone	-26.5	-13.5	-24.8	-22.5	13.0	1.7	4.0	1.0	a
Isophthalic acid	-166.4	-144.3	-151.2	-155.7	22.1	15.2	10.7	0.5	a
Terephthalic acid	-171.6	-144.4	-151.2	-155.7	27.2	20.4	15.9	0.6	a
1-Naphthoic acid	-53.3	-33.6	-45.4	-48.3	19.7	7.9	5.0	0.2	a
2-Naphthoic acid	-55.6	-35.9	-47.5	-50.5	19.7	8.1	5.1	0.4	a
Anthracene	55.2	76.8	60.4	58.9	21.6	5.2	3.7	0.5	a
Phenanthrene	49.6	73.5	57.1	55.7	23.9	7.5	6.1	0.4	b
9,10-Anthraquinone	-22.8	6.2	-12.7	-10.5	29.0	10.1	12.3	1.6	a
Phenanthraquinone	-33.3	10.4	-8.3	-6.8	43.7	25.0	26.5	1.1	a

... Continued ...

**Tabelle 3.13:** (continued)

Standardbildungsenthalpien [kcal/mol]									
Molekül	Exp.	Berechnet			Fehler			$\delta_{exp}$	Lit.
		MNDO <sub>neu</sub>	MNDO <sub>arom</sub>	MNDO	MNDO <sub>neu</sub>	MNDO <sub>arom</sub>	MNDO		
1,2-Benzanthracene- 9,10-dione	-35.6	30.8	4.6	6.5	66.4	40.2	42.1	1.3	a
				Mean error	16.77	6.84	5.32		
				Mean absolute error	17.75	9.41	9.20		
				Number of comparison	20	20	20		

a J.B.Pedley, R.D.Naylor, and S.P.Kirby, "Thermochemical Data of Organic Compounds", 2nd ed., Chapman and Hall, London, 1986. b S.G.Lias, J.E.Bartmess, J.F.Liebman, J.L.Holmes, R.D.Levin, and W.G.Mallard, "Gas Phase Ion and Neutral Thermochemistry", J.Phys.Chem.Ref.Data 17, Suppl. 1 (1988).

**Tabelle 3.14:** Nach Molekülklassen getrennte Fehler der Bindungslängen [Å] in dem Referenzdatensatz

	Bindungslänge [Å]						
	Anzahl	MSE	MUE	MSE	MUE	MSE	MUE
All molecules	242	0.010	0.020	0.014	0.021	0.001	0.014
CH compounds	113	0.009	0.014	0.012	0.014	0.005	0.011
CHN compounds	49	0.007	0.015	0.016	0.018	0.003	0.011
CHO compounds	57	0.020	0.026	0.022	0.028	0.003	0.017
CHNO compounds	5	0.024	0.024	0.027	0.027	0.014	0.015
HNO compounds	18	-0.010	0.045	-0.006	0.047	-0.033	0.037
C-H	61	0.019	0.019	0.015	0.016	0.005	0.009
C-C	69	0.008	0.014	0.012	0.015	0.006	0.013
C=C arom.	9	0.016	0.016	0.021	0.021	0.014	0.014
C=C	19	0.006	0.007	0.013	0.013	0.007	0.008
C≡C	7	-0.015	0.015	-0.006	0.006	-0.012	0.012
N-H	9	0.011	0.011	0.014	0.014	-0.002	0.007
C-N	8	-0.003	0.013	0.012	0.014	0.005	0.012
C=N	3	0.016	0.016	0.030	0.030	0.018	0.018
C≡N	6	0.007	0.008	0.023	0.023	0.008	0.008
N-N	3	-0.101	0.101	-0.084	0.084	-0.079	0.079
N=N	6	0.007	0.013	0.028	0.028	0.004	0.012
O-H	7	0.046	0.046	0.037	0.037	-0.011	0.012
C-O	8	0.023	0.023	0.030	0.030	-0.006	0.018
C=O, C≡O	15	0.036	0.036	0.043	0.043	0.017	0.017
O-O	3	-0.069	0.069	-0.071	0.071	-0.108	0.108
N=O (as in NO <sub>2</sub> Groups)	6	0.023	0.023	0.032	0.032	0.001	0.003

**Tabelle 3.15:** Nach Molekülklassen getrennte Fehler der Bindungswinkel [°] in dem Referenzdatensatz

Bindungswinkel [°]							
	Anzahl	MNDO <sub>neu</sub>		MNDO <sub>arom</sub>		MNDO	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	101	0.97	2.38	0.71	2.28	0.73	2.56
CH compounds	38	0.36	1.61	0.36	1.69	0.39	1.81
CHN compounds	22	2.75	3.36	1.87	2.82	0.45	2.03
CHO compounds	31	0.46	2.19	0.49	2.21	1.32	3.11
HNO compounds	10	0.95	3.77	0.20	3.54	0.84	4.83
HCH	15	-2.11	2.73	-2.35	2.84	-2.61	3.04
HCC	24	0.45	1.12	0.59	1.13	0.57	1.13
CCC	17	1.32	1.67	1.53	1.87	1.68	2.03
at N	19	3.24	4.36	1.94	3.59	-0.28	3.04
at O	9	2.87	4.01	2.28	3.92	6.98	7.04
XCO	12	0.18	1.50	0.20	1.49	0.49	1.86

**Tabelle 3.16:** Nach Molekülklassen getrennte Fehler der Ionisierungspotentiale [eV] in dem Referenzdatensatz

Ionisierungspotentiale [eV]							
	Anzahl	MNDO <sub>neu</sub>			MNDO		
		MSE	MUE	SD	MSE	MUE	SD
All molecules	52	-0.10	0.31	0.40	0.25	0.46	0.57
CH compounds	22	-0.10	0.22	0.28	0.27	0.38	0.47
CHN compounds	13	-0.11	0.33	0.43	0.29	0.54	0.69
CHO compounds	14	0.02	0.39	0.49	0.30	0.57	0.67
HNO compounds	3	-0.57	0.57	0.80	-0.24	0.24	0.35
Aromatic Rings	3	0.05	0.13	0.16	0.34	0.34	0.45
Triple bonds	3	-0.20	0.27	0.37	-0.08	0.31	0.39
Amines	5	0.06	0.33	0.43	0.90	0.90	1.06
Cyanides	5	-0.12	0.25	0.30	-0.07	0.30	0.37
Alcohols, Ethers	5	0.13	0.34	0.46	0.44	0.61	0.76
Carbonyl compounds	6	0.09	0.39	0.47	0.38	0.49	0.64

**Tabelle 3.17:** Nach Molekülklassen getrennte Fehler der Dipolmomente [D] in dem Referenzdatensatz

Dipolmomente [D]							
	Anzahl	MNDO <sub>neu</sub>			MNDO		
		MSE	MUE	SD	MSE	MUE	SD
All molecules	53	-0.05	0.22	0.34	-0.18	0.35	0.50
CH compounds	10	-0.18	0.22	0.29	-0.19	0.25	0.33
CHN compounds	16	-0.19	0.26	0.45	-0.40	0.54	0.75
CHO compounds	19	0.00	0.12	0.15	-0.12	0.18	0.23
HNO compounds	6	0.45	0.45	0.56	0.39	0.41	0.52
Amines	6	-0.02	0.11	0.14	0.15	0.20	0.23

... Continued ...



**Tabelle 3.17:** (continued)

	Dipolmomente [D]						
	Anzahl	MNDO <sub>neu</sub>			MNDO		
		MSE	MUE	SD	MSE	MUE	SD
Cyanides	4	-0.41	0.41	0.56	-0.79	0.79	0.99
With NN bonds	4	0.14	0.21	0.30	-0.40	0.70	0.95
Alcohols, Ethers, Water	9	-0.02	0.10	0.12	-0.14	0.15	0.19
Carbonyl compounds	6	0.00	0.15	0.23	-0.18	0.31	0.36
Acids, Esters, Anhydrides	6	-0.10	0.18	0.25	-0.24	0.28	0.45
With NO bonds	3	0.51	0.51	0.65	0.39	0.39	0.53

**Tabelle 3.18:** Nach Molekülklassen getrennte Fehler der Wellenzahlen [ $\text{cm}^{-1}$ ] in dem Referenzdatensatz

	Wellenzahlen [ $\text{cm}^{-1}$ ]						
	Anzahl	MNDO <sub>neu</sub>			MNDO		
		MSE	MUE	SD	MSE	MUE	SD
All molecules	112	153.5	206.8	268.8	200.2	246.0	306.9
CH compounds	33	142.5	182.8	219.3	175.1	203.3	248.9
CHN compounds	19	114.5	243.3	303.1	186.3	295.1	345.1
CHO compounds	41	144.7	182.1	242.9	186.3	221.4	284.0
HNO compounds	19	230.5	265.2	370.3	287.9	324.0	415.4
With NN bonds	4	265.2	265.2	367.7	301.5	301.5	381.2
With OO bonds	4	608.6	608.6	787.8	657.3	657.3	847.9

**Tabelle 3.19:** Nach Molekülklassen getrennte Fehler der Standardbildungsenthalpien [kcal/mol] von Ionen

	Standardbildungsenthalpien [kcal/mol]						
	Anzahl	MNDO <sub>neu</sub>			MNDO		
		MSE	MUE	SD	MSE	MUE	SD
<b>Kationen</b>							
All molecules	32	-8.37	11.52	16.31	-0.63	11.64	15.12
CH compounds	15	-8.32	13.69	19.47	0.91	15.29	19.10
CHN compounds	4	-10.49	10.56	16.15	3.46	8.43	10.96
CHO compounds	11	-8.27	10.08	14.43	-4.60	8.81	12.43
<b>Anionen</b>							
All molecules	24	12.38	14.69	20.33	8.78	14.41	18.99
CH compounds	7	22.51	22.70	29.95	17.10	19.51	26.41
CHN compounds	6	11.55	14.43	21.36	5.69	15.90	21.34
CHO compounds	9	9.23	9.57	13.87	6.99	9.25	12.99

**Tabelle 3.20:** Nach Molekülklassen getrennte Fehler der Standardbildungsenthalpien [kcal/mol] von Radikalen

Standardbildungsenthalpien [kcal/mol]							
	Anzahl	MNDO <sub>neu</sub>			MNDO		
		MSE	MUE	SD	MSE	MUE	SD
All molecules	50	-9.12	13.05	15.49	-9.13	11.73	14.79
CH compounds	24	-11.33	12.96	15.92	-10.16	11.02	13.87
CHN compounds	5	-3.87	14.08	18.91	-1.62	11.73	16.51
CHO compounds	5	-11.61	11.99	16.11	-10.06	10.06	14.67
CHF compounds	9	-3.71	13.27	15.32	-8.79	14.54	19.17
HNO compounds	6	-12.12	15.22	17.76	-12.54	13.69	16.77

## 3.7 Parametrisierung von Fluor für OM2 und OM3

Eine wesentliche Aufgabe der Dissertation bestand darin, die Parameterwerte von Fluor für OM2 und OM3 mit Hilfe des Programmpakets `gps1` zu bestimmen. OM1 [64, 66], OM2 [63, 67] und OM3 [68] sind neuere semiempirische Verfahren, die in ihrem Ansatz über das etablierte MNDO-Modell hinausgehen, indem sie Orthogonalisierungskorrekturen (und gewisse andere Korrekturen) in den Ein-Elektronenintegralen einführen. Die Abkürzung 'OM' steht für 'orthogonalization model', und die drei Verfahren werden im folgenden oft summarisch  $OMx$ -Methoden genannt. Bei OM1 erfolgen die Orthogonalisierungskorrekturen nur im Einzentrenteil der Ein-Elektronenmatrix (core Hamiltonian), bei OM2 und OM3 hingegen auch im Zweizentrenteil bei den Resonanzintegralen (mit gewissen Vereinfachungen in OM3, siehe unten). Theoretische Analysen weisen darauf hin, dass einige der qualitativen Mängel der auf dem MNDO-Modell beruhenden Implementierungen (MNDO, AM1, PM3) durch die fehlenden Orthogonalisierungskorrekturen bedingt sind [63, 67, 69, 70, 71, 72, 73, 74], und es ist mit den bislang durchgeführten Parametrisierungen der  $OMx$ -Methoden für die Elemente H, C, N und O in der Tat gelungen, die betreffenden Mängel weitgehend zu beheben [63, 64, 66, 67]. Deshalb ist es wünschenswert, diese Parametrisierungen auf weitere Elemente auszudehnen. Fluorparameter für OM1 existieren bereits [64], daher befasst sich dieses Kapitel nur mit der Optimierung der Fluorparameter für OM2 und OM3.

Im folgenden wird zunächst der bekannte theoretische Formalismus von OM2 und OM3 kurz dokumentiert. Danach wird der Verlauf der Parametrisierung skizziert, bevor die endgültigen Parameterwerte von Fluor vorgestellt werden. Zum Abschluss werden die OM2 und OM3 Ergebnisse für den Referenzdatensatz und für drei Validierungssätze detailliert beschrieben und mit den entsprechenden MNDO, AM1, PM3 und OM1 Ergebnissen verglichen.

### 3.7.1 Der theoretische Ansatz in OM2 und OM3

Die  $OMx$ -Methoden sind semiempirische Molekülorbital(MO)-Verfahren, die auf der NDDO-Näherung (neglect of diatomic differential overlap) [75] basieren. Die MOs werden durch Lösung der Säkulargleichungen

$$\mathbf{FC} = \mathbf{CE}. \quad (3.8)$$

bestimmt, wobei  $\mathbf{F}$ ,  $\mathbf{C}$  und  $\mathbf{E}$  die Fockmatrix, die Matrix der MO-Koeffizienten und die Diagonalmatrix der MO-Energien bezeichnen. Verglichen mit den *ab initio* Säkulargleichungen

$$\mathbf{FC} = \mathbf{SCE}, \quad (3.9)$$

fehlt aufgrund der NDDO-Näherung die Überlappungsmatrix  $\mathbf{S}$ . Bei *ab initio* Rechnungen wird eine explizite Transformation von (3.9) nach (3.8) vorgenommen, bei

der die Basis orthogonalisiert wird und die Elemente der Fockmatrix sich ändern. Da die semiempirischen Methoden direkt mit (3.8) arbeiten, müssen die Effekte dieser Orthogonalisierung in der semiempirischen Darstellung der Fockmatrixelemente implizit berücksichtigt werden. Hierzu existiert eine umfangreiche Literatur [76, 77, 78, 79, 80, 81, 82, 83, 84, 70, 85, 86, 87, 88, 89]. In den OM $x$ -Methoden wird versucht, einige dieser Orthogonalisierungseffekte explizit zu erfassen (insbesondere bei den Eielektronenintegralen), wobei dann konsistenterweise auch andere Korrekturen ähnlicher Größenordnung explizit zu behandeln sind (Durchdringungsintegrale und Rumpf-Valenz-Wechselwirkungen). Für eine ausführliche Diskussion dieser Zusammenhänge sei wiederum auf die Literatur verwiesen [64, 66]. Hier sollen im folgenden nur die grundlegenden Formeln für die OM $x$ -Methoden zusammengestellt werden, um die zu optimierenden Parameter eindeutig zu definieren.

Die Fockmatrixelemente haben aufgrund der Verwendung der NDDO-Näherung in den OM $x$ -Methoden die gleiche Form wie im MNDO-Modell [24]. Mit der Konvention, dass die Basisorbitale  $\mu$  und  $\nu$  am Atom  $A$  und die Basisorbitale  $\lambda$  und  $\sigma$  am Atom  $B$  lokalisiert sind, lauten die entsprechenden Gleichungen:

$$F_{\mu\mu} = H_{\mu\mu} + \sum_{\nu}^A P_{\nu\nu}[\langle\mu\mu|\nu\nu\rangle^s - \frac{1}{2}\langle\mu\nu|\mu\nu\rangle^s] + \sum_{B \neq A} \sum_{\lambda}^B \sum_{\sigma}^B P_{\lambda\sigma} \langle\mu\mu|\lambda\sigma\rangle^s \quad (3.10)$$

$$F_{\mu\nu} = H_{\mu\nu} + \frac{1}{2}P_{\mu\nu}[3\langle\mu\nu|\mu\nu\rangle^s - \langle\mu\mu|\nu\nu\rangle^s] + \sum_{B \neq A} \sum_{\lambda}^B \sum_{\sigma}^B P_{\lambda\sigma} \langle\mu\nu|\lambda\sigma\rangle^s \quad (3.11)$$

$$F_{\mu\lambda} = H_{\mu\lambda} - \frac{1}{2} \sum_{\nu}^A \sum_{\sigma}^B P_{\nu\sigma} \langle\mu\nu|\lambda\sigma\rangle^s \quad (3.12)$$

In diesen Gleichungen treten Elemente der Eielektronenmatrix  $\mathbf{H}$  ('core Hamiltonian') und der Dichtematrix  $\mathbf{P}$  auf sowie Zweielektronenintegrale des allgemeinen Typs  $\langle\mu\nu|\lambda\sigma\rangle$ , wobei aufgrund der NDDO-Näherung nur Einzentren- und Zweizentrensterme berücksichtigt werden. Der Superskript  $s$  steht für die semiempirische Bestimmung von Integralen, was bei den Zweielektronenintegralen im Einzentrenfall eine Anpassung an Atomspektren und im Zweizentrenfall eine Klopman-Ohno-Skalierung von analytisch berechneten Integralen (im folgenden Superskript  $a$ ) impliziert [90, 91, 92, 93, 94].

Die Elemente der Eielektronenmatrix  $\mathbf{H}$  lauten:

$$H_{\mu\mu} = U_{\mu\mu} + \sum_B [Z_B \langle\mu\mu|ss\rangle^s + V_{\mu\mu,B}(PI) + V_{\mu\mu,B}(ORT) + V_{\mu\mu,B}(ECP)] \quad (3.13)$$

$$H_{\mu\nu} = \sum_B [Z_B \langle\mu\nu|ss\rangle^s + V_{\mu\nu,B}(PI) + V_{\mu\nu,B}(ORT) + V_{\mu\nu,B}(ECP)] \quad (3.14)$$

$$H_{\mu\lambda} = \beta_{\mu\lambda} + \sum_C V_{\mu\lambda,C}(ORT) \quad (3.15)$$

Dabei bezeichnet  $U_{\mu\mu}$  die Einzentren-Einelektronenenergie eines Elektrons im Basisorbital  $\mu$  und  $Z_B$  die Rumpfladung von Atom  $B$ . Die übrigen Grössen sind in den folgenden Gleichungen definiert, welche für OM2 und OM3 gelten:

$$\langle \mu\nu | \lambda\sigma \rangle^s = f_{KO} \langle \mu\nu | \lambda\sigma \rangle^a \quad (3.16)$$

$$f_{KO} = \frac{\langle ss | ss \rangle^s}{\langle ss | ss \rangle^a} \quad (3.17)$$

$$\beta_{\mu\lambda} = \pm \frac{1}{2} (\beta_\mu^A + \beta_\lambda^B) \sqrt{R_{AB}} e^{-(\alpha_\mu^A + \alpha_\lambda^B)} R_{AB}^2 \quad (3.18)$$

$$V_{\mu\nu,B}(PI) = [V_{\mu\nu,B}^a - Z_B \langle \mu\nu | ss \rangle^a] f_{KO} \quad (3.19)$$

$$\begin{aligned} V_{\mu\nu,B}(ORT) = & -\frac{1}{2} F_1^A \sum_{\lambda}^B (S_{\mu\lambda} \beta_{\lambda\nu} + \beta_{\mu\lambda} S_{\lambda\nu}) \\ & + \frac{1}{8} F_2^A \sum_{\lambda}^B S_{\mu\lambda} S_{\lambda\nu} (H_{\mu\mu}^{loc} + H_{\nu\nu}^{loc} - 2H_{\lambda\lambda}^{loc}) \end{aligned} \quad (3.20)$$

$$\begin{aligned} V_{\mu\lambda,C}(ORT) = & -\frac{1}{2} G_1^{AB} \sum_{\rho}^C (S_{\mu\rho} \beta_{\rho\lambda} + \beta_{\mu\rho} S_{\rho\lambda}) \\ & + \frac{1}{8} G_2^{AB} \sum_{\rho}^C S_{\mu\rho} S_{\rho\lambda} (H_{\mu\mu}^{loc} + H_{\lambda\lambda}^{loc} - 2H_{\rho\rho}^{loc}) \end{aligned} \quad (3.21)$$

$$G_1^{AB} = \frac{1}{2} (G_1^A + G_1^B)$$

$$G_2^{AB} = \frac{1}{2} (G_2^A + G_2^B) \quad (3.22)$$

$$V_{\mu\nu,B}(ECP) = -(S_{\mu\alpha} G_{\alpha\nu} + G_{\mu\alpha} S_{\alpha\nu}) - S_{\mu\alpha} S_{\alpha\nu} F_{\alpha\alpha} \quad (3.23)$$

$$G_{\mu\alpha} = \frac{1}{2} (\beta_\mu^A + \beta_\alpha^B) \sqrt{R_{AB}} e^{-(\alpha_\mu^A + \alpha_\alpha^B)} R_{AB}^2 \quad (3.24)$$

Die Durchdringungsintegrale (3.19), die Orthogonalisierungskorrekturen (3.20)-(3.22) und die Rumpf-Valenz-Pseudopotentiale (3.23)-(3.24) sind Zusatzterme, durch die sich das OM $x$ -Modell von dem MNDO-Modell unterscheidet. OM3 ist eine vereinfachte Variante von OM2, bei der die jeweils zweite Summe in (3.20) und (3.21) weggelassen wird, weil diese Terme genaueren Analysen zufolge wenig zu den Orthogonalisierungskorrekturen beitragen und daher vernachlässigt werden können [95]. Formal gilt daher:

$$F_2^A = 0, G_2^A = 0 \text{ in OM3} \quad (3.25)$$

Nachdem die grundlegenden Gleichungen für OM2 und OM3 jetzt vollständig angegeben sind, sollen jetzt abschliessend noch die zu optimierenden Parameter spezifiziert werden.

OM2 und OM3 verwenden ebenso wie OM1 eine an Atomen optimierte ECP-3G-Basis von Gauss-Orbitalen [64, 66]. Um deren Kontraktion in Molekülen zu erlauben, wird eine gemeinsame Skalierung aller Basisorbitale an einem gegebenen Atom  $A$  gemäss

$$\alpha_i^A \rightarrow \alpha_i^A (\zeta^A)^2 \quad (3.26)$$

eingeführt, wobei der Skalierungsfaktor  $\zeta$  ein justierbarer Parameter ist.

Wie bei den etablierten semiempirischen Verfahren werden die Einzentrenenergien  $U_{ss}$  und  $U_{pp}$  ebenfalls als freie Parameter behandelt.

In dem empirischen Ausdruck [64, 66] für die Resonanzintegrale (3.18) treten für jedes Basisorbital  $\mu$  zwei freie Parameter  $\beta_\mu$  und  $\alpha_\mu$  auf. Bei einem Atom mit einer  $sp$ -Basis (z.B. Kohlenstoff oder Fluor) kann man in einem lokalen Koordinatensystem  $s$ ,  $p_\sigma$  und  $p_\pi$  Orbitale unterscheiden, was dann zu insgesamt sechs Parametern führt. Es hat sich darüber hinaus als zweckmässig erwiesen, in den Resonanzintegralen für X-H-Atompaare spezielle  $\beta_\mu$  und  $\alpha_\mu$  Parameter am Atom X zu nutzen, so dass schlussendlich zehn Parameter für ein Element wie Fluor resultieren.

Die Orthogonalisierungskorrekturen (3.20)-(3.22) enthalten Vorfaktoren mit insgesamt vier freien Parametern (OM2), welche die theoretisch hergeleiteten Korrekturen [63, 67] skalieren und somit deren Feinjustierung gestatten. In OM3 sind zwei dieser Parameter per definitionem gleich Null (3.25).

Schliesslich benötigen die semiempirischen Rumpf-Valenz-Pseudopotentiale [63, 67] (3.23)-(3.24) vier Parameter: den Orbitalexponent  $\zeta_\alpha$  des Rumpforbitals zur Berechnung des Rumpf-Valenz-Überlappungsintegrals in (3.23), die Rumpffnergie  $F_{\alpha\alpha}$  in (3.23) sowie die beiden freien Rumpforbital-Parameter  $\beta_\alpha$  und  $\alpha_\alpha$  in Gleichung (3.24).

Insgesamt resultieren für Fluor somit insgesamt 21 zu optimierende Parameter in OM2 und 19 in OM3.

### 3.7.2 Parametrisierungsverlauf

Wie bereits erwähnt, sind die OM2 und OM3 Parameter für die Elemente H, C, N und O bereits früher von W. Weber [63] und M. Scholten [68] bestimmt worden. Diese Parameterwerte wurden bei der Fluorparametrisierung unverändert beibehalten. Durch Extrapolation dieser bekannten Parameter entlang der Reihe C-N-O lassen sich in den meisten Fällen vernünftige Startparameter für die weitere Optimierung erraten. Darüber hinaus ist in manchen Fällen auch aufgrund der physikalischen Bedeutung der jeweiligen Parameter ein relativ enger Wertebereich vorgegeben, in dem die physikalisch sinnvollen Parameterwerte liegen müssen. Schliesslich waren vor Beginn der vorliegenden Dissertation schon erste Voroptimierungen der OM2 Parameter für Fluor durch C. Kollmar [95] mittels pp99 erfolgt. Daher standen als Ausgangspunkt für die eigenen Parametrisierungen bereits relativ gute Startsätze zur Verfügung, insbesondere für OM2.

Von zentraler Bedeutung für eine erfolgreiche Parametrisierung ist die Wahl

geeigneter Referenzdaten und Validierungsdaten. Auch hier konnte auf den Vorarbeiten von M. Kolb (OM1) und C. Kollmar (OM2) aufgebaut werden. Der bei der jetzigen Parametrisierungen eingesetzte Referenzdatensatz wurde ursprünglich von M. Kolb zusammengestellt [64] und im Laufe der Zeit je nach Bedarf vervollständigt. Er umfasst 70 neutrale geschlossenschalige Fluorverbindungen, welche die verschiedenen Bindungstypen des Fluors abdecken, die in Molekülen mit den Elementen H, C, N, O und F auftreten können. Als Referenzdaten dienten experimentelle Werte für die wichtigsten Moleküleigenschaften, wobei anfangs hauptsächlich Standardbildungsenthalpien, Bindungslängen und Bindungswinkel (jeweils mit hoher Gewichtung) sowie Dipolmomente und Ionisierungspotentiale (jeweils mit niedriger Gewichtung) Verwendung fanden. Im Laufe der Parametrisierung wurden die Referenzdaten mehrfach erweitert (z.B. durch Hinzunahme von einigen Schwingungswellenzahlen) und modifiziert (z.B. durch andere Gewichtung von Problemfällen), um die Feinjustierung der Parameter voranzutreiben. Zur Bewertung von optimierten Parametersätzen wurde zusätzlich ein Validierungssatz mit insgesamt 23 Radikalen und Ionen herangezogen, der ebenfalls auf M. Kolb zurückgeht [64]. Die beiden anderen Validierungssätze mit *ab initio* Referenzdaten (siehe unten) wurden erst relativ spät im Laufe dieser Arbeit erzeugt und dienten daher nur zur Einschätzung der letzten Parametersätze.

Angesichts der Verfügbarkeit guter Startparameter war der Suchraum bei den durchgeführten Parametrisierungen relativ eingeschränkt. Aufgrund der Erfahrungen bei der testweisen Reparametrisierung von MNDO (siehe Kap. 3.6.1) wurden daher mit dem Programm `gps1` vornehmlich systematische Suchen in Kombination mit partiellen gradientenbasierten SSQ-Minimierungen durchgeführt. Dabei wurde jeweils ein Bereich von  $\pm 10\%$  um die Startwerte herum abgesucht, typischerweise unter Verwendung folgender Optionen (siehe Kapitel 3.6.1): `NValues = 11`, `NProz = 40`, `n = 1` oder `n = 10`. Die Konvergenz dieser systematischen Suchen war generell zufrieden stellend (ähnlich wie in Kapitel 3.6.1). Da die Rechenzeiten für derartige Parametrisierungen mit dem gewählten Referenzdatensatz nicht allzu hoch waren (etwa 10 sec pro SSQ-Berechnung, vgl. Kapitel 3.6.1), wurden eine Reihe solcher Optimierungen mit unterschiedlichen Startparametern und unterschiedlichen Gewichtungen der Referenzdaten sowie auch mit unterschiedlichen Optionen durchgeführt, um eine Serie von Parametersätzen zu erzeugen, die dann anhand der erhaltenen SSQ-Werte und separater Validierungsrechnungen bewertet wurden. Erwartungsgemäss unterschieden sich die so optimierten Parametersätze oft nur graduell. Bei der endgültigen Auswahl wurde darauf geachtet, dass die Parameterwerte physikalisch sinnvoll sind und dass erkennbare Trends in der Reihe C-N-O (sofern möglich) konsistent weitergeführt werden.

Genetische Algorithmen eignen sich nach den vorangehenden Erfahrungen für derartige Feinoptimierungen weniger gut. Sie wurden daher bei der Fluorparametrisierung vornehmlich dazu benutzt, ausgehend von denselben Startwerten wie bei

der systematischen Suche wesentlich grössere Bereiche der Parameter-Hyperfläche zu durchmustern. Dabei waren Variation bis zu  $\pm 50\%$  um die Startwerte herum erlaubt. Trotz ausgedehnter Versuche konnte bei diesen stochastischen Optimierungen mit Hilfe genetischer Algorithmen keine konkurrenzfähige Alternative gefunden werden. Diese Ergebnisse stützen die Annahme, dass die systematische Suche im 'richtigen' Bereich der Parameter-Hyperfläche erfolgt ist und dass es deswegen schwierig sein dürfte, wesentlich bessere als die jetzigen Fluorparameter für OM2 und OM3 zu finden.

### 3.7.3 Optimierte Parameterwerte

Die endgültigen Fluorparameter für OM2 sowie OM3 sind in Tabelle 3.21 wiedergegeben. Man sieht auf den ersten Blick, dass einige der Parameter in OM2 und OM3 sehr ähnliche Werte annehmen. Dies gilt beispielsweise für die Einzentrenterme  $U_{ss}$  und  $U_{pp}$ , für die Parameter in den Resonanzintegralen und für die drei ersten Parameter in den effektiven Rumpfpotentialen. Dies ist nicht allzu überraschend, da die OM2 und OM3 Formalismen ähnlich sind.

Die Skalierungsfaktoren  $\zeta$  für die Basisorbitale unterscheiden sich zwar deutlich, sind aber konsistent mit den Trends, die man bei den früher bestimmten OM2 [63] und OM3 [68] Werten in der Reihe C-N-O findet. Die  $U_{ss}$  und  $U_{pp}$  Parameter haben in MNDO, AM1, PM3, OM1, OM2 und OM3 die gleiche physikalische Bedeutung, und es ist daher nicht überraschend, dass die jetzt gefundenen Werte nahe bei jenen in den anderen Verfahren liegen (insbesondere  $U_{pp}$ ).

Die einzelnen Parameter in den Resonanzintegralen (3.18) haben für sich genommen wenig Bedeutung, wichtig ist der resultierende funktionale Verlauf für die verschiedenen Wechselwirkungen. Die Abstandsabhängigkeit der Resonanzintegrale ist für alle relevanten Wechselwirkungen der Atompaare F-F, F-C und F-H in den Abbildungen 3.9, 3.10 und 3.11 aufgetragen (jeweils OM2 und OM3). Bei den Berechnungen gemäss Gleichung 3.18 wurden die entsprechenden optimierten Parameterwerte aus Tabelle 3.21 verwendet (im Falle von F-H die spezifischen Bindungsparameter und zu Vergleichszwecken auch die generischen Parameter); dabei wurden jeweils die Absolutwerte der Resonanzintegrale bestimmt, auch wenn bindende Wechselwirkungen (positive Überlappungen) vorliegen. Bei allen Abbildungen fällt sofort auf, dass die einzelnen Kurven für OM2 und OM3 einander sehr ähnlich sind. Die Maxima der einzelnen Funktionen liegen unter den üblichen Gleichgewichtsabständen, so dass im Bindungsbereich (ab ca. 1.1 Å) ein exponentiell abklingendes Verhalten beobachtet wird. Bei den homogenen Kombinationen fallen die s-s-Wechselwirkungen am schnellsten ab, gefolgt von den  $\pi-\pi$ -Wechselwirkungen. Die  $\sigma-\sigma$ -Wechselwirkungen haben die grösste Reichweite. Bemerkenswert ist bei der Abbildung 3.11, dass die Resonanzintegrale mit den generischen und spezifischen Parametern sich deutlich unterscheiden und dass im zweiten Fall die s-s- und s- $\sigma$ -Wechselwirkungen bei OM3 eine andere Abfolge haben als bei OM2 (auch wenn die absoluten Unterschiede klein sind).

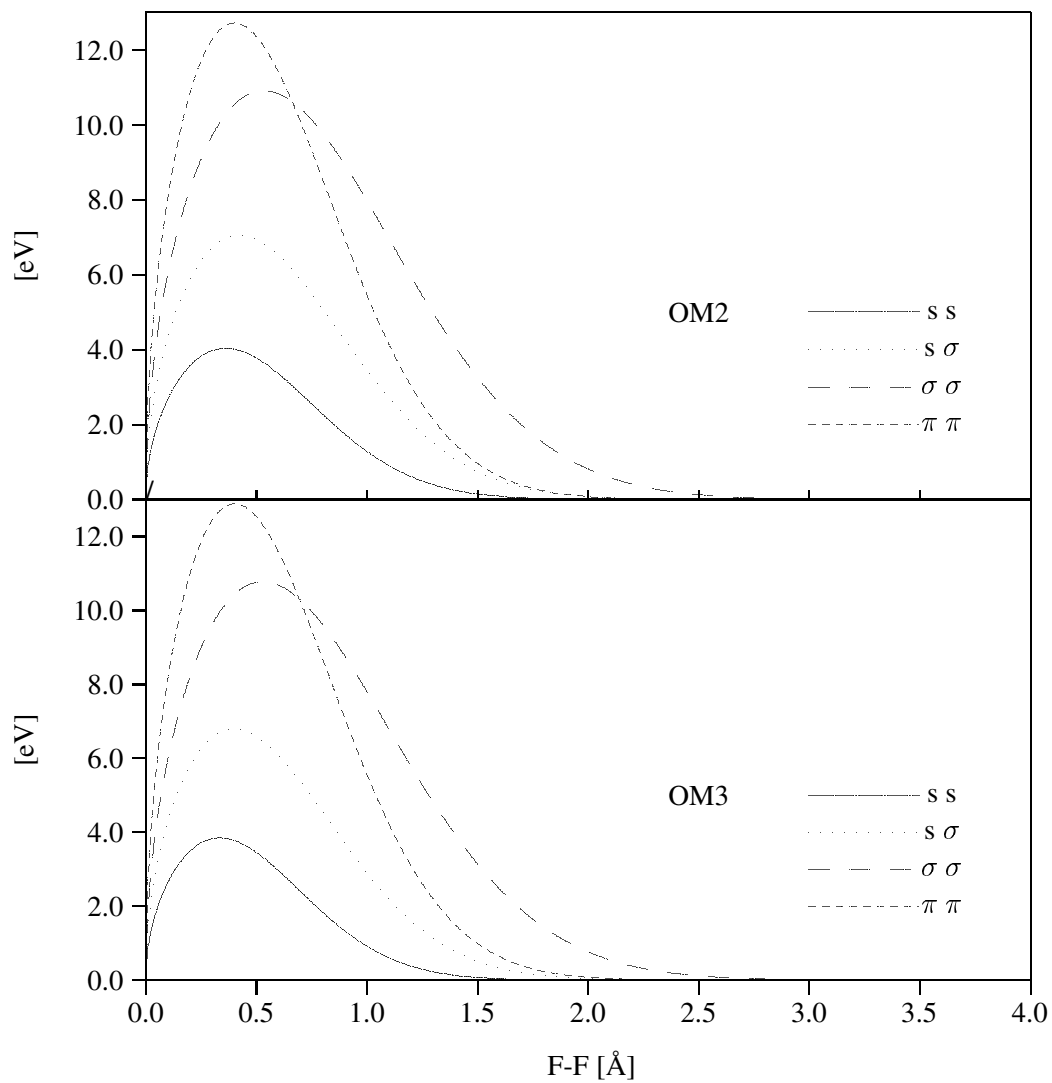


**Tabelle 3.21:** Fluorparameter für OM2 und OM3

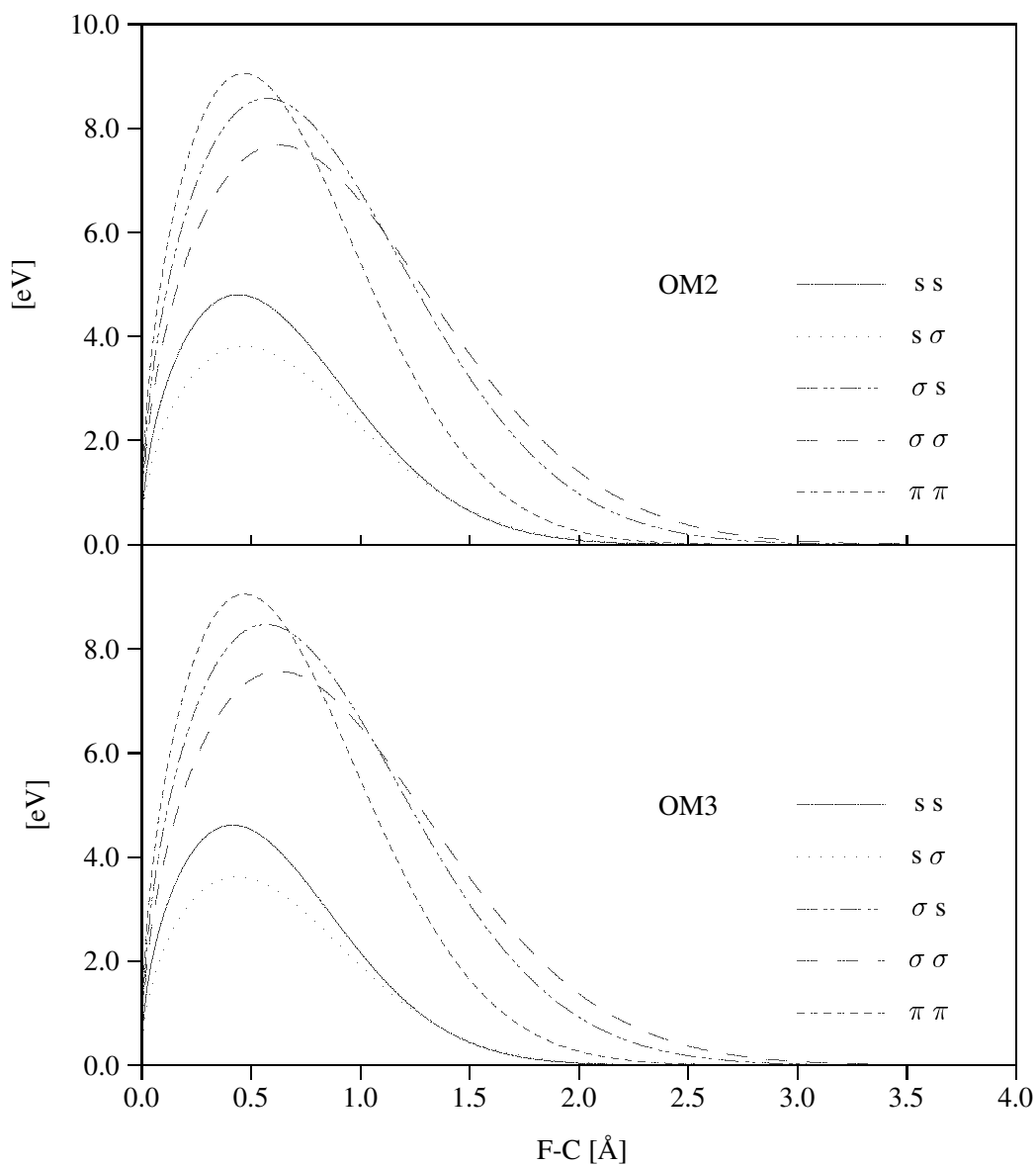
	OM2	OM3
Skalierungsfaktor		
$\zeta$ [a.u.]	1.45216726	1.20564838
Einzentrenenergien		
$U_{ss}$ [eV]	-120.62785370	-120.65477058
$U_{pp}$ [eV]	-107.27105397	-107.50304443
Resonanzintegrale		
$\beta_s$ [eV bohr <sup>-1/2</sup> ]	-6.25438426	-6.19918959
$\beta_p$ [eV bohr <sup>-1/2</sup> ]	-13.93492471	-13.82075048
$\beta_\pi$ [eV bohr <sup>-1/2</sup> ]	-18.73205761	-18.96710976
$\alpha_s$ [a.u.]	0.26624434	0.31128286
$\alpha_p$ [a.u.]	0.12261412	0.12448702
$\alpha_\pi$ [a.u.]	0.21684388	0.21596283
$\beta_s(\text{X-H})$ [eV bohr <sup>-1/2</sup> ]	-6.25104378	-8.06286638
$\beta_p(\text{X-H})$ [eV bohr <sup>-1/2</sup> ]	-13.94492971	-13.92380910
$\alpha_s(\text{X-H})$ [eV bohr <sup>-1/2</sup> ]	0.44713918	0.32623663
$\alpha_p(\text{X-H})$ [eV bohr <sup>-1/2</sup> ]	0.15648906	0.15497349
Orthogonalisierung		
$F_1$	2.11499396	1.03522357
$F_2$	1.09156321	0.00000000
$G_1$	0.31704089	0.14035077
$G_2$	0.02140504	0.00000000
Effektive Rumpfpotentiale		
$\zeta_\alpha$ [a.u.]	8.71226515	8.70367110
$F_{\alpha\alpha}$ [eV]	-685.41988599	-685.41999336
$\beta_\alpha$ [eV bohr <sup>-1/2</sup> ]	-9.17960365	-10.62649349
$\alpha_\alpha$ [a.u.]	0.99971548	0.00010713

Bei den Vorfaktoren für die Orthogonalisierungskorrekturen unterscheiden sich OM2 und OM3 per definitionem dadurch, dass zwei dieser Werte in OM3 gleich Null sind (siehe Tabelle 3.21). Die Vorfaktoren  $F_1$  und  $F_2$  für die Einzentrenterme (3.20) liegen in der Grössenordnung von Eins, so dass die entsprechenden Korrekturen etwa so gross ausfallen wie aufgrund einer Reihenentwicklung zweiter Ordnung für  $\mathbf{S}^{-1/2}$  erwartet [63, 67]. Dies trifft nicht auf die Vorfaktoren in den Korrekturen (3.21) für die Resonanzintegrale zu, die deutlich kleiner als Eins ausfallen, im Einklang mit den Trends der entsprechenden optimierten OM2 und OM3 Parameter in der Reihe C-N-O. Da die Korrekturen (3.21) hier offenbar weniger wichtig sind, sollte sich Fluor bereits auf OM1 Niveau gut beschreiben lassen.

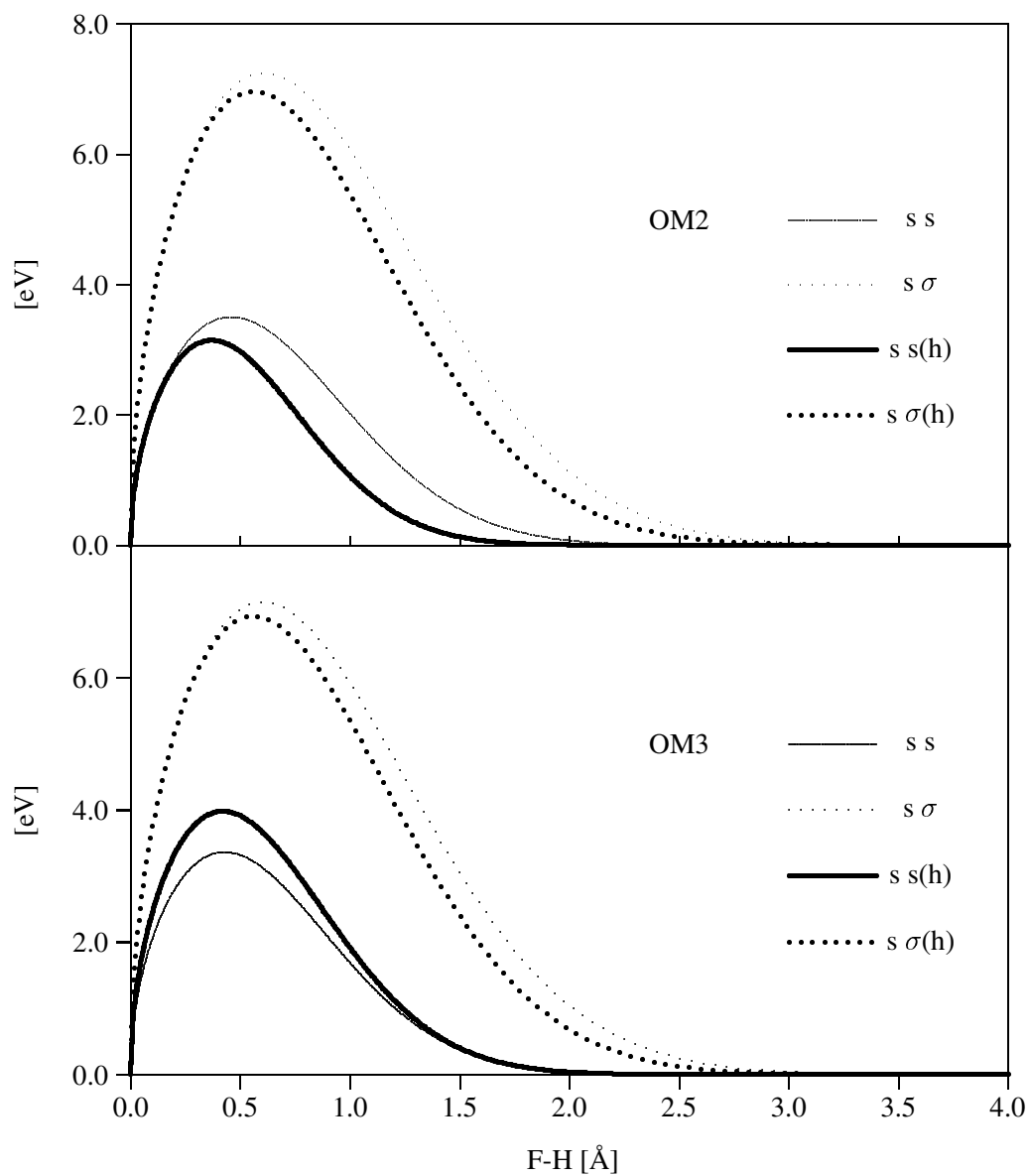
Bei den effektiven Rumpfpotentialen liegt der optimierte Wert für den Exponenten  $\zeta_\alpha$  des verwendeten STO-3G-Rumpforbitals sehr nahe an dem etablierten Slater-Wert (8.70 a.u.) und jener für die Rumpfenergie  $F_{\alpha\alpha}$  sehr nahe an dem experimentellen Wert (-685.4 eV), der sich aus dem 1s Ionisierungspotential ergibt. Es sollte allerdings betont werden, dass bei der Parametrisierung hierauf auch geachtet wurde. Die beiden anderen Parameter in den effektiven Rumpfpotentialen haben keine derart unmittelbare physikalische Bedeutung, so dass es wiederum angebracht erscheint, für eine genauere Einschätzung der optimierten Parameter die Potentiale (3.23) graphisch darzustellen. Die Abstandsverläufe dieser semiempirischen Potentiale sind in den Abbildungen 3.12 und 3.13 aufgetragen, wobei sich die erste Abbildung auf OM2 und die zweite auf OM3 bezieht. Zusätzlich sind in diesen Abbildungen jeweils gestrichelt die entsprechenden *ab initio* Kurven gezeigt, die für die verwendeten (skalierten) ECP-3G-Basisfunktionen mit einem etablierten effektiven Rumpfpotential [96, 97] analytisch berechnet wurden. Alle relevanten Potentiale (*ss*-, *sp*- und *pp*-Ladungsverteilungen) sind für die Atompaare F-F, F-C und F-H gezeigt. Man erkennt generell eine sehr gute Übereinstimmung zwischen den analytisch und semiempirisch berechneten Rumpfpotentialen, sowohl hinsichtlich des Kurvenverlaufs als auch hinsichtlich der absoluten Grösse der Potentiale. Die Potentiale sind innerhalb des Bindungsbereichs erwartungsgemäss stark repulsiv (bei OM3 etwas stärker als bei OM2), und sie fallen bei grösseren Abständen schnell gegen Null ab. Die gute Übereinstimmung mit den analytischen *ab initio* Potentialen deutet darauf hin, dass die Parametrisierung zumindest bei den Rumpfpotentialen physikalisch sinnvolle Ergebnisse geliefert hat.



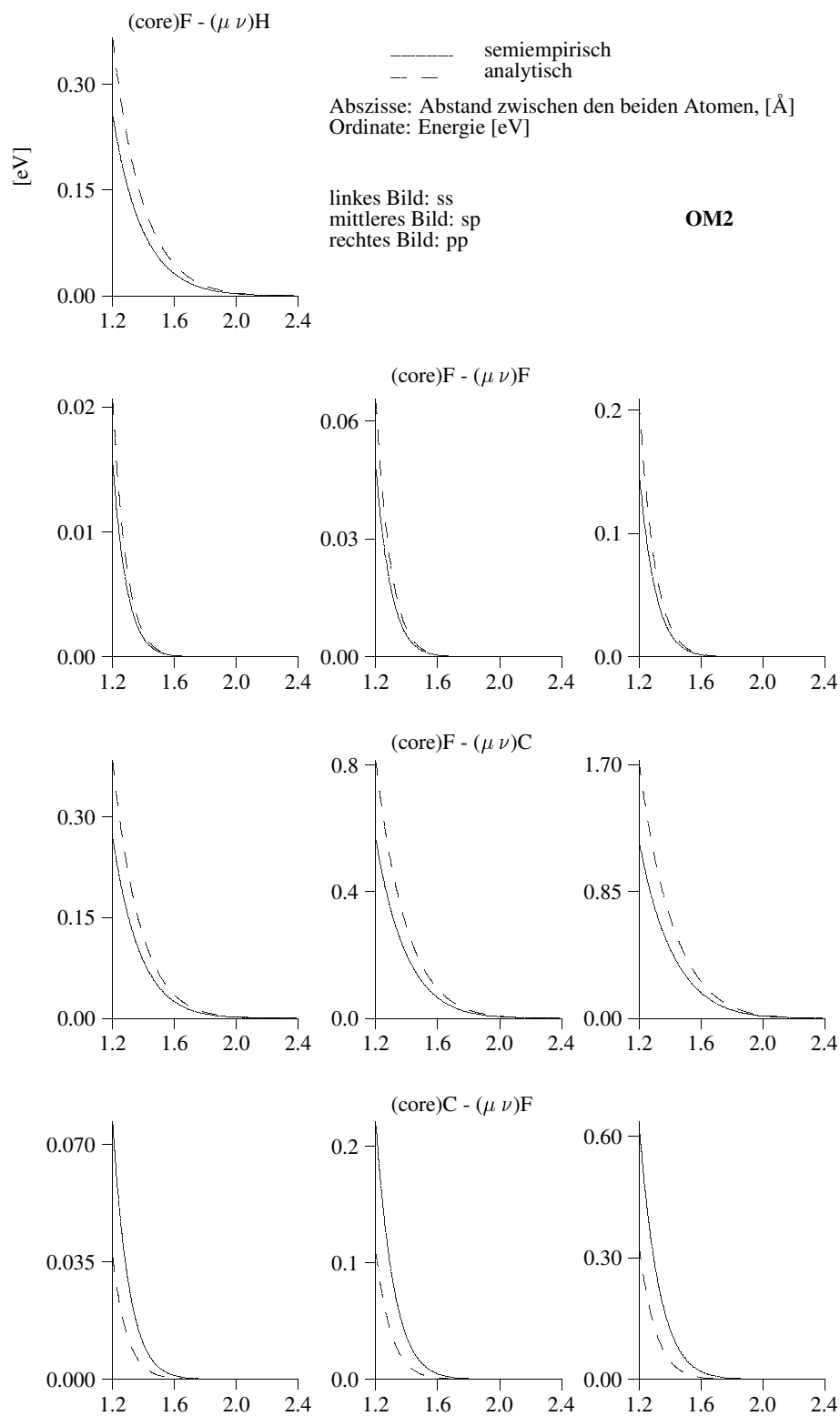
**Abbildung 3.9:** Semiempirische Resonanzintegrale berechnet mit Gleichung 3.18 und den Parameterwerten aus Tabelle 3.21. Dargestellt ist die Wechselwirkung zwischen Fluor und Fluor. Oben für OM2, unten für OM3



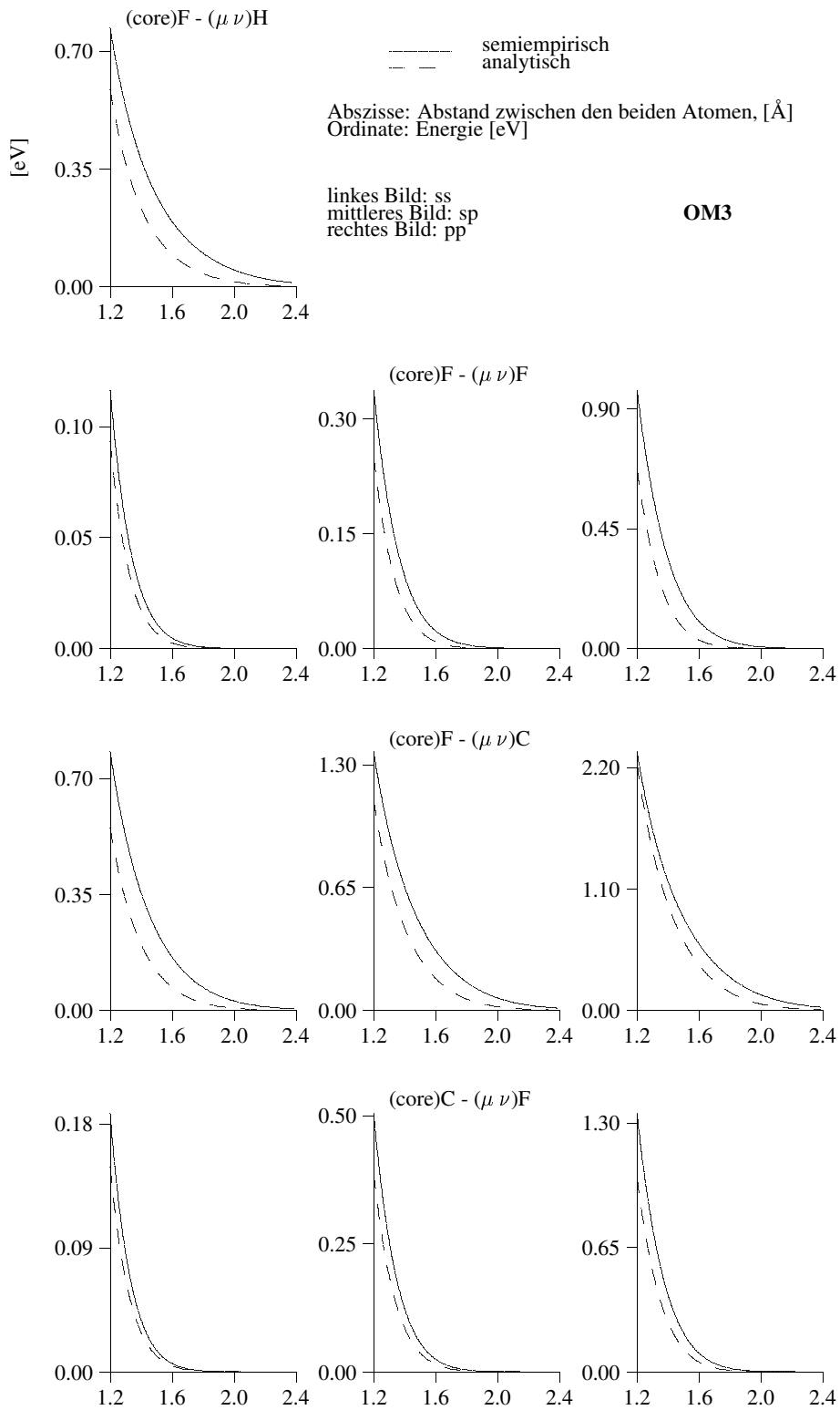
**Abbildung 3.10:** Semiempirische Resonanzintegrale berechnet mit Gleichung 3.18 und den Parameterwerten aus Tabelle 3.21. Dargestellt ist die Wechselwirkung zwischen Fluor und Kohlenstoff. Oben für OM2, unten für OM3. Die notwendigen Parameter für Kohlenstoff stammen von Weber [63] resp. Scholten [68].



**Abbildung 3.11:** Semiempirische Resonanzintegrale berechnet mit Gleichung 3.18 und den Parameterwerten aus Tabelle 3.21. Dargestellt ist die Wechselwirkung zwischen Fluor und Wasserstoff. Oben für OM2, unten für OM3. Fett sind die Resonanzintegrale gezeichnet, welche mit den speziellen F-H Parametern berechnet sind.



**Abbildung 3.12:** Rumpfpotentiale von OM2 für F-H, F-F und F-C. Die durchgezogenen Linien stellen die semiempirischen Potentiale dar. Die gestrichelten Linien geben als Referenz die entsprechenden *ab initio*-Potentiale wieder (siehe Text).



**Abbildung 3.13:** Rumpfpotentiale von OM3 für F-H, F-F und F-C. Die durchgezogenen Linien stellen die semiempirischen Potentiale dar. Die gestrichelten Linien geben als Referenz die entsprechenden *ab initio*-Potentiale wieder (siehe Text).

### 3.7.4 Ergebnisse und Diskussion

Im nachfolgenden werden die Ergebnisse für den Referenzdatensatz und drei Validierungsdatensätze besprochen. Die Moleküle in diesen Datensätzen sind sowohl mit den neuen Methoden OM1, OM2 und OM3 als auch mit den etablierten Verfahren MNDO, AM1 und PM3 berechnet worden. Auf diese Weise kann durch entsprechende Vergleiche die Leistungsfähigkeit der OM $x$ -Methoden beurteilt werden, und zwar nicht nur relativ zueinander, sondern auch relativ zu den semiempirischen Standardverfahren.

Um den Lesefluss nicht zu stören, werden jeweils nur die Statistiken der diskutierten Moleküleigenschaften direkt im Text angegeben. Die den vorgestellten Statistiken zugrunde liegenden Einzelergebnisse sind von Seite 102 bis Seite 134 aufgeführt. Diese Einzelergebnisse sind bei OM1, OM2 und OM3 stets für alle Moleküleigenschaften aufgelistet, bei MNDO, AM1 und PM3 hingegen nur für die Standardbildungsenthalpien.

Der Referenzdatensatz enthält 70 neutrale geschlossenschalige Fluorverbindungen, welche die typischen Bindungsverhältnisse des Fluors abdecken. Der Validierungssatz A umfasst 23 einfache radikalische und ionische Verbindungen. Der Validierungssatz B stammt aus einer systematischen *ab initio* Untersuchung von Bacsikay und Mitarbeitern [98], in der G3-Standardbildungsenthalpien für 127 kleine (H,C,O,F) Moleküle angegeben sind. Unsere Auswertungen zu Satz B konzentrieren sich auf die 91 fluorhaltigen Verbindungen, wobei im Falle der Standardbildungsenthalpien nicht nur mit den G3-Werten, sondern auch mit den verfügbaren (46) experimentellen Werten verglichen wird. In dem Validierungssatz C sind schließlich alle fluorhaltigen Spezies zusammengestellt, die in den etablierten G2 [99] und G3 [100] Sätzen zur Verifizierung von *ab-initio*-Verfahren dienen.

#### 3.7.4.1 Referenzdatensatz: Übersicht

Der Referenzdatensatz setzt sich aus den Molekülen zusammen, die für die Parametrisierung verwendet wurden. Allerdings waren während der Parametrisierung nicht alle Referenzdaten aktiviert, welche im folgenden diskutiert werden. Die Tabelle 3.22 gibt eine Übersicht über die wichtigsten Eigenschaften, die im Referenzdatensatz vertreten sind, und die zugehörigen mittleren absoluten Fehler bei den untersuchten semiempirischen Methoden. Die berechneten Werte zeigen den erwarteten Trend: die OM $x$  Methoden sind gesamthaft gesehen zu bevorzugen. Am deutlichsten ist deren Überlegenheit bei den Standardbildungsenthalpien, welche durchwegs besser berechnet werden als mit den herkömmlichen Methoden. Die Fehler bei OM2 und OM3 liegen bei den Fluorverbindungen allgemein in einem ähnlichen Bereich wie bei den (H,C,N,O) Verbindungen [63, 68], so z.B. für die Standardbildungsenthalpien bei 3–4 kcal/mol, für die Bindungslängen bei etwa 0.02 Å und für die Bindungswinkel bei 2°. Von den drei Standardverfahren schneidet PM3 generell am besten ab.

Ebenso interessant wie der mittlere Fehler ist die Frage, welche Moleküle am schlech-



**Tabelle 3.22:** Mittlere absolute Fehler von Grundzustandseigenschaften für den Referenzdatensatz

	Anzahl	MNDO	AM1	PM3	OM1	OM2	OM3
Stb [kcal/mol] <sup>a</sup>	48	10.81	7.75	5.87	4.88	3.45	3.73
Bindungslänge [Å]	125	0.037	0.028	0.022	0.020	0.023	0.024
Bindungswinkel [°]	69	3.00	3.16	2.75	1.97	2.22	1.78
Ip [eV] <sup>b</sup>	39	0.34	0.53	0.39	0.23	0.26	0.32
Dipolmoment [D]	39	0.38	0.31	0.29	0.27	0.31	0.25
Wellenzahlen [cm <sup>-1</sup> ]	42	225.9	140.8	140.4	167.9	188.9	173.9

<sup>a</sup> Standardbildungsenthalpien<sup>b</sup> Ionisierungspotentiale (Koopmans' Theorem)

testen beschrieben werden und wie gross deren Fehler sind. Eine Antwort auf diese Frage gibt nachfolgende Tabelle 3.23. Die angegebenen maximalen Fehler für die einzelnen Eigenschaften sind positiv (negativ), wenn der berechnete Wert grösser (kleiner) ist als der experimentelle Wert, und die Zahl in der Klammer identifiziert das zugehörige Molekül. Die 36 maximalen Fehler sind über 15 verschiedene Moleküle

**Tabelle 3.23:** Maximale Fehler von Grundzustandseigenschaften für den Referenzdatensatz. Die Zahlen in den Klammern verweisen auf das entsprechende Molekül.

	MNDO	AM1	PM3	OM1	OM2	OM3
Stb [kcal/mol] <sup>a</sup>	61.77(1)	24.41(1)	-21.69(5)	-18.70(6)	-14.02(11)	16.51(14)
Bindungslänge [Å]	-0.212(2)	-0.150(2)	-0.150(2)	-0.090(6)	-0.130(6)	-0.091(5)
Bindungswinkel [°]	10.66(3)	11.91(6)	-8.67(7)	6.19(9)	-6.72(12)	6.712(3)
Ip [eV] <sup>b</sup>	-1.21(4)	-1.94(4)	-1.49(8)	-0.71(2)	-0.93(13)	1.00(4)
Dipolmoment [D]	-1.30(2)	-1.43(2)	-1.55(2)	0.69(10)	0.83(12)	-0.64(15)
Wellenzahlen [cm <sup>-1</sup> ]	799.8(5)	503.1(4)	428.7(5)	567.0(5)	673.3(5)	616.7(5)

<sup>a</sup> Standardbildungsenthalpien<sup>b</sup> Ionisierungspotentiale

- (1) Nitrosyl trifluoride (2) Nitrosyl fluoride (3) Hypofluorous acid (4) Hydrogen fluoride (5) Fluorine (6) Tetrafluorohydrazine gauche (7) Hexafluoroacetone (8) Trifluoroamine (9) Hypofluorous acid (10) Trifluoromethane (11) Difluorocarbene (12) Formyl fluoride (13) Hexafluoroethane (14) Trifluoromethylhypofluorite (15) Trifluoroacetic acid trans

verteilt. Besonders problematisch sind Nitrosylfluorid, dessen Bindungslänge (N-F) und Dipolmoment durch MNDO, AM1 und PM3 stark unterschätzt werden, sowie das Fluormolekül, bei dem die Schwingungsfrequenz durchweg stark überschätzt wird. Auch wenn die maximalen Fehler generell recht gross ausfallen, sind sie bei den OM $x$  Methoden doch insgesamt deutlich kleiner als bei den Standardmethoden (MNDO, AM1, PM3).

### 3.7.4.2 Referenzdatensatz: Standardbildungsenthalpien

Eine detaillierte statistische Auswertung der Standardbildungsenthalpien ist in der Tabelle 3.24 für die OM $x$ -Methoden und die Standardmethoden gegeben. Die zugehörigen Einzelergebnisse sind in den Tabellen 3.41 und 3.42 beigelegt. Offensichtlich werden die Standardbildungsenthalpien mit den OM $x$ -Methoden

**Tabelle 3.24:** Nach Molekülklassen getrennte Fehler der Standardbildungsenthalpien [kcal/mol] in dem Referenzdatensatz

	Standardbildungsenthalpien [kcal/mol]						
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	48	-0.35	4.88	-0.52	3.45	-0.28	3.73
CHF compounds	39	-0.29	3.63	-0.95	3.77	-0.23	3.91
HNOF compounds	9	-0.64	10.32	1.31	2.08	-0.52	2.93
Aromatic Rings	6	-0.47	0.52	0.69	0.74	-0.12	0.23
With XF bonds <sup>a</sup>	13	0.15	9.41	0.86	3.12	-0.44	4.68
		MNDO		AM1		PM3	
All molecules	48	0.65	10.81	-2.23	7.75	1.27	5.87
CHF compounds	39	0.28	9.89	-3.21	7.37	1.69	4.96
HNOF compounds	9	2.25	14.79	2.03	9.42	-0.56	9.81
Aromatic Rings	6	-2.78	5.35	2.90	3.85	6.38	6.71
With XF bonds <sup>a</sup>	13	7.88	17.24	2.18	10.28	0.82	8.69

<sup>a</sup> X=H,N,O,F

deutlich besser berechnet werden als mit MNDO, AM1 oder PM3. Der mittlere absolute Fehler reduziert sich von 10.8 kcal/mol für MNDO auf 3.4 kcal/mol für OM2 resp. 3.7 kcal/mol für OM3. Die systematischen Fehler (MSE-Werte) sind bei den OM $x$ -Methoden generell recht klein (weniger als 1 kcal/mol) und auch bei den anderen Methoden noch akzeptabel: die Stabilität der fluorhaltigen Moleküle wird im Mittel von AM1 um 2.2 kcal/mol überschätzt und von PM3 um 1.3 kcal/mol unterschätzt. Betrachtet man die einzelnen Molekülklassen in Tabelle 3.24, so sind die Fehler bei OM2 und OM3 recht gleichförmig verteilt und insbesondere bei den (H,N,O,F) Verbindungen deutlich niedriger als in den anderen Methoden.

Bei einer individuellen Betrachtung der 48 Moleküle in den Tabellen 3.41 und 3.42 fällt zunächst auf, dass grosse Abweichungen von mehr als 10 kcal/mol bei den Standardmethoden sehr viel häufiger auftreten als bei den OM $x$ -Methoden (MNDO 19, AM1 16, PM3 10, OM1 8, OM2 3, OM3 5). Gemeinsame Problemfälle für OM2 und OM3 sind Difluorcarben und Nitrosyltrifluorid, die jeweils um 11-14 kcal/mol zu stabil berechnet werden. Es wäre denkbar, die Fehler in diesen und anderen Fällen durch eine stärkere Gewichtung noch zu verringern, aber entsprechende Versuche wurden nicht unternommen, da die erreichte Genauigkeit insgesamt zufrieden stellend ist.

Interessant sind auch Stabilitätsänderungen beim Ersatz von Wasserstoff durch

Fluor. Hierzu wurden die in der Tabelle 3.25 aufgelisteten Substitutionen untersucht. Die Änderungen der Standardbildungsenthalpien bei diesen Substitutionen

**Tabelle 3.25:** Substitutionen

H <sub>2</sub> O	$\xrightarrow{1.1}$	HOF	$\xrightarrow{1.2}$	F <sub>2</sub> O
H-C≡C-H	$\xrightarrow{2.1}$	F-C≡C-H	$\xrightarrow{2.2}$	F-C≡C-F
H-C≡N	$\xrightarrow{3.1}$	F-C≡N		
CH <sub>2</sub> O	$\xrightarrow{4.1}$	CHFO	$\xrightarrow{4.2}$	CF <sub>2</sub> O
NH <sub>3</sub>	→	...	$\xrightarrow{5.2}$	NF <sub>3</sub>

sind in in der Tabelle 3.26 dargestellt. Die experimentelle Werte werden von den einzelnen Methoden generell mit der erwarteten Genauigkeit reproduziert, die statistischen Fehler (MSE) liegen etwa in der gleichen Grössenordnung wie bei den zugrunde liegenden Standardbildungsenthalpien (Tabelle 3.24).

**Tabelle 3.26:** Änderungen der Standardbildungsenthalpien [kcal/mol] bei den Substitutionen gemäss Tabelle 3.25

Reaktion	EXP	MNDO	AM1	PM3	OM1	OM2	OM3
1.1	34.30 <sup>a,b</sup>	42.18	36.66	24.17	40.91	36.68	37.08
1.2	29.40 <sup>b,b</sup>	36.83	33.07	24.47	32.02	24.92	27.50
2.1	-29.35 <sup>a,c</sup>	-42.23	-39.58	-32.60	-26.96	-27.37	-30.76
2.2	-23.77 <sup>c,c</sup>	-36.65	-34.81	-29.69	-26.41	-30.36	-27.68
3.1	-23.70 <sup>b,b</sup>	-37.66	-35.37	-26.44	-22.37	-24.16	-26.88
4.1	-64.00 <sup>a,b</sup>	-55.92	-61.44	-54.72	-53.02	-56.03	-56.34
4.2	-55.90 <sup>b,d</sup>	-49.77	-53.33	-52.78	-49.79	-50.14	-50.43
5.2	-18.80 <sup>a,a</sup>	-27.84	-32.74	-21.33	-32.34	-24.24	-24.71
	MUE	2.41	4.47	2.14	-1.73	-0.14	0.05
	MSE	9.78	7.25	5.24	5.78	4.38	4.03

<sup>a</sup> J.B.Pedley, R.D.Naylor, and S.P.Kirby, "Thermochemical Data of Organic Compounds", 2nd ed., Chapman and Hall, London, 1986.

<sup>b</sup> M.W.Chase, C.A.Davies, J.R.Downey, D.R.Frurip, R.A.McDonald, and A.N.Syverud, "JANAF Thermochemical Tables", 3rd edition, J.Phys.Chem.Ref.Data 14, Suppl.1 (1985).

<sup>c</sup> S.Parthiban, J.M.L.Martin, and J.F.Liebman, Mol.Phys. 100, 453 (2002).

<sup>d</sup> N.L.Haworth, M.H.Smith, G.B.Bacskey, and J.C.Mackie, J.Phys.Chem. A 104, 7600 (2000).

Qualitativ sind die Daten in Tabelle 3.26 ein Mass für die unterschiedliche Energie verschiedener Bindungen (O-H vs. O-F in 1.1 und 1.2; C(sp)-H vs. C(sp)-F in 2.1, 2.2 und 3.1; C(sp<sup>2</sup>)-H vs. C(sp<sup>2</sup>)-F in 4.1 und 4.2; N-H vs. N-F in 5.2). Bei den untersuchten Beispielen sind experimentell die O-H, C(sp)-F, C(sp<sup>2</sup>)-F und N-F Bindungen jeweils um etwa 29-34, 24-29, 56-64 und 19/3 kcal/mol stabiler. Diese unterschiedlichen Bindungsstärken werden besonders von OM2 und OM3 offensichtlich gut wiedergegeben.

### 3.7.4.3 Referenzdatensatz: Geometrien

Die Tabellen 3.27 und 3.29 enthalten die Statistiken der Bindungslängen und Bindungswinkel für OM $x$  sowie für MNDO, AM1 und PM3. Die OM $x$ -Werte fundieren auf den Tabellenwerten 3.43; eine ausführliche Tabelle für MNDO, AM1 und PM3 ist nicht aufgelistet.

Generell liegen die mittleren absoluten Fehler der Bindungslängen und Bindungswinkel in einem für die semiempirischen Methoden typischen Bereich, wobei die Neuentwicklungen OM $x$  die Bindungswinkel durchwegs besser berechnen als MNDO, AM1 und PM3. Die Bindungslängen werden hingegen von PM3 ähnlich gut beschrieben wie von den OM $x$ -Methoden, während MNDO und AM1 grössere Fehler aufweisen.

Die Statistik für die **Bindungslängen** in der Tabelle 3.27 beruht auf 125 fluorhaltigen Molekülen. Sie umfasst sowohl X-F-Bindungen ( $X = \text{H, C, N}$  und  $\text{O}$ ) als auch X-Y-Bindungen ohne Beteiligung von Fluor. Die Gesamtstatistik hängt somit von der Qualität der vorgegebenen Parameter für die anderen Elemente ab, sie ist aber trotzdem relevant, wenn man die Güte der berechneten Geometrien insgesamt beurteilen will.

**Tabelle 3.27:** Nach Molekülklassen getrennte Fehler der Bindungslängen in dem Referenzdatensatz

		Bindungslängen [ $\text{\AA}$ ]					
		OM1		OM2		OM3	
	Anzahl	MSE	MUE	MSE	MUE	MSE	MUE
All molecules	125	0.001	0.020	0.006	0.023	0.010	0.024
CHF compounds	104	0.004	0.018	0.009	0.020	0.012	0.021
CHNOF compounds	3	-0.010	0.024	-0.003	0.020	0.014	0.014
HNOF compounds	17	-0.017	0.032	-0.016	0.043	-0.002	0.044
C-F	35	0.019	0.019	0.017	0.019	0.001	0.011
N-F	9	-0.025	0.026	-0.028	0.028	-0.020	0.022
O-F	3	-0.036	0.036	-0.031	0.031	-0.059	0.059
		MNDO		AM1		PM3	
All molecules	125	0.004	0.037	0.012	0.028	0.002	0.022
CHF compounds	104	0.014	0.030	0.018	0.026	0.007	0.020
CHNOF compounds	3	-0.003	0.029	0.012	0.052	-0.002	0.021
HNOF compounds	17	-0.056	0.080	-0.025	0.037	-0.024	0.037
C-F	35	-0.002	0.016	0.020	0.023	0.004	0.015
N-F	9	-0.101	0.101	-0.042	0.042	-0.044	0.044
O-F	3	-0.145	0.145	-0.059	0.059	-0.030	0.030

In der Gesamtstatistik zeigen die OM $x$ -Methoden jeweils ähnliche Fehler (MUE), die von OM1 über OM2 nach OM3 ansteigen (0.0201, 0.0227 und 0.0238  $\text{\AA}$ ). Die Fehler für PM3 sind vergleichbar (0.0218  $\text{\AA}$ ), während jene für MNDO und AM1 höher liegen (0.0366 resp. 0.0279  $\text{\AA}$ ). Wenn man nach Molekülklassen differenziert, werden die organischen (H,C,F) Fluorverbindungen deutlich besser beschrieben als

die anorganischen (H, N, O, F). Für die Bindungen zwischen Fluor und X (X = C, N, O) findet man, dass die C-F-Abstände generell zufrieden stellend reproduziert werden (am besten durch OM3), während die N-F- und speziell die O-F-Abstände systematisch und deutlich unterschätzt werden (vgl. MSE und MUE).

Qualitative Trends in den Bindungslängen werden meist richtig wiedergegeben. Dies ist in Tabelle 3.28 illustriert, in welcher einige C-F-Abstände aus Tabelle 3.43 zusammengestellt sind. Man findet die erwartete Abnahme der C-F-Bindungslänge mit zunehmendem s-Anteil in der Hybridisierung am Kohlenstoff ( $sp^3 \rightarrow sp^2 \rightarrow sp$ ), in guter Übereinstimmung mit dem Experiment.

**Tabelle 3.28:** C-F-Bindungslängen als Funktion der Hybridisierung

Hybridisierung	Molekül	Bindungslänge [Å]			
		exp	OM1	OM2	OM3
$sp^3$	Fluoroethane	1.397	1.394	1.385	1.366
$sp^2$	Fluoroethene	1.351	1.360	1.356	1.336
	1,1-Difluoroethene	1.315	1.340	1.342	1.325
	Tetrafluoroethene	1.319	1.335	1.335	1.319
$sp$	Fluoroacetylene	1.279	1.300	1.296	1.277

Die Statistik für die **Bindungswinkel** in der Tabelle 3.29 basiert wiederum auf den Geometriedaten in Tabelle 3.43. Bei den Bindungswinkeln schneiden die

**Tabelle 3.29:** Nach Molekülklassen getrennte Fehler der Bindungswinkel in dem Referenzdatensatz

		Bindungswinkel [°]					
		OM1		OM2		OM3	
	Anzahl	MSE	MUE	MSE	MUE	MSE	MUE
All molecules	69	-0.22	1.97	-0.17	2.22	-0.16	1.78
CHF compounds	56	-0.33	1.88	-0.25	2.06	-0.16	1.61
CHNOF compounds	3	1.30	1.30	1.16	2.46	0.82	1.75
HNOF compounds	9	0.52	2.44	0.36	2.91	-0.12	2.68
FXY	41	-0.49	2.12	-0.40	2.51	-0.23	1.84
		MNDO		AM1		PM3	
All molecules	69	0.98	3.00	0.72	3.16	0.76	2.75
CHF compounds	56	0.27	2.59	0.03	2.93	0.36	2.69
CHNOF compounds	3	1.77	2.89	3.76	3.76	3.41	3.41
HNOF compounds	9	4.91	5.56	3.97	4.59	2.31	3.05
FXY	41	1.97	3.15	1.50	3.52	0.95	2.90

OM $x$ -Methoden durchwegs besser ab als die Standardmethoden. In der Gesamtstatistik sind die mittleren absoluten Fehler generell um etwa 1° niedriger (OM1 2.0°, OM2 2.2°, OM3 1.8°, verglichen mit MNDO 3.1°, AM1 3.2° und PM3 2.7°). Ein ähnliches Bild ergibt sich bei den FXY Winkeln mit endständigem Fluor

(OM $x$ -Methoden 1.8-2.5°, Standardmethoden 2.9-3.5°). Die systematischen Fehler (MSE) bei den FXY Winkeln sind in den OM $x$ -Methoden allgemein recht klein (leichte Unterschätzung um 0.2-0.5°) und deutlich grösser bei den Standardmethoden (Überschätzung um 1.0-2.0°). Insgesamt sind bei den Bindungswinkeln die OM3 Ergebnisse am besten.

#### 3.7.4.4 Referenzdatensatz: Ionisierungspotentiale

Die hier vorgestellten Ionisierungspotentiale wurden nach Koopmans' Theorem als negative Orbitalenergien berechnet. Die statistischen Auswertungen finden sich in den Tabellen 3.30 und 3.31). Die zugehörigen individuellen Daten sind in den Tabellen 3.44 sowie 3.45 aufgelistet (nur für die OM $x$ -Methoden).

Bei der Einschätzung der Ergebnisse sind zwei Einschränkungen zu beachten. Einerseits stellt Koopmans' Theorem eine Näherung dar, welche Relaxationseffekte und Korrelationseffekte bei der Ionisierung vernachlässigt, so dass eine genauere Berechnung leicht verschiedene Ionisierungspotentiale liefern würde. Andererseits sind die gewählten Referenzdaten relativ heterogen: bei den ersten Ionisierungspotentialen (Tabelle 3.44) sind neben fluorbasierten Orbitalen auch andere Orbitaltypen vertreten (z.B.  $\pi$ -Orbitale), deren Energien nur wenig von den Fluorparametern abhängen, und bei den höheren Ionisierungspotentialen (Tabelle 3.45) finden sich neben einsamen Elektronenpaaren am Fluor auch andere tiefliegende Orbitale, bei denen die Population am Fluor nicht immer dominiert.

Trotz dieser Einschränkungen darf man festhalten, dass die OM $x$ -Methoden die ersten Ionisierungspotentiale von Fluorverbindungen mit einer brauchbaren Genauigkeit berechnen (MUE 0.23-0.32 eV), ähnlich gut wie bei organischen Verbindungen (H, C, N, O), während die Standardmethoden etwas höhere Fehler aufweisen ((MUE 0.34-0.53 eV). Bei den höheren Ionisierungspotentialen lassen sich in der Statistik keine klaren Trends erkennen. Dies liegt daran, dass die individuellen Fehler (Tabelle 3.45) im Bereich von 10-20 eV akzeptabel klein sind und dann jenseits von 20 eV bei gewissen 2s-Ionisierungen sehr gross werden, so dass eine Statistik hier im Grunde genommen wenig Sinn macht.

#### 3.7.4.5 Referenzdatensatz: Dipolmomente

Das Dipolmoment ist eine experimentell leicht zugängliche Grösse, welche unmittelbar von der Ladungsverteilung abhängt und sich daher gut zur Bewertung der berechneten Elektronendichteverteilungen eignen sollte. Die hier vorgestellten Dipolmomente wurden über die ZDO-Näherung berechnet und enthalten damit Beiträge von den atomaren Nettoladungen und den atomaren Dipolen (sp-Hybridisierung). Die Statistik 3.32 basiert auf den individuellen Dipolmomenten aus Tabelle 3.46 (nur für die OM $x$ -Methoden angegeben). Die mittleren absoluten Fehler liegen bei den OM $x$ -Methoden (0.25–0.31 D) etwas niedriger als bei den herkömmlichen Methoden (0.29–0.38 D), aber die Unterschiede sind nicht allzu gross. Die systematischen Fehler

**Tabelle 3.30:** Nach Molekülklassen getrennte Fehler der ersten Ionisierungspotentiale [eV] in dem Referenzdatensatz

Ionisierungspotentiale [eV]							
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	39	0.00	0.23	-0.03	0.26	0.18	0.32
CHF compounds	29	0.10	0.20	-0.06	0.25	0.17	0.29
HNOF compounds	9	-0.27	0.30	0.08	0.29	0.27	0.42
Aromatic rings	7	0.39	0.39	0.28	0.28	0.41	0.41
With XF bonds <sup>a</sup>	10	-0.22	0.30	0.11	0.29	0.30	0.43
Methylen-Substituierte Moleküle							
	Anzahl	MNDO		AM1		PM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	39	0.09	0.34	-0.40	0.53	0.00	0.39
CHF compounds	29	0.19	0.33	-0.30	0.47	0.18	0.32
HNOF compounds	9	-0.19	0.38	-0.70	0.70	-0.50	0.57
Aromatic rings	7	0.36	0.36	0.18	0.20	0.49	0.49
With XF bonds <sup>a</sup>	10	-0.11	0.40	-0.63	0.63	-0.39	0.57

<sup>a</sup> X=H,N,O,F**Tabelle 3.31:** Nach Molekülklassen getrennte Fehler der höheren Ionisierungspotentiale [eV] in dem Referenzdatensatz

höhere Ionisierungspotentiale [eV]							
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	26	-0.76	1.51	-0.86	1.83	-0.34	2.09
CHF compounds	19	-1.02	1.79	-1.36	2.10	-0.82	2.41
HNOF compounds	7	-0.05	0.75	0.48	1.07	0.97	1.23
With XF bonds	7	-0.05	0.75	0.48	1.07	0.97	1.23
Methylen-Substituierte Moleküle							
	Anzahl	MNDO		AM1		PM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	26	1.02	1.53	0.82	2.43	-1.94	2.20
CHF compounds	19	1.26	1.85	1.24	2.84	-2.24	2.54
HNOF compounds	7	0.37	0.67	-0.31	1.33	-1.12	1.27
With XF bonds	7	0.37	0.67	-0.31	1.33	-1.12	1.27

**Tabelle 3.32:** Nach Molekülklassen getrennte Fehler der Dipolmomente [D] in dem Referenzdatensatz

Dipolmoment [D]							
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	39	0.10	0.27	0.05	0.31	0.01	0.25
CHF compounds	30	0.16	0.31	0.12	0.33	0.00	0.24
HNOF compounds	8	-0.13	0.15	-0.21	0.26	0.04	0.29
Methode							
	Anzahl	MNDO		AM1		PM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	39	0.09	0.38	-0.09	0.31	-0.09	0.29
CHF compounds	30	0.18	0.41	-0.05	0.29	-0.06	0.25
HNOF compounds	8	-0.20	0.29	-0.22	0.44	-0.22	0.47

(MSE) sind bei OM3 sehr klein und bewegen sich ansonsten im Bereich von 0.05–0.10 D. Klare Trends bei den einzelnen Molekülklassen sind nicht zu erkennen, daher wird auf eine detailliertere statistische Aufschlüsselung verzichtet. Bei den insgesamt 39 Referenzmolekülen treten Fehler von mehr als 0.5 D bei OM3 (2) seltener auf als bei OM1 (5) oder OM2 (8), im Einklang mit dem statistischen Befund, dass OM3 die besten Dipolmomente liefert.

### 3.7.4.6 Referenzdatensatz: Schwingungen

Die Statistik 3.33 beruht auf den individuellen Wellenzahlen für die Grundschnwingungen der Referenzmoleküle, die in Tabelle 3.47 angegeben sind (nur für die OM $x$ -Methoden). Die Schwingungsanalyse erfolgte auf harmonischem Niveau, unter Ver-

**Tabelle 3.33:** Nach Molekülklassen getrennte Fehler der Schwingungswellenzahlen [ $\text{cm}^{-1}$ ] in dem Referenzdatensatz

Wellenzahlen [ $\text{cm}^{-1}$ ]							
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	42	76.9	163.6	85.7	190.0	76.6	164.4
CHF compounds	33	40.5	139.0	46.0	158.3	40.7	141.1
HNOF compounds	9	210.5	253.8	231.3	305.1	208.1	249.7
Methode							
	Anzahl	MNDO		AM1		PM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	42	220.0	225.9	98.1	140.8	112.4	140.4
CHF compounds	33	150.2	157.7	61.5	111.2	98.3	128.2
HNOF compounds	9	476.1	476.1	232.1	249.1	164.2	185.3

nachlässigung der Anharmonizitäten. Man erwartet daher, dass die berechneten harmonischen Frequenzen im allgemeinen grösser sind als die gemessenen anharmonischen Frequenzen. Ein solcher systematischer Fehler wird in der Tat bei allen untersuchten Methoden beobachtet (positive MSE-Werte in Tabelle 3.33). Die mittleren absoluten Fehler (MUE) liegen bei den OM $x$ -Methoden im Bereich von



164-190  $\text{cm}^{-1}$ , niedriger als bei MNDO (226  $\text{cm}^{-1}$ ), aber höher als bei AM1 und PM3 (ca. 140  $\text{cm}^{-1}$ ).

Betrachtet man die Einzelergebnisse in Tabelle 3.47 etwas genauer, so stellt man fest, dass die OM $x$ -Methoden die Wellenzahlen der X-F-Streckschwingungen (X = N, O, F) stark überschätzen, so z.B. bei Trifluoramin um 281-376  $\text{cm}^{-1}$ , bei Difluoroxid um 396-527  $\text{cm}^{-1}$  und beim Fluormolekül sogar um 567-673  $\text{cm}^{-1}$ . Dies hängt natürlich mit der starken Unterschätzung der entsprechenden Bindungslängen zusammen und ist ein systematischer Fehler, der auch bei den Standardverfahren auftritt (bei MNDO noch ausgeprägter, bei PM3 weniger stark).

Anhand der in Tabelle 3.47 ausgedruckten Nullen erkennt man einen weiteren qualitativen Fehler: Difluoracetylen ist für OM1, OM2 und OM3 in der angenommenen linearen Struktur kein Minimum, das gleiche gilt bei Tetrafluorethylen in der  $D_{2h}$  Struktur für OM1 und OM2. Uneingeschränkte Optimierungen zeigen in diesen Fällen Symmetrierniedrigungen jeweils zu  $C_{2h}$  (mit FCCF Diederwinkeln von 33-43° in Tetrafluorethylen und CCF Winkeln von 142-155° in Difluoracetylen). Diese Verzerrungen sind konsistent mit der Überschätzung der Stabilität von Mono- und Difluorcarben (siehe Tabelle 3.41). Auch wenn die entsprechenden energetischen Unterschiede gering sind, so signalisieren sie doch einen unangenehmen qualitativen Fehler, der bei den Standardverfahren MNDO, AM1 und PM3 nicht vorkommt. Dies trägt natürlich dazu bei, dass die oben diskutierten statistischen Fehler ungewohnt sind: im Gegensatz zu den Fluorverbindungen sind bei organischen Verbindungen (H, C, H, O) die OM $x$ -Methoden den Standardmethoden auch bei den Schwingungswellenzahlen überlegen.

#### 3.7.4.7 Validierungssatz A

Der erste Validierungssatz umfasst fluorhaltige Radikale, Kationen und Anionen. Keine dieser Spezies war im Referenzdatensatz enthalten. Die Validierung geht der Frage nach, wie gut die Standardbildungsenthalpien dieser radikalischen und ionischen Verbindungen durch die verschiedenen semiempirischen Methoden wiedergegeben werden, deren Parameter anhand geschlossenschaliger neutraler Moleküle optimiert worden sind.

Die statistische Auswertung ist in Tabelle 3.34 zusammengefasst, die zugrunde liegenden Standardbildungsenthalpien sind in den Tabellen 3.48 und 3.49 aufgeführt. Man erkennt sofort, dass die mittleren absoluten Fehler durchweg etwas höher liegen als bei dem Referenzdatensatz (vgl. Table 3.24), aber zumindest im Falle der OM $x$ -Methoden verbleiben sie in der gleichen Größenordnung. Die niedrigsten Fehler findet man für OM1 und OM3 (etwa 5.5 kcal/mol), gefolgt von OM2 und PM3 (knapp 8 kcal/mol), während MNDO und AM1 weniger gut abschneiden (14-17 kcal/mol). Dies liegt zum grossen Teil an systematischen Fehlern (MSE): die Radikale und Ionen in Satz A werden mit allen verwendeten Methoden im Mittel zu stabil berechnet, aber dieser Trend ist bei MNDO und AM1 besonders ausgeprägt.

Die maximalen Fehler der einzelnen Methoden sind in Tabelle 3.35 zusammen-

**Tabelle 3.34:** Nach Molekülklassen getrennte Fehler der Standardbildungsenthalpien [kcal/mol] für Radikale und Ionen

Standardbildungsenthalpien [kcal/mol]							
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	23	-2.23	5.49	-4.59	7.92	-2.94	5.38
CHF compounds	18	-4.59	5.10	-7.56	8.43	-4.90	5.73
CHNOF compounds	5	6.26	6.88	6.09	6.09	4.13	4.13
	Anzahl	MNDO		AM1		PM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	23	-9.92	14.23	-16.40	16.97	-2.34	7.76
CHF compounds	18	-9.39	12.55	-18.91	18.98	-2.55	8.18
CHNOF compounds	5	-11.82	20.27	-7.34	9.75	-1.59	6.28

gefasst. Es ist offensichtlich, dass die maximalen Fehler generell sehr gross sind.

**Tabelle 3.35:** Maximale Fehler der Standardbildungsenthalpien [kcal/mol] in Validierungssatz A

Methode	Molekül	Exp	Berechnet	Fehler
MNDO	CHF <sub>2</sub>	-57.0	-88.3	31.3
AM1	CH <sub>2</sub> F <sub>2</sub> <sup>+</sup>	185.0	151.6	33.4
PM3	CF <sub>3</sub> <sup>-</sup>	-154.4	-178.8	24.4
OM1	CF <sub>3</sub> <sup>-</sup>	-154.4	-168.9	14.5
OM2	CF <sub>3</sub> <sup>-</sup>	-154.4	-177.5	23.1
OM3	CF <sub>3</sub> <sup>-</sup>	-154.4	-173.8	19.4

Sie fallen bei OM1 und OM3 etwas kleiner aus als bei den anderen Methoden. Besonders problematisch ist offenbar das Anion CF<sub>3</sub><sup>-</sup>, was nicht verwundert, da kleine Anionen mit semiempirischen Methoden generell schwer zu beschreiben sind.

### 3.7.4.8 Validierungssatz B

Der Validierungssatz B entstammt einer *ab initio* Studie [98] auf G3-Niveau [100], welche sich hauptsächlich mit der Stabilität von Fluorkohlenwasserstoffen befasst. Insgesamt werden dort die G3-Standardbildungsenthalpien für insgesamt 127 kleine (H,C,O,F) Moleküle angegeben, davon enthalten 91 Fluor. Die Inputdatei für Satz B enthält die in der G3-Studie optimierten MP2/6-31G\* Geometrien; die entsprechenden kartesischen Koordinaten wurden von G.B. Bacskay<sup>1</sup> erhalten. Aus diesen Koordinaten wurden die relevanten geometrischen Referenzdaten (Bindungslängen und Bindungswinkel) durch ein hierfür geschriebenes Hilfsprogramm automatisch

<sup>1</sup>G.B. Bacskay, School of Chemistry, University of Sydney, NSW 2006 Australia. Email: bacskay@chem.usyd.edu.au

erzeugt, während die energetischen Referenzdaten (G3-Standardbildungsenthalpien und verfügbare experimentelle Standardbildungsenthalpien) manuell aus der Publikation [98] in die Inputdatei übertragen wurden. Der Einfachheit halber sind alle Moleküle in der Inputdatei und in den betreffenden Tabellen mit Nummern versehen.

Für die 91 fluorhaltigen Verbindungen sind die berechneten **Standardbildungsenthalpien** gemeinsam mit den G3-Referenzwerten [98] in den Tabellen 3.50 und 3.51 zusammengestellt. Die zugehörige Statistik findet sich in Tabelle 3.36. Einige der

**Tabelle 3.36:** Nach Molekülklassen getrennte Fehler der Standardbildungsenthalpien [kcal/mol] im Validierungssatz B, bezogen auf G3-Referenzwerte

		Standardbildungsenthalpien [kcal/mol]					
		OM1		OM2		OM3	
	Anzahl	MSE	MUE	MSE	MUE	MSE	MUE
All molecules	91	-3.03	6.98	-4.56	7.14	-2.41	7.15
singlets	51	-0.03	5.73	-1.82	5.54	0.69	5.72
doublets	40	-6.86	8.58	-8.05	9.18	-6.35	8.97
		MNDO		AM1		PM3	
All molecules	91	-6.49	11.13	-9.73	11.01	-4.08	7.76
singlets	51	-0.29	7.93	-4.83	6.72	0.90	5.17
doublets	40	-14.40	15.21	-15.98	16.47	-10.42	11.07

Moleküle im Validierungssatz B kommen auch im Referenzdatensatz vor, aber mehrheitlich handelt es sich um neue Moleküle, darunter viele reaktive Spezies und instabile Intermediate mit ungewöhnlicher Elektronenstruktur, so dass die Vergleiche einen anspruchsvollen Test für die Zuverlässigkeit der semiempirischen Parametrisierungen darstellen. Die G3-Standardbildungsenthalpien sollten auf etwa 2 kcal/mol genau sein und können daher anstelle experimenteller Referenzdaten verwendet werden.

In der Gesamtstatistik liegen die mittleren absoluten Fehler für die OM $x$ -Methoden bei ca. 7 kcal/mol, für PM3 bei knapp 8 kcal/mol und für AM1 sowie MNDO bei ca. 11 kcal/mol. Aufschlussreicher sind die Vergleiche für die Untergruppen der 51 Singulets (abgeschlossenschalige Neutralmoleküle) und 40 Dubletts (Radikale) in Satz B. Bei den Singulets sind die Qualitätsunterschiede zwischen den einzelnen Methoden relativ gering (MUE-Werte: 5-6 kcal/mol für die OM $x$ -Methoden und 5-8 kcal/mol für die Standardverfahren), und die systematischen Fehler sind relativ klein (MSE-Werte: betragsmässig kleiner als 2 kcal/mol, mit Ausnahme von AM1). Dagegen sind bei den Dubletts die OM $x$ -Methoden merklich überlegen (MUE-Werte von 8-9 kcal/mol gegenüber 11-16 kcal/mol für die Standardverfahren). Die Stabilität der Dubletts wird generell überschätzt, jedoch in den OM $x$ -Methoden weniger stark als in den Standardverfahren (6-8 vs. 10-16 kcal/mol). Insgesamt darf man festhalten, dass die OM $x$ -Methoden bei diesen Vergleichen etwa gleichwertig sind und dass ihre Überlegenheit gegenüber den Standardverfahren hauptsächlich durch die bessere Behandlung der Radikale bedingt ist.

Für 46 der 91 fluorhaltigen Verbindungen im Validierungssatz B sind auch die experimentellen Standardbildungsenthalpien bekannt [98]. Die statistische Auswertung bezüglich dieser experimentellen Daten ist in Tabelle 3.37 gegeben. Die Einzelergebnisse sind nicht gesondert abgedruckt, da die berechneten Werte für alle Moleküle bereits in den Tabellen 3.50 und 3.51 aufgelistet sind, während die verfügbaren experimentellen Werte [98] in Tabelle 3.50 angegeben sind.

**Tabelle 3.37:** Nach Molekülklassen getrennte Fehler der Standardbildungsenthalpien [kcal/mol] im Validierungssatz B, bezogen auf experimentelle Referenzdaten

		Standardbildungsenthalpien [kcal/mol]					
		OM1		OM2		OM3	
	Anzahl	MSE	MUE	MSE	MUE	MSE	MUE
All molecules	46	-3.03	4.89	-4.89	5.78	-2.58	5.61
singlets	27	0.09	3.00	-1.82	3.34	0.70	4.07
doublets	19	-7.47	7.58	-9.25	9.25	-7.24	7.81
		MNDO		AM1		PM3	
All molecules	46	-8.74	13.16	-12.48	13.12	-5.21	8.27
singlets	27	-1.15	8.68	-6.26	7.35	-0.16	5.07
doublets	19	-19.52	19.52	-21.31	21.31	-12.39	12.82

Im wesentlichen führt diese zweite statistische Auswertung für Satz B zu analogen Ergebnissen wie die erste (vgl. die Tabellen 3.36 und 3.37). Da im zweiten Fall eine Reihe von 'schwierigen' reaktiven Spezies wegfällt, für die keine experimentellen Daten vorhanden sind, verringern sich die statistischen Fehler in der Regel merklich: so liegen beispielsweise bei den OM $x$ -Methoden die mittleren absoluten Fehler für alle Verbindungen jetzt bei 5-6 kcal/mol (statt vorher bei ca. 7 kcal/mol). Die Aussagen zu der relativen Güte der einzelnen Methoden, zu den Unterschieden zwischen Singulets und Dubletts sowie zu systematischen Fehlern bleiben jedoch qualitativ unverändert und werden deshalb hier nicht wiederholt. Interessant sind noch die maximalen Fehler, die bei den Vergleichen mit den experimentellen Referenzdaten auftreten: Sie betragen bei den OM $x$  Methoden 11-13 kcal/mol für die Singulets und 14-17 kcal/mol für die Dubletts, die entsprechenden Zahlen bei den Standardverfahren belaufen sich auf 17-26 und 20-32 kcal/mol.

Da für den Validierungssatz B auch die optimierten **Molekülgeometrien** auf MP2/6-31G\* Niveau verfügbar sind (siehe oben), bietet es sich an, die Qualität der semiempirischen Ergebnisse auch in dieser Hinsicht zu prüfen. Die berechneten Bindungslängen und Bindungswinkel sind gemeinsam mit den *ab initio* Referenzdaten in der Tabelle 3.52 zusammengestellt (nur für OM1, OM2 und OM3). Die entsprechenden Statistiken sind für die Bindungslängen in Tabelle 3.38 und für die Bindungswinkel in Tabelle 3.39 zu finden.

Die angegebene Statistik der Bindungslängen ist von C-F-Bindungen dominiert (208 von 214). Der mittlere absolute Fehler liegt bei OM1, OM2, OM3 und PM3 bei 0.012-

**Tabelle 3.38:** Nach Molekülklassen getrennte Fehler der Bindungslängen für den Validierungssatz B, bezogen auf MP2/6-31G\* Referenzdaten

		Bindungslängen [Å]					
		OM1		OM2		OM3	
	Anzahl	MSE	MUE	MSE	MUE	MSE	MUE
All molecules	214	0.009	0.012	0.006	0.012	-0.011	0.014
CHF compounds	99	0.006	0.007	0.003	0.008	-0.015	0.015
CF compounds	46	0.010	0.012	0.011	0.013	-0.006	0.008
CHFO compounds	46	0.012	0.021	0.006	0.016	-0.011	0.021
CFO compounds	23	0.011	0.016	0.011	0.015	-0.006	0.009
C-F	208	0.009	0.010	0.006	0.010	-0.011	0.012
F-O	6	0.006	0.094	-0.002	0.066	-0.022	0.094
		MNDO		AM1		PM3	
All molecules	214	-0.017	0.023	0.004	0.018	-0.011	0.014
CHF compounds	99	-0.019	0.023	0.002	0.016	-0.014	0.016
CF compounds	46	-0.004	0.014	0.012	0.020	-0.003	0.009
CHFO compounds	46	-0.031	0.034	-0.006	0.019	-0.016	0.017
CFO compounds	23	-0.005	0.021	0.012	0.022	-0.002	0.009
C-F	208	-0.012	0.019	0.006	0.017	-0.010	0.013
F-O	6	-0.169	0.169	-0.066	0.066	-0.029	0.029

0.014 Å und ist nur bei AM1 und MNDO mit 0.018-0.023 Å grösser. Signifikant höhere Abweichungen treten bei den wenigen F-O-Bindungen in Satz B auf, analog wie auch im Referenzdatensatz.

Bei den Bindungswinkeln sind im Validierungssatz B insgesamt 212 Vergleiche möglich, davon betreffen 125 die CCF Winkel. Insgesamt erweisen sich hier die OM $x$ -Methoden (MUE-Werte: 1.8-2.2°) als überlegen gegenüber den Standardverfahren (MUE-Werte: 3.2-3.6°). Dies steht wiederum im Einklang mit entsprechenden Befunden bei dem Referenzdatensatz.

Zusammenfassend kann für den Validierungssatz B festgehalten werden, dass die OM $x$ -Methoden sich auch für diese eher atypischen Testmoleküle gut bewähren. Sie reproduzieren die *ab initio* Referenzdaten insgesamt besser als die Standardverfahren MNDO, AM1 und PM3, insbesondere bei den Standardbildungsenthalpien und Bindungswinkeln. Untereinander sind die OM $x$ -Methoden bei Satz B nahezu gleichwertig, mit einer leichten Präferenz für OM2.

### 3.7.4.9 Validierungssatz C

Zur Beurteilung von *ab initio* Methoden und Dichtefunktionalmethoden haben sich in den letzten Jahre die Testsätze etabliert, welche für die Validierung der Gaussian-2 [99] (G2) und der Gaussian-3 [100] (G3) Methoden erstellt wurden. Sie zeichnen sich durch eine hohe Genauigkeit der experimentellen Referenzwerte aus. So werden Referenzdaten nur dann akzeptiert, wenn die experimentelle Unsicherheit bei Stan-

**Tabelle 3.39:** Nach Molekülklassen getrennte Fehler der Bindungswinkel für den Validierungssatz B, bezogen auf MP2/6-31G\* Referenzdaten

	Bindungswinkel [°]						
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
All molecules	212	0.25	1.73	0.56	2.18	1.00	1.78
CHF compounds	98	0.14	1.73	0.71	1.97	1.10	1.82
CF compounds	39	0.37	1.73	0.61	2.59	0.97	1.91
CHFO compounds	52	0.07	1.92	0.37	2.11	0.79	1.75
CFO compounds	23	0.94	1.36	0.23	2.57	1.13	1.47
CCF	125	0.42	1.52	1.20	2.15	1.51	1.80
COF	6	3.29	3.29	3.05	3.05	3.95	3.95
FCF	31	-2.05	2.10	-2.61	2.69	-1.51	1.60
FCO	23	-0.03	1.73	-0.38	1.85	0.38	1.53
FCH	27	1.66	1.96	1.45	1.86	1.40	1.63
		MNDO		AM1		PM3	
All molecules	212	2.64	3.24	2.67	3.63	2.13	3.16
CHF compounds	98	3.24	3.68	3.27	3.95	2.61	3.34
CF compounds	39	1.73	2.42	2.08	3.34	1.59	2.78
CHFO compounds	52	2.64	3.38	2.41	3.59	1.85	3.36
CFO compounds	23	1.67	2.48	1.69	2.84	1.60	2.57
CCF	125	3.31	3.53	3.98	4.08	3.32	3.42
COF	6	11.21	11.21	8.29	8.29	6.67	6.67
FCF	31	0.14	1.57	-1.10	2.53	-1.13	2.39
FCO	23	0.46	2.59	-0.48	2.35	-0.64	2.50
FCH	27	2.38	2.61	2.36	2.83	1.69	2.61

Standardbildungsenthalpien weniger als 1 kcal/mol und bei Ionisierungspotentialen (IP) oder Elektronenaffinitäten (EA) weniger als 0.05 eV beträgt. Der Validierungssatz C enthält die fluorhaltigen Verbindungen aus den G2 und G3 Testsätzen. Damit bestehen notwendigerweise Überlappungen mit dem Referenzdatensatz und auch mit den beiden anderen Validierungssätzen, aber es wurde trotzdem als sinnvoll erachtet, die G2- und G3-analogen Auswertungen vorzunehmen, um Vergleiche mit anderen Methoden zu erleichtern.

Die Tabelle 3.40 zeigt die statistische Auswertung hinsichtlich der Standardbildungsenthalpien, Ionisationsenergien und Elektronenaffinitäten. Die zugehörigen Einzelergebnisse sind für die OM $x$ -Methoden in den Tabellen 3.54 bis 3.56 aufgeführt. Die mittleren absoluten Fehler der Standardbildungsenthalpien liegen im üblichen Bereich, wobei die OM $x$ -Methoden (3.5-5.5 kcal/mol) erwartungsgemäss besser abschneiden als die Standardverfahren (8.3-8.8 kcal/mol). Die wenigen Ionisierungsenergien und Elektronenaffinitäten im G2-Satz betreffen sehr kleine und für semiempirische Anwendungen atypische Spezies (siehe Tabelle 3.55 und 3.56). Von daher ist es eine positive Überraschung, dass die mittleren absoluten Fehler für OM2 und OM3 klein bleiben (2.9-4.8 kcal/mol), während sie bei den anderen Verfahren auf deutlich höhere Werte ansteigen (10.2-28.4 kcal/mol).

**Tabelle 3.40:** Statistische Auswertung der Standardbildungsenthalpien, Ionisierungsenergien und Elektronenaffinitäten im Validierungssatz C

Standardbildungsenthalpie, IP und EA [kcal/mol]							
	Anzahl	OM1		OM2		OM3	
		MSE	MUE	MSE	MUE	MSE	MUE
Heats of formation							
All molecules	17	2.08	5.45	0.79	3.48	1.17	4.05
G2 molecules	12	3.45	6.38	1.61	3.44	1.74	3.86
G3 molecules	5	-1.22	3.23	-1.19	3.56	-0.20	4.52
Ionization energies	6	-11.48	12.28	-4.76	4.76	-0.04	4.10
Electron affinities	4	-15.19	15.19	-0.86	3.87	1.52	2.87
		MNDO		AM1		PM3	
Heats of formation							
All molecules	17	2.50	8.62	-3.91	8.82	0.57	8.34
G2 molecules	12	3.24	7.62	-4.89	7.89	-0.09	7.46
G3 molecules	5	0.73	11.03	-1.54	11.05	2.15	10.47
Ionization energies	6	-13.66	16.68	-28.37	28.37	-9.37	10.16
Electron affinities	4	-13.51	19.35	-20.02	23.56	1.97	16.64

#### 3.7.4.10 Zusammenfassung

Die durchgeführte Vergleiche belegen deutlich, dass die OM $x$ -Methoden bei der Berechnung von Fluorverbindungen den Standardverfahren MNDO, AM1 und PM3 generell überlegen sind. Dies zeigt sich nicht nur anhand der statistischen Fehler für verschiedene Moleküleigenschaften und Molekülklassen, sondern auch an der geringeren Häufigkeit von Ausreißern. Untereinander sind die OM $x$ -Methoden in vielerlei Hinsicht relativ gleichwertig, mit einer leichten Präferenz für OM2 und OM3, weil OM1 im Validierungssatz C einige grössere Fehler aufweist. Die mit OM2 und OM3 erzielte Genauigkeit ist bei Fluorverbindungen und bei den früher parametrisierten organischen (H, C, N, O) Verbindungen vergleichbar gut. In diesem Sinne ist die durchgeführte Fluorparametrisierung erfolgreich und kann für Anwendungen freigegeben werden.

#### 3.7.5 Tabellen mit Einzelergebnissen

Die Tabellen 3.41 - 3.56 dokumentieren die erhaltenen Ergebnisse für die Fluorverbindungen. Sie basieren auf automatisch erzeugten Dateien und enthalten daher eine Reihe von Bezeichnungen in englischer Sprache.

**Tabelle 3.41:** Standardbildungsenthalpien für den Referenzdatensatz: OM1, OM2 und OM3 Ergebnisse.

Molekül	Standardbildungsenthalpien [kcal/mol]								$\delta_{exp}$	Lit.
	Exp.	Berechnet			Fehler					
		OM1	OM2	OM3	OM1	OM2	OM3			
Fluorine	0.0	12.3	0.6	-1.7	12.3	0.6	-1.7	0.0	a	
Hydrogen fluoride	-65.1	-69.7	-63.3	-64.9	-4.6	1.8	0.2	0.2	a	
Methylfluoride	-56.0	-54.8	-60.4	-57.3	1.2	-4.4	-1.3	7.0	a	
Difluoromethane	-108.1	-104.7	-110.9	-106.6	3.4	-2.8	1.5	0.2	b	
Trifluoromethane	-166.2	-160.6	-166.8	-162.0	5.6	-0.6	4.2	0.6	b	
Tetrafluoromethane	-223.1	-217.2	-220.5	-215.8	5.9	2.6	7.3	0.3	b	
Fluoroethane	-63.0	-65.6	-67.5	-65.8	-2.6	-4.5	-2.8	0.5	c	
1,1-Difluoroethane	-118.8	-119.0	-121.8	-119.4	-0.2	-3.0	-0.6	2.0	b	
1,1,1-Trifluoroethane	-178.0	-176.9	-178.4	-176.3	1.1	-0.4	1.7	0.4	b	
Hexafluoroethane	-321.3	-316.2	-317.5	-310.5	5.1	3.8	10.8	1.3	b	
i-Propylfluoride	-70.1	-74.5	-76.2	-75.1	-4.4	-6.1	-5.0	0.4	b	
Octafluorocyclobutane	-368.7	-379.9	-380.0	-372.9	-11.2	-11.3	-4.2	2.6	b	
Fluoroethene	-33.2	-33.1	-32.3	-32.2	0.1	0.9	1.0	0.4	b	
1,2-Difluoroethene cis	-71.0	-73.0	-72.2	-71.7	-2.0	-1.2	-0.7	2.4	c	
1,2-Difluoroethene trans	-70.0	-74.7	-74.3	-72.7	-4.7	-4.3	-2.7	2.4	c	
1,1-Difluoroethene	-80.1	-80.3	-78.4	-79.9	-0.2	1.7	0.2	1.1	b	
Trifluoroethene	-117.2	-117.6	-116.4	-115.9	-0.4	0.8	1.3	2.0	b	
Tetrafluoroethene	-157.5	-157.6	-156.3	-155.9	-0.1	1.2	1.6	1.2	b	
Fluorobenzene	-27.7	-28.8	-27.3	-28.0	-1.1	0.4	-0.3	0.3	b	
1,2-Difluorobenzene	-70.2	-70.5	-69.1	-69.9	-0.3	1.1	0.3	0.2	b	
1,3-Difluorobenzene	-73.9	-74.2	-72.9	-74.0	-0.3	1.0	-0.1	0.2	b	
1,4-Difluorobenzene	-73.3	-73.9	-72.5	-73.5	-0.6	0.8	-0.2	0.2	b	
Pentafluorobenzene	-192.8	-192.6	-191.8	-192.7	0.2	1.0	0.1	0.4	b	
Hexafluorobenzene	-228.3	-229.0	-228.4	-228.8	-0.7	-0.1	-0.5	0.4	b	
Fluoroacetylene	25.1	26.7	26.9	23.9	1.5	1.7	-1.2	0.5	d	
Difluoroacetylene	1.4	4.0	4.4	-2.6	2.7	3.0	-4.0	0.5	d	
Hypofluorous acid	-23.5	-15.8	-19.6	-20.3	7.7	3.9	3.2	1.0	a	
Fluoroethanol gauche	-100.0	-101.0	-100.7	-100.1	-1.0	-0.7	-0.1		c	
1,1,1-Trifluoroethane	-212.3	-212.4	-209.2	-206.4	-0.1	3.1	5.9	0.2	a	
Trifluoromethyl-										
hypofluorite	-182.8	-167.0	-174.0	-166.3	15.8	8.8	16.5	3.1	a	
Difluorine monoxide	5.9	16.2	5.3	7.2	10.3	-0.6	1.3	0.4	a	
Formyl fluoride	-90.0	-87.3	-86.3	-88.2	2.7	3.7	1.8	3.6	a	
Acetyl fluoride	-105.7	-99.8	-100.0	-103.0	5.9	5.7	2.7	0.8	b	
Carbonyl fluoride	-145.9	-137.1	-136.5	-138.7	8.8	9.4	7.2		e	
Trifluoroacetic acid	-246.5	-244.7	-242.3	-239.7	1.8	4.2	6.8	0.4	b	
p-Fluorobenzoic acid	-118.2	-114.8	-116.9	-117.4	3.4	1.3	0.8	0.5	b	
Trifluoroamine	-29.8	-36.1	-30.1	-35.9	-6.3	-0.3	-6.1	0.2	b	
(Z)-Difluorodiazene	16.4	4.5	13.5	11.7	-11.9	-2.9	-4.7	1.2	a	
(E)-Difluorodiazene	19.4	6.0	22.7	22.4	-13.4	3.3	3.0	1.2	a	
Tetrafluorohydrazine	-8.0	-26.7	-6.3	-6.4	-18.7	1.7	1.6		c	
Fluorocyanide	8.6	2.2	1.9	-1.5	-6.4	-6.7	-10.1	4.0	a	
Trifluoroacetonitril	-119.0	-126.7	-125.4	-125.1	-7.7	-6.4	-6.1	0.7	b	
Fluorocarbene	34.2	26.1	24.9	24.4	-8.1	-9.3	-9.8	3.0	f	
Difluorocarbene	-43.8	-56.0	-57.8	-55.7	-12.2	-14.0	-12.0	0.5	g	

... Continued ...



**Tabelle 3.41:** (continued)

Standardbildungsenthalpien [kcal/mol]									
Molekül	Exp.	Berechnet			Fehler			$\delta_{exp}$	Lit.
		OM1	OM2	OM3	OM1	OM2	OM3		
Nitrosyl fluoride	-15.7	-20.9	-10.4	-21.6	-5.2	5.3	-5.9	0.4	a
Nitryl fluoride	-26.0	-22.6	-25.7	-29.3	3.4	0.3	-3.3	5.0	a
Fluorine nitrate	2.5	12.4	2.6	4.3	9.9	0.1	1.8	0.5	b
Nitrosyl trifluoride	-39.0	-36.3	-49.9	-50.7	2.7	-10.9	-11.7	5.0	a
Mean error					-0.27	-0.36	-0.26		
Mean absolute error					4.92	3.42	3.20		
Number of comparison					48	48	48		

a M.W.Chase, C.A.Davies, J.R.Downey, D.R.Frurip, R.A.McDonald, and A.N.Syverud, "JANAF Thermochemical Tables", 3rd edition, J.Phys.Chem.Ref.Data 14, Suppl.1 (1985). b J.B.Pedley, R.D.Naylor, and S.P.Kirby, "Thermochemical Data of Organic Compounds", 2nd ed., Chapman and Hall, London, 1986. c S.G.Lias, J.E.Bartmess, J.F.Liebman, J.L.Holmes, R.D.Levin, and W.G.Mallard, "Gas Phase Ion and Neutral Thermochemistry", J.Phys.Chem.Ref.Data 17, Suppl. 1 (1988). d S.Parthiban, J.M.L.Martin, and J.F.Liebman, Mol.Phys. 100, 453 (2002). e N.L.Haworth, M.H.Smith, G.B.Bacskey, and J.C.Mackie, J.Phys.Chem. A 104, 7600 (2000). f J.C.Poutsma, J.A.Paulimo, and R.R.Squires, J.Phys.Chem. A 101, 5327 (1997). g J.Demaison, L.Margules, J.M.L.Martin, and J.E.Boggs, Phys.Chem.Chem.Phys. 4, 3282 (2002).

**Tabelle 3.42:** Standardbildungsenthalpien für den Referenzdatensatz: MNDO, AM1 und PM3 Ergebnisse.

Standardbildungsenthalpien [kcal/mol]									
Molekül	Exp.	Berechnet			Fehler			$\delta_{exp}$	Lit.
		MNDO	AM1	PM3	MNDO	AM1	PM3		
Fluorine	0.0	7.3	-22.5	-21.7	7.3	-22.5	-21.7	0.0	a
Hydrogen fluoride	-65.1	-59.7	-74.3	-62.7	5.4	-9.2	2.4	0.2	a
Methylfluoride	-56.0	-60.9	-61.0	-53.8	-4.9	-5.0	2.2	7.0	a
Difluoromethane	-108.1	-111.8	-116.1	-103.8	-3.7	-8.0	4.3	0.2	b
Trifluoromethane	-166.2	-163.8	-172.5	-162.0	2.4	-6.3	4.2	0.6	b
Tetrafluoromethane	-223.1	-214.2	-225.7	-225.1	8.9	-2.6	-2.0	0.3	b
Fluoroethane	-63.0	-65.1	-66.3	-60.2	-2.1	-3.3	2.8	0.5	c
1,1-Difluoroethane	-118.8	-113.4	-118.6	-111.9	5.4	0.2	6.9	2.0	b
1,1,1-Trifluoroethane	-178.0	-164.4	-172.6	-172.3	13.6	5.4	5.7	0.4	b
Hexafluoroethane	-321.3	-299.6	-313.2	-317.8	21.7	8.1	3.5	1.3	b
i-Propylfluoride	-70.1	-66.6	-69.8	-66.8	3.5	0.3	3.3	0.4	b
Octafluorocyclobutan	-368.7	-363.7	-367.2	-379.2	5.0	1.5	-10.5	2.6	b
Fluoroethene	-33.2	-34.5	-34.0	-28.6	-1.3	-0.8	4.6	0.4	b
1,2-Difluoroethene cis	-71.0	-83.6	-82.7	-73.0	-12.6	-11.7	-2.0	2.4	c
1,2-Difluoroethene trans	-70.0	-83.1	-82.7	-71.5	-13.1	-12.7	-1.5	2.4	c
1,1-Difluoroethene	-80.1	-84.1	-84.3	-79.6	-4.0	-4.2	0.5	1.1	b
Trifluoroethene	-117.2	-131.1	-130.6	-121.5	-13.9	-13.4	-4.3	2.0	b
Tetrafluoroethene	-157.5	-175.7	-175.1	-168.2	-18.2	-17.6	-10.7	1.2	b

... Continued ...

Tabelle 3.42: (continued)

Standardbildungsenthalpien [kcal/mol]									
Molekül	Exp.	Berechnet			Fehler			$\delta_{exp}$	Lit.
		MNDO	AM1	PM3	MNDO	AM1	PM3		
Fluorobenzene	-27.7	-25.2	-23.3	-20.2	2.5	4.4	7.5	0.3	b
1,2-Difluorobenzene	-70.2	-70.6	-66.4	-63.1	-0.4	3.8	7.1	0.2	b
1,3-Difluorobenzene	-73.9	-70.9	-67.9	-63.3	3.0	6.0	10.6	0.2	b
1,4-Difluorobenzene	-73.3	-71.0	-68.0	-63.3	2.3	5.3	10.0	0.2	b
Pentafluorobenzene	-192.8	-201.7	-192.0	-188.6	-8.9	0.8	4.2	0.4	b
Hexafluorobenzene	-228.3	-243.4	-231.2	-229.3	-15.1	-2.9	-1.0	0.4	b
Fluoroacetylene	25.1	15.7	15.2	18.1	-9.5	-9.9	-7.0	0.5	d
Difluoroacetylene	1.4	-21.0	-19.6	-11.6	-22.4	-21.0	-13.0	0.5	d
Hypofluorous acid	-23.5	-18.6	-22.6	-29.2	4.9	0.9	-5.7	1.0	a
Fluoroethanol gauche	-100.0	-107.8	-111.5	-99.8	-7.8	-11.5	0.2		c
1,1,1-Trifluoroethane	-212.3	-207.2	-214.1	-209.1	5.1	-1.8	3.2	0.2	a
Trifluoromethyl-									
hypofluorite	-182.8	-163.3	-177.7	-187.3	19.5	5.1	-4.5	3.1	a
Difluorine monoxide	5.9	18.2	10.5	-4.8	12.3	4.6	-10.7	0.4	a
Formyl fluoride	-90.0	-88.8	-92.9	-88.8	1.2	-2.9	1.2	3.6	a
Acetyl fluoride	-105.7	-96.5	-98.8	-98.7	9.2	6.9	7.0	0.8	b
Carbonyl fluoride	-145.9	-138.6	-146.3	-141.6	7.3	-0.4	4.3		e
Trifluoroacetic acid	-246.5	-238.1	-242.7	-244.0	8.4	3.8	2.5	0.4	b
p-Fluorobenzoic acid	-118.2	-111.8	-113.0	-109.5	6.4	5.2	8.7	0.5	b
Trifluoroamine	-29.8	-34.2	-40.0	-24.4	-4.4	-10.2	5.4	0.2	b
(Z)-Difluorodiazene	16.4	-2.3	20.8	28.0	-18.7	4.4	11.6	1.2	a
(E)-Difluorodiazene	19.4	2.4	31.3	29.2	-17.0	11.9	9.8	1.2	a
Tetrafluorohydrazine	-8.0	-19.7	5.8	-0.5	-11.7	13.8	7.5		c
Fluorocyanide	8.6	-2.3	-4.4	6.5	-10.9	-13.0	-2.1	4.0	a
Trifluoroacetonitrile	-119.0	-113.2	-119.5	-115.0	5.8	-0.5	4.0	0.7	b
Fluorocarbene	34.2	21.5	20.9	35.3	-12.7	-13.3	1.1	3.0	f
Difluorocarbene	-43.8	-65.2	-68.0	-49.1	-21.4	-24.2	-5.3	0.5	g
Nitrosyl fluoride	-15.7	-24.8	-26.5	-3.3	-9.1	-10.8	12.4	0.4	a
Nitryl fluoride	-26.0	0.7	-21.3	-25.6	26.7	4.7	0.4	5.0	a
Fluorine nitrate	2.5	28.0	13.7	-6.1	25.5	11.2	-8.6	0.5	b
Nitrosyl trifluoride	-39.0	22.8	-14.6	-26.6	61.8	24.4	12.4	5.0	a
Mean error					0.65	-2.23	1.27		
Mean absolute error					10.81	7.75	5.87		
Number of comparison					48	48	48		

a - g gleiche Referenzen wie in der Tabelle 3.41

Tabelle 3.43: Bindungslängen [Å] und Bindungswinkel [°] für den Referenzdatensatz.

Bindungslängen [Å] und Bindungswinkel [°]										
Molekül	Symmetrie	Variable	Exp.	Berechnet			Fehler			Lit.
				OM1	OM2	OM3	OM1	OM2	OM3	
Fluorine	$D_{\infty h}$	F-F	1.412	1.365	1.365	1.321	-0.047	-0.047	-0.091	a
Hydrogen-fluoride	$C_{\infty v}$	H-F	0.917	0.962	0.968	0.948	0.045	0.051	0.031	a

... Continued ...

Tabelle 3.43: (continued)

Molekül	Symmetrie	Variable	Bindungslängen [Å] und Bindungswinkel [°]							Lit.
			Exp	Berechnet			Fehler			
				OM1	OM2	OM3	OM1	OM2	OM3	
Methylfluoride	$C_{3v}$	C-F	1.383	1.390	1.374	1.356	0.007	-0.009	-0.027	b
		C-H	1.100	1.106	1.111	1.120	0.006	0.011	0.020	b
Difluoromethane	$C_{2v}$	HCF	108.3	108.5	108.5	109.5	0.2	0.2	1.2	b
		C-F	1.351	1.377	1.366	1.350	0.026	0.015	-0.001	b
		C-H	1.084	1.122	1.133	1.144	0.038	0.049	0.060	b
		HCH	112.8	113.1	113.0	111.4	0.3	0.2	-1.4	b
Trifluoromethane	$C_{3v}$	FCF	108.5	103.0	102.9	104.8	-5.5	-5.6	-3.7	b
		C-F	1.328	1.363	1.359	1.344	0.035	0.031	0.016	b
		C-H	1.091	1.130	1.143	1.156	0.039	0.052	0.065	b
Tetrafluoromethane	$T_d$	HCF	108.6	113.5	113.6	112.7	4.9	5.0	4.1	b
		C-F	1.319	1.346	1.347	1.334	0.027	0.028	0.015	b
Fluoroethane	$C_s$	C-H	1.091	1.090	1.097	1.102	-0.001	0.006	0.011	b
		H3CC	108.6	109.7	109.5	110.1	1.1	0.9	1.5	b
		H2CC	110.2	109.7	109.5	110.1	-0.5	-0.7	-0.1	b
		C-C	1.502	1.509	1.511	1.516	0.007	0.009	0.014	b
		C-F	1.397	1.394	1.385	1.366	-0.003	-0.012	-0.031	b
		H1CC	113.6	111.3	112.0	111.5	-2.3	-1.6	-2.1	b
		C-C	1.535	1.523	1.521	1.529	-0.012	-0.014	-0.006	c
1,2-Difluoroethane	$C_2$	C-F	1.394	1.391	1.382	1.363	-0.003	-0.012	-0.031	c
		HCC	111.0	110.8	111.8	111.2	-0.2	0.8	0.2	b
		CCF	110.3	107.1	107.4	109.2	-3.2	-2.9	-1.1	b
1,1,1-Trifluoroethane	$C_{3v}$	C-C	1.494	1.506	1.512	1.515	0.012	0.018	0.021	b
		C-F	1.340	1.362	1.365	1.351	0.022	0.025	0.010	b
		C-H	1.081	1.087	1.091	1.095	0.006	0.010	0.014	b
		FCC	111.9	112.9	113.9	113.2	1.0	2.0	1.3	b
		HCC	112.0	109.3	109.0	109.7	-2.7	-3.0	-2.3	b
Hexafluoroethane	$D_{3d}$	C-C	1.545	1.567	1.560	1.592	0.022	0.015	0.047	c
		C-F	1.326	1.352	1.354	1.339	0.026	0.028	0.013	c
		CCF	109.8	111.4	112.4	111.6	1.6	2.6	1.8	c
1,1-Difluorocyclopane	$C_{2v}$	C-H	1.082	1.080	1.083	1.090	-0.002	0.001	0.008	b
		$C_2$ -C3	1.553	1.481	1.491	1.496	-0.072	-0.062	-0.057	b
		$C_1$ - $C_2$	1.464	1.509	1.523	1.524	0.045	0.059	0.060	b
		FCF	108.3	108.7	107.0	108.2	0.4	-1.3	-0.1	b
Hexafluorocyclopropane	$D_{3h}$	C-C	1.505	1.535	1.547	1.559	0.030	0.042	0.054	c
		C-F	1.314	1.344	1.346	1.329	0.030	0.032	0.015	c
		FCF	112.2	109.8	108.4	109.9	-2.4	-3.8	-2.3	c
Octafluorocyclobutane	$D_{2d}$	C-C	1.566	1.582	1.579	1.598	0.016	0.013	0.032	c
		C-F	1.333	1.349	1.351	1.335	0.016	0.018	0.002	c
		FCF	109.9	107.3	105.6	107.0	-2.6	-4.3	-2.9	c
Fluoroethene	$C_s$	C=C	1.330	1.316	1.333	1.336	-0.014	0.003	0.006	b
		H1-C	1.107	1.092	1.101	1.112	-0.015	-0.006	0.005	b
		H2-C	1.108	1.074	1.077	1.083	-0.034	-0.031	-0.025	b
		H3-C	1.097	1.074	1.077	1.082	-0.023	-0.020	-0.015	b
		C-F	1.351	1.360	1.356	1.336	0.009	0.005	-0.015	b
		H1CC	130.8	127.8	128.3	127.1	-3.0	-2.5	-3.7	b
		H3CC	120.4	121.1	121.8	122.3	0.7	1.4	1.9	b
		H2CC	118.7	121.3	120.7	121.1	2.6	2.0	2.4	b
		CCF	121.5	117.5	118.0	119.3	-4.0	-3.5	-2.2	b
		C=C	1.330	1.327	1.344	1.349	-0.003	0.014	0.019	b
		H-C	1.099	1.087	1.095	1.106	-0.012	-0.004	0.007	b
1,2-Difluoroethene	$C_{2v}$	C-F	1.342	1.353	1.348	1.329	0.011	0.006	-0.013	b
		CCH	124.1	125.8	126.6	124.4	1.7	2.5	0.3	b
		CCF	122.0	118.3	118.3	120.4	-3.7	-3.7	-1.6	b
		C=C	1.329	1.326	1.342	1.348	-0.003	0.013	0.019	c
		H-C	1.080	1.086	1.094	1.104	0.006	0.014	0.024	c
1,2-Difluoroethene	$C_{2h}$	C-F	1.344	1.356	1.350	1.330	0.012	0.006	-0.014	c
		CCH	129.3	127.3	127.1	125.5	-2.0	-2.2	-3.8	c
		CCF	119.3	116.3	117.4	118.8	-3.0	-1.9	-0.5	c
		CCH	129.3	127.3	127.1	125.5	-2.0	-2.2	-3.8	c

... Continued ...

Tabelle 3.43: (continued)

Molekül	Symmetrie	Variable	Bindungslängen [Å] und Bindungswinkel [°]							Lit.
			Exp	Berechnet			Fehler			
				OM1	OM2	OM3	OM1	OM2	OM3	
1,1-Difluoroethene	$C_{2v}$	C=C	1.340	1.316	1.335	1.336	-0.024	-0.005	-0.004	b
		H-C	1.091	1.070	1.072	1.077	-0.021	-0.019	-0.014	b
		C-F	1.315	1.340	1.342	1.325	0.025	0.027	0.010	b
		CCH	119.0	120.7	120.6	121.2	1.7	1.6	2.2	b
Trifluoroethene	$C_s$	CCF	124.7	125.6	126.4	125.9	0.9	1.7	1.2	b
		C=C	1.341	1.330	1.347	1.354	-0.011	0.006	0.013	b
		C-H	1.100	1.083	1.089	1.099	-0.017	-0.011	-0.001	b
		C-F3	1.316	1.339	1.340	1.323	0.023	0.024	0.007	b
		CCF5	120.0	116.9	117.4	119.5	-3.1	-2.6	-0.5	b
Tetrafluoroethene	$D_{2h}$	CCF3	123.1	123.3	124.4	123.6	0.2	1.3	0.5	b
		CCH	124.0	125.4	125.7	123.4	1.4	1.7	-0.6	b
		C=C	1.311	1.334	1.349	1.360	0.023	0.038	0.049	c
		C-F	1.319	1.335	1.335	1.319	0.016	0.016	0.000	c
Hexafluorocyclobutene	$C_{2v}$	CCF	123.8	124.0	124.7	124.2	0.2	0.9	0.4	c
		$C_1=C_2$	1.348	1.340	1.362	1.366	-0.008	0.014	0.018	d
		$C_2-C_3$	1.506	1.530	1.537	1.542	0.024	0.031	0.036	d
1,1-Difluoroallene	$C_{2v}$	$C_3-C_4$	1.591	1.598	1.592	1.611	0.007	0.001	0.020	d
		$C_1-F$	1.312	1.325	1.325	1.306	0.013	0.013	-0.006	d
		$C_3-F$	1.339	1.350	1.354	1.338	0.011	0.015	-0.001	d
		$C_1C_2F$	136.6	133.6	133.8	134.0	-3.0	-2.8	-2.6	d
		$C_2C_3C_4$	85.4	85.2	85.7	85.4	-0.2	0.3	0.0	d
		$C_2=C_3$	1.306	1.291	1.308	1.312	-0.015	0.002	0.006	c
		C-H	1.086	1.086	1.088	1.097	0.000	0.002	0.011	c
		C-F	1.323	1.340	1.341	1.325	0.017	0.018	0.002	c
		$C_1=C_2$	1.302	1.301	1.317	1.324	-0.001	0.015	0.022	c
		CCH	121.1	120.6	121.0	121.4	-0.5	-0.1	0.3	c
Fluorobenzene	$C_{2v}$	CCF	124.9	124.9	125.7	125.2	0.0	0.8	0.3	c
		C-F	1.354	1.356	1.360	1.340	0.002	0.006	-0.014	b
		C-H	1.081	1.079	1.088	1.095	-0.002	0.007	0.014	b
		$C_3=C_4$	1.383	1.383	1.395	1.401	0.000	0.012	0.018	b
		$C_2=C_3$	1.395	1.380	1.392	1.396	-0.015	-0.003	0.001	b
		$C_3=C_4$	1.397	1.382	1.395	1.400	-0.015	-0.002	0.003	b
		$C_6C_1C_2$	123.4	123.7	124.0	123.0	0.3	0.6	-0.4	b
		$C_1C_2C_3$	117.9	117.4	117.2	117.9	-0.5	-0.7	0.0	b
1,3-Difluorobenzene	$C_{2v}$	$C_1=C_2$	1.389	1.382	1.394	1.400	-0.007	0.005	0.011	b
		$C_4=C_5$	1.395	1.381	1.393	1.398	-0.014	-0.002	0.003	b
		$C_3=C_4$	1.383	1.385	1.397	1.403	0.002	0.014	0.020	b
		$C_2-H$	1.082	1.075	1.081	1.088	-0.007	-0.001	0.006	b
Hexafluorobenzene	$D_{6h}$	C=C	1.401	1.392	1.404	1.413	-0.009	0.003	0.012	c
		C-F	1.325	1.340	1.341	1.323	0.015	0.016	-0.002	c
Difluorobenzocyclopropene	$C_{2v}$	$C_1-C_2$	1.520	1.465	1.478	1.480	-0.055	-0.042	-0.040	d
		$C_2=C_7$	1.339	1.379	1.408	1.408	0.040	0.069	0.069	d
		$C_2=C_3$	1.355	1.361	1.371	1.379	0.006	0.016	0.024	d
		$C_3=C_4$	1.423	1.395	1.406	1.409	-0.028	-0.017	-0.014	d
		$C_4=C_5$	1.411	1.387	1.399	1.405	-0.024	-0.012	-0.006	d
Fluoroacetylene	$C_{\infty v}$	C-F	1.329	1.350	1.356	1.339	0.021	0.027	0.010	d
		FCF	109.4	108.4	106.3	107.6	-1.0	-3.1	-1.8	d
		C-F	1.279	1.300	1.296	1.277	0.021	0.017	-0.002	c
		$C\equiv C$	1.198	1.206	1.223	1.227	0.008	0.025	0.029	c
Hypofluorous acid	$C_s$	O-H	0.964	0.974	1.010	1.031	0.010	0.046	0.067	c
		O-F	1.442	1.395	1.398	1.369	-0.047	-0.044	-0.073	c
		HOF	97.2	103.4	102.3	103.9	6.2	5.1	6.7	c
Trifluoromethylhypofluorite	$C_s$	C-O	1.395	1.393	1.388	1.429	-0.002	-0.007	0.034	c
		C-F	1.319	1.346	1.345	1.330	0.027	0.026	0.011	c
		O-F	1.421	1.387	1.397	1.367	-0.034	-0.024	-0.054	c
		COF	104.8	106.8	106.6	107.4	2.0	1.8	2.6	c
Difluorine monoxid	$C_{2v}$	FCF	109.4	107.7	108.0	108.2	-1.7	-1.4	-1.2	c
		O-F	1.405	1.377	1.383	1.355	-0.028	-0.023	-0.051	c
		FOF	103.1	101.7	101.2	102.0	-1.4	-1.9	-1.1	c

... Continued ...

Tabelle 3.43: (continued)

Bindungslängen [Å] und Bindungswinkel [°]										
Molekül	Symmetrie	Variable	Exp	Berechnet			Fehler			Lit.
				OM1	OM2	OM3	OM1	OM2	OM3	
Formylfluoride	$C_s$	C=O	1.188	1.190	1.212	1.216	0.002	0.024	0.028	b
		C-F	1.346	1.369	1.360	1.343	0.023	0.014	-0.003	b
		OCF	122.3	116.3	115.6	117.1	-6.0	-6.7	-5.2	b
Acetylfluoride	$C_s$	C=O	1.181	1.192	1.217	1.222	0.011	0.036	0.041	c
		C-F	1.348	1.370	1.368	1.351	0.022	0.020	0.003	c
		CCO	128.3	131.5	130.9	130.5	3.2	2.6	2.2	c
		CCF	110.3	111.6	113.2	112.4	1.3	2.9	2.1	c
Carbonylfluoride	$C_{2v}$	C=O	1.172	1.186	1.211	1.216	0.014	0.039	0.044	b
		C-F	1.316	1.338	1.337	1.322	0.022	0.021	0.006	b
		FCO	126.2	126.3	126.4	126.4	0.1	0.2	0.2	b
Hexafluoroacetone	$C_{2v}$	C-C	1.549	1.555	1.546	1.569	0.006	-0.003	0.020	c
		C=O	1.246	1.190	1.211	1.216	-0.056	-0.035	-0.030	c
		C-F	1.335	1.351	1.352	1.337	0.016	0.017	0.002	c
		CCC	121.4	115.3	117.0	115.6	-6.1	-4.4	-5.8	c
		FCC	110.3	112.7	113.0	112.5	2.4	2.7	2.2	c
Trifluoroamine	$C_{3v}$	N-F	1.365	1.370	1.366	1.373	0.005	0.001	0.008	c
		FNF	102.3	101.1	100.2	99.2	-1.2	-2.1	-3.1	c
		N=N	1.214	1.201	1.188	1.270	-0.013	-0.026	0.056	c
(Z)-Difluorodiazan	$C_{2v}$	N-F	1.384	1.376	1.381	1.385	-0.008	-0.003	0.001	c
		NNF	114.5	113.6	115.0	115.8	-0.9	0.5	1.3	c
		N=N	1.231	1.207	1.204	1.287	-0.024	-0.027	0.056	c
(E)-Difluorodiazan	$C_{2h}$	N-F	1.396	1.369	1.371	1.381	-0.027	-0.025	-0.015	c
		NNF	105.5	108.7	109.4	106.6	3.2	3.9	1.1	c
		N-N	1.492	1.402	1.363	1.472	-0.090	-0.129	-0.020	c
Tetrafluorohydrazine	$C_2$	N-F	1.372	1.366	1.366	1.373	-0.006	-0.006	0.001	c
		NNF	101.4	105.3	106.7	103.5	3.9	5.3	2.1	c
		N-F	1.400	1.378	1.371	1.380	-0.022	-0.029	-0.020	c
Difluoroamine	$C_s$	FNF	102.9	100.5	99.9	98.9	-2.4	-3.0	-4.0	c
		C-N	1.449	1.483	1.457	1.484	0.034	0.008	0.035	c
		C-H	1.091	1.099	1.104	1.104	0.008	0.013	0.013	c
N,N-Difluoromethylamine	$C_s$	N-F	1.413	1.379	1.385	1.396	-0.034	-0.028	-0.017	c
		CNF	104.6	105.6	106.0	105.3	1.0	1.4	0.7	c
		H1CN	110.4	111.7	111.9	113.4	1.3	1.5	3.0	c
		H2CN	106.2	108.9	108.9	108.2	2.7	2.7	2.0	c
		H1CNF	52.9	53.0	52.1	51.5	0.1	-0.8	-1.4	c
		C-N	1.426	1.440	1.462	1.468	0.014	0.036	0.042	c
		N=N	1.293	1.218	1.212	1.299	-0.075	-0.081	0.006	c
		C-F	1.315	1.342	1.339	1.329	0.027	0.024	0.014	c
		NCN	54.0	50.0	49.0	52.5	-4.0	-5.0	-1.5	c
Difluorodiazacyclopropane	$C_{2v}$	FCF	112.0	113.8	114.4	113.8	1.8	2.4	1.8	c
		N=N	1.235	1.199	1.192	1.264	-0.036	-0.043	0.029	b
		C-N	1.460	1.478	1.489	1.515	0.018	0.029	0.055	b
		C-F	1.321	1.347	1.347	1.335	0.026	0.026	0.014	b
		NNC	113.4	117.9	116.8	115.6	4.5	3.4	2.2	b
Hexafluorodimethylimine	$C_{2h}$	NCF	109.4	109.2	109.3	108.8	-0.2	-0.1	-0.6	b
		C≡N2	1.158	1.157	1.172	1.204	-0.001	0.014	0.046	c
		N1-C	1.386	1.405	1.395	1.446	0.019	0.009	0.060	c
F2N-CN	$C_s$	N-F	1.399	1.374	1.375	1.382	-0.025	-0.024	-0.017	c
		N1CN2	173.9	173.8	174.1	172.5	-0.1	0.2	-1.4	c
		CN1F	105.4	106.6	106.8	105.5	1.2	1.4	0.1	c
		C≡N	1.159	1.154	1.171	1.202	-0.005	0.012	0.043	c
		C-F	1.262	1.297	1.291	1.279	0.035	0.029	0.017	c
Fluorocyanide	$C_{\infty v}$	C≡N	1.153	1.154	1.168	1.199	0.001	0.015	0.046	d
		C-C	1.461	1.484	1.483	1.499	0.023	0.022	0.038	d
		C-F	1.335	1.350	1.352	1.338	0.015	0.017	0.003	d
		FCC	111.4	110.8	111.7	111.2	-0.6	0.3	-0.2	d
Trifluoroacetone	$C_{3v}$	C-F	1.310	1.329	1.334	1.321	0.019	0.024	0.011	d
		C-N	1.333	1.328	1.343	1.366	-0.005	0.010	0.033	d
		CNC	113.0	115.0	112.4	114.1	2.0	-0.6	1.1	d

... Continued ...

Tabelle 3.43: (continued)

Bindungslängen [Å] und Bindungswinkel [°]										
Molekül	Symmetrie	Variable	Exp	Berechnet			Fehler			Lit.
				OM1	OM2	OM3	OM1	OM2	OM3	
Fluorocarbene	$C_s$	C-F	1.305	1.320	1.312	1.297	0.015	0.007	-0.008	b
		C-H	1.138	1.141	1.176	1.179	0.003	0.038	0.041	b
		HCF	104.1	105.0	105.1	104.9	0.9	1.0	0.8	b
Difluorocarbene	$C_{2v}$	C-F	1.300	1.333	1.331	1.317	0.033	0.031	0.017	b
		FCF	104.9	99.9	100.4	101.5	-5.0	-4.5	-3.4	b
Nitrosyl fluoride	$C_s$	N-F	1.517	1.432	1.414	1.434	-0.085	-0.103	-0.083	b
		N=O	1.132	1.155	1.185	1.194	0.023	0.053	0.062	b
		ONF	109.9	107.3	105.8	105.5	-2.6	-4.1	-4.4	b
Nitrosyl trifluoride	$C_{3v}$	N=O	1.158	1.201	1.231	1.235	0.043	0.073	0.077	c
		N-F	1.431	1.411	1.394	1.396	-0.020	-0.037	-0.035	c
		ONF	117.1	117.0	116.8	117.3	-0.1	-0.3	0.2	c
CF <sub>3</sub> NO	$C_s$	C-N	1.546	1.523	1.512	1.568	-0.023	-0.034	0.022	c
		N=O	1.197	1.169	1.198	1.206	-0.028	0.001	0.009	c
		C-F	1.326	1.348	1.352	1.338	0.022	0.026	0.012	c
		CNO	113.2	113.5	111.2	111.9	0.3	-2.0	-1.3	c
		NCF	109.0	109.7	110.7	108.9	0.7	1.7	-0.1	c
		NCF	109.0	111.8	112.8	112.8	2.8	3.8	3.8	c
Bindungslänge										
Mean error							0.0007	0.0057	0.0101	
Mean absolute error							0.0201	0.0227	0.0238	
Number of comparison							125	125	125	
Bindungswinkel										
Mean error							-0.22	-0.17	-0.16	
Mean absolute error							1.97	2.22	1.78	
Number of bond angles							69	69	69	

a K.P.Huber and G.Herzberg, "Molecular Spectra and Structure IV. Constants of Diatomic Molecules", van Nostrand, New York, 1979. b J.H.Callomon, E.Hirota, T.Iijima, K.Kuchitsu, and W.J.Lafferty, "Structure Data of Free Polyatomic Molecules", Landolt-Boernstein, New Series, Vol. II/15, Springer-Verlag, Berlin, 1987. c J.H.Callomon, E.Hirota, K.Kuchitsu, W.J.Lafferty, A.G.Maki, and C.S.Pote, "Structure Data of Free Polyatomic Molecules", Landolt-Boernstein, New Series, Vol. II/7, Springer-Verlag, Berlin, 1976. d Taken from the experimental data quoted in: M.J.S. Dewar, H.S. Rzepa, J.Am.Chem.Soc. 100, 58 (1978).

Tabelle 3.44: Ionisierungspotentiale [eV] für den Referenzdatensatz.

Ionisierungspotentiale (eV)									
Molekül	Exp.	Berechnet			Fehler			Lit.	
		OM1	OM2	OM3	OM1	OM2	OM3		
Fluorine	15.83	15.92	16.32	16.36	0.09	0.49	0.53	a	
Hydrogen fluoride	16.03	15.87	16.45	17.03	-0.16	0.42	1.00	a	
Methylfluoride	13.31	13.40	13.07	13.41	0.09	-0.24	0.10	a	
Difluoromethane	13.17	13.30	12.98	13.23	0.13	-0.19	0.06	a	
Trifluoromethane	14.67	14.43	14.17	14.45	-0.24	-0.50	-0.22	a	
Tetrafluoromethane	16.23	16.48	16.27	16.89	0.25	0.04	0.66	a	
Fluoroethane	12.43	12.46	12.01	12.28	0.03	-0.42	-0.15	a	
1,1-Difluoroethane	12.80	12.91	12.30	12.70	0.11	-0.50	-0.10	a	
1,1,1-Trifluoroethane	13.80	13.96	13.23	13.79	0.16	-0.57	-0.01	a	
Hexafluoroethane	14.60	14.24	13.67	13.92	-0.36	-0.93	-0.68	a	
... Continued ...									

**Tabelle 3.44:** (continued)

Ionisierungspotentiale (eV)								
Molekül	Exp.	Berechnet			Fehler			Lit.
		OM1	OM2	OM3	OM1	OM2	OM3	
Fluoroethene	10.58	10.66	10.59	10.85	0.08	0.01	0.27	a
1,2-Difluoroethene cis	10.43	10.46	10.38	10.56	0.03	-0.05	0.13	a
1,2-Difluoroethene trans	10.38	10.52	10.43	10.61	0.14	0.05	0.23	a
1,1-Difluoroethene	10.72	10.71	10.62	10.84	-0.01	-0.10	0.12	a
Trifluoroethene	10.53	10.51	10.40	10.53	-0.02	-0.13	0.00	a
Tetrafluoroethene	10.52	10.50	10.38	10.45	-0.02	-0.14	-0.07	a
Fluorobenzene	9.19	9.68	9.58	9.85	0.49	0.39	0.66	a
1,2-Difluorobenzene	9.68	9.83	9.71	9.92	0.15	0.03	0.24	a
1,3-Difluorobenzene	9.68	9.91	9.79	10.02	0.23	0.11	0.34	a
1,4-Difluorobenzene	9.15	9.71	9.58	9.80	0.56	0.43	0.65	a
Hexafluorobenzene	10.12	10.69	10.50	10.60	0.57	0.38	0.48	a
Fluoroacetylene	11.30	11.31	11.33	11.52	0.01	0.03	0.22	a
Hypofluorous acid	13.00	12.79	13.23	13.44	-0.21	0.23	0.44	a
1,1,1-Trifluoroethane	11.74	12.06	12.04	12.34	0.32	0.30	0.60	a
Trifluoromethylhypofluorite	13.64	13.90	14.01	14.20	0.26	0.37	0.56	a
Difluorine monoxide	13.26	13.25	13.56	13.78	-0.01	0.30	0.52	a
Acetyl fluoride	11.80	11.79	11.86	12.30	-0.01	0.06	0.50	a
Carbonyl fluoride	13.60	13.21	13.39	13.86	-0.39	-0.21	0.26	a
Hexafluoroacetone	12.09	12.11	12.03	12.16	0.02	-0.06	0.07	a
Trifluoroamine	13.73	13.25	13.29	13.68	-0.48	-0.44	-0.05	a
(E)-Difluorodiazene	12.80	12.46	12.47	12.45	-0.34	-0.33	-0.35	b
Tetrafluorohydrazine	12.84	12.90	12.87	12.59	0.06	0.03	-0.25	a
Difluorodiazacyclopr	11.78	11.67	11.41	11.58	-0.11	-0.37	-0.20	a
Hexafluorodimethyldiimine cis	11.35	10.96	11.20	10.98	-0.39	-0.15	-0.37	a
2,4,6-Trifluorotriazine	12.00	12.33	12.29	12.15	0.33	0.29	0.15	a
Pentafluoropyridine	10.27	10.65	10.59	10.65	0.38	0.32	0.38	a
Difluorocarbene	12.27	11.93	12.18	12.12	-0.34	-0.09	-0.15	a
Nitrosyl fluoride	12.94	12.23	12.80	13.00	-0.71	-0.14	0.06	a
Nitryl fluoride	13.51	12.84	13.69	14.08	-0.67	0.18	0.57	a
Mean error					0.00	-0.03	0.18	
Mean absolute error					0.23	0.26	0.32	
Number of comparison					39	39	39	

a Taken from the experimental data quoted in: M.J.S. Dewar, H.S. Rzepa, J.Am.Chem.Soc. 100, 58 (1978). b S.G.Lias, J.E.Bartmess, J.F.Liebman, J.L.Holmes, R.D.Levin, and W.G.Mallard, "Gas Phase Ion and Neutral Thermochemistry", J.Phys.Chem.Ref.Data 17, Suppl. 1 (1988).

**Tabelle 3.45:** Höhere Ionisierungspotentiale für den Referenzdatensatz.

Höhere Ionisierungspotentiale [eV]									
Molekül	Ionisierung	Exp.	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
Fluorine	$1\pi_u$	18.80	19.00	19.74	20.54	0.20	0.94	1.74	a
... Continued ...									

Tabelle 3.45: (continued)

Höhere Ionisierungspotentiale [eV]									
Molekül	Ionisierung	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
Methylfluoride	$3\sigma_g^+$	21.00	19.74	20.01	21.18	-1.26	-0.99	0.18	a
	$1e$	16.55	16.98	16.34	17.46	0.43	-0.21	0.91	a
	$5a_1$	17.56	17.25	17.49	18.19	-0.31	-0.07	0.63	a
	$4a_1$	23.48	25.51	24.50	24.37	2.03	1.02	0.89	a
Tetrafluoromethane	$3a_1$	38.41	31.21	29.12	29.14	-7.20	-9.29	-9.27	a
	$4t_2$	17.41	17.21	17.58	17.94	-0.20	0.17	0.53	a
	$1e$	18.43	17.54	17.96	18.31	-0.89	-0.47	-0.12	a
	$3t_2$	22.14	21.56	21.51	22.70	-0.58	-0.63	0.56	a
	$4a_1$	25.11	25.85	26.85	27.14	0.74	1.74	2.03	a
	$2t_2$	40.30	32.74	30.65	30.69	-7.56	-9.65	-9.61	a
	$3a_1$	43.81	35.33	32.49	32.55	-8.48	-11.32	-11.26	a
Tetrafluoroethene	$6a_g$	15.95	15.62	15.08	15.61	-0.33	-0.87	-0.34	a
	$4b_{1u}$	16.40	16.76	16.64	17.34	0.36	0.24	0.94	a
	$4b_{2g}$	16.60	16.84	16.94	17.49	0.24	0.34	0.89	a
	$1a_u$	16.90	17.16	17.59	18.01	0.26	0.69	1.11	a
	$1b_{3g}$	17.60	17.19	17.63	18.05	-0.41	0.03	0.45	a
	$5b_{3u}$	18.21	17.53	17.98	18.46	-0.68	-0.23	0.25	a
	$3b_{1u}$	19.46	21.11	20.70	21.81	1.65	1.24	2.35	a
	$5a_g$	21.00	21.51	20.88	22.23	0.51	-0.12	1.23	a
	$4b_{3u}$	22.50	23.63	24.11	24.81	1.13	1.61	2.31	a
	(E)-Difluorodiazen	$2a_u$	14.10	14.23	14.49	14.12	0.13	0.39	0.02
$1a_u$		18.70	19.07	19.76	20.13	0.37	1.06	1.43	a
$5b_u$		19.80	19.13	19.66	20.07	-0.67	-0.14	0.27	a
$5a_g$		21.00	20.11	20.06	20.08	-0.89	-0.94	-0.92	a
$4b_u$		22.70	24.46	25.75	26.74	1.76	3.05	4.04	a
Mean error					-0.76	-0.86	-0.34		
Mean absolute error					1.51	1.83	2.09		
Number of comparison					26	26	26		

a Taken from the experimental data quoted in: M.J.S. Dewar, H.S. Rzepa, J.Am.Chem.Soc. 100, 58 (1978).

Tabelle 3.46: Dipolmomente für den Referenzdatensatz.

Dipolmomente (D)									
Molekül	Exp.	Berechnet			Fehler			Lit.	
		OM1	OM2	OM3	OM1	OM2	OM3		
Hydrogen fluoride	1.83	1.60	1.53	1.53	-0.23	-0.30	-0.30	a	
Methylfluoride	1.86	1.93	1.70	1.61	0.07	-0.16	-0.25	b	
Difluoromethane	1.98	2.35	2.05	1.84	0.37	0.07	-0.14	b	
Trifluoromethane	1.65	2.34	2.07	1.78	0.69	0.42	0.13	b	
Fluoroethane	1.96	2.04	1.84	1.79	0.08	-0.12	-0.17	a	
1,1-Difluoroethane	2.30	2.60	2.40	2.30	0.30	0.10	0.00	a	
... Continued ...									



Tabelle 3.46: (continued)

Molekül	Dipolmomente (D)							Lit.
	Exp.	Berechnet			Fehler			
		OM1	OM2	OM3	OM1	OM2	OM3	
1,2-Difluoroethane gauche	2.67	2.64	2.33	2.24	-0.03	-0.34	-0.43	a
1,1,1-Trifluoroethane	2.35	2.74	2.68	2.57	0.39	0.33	0.22	b
1,1-Difluorocycloproane	2.32	2.19	2.05	1.88	-0.13	-0.27	-0.44	b
Fluoroethene	1.43	1.73	1.58	1.45	0.30	0.15	0.02	a
1,2-Difluoroethene cis	2.42	3.09	2.80	2.52	0.67	0.38	0.10	a
1,1-Difluoroethene	1.37	1.60	1.53	1.36	0.23	0.16	-0.01	b
Trifluoroethene	1.30	1.77	1.65	1.43	0.47	0.35	0.13	a
1,1-Difluoroallene	2.07	2.21	2.14	2.05	0.14	0.07	-0.02	a
Fluorobenzene	1.66	1.89	1.80	1.63	0.23	0.14	-0.03	a
1,2-Difluorobenzene	2.59	3.15	3.01	2.72	0.56	0.42	0.13	a
1,3-Difluorobenzene	1.51	1.85	1.78	1.62	0.34	0.27	0.11	a
Difluorobenzocycloprene	3.57	3.15	3.24	3.17	-0.42	-0.33	-0.40	a
Fluoroacetylene	0.72	1.18	1.20	0.93	0.46	0.48	0.21	b
H-CC-CO-F	2.98	3.28	3.58	3.43	0.30	0.60	0.45	b
Hypofluorous acid	2.23	2.02	1.87	1.97	-0.21	-0.36	-0.26	b
Trifluoromethylhypofluorite	0.33	0.15	0.50	0.59	-0.18	0.17	0.26	a
Difluorine monoxide	0.30	0.34	0.49	0.23	0.04	0.19	-0.07	a
Formyl fluoride	2.02	2.58	2.85	2.51	0.56	0.83	0.49	a
Acetyl fluoride	2.96	3.28	3.60	3.37	0.32	0.64	0.41	a
Carbonyl fluoride	0.95	0.86	1.34	1.20	-0.09	0.39	0.25	a
Hexafluoroacetone	0.65	0.05	0.42	0.39	-0.60	-0.23	-0.26	a
Trifluoroacetic acid	2.28	2.08	1.67	1.64	-0.20	-0.61	-0.64	a
Trifluoroamine	0.24	0.23	0.28	0.86	-0.01	0.04	0.62	a
(Z)-Difluorodiazene	0.16	0.02	0.09	0.49	-0.14	-0.07	0.33	a
Difluoroamine	1.92	1.93	1.71	1.68	0.01	-0.21	-0.24	a
N,N-Difluoromethylamine	2.57	2.54	2.60	3.00	-0.03	0.03	0.43	a
Difluorodiazacyclopropane	0.00	0.46	0.73	0.29	0.46	0.73	0.29	a
Fluorocyanide	2.12	1.74	1.55	1.65	-0.38	-0.57	-0.47	b
Trifluoroacetoneitrile	1.26	1.01	0.75	0.86	-0.25	-0.51	-0.40	b
Pentafluoropyridine	0.98	1.04	1.08	1.05	0.06	0.10	0.07	b
Difluorocarbene	0.47	0.56	0.50	0.30	0.09	0.03	-0.17	b
Nitrosyl fluoride	1.81	1.28	1.18	1.70	-0.53	-0.63	-0.11	a
Nitryl fluoride	0.47	0.49	0.17	0.83	0.02	-0.30	0.36	a
Mean error					0.10	0.05	0.01	
Mean absolute error					0.27	0.31	0.25	
Number of comparison					39	39	39	

a Taken from the experimental data quoted in: M.J.S. Dewar, H.S. Rzepa, J.Am.Chem.Soc. 100, 58 (1978). b J.Demaison, A.Dubrulle, W.Huettner, and E.Tiemann, "Molecular Constants", Landolt-Boernstein, New Series, Vol.II/14, Springer-Verlag, Berlin, 1982.

**Tabelle 3.47:** Schwingungswellenzahlen [ $\text{cm}^{-1}$ ] für den Referenzdatensatz.

Schwingungswellenzahlen ( $\text{cm}^{-1}$ )										
Molekül	Vibration	Exp	Berechnet			Fehler			Lit.	
			OM1	OM2	OM3	OM1	OM2	OM3		
Fluorine	F-F str	894	1461	1567	1511	567	673	617	a	
Hydrogen fluoride	H-F str	3956	3803	3816	3916	-153	-139	-39	a	
Tetrafluoromethane	C-F s-str $a_1$	909	1037	1083	1070	128	174	161	b	
	FCF d-def $e$	435	353	319	328	-82	-116	-107	b	
	C-F d-str $f_2$	1281	1467	1485	1461	186	204	180	b	
	FCF d-def $f_2$	632	516	475	497	-116	-157	-135	b	
Hexafluoroethane	CC str $a_{1g}$	1228	1471	1542	1469	243	314	241	b	
	CF3 s-str $a_{1g}$	807	861	875	864	54	68	57	b	
	CF3 s-def $a_{1g}$	348	325	318	320	-23	-30	-28	b	
	Torsion $a_{1u}$	68	28	13	18	-40	-55	-50	b	
	CF3 s-str $a_{2u}$	1117	1231	1294	1254	114	177	137	b	
	CF3 s-def $a_{2u}$	714	601	566	586	-113	-148	-128	b	
	CF3 d-str $e_g$	1250	1417	1416	1400	167	166	150	b	
	CF3 d-def $e_g$	619	444	408	420	-175	-211	-199	b	
	CF3 rock $e_u$	372	327	310	309	-45	-62	-63	b	
	CF3 d-str $e_u$	1251	1461	1481	1463	210	230	212	b	
	CF3 d-def $e_u$	520	542	526	531	22	6	11	b	
	CF3 rock $e_u$	220	179	168	178	-41	-52	-42	b	
	Tetrafluoroethene	CC str $a_g$	1872	2125	2127	2069	253	255	197	b
		CF2 s-str $a_g$	778	917	940	928	139	162	150	b
CF2 scis $a_g$		394	351	336	342	-43	-58	-52	b	
CF2 twist $a_u$		190	189	185	177	-1	-5	-13	b	
CF2 a-str $b_{1g}$		1340	1547	1546	1531	207	206	191	b	
CF2 rock $b_{1g}$		551	519	517	517	-32	-34	-34	b	
CF2 wag $b_{1u}$		406	384	394	396	-22	-12	-10	b	
CF2 wag $b_{2g}$		508	0	0	96	-508	-508	-412	b	
CF2 a-str $b_{2u}$		1337	1573	1596	1582	236	259	245	b	
CF2 rock $b_{2u}$		218	180	171	182	-38	-47	-36	b	
CF2 s-str $b_{3u}$		1186	1368	1428	1401	182	242	215	b	
CF2 scis $b_{3u}$		558	480	457	471	-78	-101	-87	b	
Difluoroacetylene		CF str $\sigma_g$	2450	2902	2881	2848	452	431	398	c
		CC str $\sigma_g$	794	915	945	935	121	151	141	c
	CF str $\sigma_u$	1349	1597	1671	1663	248	322	314	c	
	CF bend $\pi_g$	258	0	0	0	-258	-258	-258	c	
Difluorine monoxide	CF bend $\pi_u$	282	269	287	281	-13	5	-1	c	
	O-F s-str $a_1$	928	1324	1445	1355	396	517	427	b	
	FOF bend $a_1$	461	491	420	464	30	-41	3	b	
Trifluoroamine	O-F a-str $b_2$	831	1270	1358	1267	439	527	436	b	
	N-F s-str $a_1$	1032	1313	1374	1356	281	342	324	b	
	FNF s-def $a_1$	647	621	565	572	-26	-82	-75	b	
	N-F d-str $e$	907	1283	1263	1161	376	356	254	b	
	FNF d-def $e$	492	476	422	419	-16	-70	-73	b	
Mean error						76.9	85.7	76.6		
Mean absolute error						163.6	189.8	164.4		

... Continued ...

**Tabelle 3.47:** (continued)

Schwingungswellenzahlen (cm <sup>-1</sup> )									
Molekül	Vibration	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
Number of comparison						42	42	42	

a K.P.Huber and G.Herzberg, "Molecular Spectra and Structure IV. Constants of Diatomic Molecules", van Nostrand, New York, 1979. b T. Shimanouchi, "Tables of Molecular Vibrational Frequencies", Vol. I, NSRDS-NBS 39, U.S. Government Printing Office, Washington, 1972. c H.Buerger, W.Schneider, S.Sommer, W.Thiel, and H.Willner, J.Chem.Phys. 95, 5660 (1991).

**Tabelle 3.48:** Standardbildungsenthalpien für den Validierungssatz A: OM1, OM2 und OM3 Ergebnisse.

Standardbildungsenthalpien [kcal/mol]								
Molekül	G3	Berechnet			Fehler			Lit.
		OM1	OM2	OM3	OM1	OM2	OM3	
CH3F+	228.0	230.0	223.2	229.5	2.0	-4.8	1.5	a
CH2F2+	185.0	184.6	173.7	180.1	-0.4	-11.3	-4.9	a
CHF3+	154.0	152.1	137.2	143.8	-1.9	-16.8	-10.2	a
CH2F+	199.0	194.6	193.7	199.2	-4.4	-5.3	0.2	a
CHF2+	146.0	138.7	136.8	141.5	-7.3	-9.2	-4.5	a
CF3+	95.4	94.4	93.3	96.4	-1.0	-2.1	1.0	a
CH3CHF+	157.0	159.6	158.2	161.8	2.6	1.2	4.8	a
CH3CF2+	109.0	108.0	106.6	108.4	-1.0	-2.4	-0.6	a
CH2FCF2+	81.0	78.0	74.9	80.4	-3.0	-6.1	-0.6	a
CF3CH2+	120.0	110.0	111.2	118.7	-10.0	-8.8	-1.3	a
NF2+	275.0	273.4	281.6	268.6	-1.6	6.6	-6.4	a
CH2F	-8.0	-11.0	-13.6	-13.1	-3.0	-5.6	-5.1	2.0 a
CHF2	-57.0	-65.1	-68.0	-67.5	-8.1	-11.0	-10.5	1.0 a
CF3	-112.4	-120.4	-123.2	-122.6	-8.0	-10.8	-10.2	1.0 b
CF3CH2	-124.0	-126.6	-127.3	-124.3	-2.6	-3.3	-0.3	2.0 a
NF2	10.1	9.3	16.4	10.9	-0.8	6.3	0.8	1.9 b
FO	26.0	35.5	30.4	30.3	9.5	4.4	4.3	10.0 b
FO2	3.0	14.3	7.1	10.6	11.3	4.1	7.6	5.0 b
SINGLET HCF	34.2	26.1	24.9	24.4	-8.1	-9.3	-9.8	3.0 c
SINGLET CF2	-43.8	-56.0	-57.8	-55.7	-12.2	-14.0	-12.0	0.5 d
TRIPLET NF	59.5	58.7	64.0	60.2	-0.8	4.5	0.7	7.9 b
CF3-	-154.4	-168.9	-177.5	-173.8	-14.5	-23.1	-19.4	2.2 e
NF2-	-20.5	-8.4	-9.4	-13.2	12.1	11.1	7.3	4.5 e
Mean error					-2.23	-4.59	-2.94	
Mean absolute error					5.49	7.92	5.38	
Number of comparison					23	23	23	

a S.G.Lias, J.E.Bartmess, J.F.Liebman, J.L.Holmes, R.D.Levin, and W.G.Mallard, "Gas Phase Ion and Neutral Thermochemistry", J.Phys.Chem.Ref.Data 17, Suppl. 1 (1988). b M.W.Chase,

C.A.Davies, J.R.Downey, D.R.Frurip, R.A.McDonald, and A.N.Syverud, "JANAF Thermochemical Tables", 3rd edition, J.Phys.Chem.Ref.Data 14, Suppl.1 (1985). c J.C.Poutsma, J.A.Paulimo, and R.R.Squires, J.Phys.Chem. A 101, 5327 (1997). d J.Demaison, L.Margules, J.M.L.Martin, and J.E.Boggs, Phys.Chem.Chem.Phys. 4, 3282 (2002). e W.G.Mallard and P.J.Linstrom (eds.), "NIST Chemistry WebBook", NIST Standard Reference Data Base 69, February 2000, National Institute of Standards and Technology, Gaithersburg MD, 20899 (<http://webbook.nist.gov>).

**Tabelle 3.49:** Standardbildungsenthalpien für den Validierungssatz A: MNDO, AM1 und PM3 Ergebnisse.

Standardbildungsenthalpien [kcal/mol]								
Molekül	Exp	Berechnet			Fehler			Lit.
		MNDO	AM1	PM3	MNDO	AM1	PM3	
CH3F+	228.0	223.2	204.0	228.2	-4.8	-24.0	0.2	a
CH2F2+	185.0	178.4	151.6	180.4	-6.6	-33.4	-4.6	a
CHF3+	154.0	158.7	121.2	149.4	4.7	-32.8	-4.6	a
CH2F+	199.0	182.8	180.4	200.3	-16.2	-18.6	1.3	a
CHF2+	146.0	132.4	121.9	145.5	-13.6	-24.1	-0.5	a
CF3+	95.4	100.9	82.1	99.6	5.5	-13.3	4.2	a
CH3CHF+	157.0	165.0	157.6	173.6	8.0	0.6	16.6	a
CH3CF2+	109.0	116.6	105.3	122.2	7.6	-3.7	13.2	a
CH2FCF2+	81.0	82.4	69.1	92.7	1.4	-11.9	11.7	a
CF3CH2+	120.0	121.2	114.4	122.4	1.2	-5.6	2.4	a
NF2+	275.0	264.2	256.7	275.1	-10.8	-18.3	0.1	a
CH2F	-8.0	-33.3	-31.1	-19.9	-25.3	-23.1	-11.9	2.0 a
CHF2	-57.0	-88.3	-89.5	-75.3	-31.3	-32.5	-18.3	1.0 a
CF3	-112.4	-137.0	-142.8	-132.1	-24.6	-30.4	-19.7	1.0 b
CF3CH2	-124.0	-129.6	-131.4	-131.2	-5.6	-7.4	-7.2	2.0 a
NF2	10.1	-14.8	-6.4	11.9	-24.9	-16.5	1.8	1.9 b
FO	26.0	21.7	22.6	21.2	-4.3	-3.4	-4.8	10.0 b
FO2	3.0	24.1	9.0	12.9	21.1	6.0	9.9	5.0 b
SINGLET HCF	34.2	21.5	20.9	35.3	-12.7	-13.3	1.1	3.0 c
SINGLET CF2	-43.8	-65.2	-68.0	-49.1	-21.4	-24.2	-5.3	0.5 d
TRIPLET NF	59.5	31.6	41.4	55.1	-27.9	-18.1	-4.4	7.9 b
CF3-	-154.4	-178.8	-178.8	-178.8	-24.4	-24.4	-24.4	2.2 e
NF2-	-20.5	-43.6	-25.2	-31.0	-23.1	-4.7	-10.5	4.5 e
Mean error					-9.92	-16.40	-2.34	
Mean absolute error					14.23	16.97	7.76	
Number of comparison					23	23	23	

a - e gleiche Referenzen wie in der Tabelle 3.48

**Tabelle 3.50:** Standardbildungsenthalpien für den Validierungssatz B: OM1, OM2 und OM3 Ergebnisse.

Molekül	Standardbildungsenthalpien [kcal/mol]									
	G3	Exp.	Berechnet			Fehler (OM $x$ -G3)			Lit.	
			OM1	OM2	OM3	OM1	OM2	OM3	G3	Exp.
2: CH3F	-56.9	-55.6±2.0	-54.8	-60.4	-57.3	2.1	-3.5	-0.4	a	b
3: CH2F2	-108.4	-108.1±0.4	-104.7	-110.9	-106.6	3.7	-2.5	1.8	a	b
4: CHF3	-167.1	-166.7±0.6	-160.6	-166.8	-162.0	6.5	0.3	5.1	a	b
5: CF4	-223.9	-223.0±0.4	-217.2	-220.5	-215.8	6.7	3.4	8.1	a	b
7: CH2F	-7.7	-7.8±2.0	-11.0	-13.6	-13.1	-3.3	-5.9	-5.4	a	
8: CHF2	-58.6	-59.2±2.0	-65.1	-68.0	-67.5	-6.5	-9.4	-8.9	a	c
9: CF3	-112.2		-120.4	-123.2	-122.6	-8.2	-11.0	-10.4	a	
11: CHF	34.8	34.2±3.0	26.1	24.9	24.4	-8.7	-9.9	-10.4	a	c
12: CF2	-46.6	-44.6±0.0	-56.0	-57.8	-55.7	-9.4	-11.2	-9.1	a	d
14: CF	58.0	59.4±0.3	53.0	52.3	52.7	-5.0	-5.7	-5.3	a	f
16: CHFO	-92.0	-91.6±1.7	-87.3	-86.3	-88.2	4.7	5.7	3.8	a	g
17: CF2O	-145.7		-137.1	-136.5	-138.7	8.6	9.2	7.0	a	
19: CFO	-42.7	-38.5±1.7	-52.5	-50.7	-54.5	-9.8	-8.0	-11.8	a	h
21: CH2FOH	-101.9		-102.0	-103.2	-100.6	-0.1	-1.3	1.3	a	
22: CHF2OH	-161.6		-159.3	-162.5	-157.8	2.3	-0.9	3.8	a	
23: CF3OH	-218.3	-217.7±2.0	-214.1	-215.1	-209.8	4.2	3.2	8.5	a	g
24: CH3OF	-21.5	-17.3±3.0	-15.8	-22.1	-19.8	5.7	-0.6	1.7	a	i
25: CH2FOF	-71.2		-62.7	-71.3	-66.4	8.5	-0.1	4.8	a	
26: CHF2OF	-123.9		-114.5	-123.5	-117.0	9.4	0.4	6.9	a	
27: CF3OF	-178.0	-173.0±2.0	-167.0	-174.0	-166.3	11.0	4.0	11.7	a	g
29: CH2FO	-48.9		-45.3	-49.9	-47.0	3.6	-1.0	1.9	a	
30: CHF2O	-98.0		-95.2	-99.0	-93.2	2.8	-1.0	4.8	a	
31: CF3O	-151.2	-149.2±2.0	-148.1	-150.5	-143.8	3.1	0.7	7.4	a	g
33: CHFOH	-53.0		-66.5	-64.2	-63.2	-13.5	-11.2	-10.2	a	
34: CF2OH	-108.7		-122.4	-122.6	-120.3	-13.7	-13.9	-11.6	a	
35: CH2OF	26.1		-7.8	-17.0	-12.0	-33.9	-43.1	-38.1	a	
36: CHFOF	-15.3		-24.3	-28.8	-24.6	-9.0	-13.5	-9.3	a	
38: CF3OOH	-193.1		-188.8	-193.9	-185.7	4.3	-0.8	7.4	a	
40: CF3OO	-152.9	-144.0±3.0	-150.6	-157.5	-147.8	2.3	-4.6	5.1	a	i
42: FCOOH	-146.9		-140.4	-136.4	-137.1	6.5	10.5	9.8	a	
44: FCOO-B2	-86.5		-68.4	-70.0	-66.5	18.1	16.5	20.0	a	
46: CF2OHOH	-212.3		-209.7	-210.6	-203.5	2.6	1.7	8.8	a	
48: OCF2OH	-146.9		-145.6	-145.5	-139.1	1.3	1.4	7.8	a	
50: CH3CH2F	-65.7	-66.1±1.0	-65.6	-67.5	-65.8	0.1	-1.8	-0.1	a	j
51: CH2FCH2F	-107.3	-103.7±2.8	-108.9	-111.3	-106.9	-1.6	-4.0	0.4	a	k
52: CH3CHF2	-121.3	-119.7±1.5	-119.0	-121.8	-119.4	2.3	-0.5	1.9	a	c
53: CHF2CH2F	-161.1	-158.9±1.0	-160.7	-164.1	-159.0	0.4	-3.0	2.1	a	l
54: CH3CF3	-181.3	-178.2±0.4	-176.9	-178.4	-176.3	4.4	2.9	5.0	a	c
55: CHF2CHF2	-212.5	-209.8±4.2	-210.0	-213.2	-206.0	2.5	-0.7	6.5	a	k
56: CH2FCF3	-219.0	-214.1±2.0	-215.5	-217.3	-212.3	3.5	1.7	6.7	a	c
57: CHF2CF3		-264.0±1.1								c
58: CF3CF3		-320.9±1.5								c
60: CH2FCH2	-14.7	-14.2±2.0	-19.6	-20.7	-19.2	-4.9	-6.0	-4.5	a	f
61: CH3CHF	-18.2	-16.8±2.0	-25.5	-25.9	-27.1	-7.3	-7.7	-8.9	a	f
62: CH2FCHF	-58.1	-57.0±3.0	-67.6	-68.8	-67.5	-9.5	-10.7	-9.4	a	k

... Continued ...

Molekül	Standardbildungsenthalpien [kcal/mol]									
	G3	Exp.	Berechnet			Fehler (OM $x$ -G3)			Lit.	
			OM1	OM2	OM3	OM1	OM2	OM3	G3	Exp.
63: CHF <sub>2</sub> CH <sub>2</sub>	-68.3	-68.3±3.6	-70.8	-72.6	-70.0	-2.5	-4.3	-1.7	a	k
64: CH <sub>3</sub> CF <sub>2</sub>	-71.9	-72.3±2.0	-81.2	-81.9	-83.2	-9.3	-10.0	-11.3	a	m
65: CH <sub>2</sub> FCF <sub>2</sub>	-110.0	-107.5±3.6	-120.0	-121.4	-119.5	-10.0	-11.4	-9.5	a	k
66: CHF <sub>2</sub> CHF	-110.6	-109.0±3.6	-117.1	-119.2	-115.8	-6.5	-8.6	-5.2	a	k
67: CF <sub>3</sub> CH <sub>2</sub>	-127.3	-123.6±1.0	-126.6	-127.3	-124.3	0.7	0.0	3.0	a	m
68: CF <sub>3</sub> CHF	-168.3	-162.7±2.3	-171.3	-172.1	-168.8	-3.0	-3.8	-0.5	a	n
69: CHF <sub>2</sub> CF <sub>2</sub>	-160.3	-158.9±4.5	-168.1	-169.9	-165.9	-7.8	-9.6	-5.6	a	k
70: CF <sub>3</sub> CF <sub>2</sub>		-213.0±1.0								m
72: CH <sub>2</sub> CHF	-34.4	-33.5±0.6	-33.1	-32.3	-32.2	1.3	2.1	2.2	a	e
73: CHFCHF-Z	-73.8	-71.0±2.4	-74.7	-74.3	-72.7	-0.9	-0.5	1.1	a	o
74: CHFCHF-E	-74.5	-70.0±2.4	-73.0	-72.2	-71.7	1.5	2.3	2.8	a	o
75: CH <sub>2</sub> CF <sub>2</sub>	-84.5	-80.4±1.0	-80.3	-78.4	-79.9	4.2	6.1	4.6	a	e
76: CHF <sub>2</sub> CF <sub>2</sub>	-120.1	-117.4±2.2	-117.6	-116.4	-115.9	2.5	3.7	4.2	a	e
77: CF <sub>2</sub> CF <sub>2</sub>	-162.3		-158.3	-158.5	-155.9	4.0	3.8	6.4	a	
79: CH <sub>2</sub> FCH	40.7		32.6	30.9	30.2	-8.1	-9.8	-10.5	a	
81: CF <sub>3</sub> CH	-58.2		-61.9	-60.9	-57.8	-3.7	-2.7	0.4	a	
82: CH <sub>3</sub> CF	17.9		-1.8	-3.9	-3.8	-19.7	-21.8	-21.7	a	
83: CH <sub>2</sub> FCF	-21.7		-39.1	-41.9	-38.4	-17.4	-20.2	-16.7	a	
84: CHF <sub>2</sub> CF	-67.7		-85.6	-87.4	-82.7	-17.9	-19.7	-15.0	a	
85: CF <sub>3</sub> CF	-124.0		-138.9	-138.8	-133.1	-14.9	-14.8	-9.1	a	
87: CHFCH-Z	28.7		26.7	27.1	26.9	-2.0	-1.6	-1.8	a	
88: CHFCH-E	28.3		27.5	28.4	27.7	-0.8	0.1	-0.6	a	
89: CH <sub>2</sub> CF	26.0		19.6	20.2	17.2	-6.4	-5.8	-8.8	a	
90: CHF <sub>2</sub> CF-Z	-10.8		-16.9	-16.7	-18.4	-6.1	-5.9	-7.6	a	
91: CHF <sub>2</sub> CF-E	-10.4		-16.9	-17.0	-18.1	-6.5	-6.6	-7.7	a	
92: CF <sub>2</sub> CH	-18.7		-16.1	-14.9	-16.3	2.6	3.8	2.4	a	
93: CF <sub>2</sub> CF	-54.0	-45.9±2.0	-57.3	-56.9	-58.7	-3.3	-2.9	-4.7	a	p
95: CH <sub>2</sub> FC	78.8		67.6	65.1	66.5	-11.2	-13.7	-12.3	a	
96: CHF <sub>2</sub> C	36.1		24.0	22.0	25.3	-12.1	-14.1	-10.8	a	
97: CF <sub>3</sub> C	-18.4		-27.3	-26.5	-21.9	-8.9	-8.1	-3.5	a	
99: HCCF	24.8	30.0±5.3	26.7	26.9	23.9	1.9	2.1	-0.9	a	d
100: FCCF	0.0	5.0±5.0	4.0	4.4	-2.6	4.0	4.4	-2.6	a	d
102: CHFC	71.5		59.2	58.6	61.0	-12.3	-12.9	-10.5	a	
103: CF <sub>2</sub> C	30.4		23.1	22.2	26.5	-7.3	-8.2	-3.9	a	
105: CCF	109.3	110.5±5.3	101.8	94.0	97.7	-7.5	-15.3	-11.6	a	k
107: CHF <sub>2</sub> CO	-38.8		-42.5	-37.8	-42.5	-3.7	1.0	-3.7	a	
108: CF <sub>2</sub> CO	-75.0		-77.5	-96.6	-78.0	-2.5	-21.6	-3.0	a	
110: CFCO	19.3		-9.9	-7.8	-7.3	-29.2	-27.1	-26.6	a	
112: CH <sub>2</sub> FCHO	-80.2		-88.2	-87.8	-87.8	-8.0	-7.6	-7.6	a	
113: CHF <sub>2</sub> CHO	-130.4		-136.5	-135.7	-133.6	-6.1	-5.3	-3.2	a	
114: CF <sub>3</sub> CHO	-186.5		-190.9	-188.1	-186.3	-4.4	-1.6	0.2	a	
115: CH <sub>3</sub> CFO	-105.8		-99.8	-100.0	-103.0	6.0	5.8	2.8	a	
116: CH <sub>2</sub> FCFO	-143.6		-139.0	-139.2	-139.7	4.6	4.4	3.9	a	
117: CHF <sub>2</sub> CFO	-190.7		-186.6	-187.1	-186.2	4.1	3.6	4.5	a	
118: CF <sub>3</sub> CFO	-246.3		-239.5	-238.5	-237.3	6.8	7.8	9.0	a	
120: CH <sub>2</sub> FCO	-41.9		-59.3	-57.4	-58.7	-17.4	-15.5	-16.8	a	
121: CHF <sub>2</sub> CO	-90.3		-107.2	-105.1	-104.1	-16.9	-14.8	-13.8	a	
122: CF <sub>3</sub> CO	-145.6		-162.4	-158.3	-157.5	-16.8	-12.7	-11.9	a	

... Continued ...

Standardbildungsenthalpien [kcal/mol]										
Molekül	G3	Exp.	Berechnet			Fehler (OM $x$ -G3)			Lit.	
			OM1	OM2	OM3	OM1	OM2	OM3	G3	Exp.
125: C3F6	-277.6	-275.3±1.1	-273.3	-272.8	-269.4	4.3	4.8	8.2	a	q
Mean error						-3.03	-4.56	-2.41		
Mean absolute error						6.98	7.14	7.15		
Number of comparison						91	91	91		

a N.L.Haworth, M.H.Smith, G.B.Bacskay, and J.C.Mackie, *J.Phys.Chem. A*, 104, 7600 (2000). b V.P.Kolesov, *Russ.Chem.Rev.* 47, 599 (1978). c D.F.McMillen, D.M.Golden, *Annu.Rev.Phys.Chem.*, 33, 493 (1982). d J.C.Poutsma, J.A.Paulimo, and R.R.Squires, *J.Phys.Chem. A* 101, 5327 (1997). e A.S.Rodgers, *ACS Symp.Ser.*, 66, 296 (1978). f D.A.Dixon, D.Feller, G.Sandrone, *J.Phys.Chem. A*, 103, 4744 (1999). g W.F.Schneider and T.J.Wallington, 98, 7448 (1994). h V.D.Knyazev, A.Bencsura, I.R.Slagle, *J.Phys.Chem.*, 101, 849 (1997). i L.Batt, R.Walsh, *Int.J.Chem.Kinet.*, 14, 933 (1982). j C.W.Bauschlicher, Jr., Ricca, A. *Chem.Phys.Lett.*, 315, 449 (1999). k J.Berkowitz, G.B.Ellison, D.Gutman, *J.Phys.Chem.*, 98, 2744 (1994). l L.Gurvich, I.V.Veyts, C.B.Alcock, Eds. *Thermodynamic Properties of Individual Substances*; Hemisphere Publishing Corp.: Bristol PA (1991). m R.K.Lengel, R.N.Zare, *J.Am.Chem.Soc.*, 100, 7495 (1978). n K.Sendt, G.B.Bacskay, *J.Chem.Phys.*, 112, 2227 (2000). o D.L.Baulch, R.A.Cox, R.F.Hampson, Jr., J.A.Kerr, J.Troe, R.T.Watson, *J.Phys.Chem.Ref.Data*, 13, 1259 (1984). p S.S.Chen, R.C.Wilhoit, B.J.Zwolinski, *J.Phys.Chem.Ref.Data*, 6, 105 (1977) q T.S.Papina, V.P.Kolesov, Yu.G.Golovanova, *Russ.J.Phys.Chem.*, 61, 1168 (19

kleinere Konvergenzkriterien der Geometrie für das Molekül FCOO-B2 (44) bei OM1.

**Tabelle 3.51:** Standardbildungsenthalpien für den Validierungssatz B: MNDO, AM1 und PM3 Ergebnisse.

Standardbildungsenthalpien [kcal/mol]								
Molekül	G3	Berechnet			Fehler			Lit.
		MNDO	AM1	PM3	MNDO	AM1	PM3	
2: CH3F	-56.9	-60.9	-61.0	-53.8	-4.0	-4.1	3.1	a
3: CH2F2	-108.4	-111.8	-116.1	-103.8	-3.4	-7.7	4.6	a
4: CHF3	-167.1	-163.8	-172.5	-162.0	3.3	-5.4	5.1	a
5: CF4	-223.9	-214.2	-225.7	-225.1	9.7	-1.8	-1.2	a
7: CH2F	-7.7	-33.3	-31.1	-19.9	-25.6	-23.4	-12.2	a
8: CHF2	-58.6	-88.3	-89.5	-75.3	-29.7	-30.9	-16.7	a
9: CF3	-112.2	-137.0	-142.8	-132.1	-24.8	-30.6	-19.9	a
11: CHF	34.8	21.5	20.9	35.3	-13.3	-13.9	0.5	a
12: CF2	-46.6	-65.2	-68.0	-49.1	-18.6	-21.4	-2.5	a
14: CF	58.0	38.6	38.0	54.0	-19.4	-20.0	-4.0	a
16: CHFO	-92.0	-88.8	-92.9	-88.8	3.2	-0.9	3.2	a
17: CF2O	-145.7	-138.6	-146.3	-141.6	7.1	-0.6	4.1	a
19: CFO	-42.7	-50.0	-55.9	-55.0	-7.3	-13.2	-12.3	a
21: CH2FOH	-101.9	-109.9	-115.0	-102.6	-8.0	-13.1	-0.7	a
22: CHF2OH	-161.6	-164.4	-174.8	-161.2	-2.8	-13.2	0.4	a
23: CF3OH	-218.3	-215.0	-227.0	-222.7	3.3	-8.7	-4.4	a

... Continued ...

Tabelle 3.51: (continued)

Molekül	Standardbildungsenthalpien [kcal/mol]							Lit.
	G3	Berechnet			Fehler			
		MNDO	AM1	PM3	MNDO	AM1	PM3	
24: CH3OF	-21.5	-11.8	-16.9	-24.8	9.7	4.6	-3.3	a
25: CH2FOF	-71.2	-63.1	-72.0	-73.9	8.1	-0.8	-2.7	a
26: CHF2OF	-123.9	-115.5	-127.9	-130.2	8.4	-4.0	-6.3	a
27: CF3OF	-178.0	-163.3	-177.7	-187.3	14.7	0.3	-9.3	a
29: CH2FO	-48.9	-50.5	-58.0	-55.2	-1.6	-9.1	-6.3	a
30: CHF2O	-98.0	-101.5	-108.7	-105.4	-3.5	-10.7	-7.4	a
31: CF3O	-151.2	-152.3	-161.0	-162.8	-1.1	-9.8	-11.6	a
33: CHFOH	-53.0	-86.4	-84.3	-77.2	-33.4	-31.3	-24.2	a
34: CF2OH	-108.7	-138.7	-141.6	-133.8	-30.0	-32.9	-25.1	a
35: CH2OF	26.1	16.9	16.6	6.2	-9.2	-9.5	-19.9	a
36: CHFOF	-15.3	-36.4	-38.9	-41.4	-21.1	-23.6	-26.1	a
38: CF3OOH	-193.1	-189.4	-195.8	-201.8	3.7	-2.7	-8.7	a
40: CF3OO	-152.9	-157.0	-168.4	-156.9	-4.1	-15.5	-4.0	a
42: FCOOH	-146.9	-143.1	-149.2	-145.3	3.8	-2.3	1.6	a
44: FCOO-B2	-86.5	-74.1	-77.0	-82.2	12.4	9.5	4.3	a
46: CF2OHOH	-212.3	-214.0	-225.6	-218.1	-1.7	-13.3	-5.8	a
48: OCF2OH	-146.9	-153.8	-161.9	-160.9	-6.9	-15.0	-14.0	a
50: CH3CH2F	-65.7	-65.1	-66.3	-60.2	0.6	-0.6	5.5	a
51: CH2FCH2F	-107.3	-109.7	-114.3	-100.5	-2.4	-7.0	6.8	a
52: CH3CHF2	-121.3	-113.4	-118.6	-111.9	7.9	2.7	9.4	a
53: CHF2CH2F	-161.1	-157.5	-165.4	-152.0	3.6	-4.3	9.1	a
54: CH3CF3	-181.3	-164.4	-172.6	-172.3	16.9	8.7	9.0	a
55: CHF2CHF2	-212.5	-204.6	-215.0	-202.0	7.9	-2.5	10.5	a
56: CH2FCF3	-219.0	-207.5	-217.1	-211.8	11.5	1.9	7.2	a
60: CH2FCH2	-14.7	-32.0	-29.5	-23.7	-17.3	-14.8	-9.0	a
61: CH3CHF	-18.2	-43.0	-39.6	-32.1	-24.8	-21.4	-13.9	a
62: CH2FCHF	-58.1	-87.0	-86.3	-71.5	-28.9	-28.2	-13.4	a
63: CHF2CH2	-68.3	-79.6	-79.6	-73.2	-11.3	-11.3	-4.9	a
64: CH3CF2	-71.9	-96.1	-94.0	-86.7	-24.2	-22.1	-14.8	a
65: CH2FCF2	-110.0	-138.7	-138.5	-125.5	-28.7	-28.5	-15.5	a
66: CHF2CHF	-110.6	-134.0	-135.3	-121.4	-23.4	-24.7	-10.8	a
67: CF3CH2	-127.3	-129.6	-131.4	-131.2	-2.3	-4.1	-3.9	a
68: CF3CHF	-168.3	-183.1	-186.2	-178.9	-14.8	-17.9	-10.6	a
69: CHF2CF2	-160.3	-183.9	-185.7	-172.4	-23.6	-25.4	-12.1	a
72: CH2CHF	-34.4	-34.5	-34.0	-28.6	-0.1	0.4	5.8	a
73: CHFCHF-Z	-73.8	-83.1	-82.7	-71.5	-9.3	-8.9	2.3	a
74: CHFCHF-E	-74.5	-83.6	-82.7	-73.0	-9.1	-8.2	1.5	a
75: CH2CF2	-84.5	-84.1	-84.3	-79.6	0.4	0.2	4.9	a
76: CHF2CF2	-120.1	-131.1	-130.6	-121.5	-11.0	-10.5	-1.4	a
77: CF2CF2	-162.3	-175.7	-175.1	-168.2	-13.4	-12.8	-5.9	a
79: CH2FCH	40.7	43.7	39.8	47.7	3.0	-0.9	7.0	a
81: CF3CH	-58.2	-48.0	-51.9	-48.2	10.2	6.3	10.0	a
82: CH3CF	17.9	8.2	5.5	13.8	-9.7	-12.4	-4.1	a
83: CH2FCF	-21.7	-34.1	-37.5	-21.7	-12.4	-15.8	0.0	a
84: CHF2CF	-67.7	-82.2	-87.1	-70.9	-14.5	-19.4	-3.2	a
85: CF3CF	-124.0	-131.6	-137.4	-126.4	-7.6	-13.4	-2.4	a

... Continued ...



Tabelle 3.51: (continued)

Molekül	Standardbildungsenthalpien [kcal/mol]							
	G3	Berechnet			Fehler			Lit.
		MNDO	AM1	PM3	MNDO	AM1	PM3	
87: CHFCH-Z	28.7	16.6	19.2	22.4	-12.1	-9.5	-6.3	a
88: CHFCH-E	28.3	16.6	19.2	22.4	-11.7	-9.1	-5.9	a
89: CH2CF	26.0	12.1	12.1	18.6	-13.9	-13.9	-7.4	a
90: CHFCH-Z	-10.8	-34.1	-33.1	-21.3	-23.3	-22.3	-10.5	a
91: CHFCH-E	-10.4	-33.4	-31.8	-19.0	-23.0	-21.4	-8.6	a
92: CF2CH	-18.7	-30.5	-26.9	-24.5	-11.8	-8.2	-5.8	a
93: CF2CF	-54.0	-78.2	-76.8	-66.1	-24.2	-22.8	-12.1	a
95: CH2FC	78.8	79.4	75.8	80.0	0.6	-3.0	1.2	a
96: CHF2C	36.1	34.1	30.7	36.7	-2.0	-5.4	0.6	a
97: CF3C	-18.4	-15.2	-18.1	-16.7	3.2	0.3	1.7	a
99: HCCF	24.8	15.7	15.2	18.1	-9.1	-9.6	-6.7	a
100: FCCF	0.0	-21.0	-19.6	-11.6	-21.0	-19.6	-11.6	a
102: CHFC	71.5	79.7	74.6	78.5	8.2	3.1	7.0	a
103: CF2C	30.4	34.5	31.9	36.7	4.1	1.5	6.3	a
105: CCF	109.3	108.9	103.0	114.6	-0.4	-6.3	5.3	a
107: CHF2CO	-38.8	-50.3	-46.5	-45.1	-11.5	-7.7	-6.3	a
108: CF2CO	-75.0	-92.5	-86.9	-88.4	-17.5	-11.9	-13.4	a
110: CFCO	19.3	-4.8	-4.3	-0.2	-24.1	-23.6	-19.5	a
112: CH2FCHO	-80.2	-86.4	-88.4	-83.8	-6.2	-8.2	-3.6	a
113: CHF2CHO	-130.4	-133.4	-136.7	-132.0	-3.0	-6.3	-1.6	a
114: CF3CHO	-186.5	-183.1	-187.5	-190.4	3.4	-1.0	-3.9	a
115: CH3CFO	-105.8	-96.5	-98.8	-98.7	9.3	7.0	7.1	a
116: CH2FCFO	-143.6	-138.6	-142.1	-135.8	5.0	1.5	7.8	a
117: CHF2CFO	-190.7	-183.7	-189.6	-183.8	7.0	1.1	6.9	a
118: CF3CFO	-246.3	-230.7	-237.6	-239.6	15.6	8.7	6.7	a
120: CH2FCO	-41.9	-56.5	-54.9	-55.7	-14.6	-13.0	-13.8	a
121: CHF2CO	-90.3	-100.8	-100.8	-102.2	-10.5	-10.5	-11.9	a
122: CF3CO	-145.6	-149.2	-151.5	-160.8	-3.6	-5.9	-15.2	a
125: C3F6	-277.6	-272.4	-277.5	-276.4	5.2	0.1	1.2	a
Mean error					-6.49	-9.73	-4.08	
Mean absolute error					11.13	11.01	7.76	
Number of comparison					91	91	91	

a N.L.Haworth, M.H.Smith, G.B.Bacskey, and J.C.Mackie, J.Phys.Chem. A, 104, 7600 (2000).

Tabelle 3.52: Bindungslängen [Å] und Bindungswinkel [°] für Validierungssatz B.

Molekül	Variable	Bindungslängen [Å] und Bindungswinkel [°]							
		G3	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
2: CH3F	F-C	1.390	1.390	1.375	1.356	0.000	-0.016	-0.034	a
	H-C-F	109.1	108.4	108.5	109.5	-0.7	-0.6	0.4	a
3: CH2F2	F-C	1.364	1.377	1.366	1.350	0.012	0.002	-0.014	a

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]										
Molekül	Variable	Exp	Berechnet			Fehler			Lit.	
			OM1	OM2	OM3	OM1	OM2	OM3		
4: CHF <sub>3</sub>	F-C	1.364	1.377	1.366	1.350	0.012	0.002	-0.014	a	
	F-C-F	108.9	103.0	102.9	104.8	-5.8	-6.0	-4.1	a	
	H-C-F	108.8	110.1	110.2	110.1	1.3	1.4	1.3	a	
	F-C	1.343	1.363	1.358	1.344	0.020	0.015	0.001	a	
	F-C	1.343	1.363	1.359	1.344	0.020	0.016	0.001	a	
	F-C	1.343	1.363	1.359	1.344	0.020	0.016	0.001	a	
	F-C-H	110.5	113.5	113.6	112.7	3.0	3.1	2.2	a	
5: CF <sub>4</sub>	F-C-F	108.5	105.2	105.1	106.1	-3.3	-3.4	-2.4	a	
	F-C-F	108.4	105.2	105.1	106.1	-3.3	-3.4	-2.4	a	
	C-F	1.329	1.346	1.347	1.334	0.017	0.018	0.004	a	
	F-C	1.329	1.346	1.347	1.334	0.017	0.018	0.005	a	
	F-C	1.329	1.346	1.347	1.334	0.017	0.018	0.005	a	
	F-C	1.329	1.346	1.347	1.334	0.017	0.018	0.005	a	
	F-C-F	109.5	109.5	109.5	109.5	0.0	0.0	0.0	a	
7: CH <sub>2</sub> F <sub>2</sub>	F-C-F	109.5	109.5	109.5	109.5	0.0	0.0	0.0	a	
	F-C-F	109.5	109.5	109.5	109.5	0.0	0.0	0.0	a	
	F-C	1.350	1.343	1.330	1.313	-0.007	-0.019	-0.037	a	
	H-C-F	114.3	114.7	115.6	115.6	0.4	1.4	1.4	a	
	F-C	1.337	1.341	1.335	1.319	0.004	-0.001	-0.018	a	
	F-C	1.337	1.341	1.335	1.319	0.004	-0.001	-0.018	a	
	H-C-F	113.6	117.3	117.3	117.0	3.7	3.7	3.4	a	
8: CHF <sub>2</sub>	F-C-H	113.6	117.3	117.3	117.0	3.7	3.7	3.4	a	
	F-C	1.327	1.335	1.335	1.319	0.008	0.008	-0.008	a	
	F-C	1.327	1.335	1.335	1.319	0.008	0.008	-0.008	a	
	F-C	1.327	1.335	1.335	1.319	0.008	0.008	-0.008	a	
	F-C-F	111.2	110.6	111.0	111.9	-0.6	-0.2	0.7	a	
	F-C-F	111.2	110.6	111.0	111.9	-0.6	-0.2	0.7	a	
	F-C	1.319	1.320	1.312	1.297	0.002	-0.007	-0.022	a	
9: CF <sub>3</sub>	F-C-H	102.0	105.0	105.1	104.9	3.0	3.1	3.0	a	
	F-C	1.313	1.333	1.331	1.317	0.020	0.018	0.004	a	
	F-C	1.313	1.333	1.331	1.317	0.020	0.018	0.004	a	
	F-C-F	104.2	99.9	100.4	101.5	-4.3	-3.8	-2.7	a	
	F-C	1.330	1.295	1.292	1.278	-0.035	-0.038	-0.052	a	
	F-C	1.352	1.369	1.360	1.343	0.017	0.008	-0.009	a	
	F-C-H	109.1	113.1	113.4	111.9	3.9	4.2	2.8	a	
11: CHF	O-C-F	123.2	116.3	115.6	117.1	-6.9	-7.6	-6.1	a	
	F-C	1.326	1.338	1.337	1.322	0.012	0.011	-0.004	a	
	F-C	1.326	1.338	1.337	1.322	0.012	0.011	-0.004	a	
	F-C-F	107.5	107.4	107.2	107.2	-0.1	-0.2	-0.3	a	
	O-C-F	126.3	126.3	126.4	126.4	0.0	0.1	0.2	a	
	F-C	1.340	1.334	1.335	1.312	-0.006	-0.005	-0.028	a	
	O-C-F	128.0	128.0	128.5	132.0	0.0	0.5	4.0	a	
12: CF <sub>2</sub>	F-C	1.390	1.398	1.387	1.368	0.008	-0.003	-0.022	a	
	F-C-H	108.5	109.3	109.6	109.7	0.8	1.1	1.3	a	
	H-C-F	106.5	107.6	107.7	107.9	1.0	1.2	1.4	a	
	F-C	1.366	1.382	1.377	1.360	0.016	0.011	-0.006	a	
	14: CF	16: CHFO	17: CF <sub>2</sub> O	19: CFO	21: CH <sub>2</sub> FOH	22: CHF <sub>2</sub> OH				

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
23: CF3OH	F-C	1.366	1.382	1.377	1.360	0.016	0.011	-0.006	a
	F-C-H	109.8	112.8	113.0	112.2	2.9	3.1	2.4	a
	F-C-F	105.7	101.9	102.1	103.2	-3.9	-3.7	-2.5	a
	F-C	1.330	1.345	1.344	1.330	0.014	0.013	-0.001	a
	F-C	1.351	1.364	1.364	1.350	0.014	0.014	-0.001	a
	F-C	1.351	1.364	1.364	1.350	0.014	0.013	-0.001	a
	F-C-O	108.3	110.0	108.8	109.5	1.7	0.5	1.2	a
24: CH3OF	F-C-F	108.8	108.4	108.8	108.7	-0.4	0.0	-0.1	a
	F-C-F	106.5	105.7	106.1	106.1	-0.7	-0.3	-0.4	a
	F-O	1.450	1.400	1.415	1.386	-0.050	-0.035	-0.064	a
	F-O-C	103.1	106.3	106.7	107.2	3.2	3.6	4.1	a
25: CH2FOF	F-C	1.371	1.387	1.378	1.358	0.016	0.006	-0.013	a
	F-O	1.451	1.398	1.412	1.383	-0.054	-0.040	-0.068	a
26: CHF2OF	F-C-H	109.7	109.4	109.5	109.9	-0.3	-0.2	0.2	a
	H-C-F	108.8	108.9	109.3	109.4	0.1	0.4	0.6	a
	F-O-C	104.2	107.3	107.4	108.0	3.1	3.2	3.8	a
	F-C	1.347	1.368	1.362	1.344	0.022	0.015	-0.002	a
	F-C	1.347	1.368	1.362	1.344	0.021	0.015	-0.002	a
	F-O	1.446	1.390	1.401	1.372	-0.055	-0.044	-0.073	a
	F-C-H	111.1	113.2	113.6	113.1	2.1	2.4	1.9	a
27: CF3OF	F-C-F	107.8	103.6	103.9	105.1	-4.2	-3.9	-2.7	a
	F-O-C	104.3	107.7	107.8	109.0	3.5	3.5	4.7	a
	F-C	1.330	1.346	1.345	1.330	0.015	0.014	-0.001	a
	F-C	1.330	1.351	1.351	1.335	0.021	0.021	0.005	a
	F-C	1.330	1.351	1.351	1.335	0.021	0.021	0.005	a
	F-O	1.442	1.387	1.397	1.367	-0.054	-0.045	-0.074	a
	F-C-O	103.5	107.1	106.1	106.2	3.6	2.6	2.7	a
29: CH2FO	F-C-F	110.3	109.1	109.5	109.8	-1.2	-0.8	-0.5	a
	F-C-F	108.6	107.7	108.0	108.2	-1.0	-0.7	-0.4	a
	F-O-C	102.9	106.8	106.6	107.4	3.9	3.7	4.5	a
	F-C	1.384	1.406	1.393	1.375	0.022	0.009	-0.009	a
	F-C-O	114.0	108.1	107.3	110.3	-5.9	-6.7	-3.8	a
	H-C-F	108.3	108.1	108.5	108.3	-0.2	0.2	0.0	a
	F-C	1.357	1.377	1.369	1.353	0.020	0.012	-0.004	a
30: CHF2O	F-C	1.357	1.377	1.369	1.353	0.020	0.012	-0.004	a
	F-C-O	111.9	108.5	109.4	111.8	-3.3	-2.5	-0.1	a
	F-C-F	106.9	102.6	103.4	104.4	-4.3	-3.5	-2.5	a
	H-C-F	110.3	112.7	112.8	111.6	2.4	2.5	1.3	a
	F-C	1.336	1.350	1.346	1.333	0.013	0.010	-0.003	a
	F-C	1.337	1.356	1.354	1.339	0.019	0.017	0.003	a
	F-C	1.337	1.356	1.354	1.340	0.019	0.017	0.003	a
31: CF3O	F-C-O	105.8	109.2	108.2	108.6	3.4	2.4	2.8	a
	F-C-F	109.9	108.8	109.5	109.4	-1.1	-0.4	-0.4	a
	F-C-F	107.7	106.5	107.1	107.0	-1.2	-0.7	-0.7	a
	F-C	1.345	1.349	1.343	1.323	0.004	-0.002	-0.022	a
	F-C-O	109.3	107.2	104.7	107.6	-2.1	-4.6	-1.7	a
	F-C	1.345	1.349	1.343	1.323	0.004	-0.002	-0.022	a
	F-C-O	109.3	107.2	104.7	107.6	-2.1	-4.6	-1.7	a

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
34: CF2OH	F-C-O-H	179.1	180.3	185.2	180.7	1.2	6.1	1.6	a
	F-C	1.330	1.338	1.340	1.320	0.008	0.009	-0.011	a
	F-C	1.353	1.359	1.359	1.340	0.006	0.006	-0.013	a
	F-C-O	109.8	111.5	110.0	111.5	1.7	0.2	1.7	a
35: CH2OF	F-C-O	113.6	113.6	113.5	114.2	0.0	-0.1	0.6	a
	F-O	1.456	1.757	1.651	1.672	0.301	0.195	0.216	a
	F-O-C	105.2	107.2	106.1	107.7	2.0	1.0	2.6	a
36: CHFOF	F-O	1.441	1.391	1.402	1.373	-0.050	-0.039	-0.068	a
	F-C	1.337	1.344	1.335	1.315	0.007	-0.003	-0.023	a
38: CF3OOH	F-O-C	102.6	106.8	105.9	106.6	4.2	3.3	4.0	a
	F-C-O	106.3	106.1	104.8	107.3	-0.3	-1.5	0.9	a
	F-C	1.333	1.349	1.348	1.334	0.016	0.016	0.001	a
	F-C	1.333	1.350	1.352	1.334	0.018	0.019	0.001	a
40: CF3OO	F-C	1.342	1.359	1.362	1.344	0.017	0.020	0.002	a
	F-C-O	105.1	107.7	107.8	107.2	2.6	2.8	2.2	a
	F-C-O	112.8	113.2	113.2	112.6	0.4	0.4	-0.2	a
	F-C-O	111.8	111.5	111.2	111.1	-0.4	-0.7	-0.7	a
	F-C	1.326	1.344	1.345	1.328	0.018	0.019	0.002	a
	F-C	1.328	1.347	1.348	1.330	0.020	0.020	0.003	a
42: FCOOH	F-C	1.328	1.347	1.348	1.330	0.020	0.020	0.003	a
	F-C-O	105.0	107.1	107.1	106.5	2.0	2.1	1.5	a
	F-C-O	110.8	111.4	111.2	110.6	0.5	0.4	-0.3	a
	F-C-O	110.8	111.4	111.2	110.6	0.5	0.3	-0.2	a
	F-C	1.335	1.344	1.339	1.323	0.009	0.004	-0.012	a
	F-C-O	124.1	123.4	124.2	124.2	-0.7	0.0	0.0	a
44: FCOO-B2	O-C-F	107.5	108.3	107.0	108.1	0.9	-0.5	0.6	a
	F-C	1.324	1.346	1.336	1.323	0.022	0.012	-0.002	a
46: CF2OHOH	O-C-F	121.3	121.8	123.6	122.6	0.5	2.3	1.3	a
	F-C	1.375	1.385	1.369	1.367	0.010	-0.006	-0.008	a
	F-C	1.375	1.385	1.369	1.367	0.010	-0.006	-0.008	a
48: OCF2OH	F-C-F	103.6	102.1	104.7	102.8	-1.5	1.1	-0.8	a
	O-C-F	111.4	110.9	110.0	111.2	-0.5	-1.4	-0.2	a
	F-C	1.354	1.375	1.366	1.355	0.020	0.012	0.001	a
	F-C	1.365	1.374	1.374	1.355	0.009	0.009	-0.010	a
	F-C-F	105.8	103.0	104.4	103.9	-2.7	-1.4	-1.9	a
50: CH3CH2F	O-C-F	110.1	111.8	112.5	112.5	1.7	2.4	2.3	a
	F-C	1.399	1.394	1.385	1.366	-0.005	-0.014	-0.033	a
	F-C-C	109.2	107.0	107.3	109.0	-2.1	-1.8	-0.1	a
51: CH2FCH2F	F-C	1.395	1.390	1.380	1.360	-0.005	-0.015	-0.035	a
	F-C	1.395	1.390	1.380	1.360	-0.005	-0.015	-0.035	a
	F-C-C	107.6	105.9	107.0	108.5	-1.7	-0.7	0.8	a
52: CH3CHF2	F-C-C	107.6	105.9	107.0	108.5	-1.7	-0.7	0.8	a
	F-C	1.373	1.378	1.376	1.359	0.005	0.003	-0.014	a
	F-C	1.373	1.378	1.375	1.359	0.005	0.002	-0.014	a
	F-C-C	109.7	109.0	109.5	110.2	-0.7	-0.1	0.5	a
	F-C-C	109.7	109.0	109.6	110.2	-0.7	-0.1	0.5	a

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
53: CHF <sub>2</sub> CH <sub>2</sub> F	F-C	1.388	1.389	1.379	1.360	0.001	-0.009	-0.029	a
	F-C	1.364	1.376	1.373	1.356	0.012	0.008	-0.009	a
	F-C	1.371	1.376	1.373	1.356	0.005	0.002	-0.015	a
	F-C-C	108.3	106.0	106.8	108.4	-2.3	-1.4	0.2	a
	F-C-C	109.5	108.6	109.6	110.2	-0.9	0.0	0.7	a
	F-C-C	107.8	108.5	109.6	110.2	0.8	1.8	2.4	a
54: CH <sub>3</sub> CF <sub>3</sub>	F-C	1.353	1.362	1.365	1.351	0.010	0.013	-0.002	a
	F-C	1.353	1.362	1.365	1.351	0.010	0.013	-0.002	a
	F-C	1.353	1.362	1.365	1.351	0.010	0.013	-0.002	a
	F-C-C	111.6	112.9	113.9	113.2	1.3	2.2	1.6	a
	F-C-C	111.6	112.9	113.9	113.3	1.3	2.2	1.6	a
55: CHF <sub>2</sub> CHF <sub>2</sub>	F-C	1.362	1.371	1.368	1.350	0.009	0.005	-0.012	a
	F-C	1.362	1.371	1.368	1.350	0.009	0.005	-0.012	a
	F-C	1.362	1.371	1.367	1.350	0.009	0.005	-0.012	a
	F-C	1.362	1.371	1.368	1.350	0.009	0.005	-0.012	a
	F-C-C	108.2	107.4	108.6	108.9	-0.8	0.4	0.7	a
	F-C-C	108.2	107.5	108.6	108.9	-0.7	0.4	0.7	a
	F-C-C	108.2	107.5	108.6	108.9	-0.7	0.4	0.7	a
56: CH <sub>2</sub> FCF <sub>3</sub>	F-C	1.382	1.380	1.370	1.351	-0.002	-0.012	-0.031	a
	F-C	1.352	1.363	1.365	1.350	0.011	0.013	-0.002	a
	F-C	1.344	1.357	1.360	1.344	0.012	0.015	0.000	a
	F-C	1.344	1.357	1.360	1.344	0.012	0.015	0.000	a
	F-C-C	108.4	106.4	107.3	109.2	-2.0	-1.1	0.7	a
	F-C-C	109.1	110.8	112.0	111.1	1.7	2.9	2.0	a
	F-C-C	111.6	113.2	114.1	113.5	1.7	2.5	1.9	a
	F-C-C	111.6	113.2	114.1	113.5	1.7	2.5	1.9	a
57: CHF <sub>2</sub> CF <sub>3</sub>	F-C	1.338	1.351	1.354	1.339	0.014	0.017	0.001	a
	F-C	1.344	1.358	1.360	1.344	0.014	0.016	0.000	a
	F-C	1.344	1.358	1.360	1.344	0.014	0.016	0.000	a
	F-C	1.357	1.366	1.362	1.345	0.008	0.004	-0.013	a
	F-C	1.357	1.366	1.362	1.345	0.009	0.004	-0.013	a
	F-C-C	111.0	113.8	114.5	113.9	2.8	3.5	2.9	a
	F-C-C	109.6	111.1	112.2	111.3	1.4	2.5	1.6	a
	F-C-C	109.6	111.1	112.2	111.3	1.5	2.6	1.6	a
	F-C-C	108.2	107.8	108.7	109.2	-0.4	0.6	1.1	a
	F-C-C	108.2	107.8	108.7	109.2	-0.4	0.6	1.1	a
58: CF <sub>3</sub> CF <sub>3</sub>	F-C	1.339	1.352	1.354	1.339	0.014	0.016	0.000	a
	F-C	1.339	1.352	1.354	1.339	0.014	0.016	0.000	a
	F-C	1.339	1.352	1.354	1.339	0.014	0.016	0.000	a
	F-C	1.339	1.352	1.354	1.339	0.014	0.016	0.000	a
	F-C	1.339	1.352	1.354	1.339	0.014	0.016	0.000	a
	F-C	1.339	1.352	1.354	1.339	0.014	0.016	0.000	a
	F-C-C	109.7	111.4	112.4	111.5	1.7	2.7	1.9	a
	F-C-C	109.7	111.4	112.4	111.5	1.7	2.7	1.9	a

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
60: CH <sub>2</sub> FCH <sub>2</sub>	F-C-C	109.7	111.4	112.4	111.6	1.7	2.7	1.9	a
	F-C-C	109.7	111.4	112.4	111.5	1.7	2.7	1.9	a
	F-C-C	109.7	111.4	112.4	111.6	1.7	2.7	1.9	a
	F-C-C	109.7	111.4	112.4	111.5	1.7	2.7	1.9	a
	F-C	1.398	1.394	1.386	1.366	-0.004	-0.012	-0.032	a
61: CH <sub>3</sub> CHF	F-C-C	109.9	108.2	108.3	110.3	-1.7	-1.6	0.4	a
	F-C	1.361	1.351	1.347	1.328	-0.010	-0.014	-0.033	a
62: CH <sub>2</sub> FCHF	F-C-C	114.4	112.7	113.3	114.3	-1.8	-1.1	-0.1	a
	F-C	1.356	1.346	1.341	1.321	-0.010	-0.015	-0.035	a
63: CHF <sub>2</sub> CH <sub>2</sub>	F-C	1.392	1.391	1.382	1.363	-0.001	-0.009	-0.028	a
	F-C-C	113.0	112.8	114.2	115.5	-0.3	1.2	2.4	a
	F-C-C	108.1	107.0	107.7	109.8	-1.1	-0.3	1.7	a
	F-C	1.374	1.378	1.375	1.358	0.005	0.001	-0.016	a
	F-C	1.374	1.378	1.375	1.358	0.005	0.001	-0.016	a
64: CH <sub>3</sub> CF <sub>2</sub>	F-C-C	110.0	109.8	110.4	111.5	-0.1	0.4	1.5	a
	F-C-C	110.0	109.9	110.4	111.5	-0.1	0.4	1.5	a
	F-C	1.347	1.346	1.349	1.331	-0.001	0.001	-0.016	a
	F-C	1.347	1.346	1.349	1.331	-0.001	0.001	-0.016	a
65: CH <sub>2</sub> FCF <sub>2</sub>	F-C-C	114.4	115.9	116.2	116.3	1.5	1.8	1.9	a
	F-C-C	114.4	115.9	116.2	116.3	1.5	1.8	1.9	a
	F-C	1.338	1.337	1.341	1.325	0.000	0.003	-0.013	a
	F-C	1.344	1.344	1.345	1.326	0.000	0.001	-0.018	a
	F-C	1.384	1.383	1.372	1.354	-0.001	-0.012	-0.030	a
66: CHF <sub>2</sub> CHF	F-C-C	114.3	116.8	117.1	117.1	2.5	2.8	2.8	a
	F-C-C	112.4	114.2	115.1	115.8	1.8	2.7	3.4	a
	F-C-C	108.8	106.8	107.4	108.7	-2.0	-1.4	-0.1	a
	F-C	1.346	1.339	1.333	1.313	-0.007	-0.013	-0.033	a
	F-C	1.368	1.378	1.369	1.359	0.010	0.001	-0.009	a
	F-C	1.368	1.373	1.376	1.352	0.005	0.008	-0.015	a
	F-C-C	113.7	112.6	114.7	115.6	-1.2	1.0	1.9	a
	F-C-C	107.7	108.7	108.7	110.0	1.0	0.9	2.3	a
67: CF <sub>3</sub> CH <sub>2</sub>	F-C-C	109.9	107.9	109.6	109.3	-2.1	-0.3	-0.6	a
	F-C	1.358	1.363	1.365	1.351	0.005	0.008	-0.007	a
	F-C	1.351	1.362	1.364	1.350	0.010	0.013	-0.002	a
	F-C	1.351	1.361	1.364	1.350	0.010	0.013	-0.002	a
	F-C-C	112.2	112.3	113.3	112.6	0.1	1.2	0.4	a
68: CF <sub>3</sub> CHF	F-C-C	111.5	112.9	113.8	113.3	1.5	2.4	1.8	a
	F-C-C	111.5	113.1	113.9	113.4	1.6	2.5	1.9	a
	F-C	1.353	1.362	1.365	1.351	0.009	0.012	-0.002	a
	F-C	1.348	1.359	1.361	1.345	0.011	0.013	-0.003	a
	F-C	1.344	1.355	1.357	1.342	0.011	0.013	-0.002	a
	F-C	1.340	1.332	1.326	1.307	-0.007	-0.014	-0.033	a
	F-C-C	112.4	112.7	113.8	112.8	0.3	1.5	0.5	a
	F-C-F	107.8	106.1	105.1	105.9	-1.6	-2.7	-1.9	a
F-C-F	108.9	106.7	105.8	106.6	-2.2	-3.1	-2.3	a	
F-C-C	113.8	113.1	115.0	116.2	-0.8	1.2	2.4	a	

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
69: CHF <sub>2</sub> CF <sub>2</sub>	F-C	1.335	1.337	1.338	1.320	0.002	0.003	-0.015	a
	F-C	1.335	1.337	1.338	1.320	0.002	0.003	-0.015	a
	F-C	1.360	1.369	1.364	1.347	0.009	0.004	-0.013	a
	F-C	1.360	1.369	1.364	1.347	0.009	0.004	-0.013	a
	F-C-C	112.9	115.0	115.9	115.7	2.0	3.0	2.8	a
	F-C-C	112.9	114.9	115.9	115.7	2.0	2.9	2.8	a
	F-C-C	108.3	108.0	108.9	109.4	-0.3	0.6	1.0	a
	F-C-C	108.3	108.0	108.9	109.4	-0.3	0.6	1.0	a
70: CF <sub>3</sub> CF <sub>2</sub>	F-C	1.347	1.359	1.362	1.346	0.012	0.015	-0.001	a
	F-C	1.330	1.331	1.332	1.314	0.001	0.001	-0.016	a
	F-C	1.330	1.331	1.332	1.315	0.001	0.001	-0.016	a
	F-C	1.340	1.354	1.355	1.340	0.014	0.015	0.000	a
	F-C	1.340	1.354	1.355	1.340	0.014	0.015	0.000	a
	F-C-C	111.2	112.5	113.6	112.9	1.3	2.5	1.8	a
	F-C-C	113.0	115.5	116.5	116.5	2.5	3.6	3.6	a
	F-C-C	113.0	115.5	116.5	116.5	2.5	3.6	3.5	a
72: CH <sub>2</sub> CHF	F-C	1.354	1.360	1.356	1.336	0.006	0.002	-0.018	a
	F-C-C	122.0	117.5	118.0	119.3	-4.5	-4.0	-2.7	a
	F-C	1.353	1.356	1.350	1.330	0.003	-0.002	-0.022	a
	F-C	1.353	1.356	1.350	1.330	0.003	-0.002	-0.023	a
73: CHFCHF-Z	F-C-H	114.8	116.3	115.6	115.6	1.5	0.7	0.8	a
	F-C-C	119.8	116.3	117.4	118.9	-3.5	-2.4	-0.9	a
	F-C	1.348	1.353	1.349	1.329	0.005	0.000	-0.019	a
	F-C	1.348	1.353	1.349	1.329	0.005	0.000	-0.019	a
74: CHFCHF-E	F-C-H	114.8	115.9	115.1	115.2	1.1	0.3	0.4	a
	F-C-C	122.0	118.3	118.3	120.4	-3.8	-3.7	-1.6	a
	F-C	1.330	1.340	1.342	1.325	0.010	0.012	-0.005	a
	F-C	1.330	1.340	1.342	1.325	0.010	0.012	-0.005	a
75: CH <sub>2</sub> CF <sub>2</sub>	F-C-C	125.0	125.6	126.4	125.9	0.6	1.4	0.9	a
	F-C-C	125.0	125.6	126.4	125.9	0.6	1.4	0.9	a
	F-C	1.351	1.351	1.346	1.326	0.001	-0.005	-0.025	a
	F-C	1.329	1.339	1.340	1.323	0.010	0.011	-0.006	a
76: CHF <sub>2</sub> CF <sub>2</sub>	F-C	1.324	1.335	1.336	1.320	0.011	0.012	-0.004	a
	F-C-H	116.5	117.7	117.0	117.1	1.2	0.5	0.6	a
	F-C-C	122.6	123.3	124.4	123.6	0.6	1.7	0.9	a
	F-C-C	124.9	126.5	126.9	126.6	1.6	2.0	1.7	a
	F-C	1.327	1.335	1.337	1.319	0.009	0.010	-0.008	a
	F-C	1.327	1.335	1.337	1.319	0.008	0.010	-0.008	a
77: CF <sub>2</sub> CF <sub>2</sub>	F-C	1.327	1.335	1.337	1.319	0.009	0.010	-0.008	a
	F-C	1.327	1.335	1.337	1.319	0.008	0.010	-0.008	a
	F-C	1.327	1.335	1.337	1.319	0.009	0.010	-0.008	a
	F-C	1.327	1.335	1.337	1.319	0.008	0.010	-0.008	a
	F-C-C	123.0	121.7	120.4	124.2	-1.3	-2.6	1.2	a
	F-C-F	114.0	109.5	106.6	111.6	-4.5	-7.4	-2.4	a
	F-C-C	123.0	121.6	120.4	124.2	-1.4	-2.6	1.2	a
	F-C-C	123.0	121.7	120.4	124.2	-1.3	-2.6	1.2	a

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
79: CH <sub>2</sub> FCH	F-C	1.405	1.403	1.395	1.373	-0.001	-0.010	-0.032	a
	F-C-C	119.6	113.9	115.0	116.2	-5.6	-4.6	-3.3	a
81: CF <sub>3</sub> CH	F-C	1.350	1.370	1.375	1.356	0.020	0.025	0.006	a
	F-C	1.337	1.358	1.361	1.344	0.021	0.023	0.007	a
	F-C	1.367	1.364	1.362	1.351	-0.003	-0.005	-0.016	a
	F-C-C	117.7	116.8	118.5	115.1	-0.9	0.8	-2.5	a
	F-C-C	115.3	113.1	112.3	113.5	-2.1	-3.0	-1.8	a
	F-C-C	98.7	109.5	111.6	111.2	10.8	12.9	12.5	a
82: CH <sub>3</sub> CF	F-C	1.333	1.335	1.339	1.321	0.001	0.006	-0.012	a
	F-C-C	105.9	106.2	106.3	107.0	0.3	0.4	1.1	a
83: CH <sub>2</sub> FCF	F-C	1.332	1.331	1.333	1.314	-0.002	0.001	-0.018	a
	F-C	1.381	1.391	1.379	1.359	0.010	-0.002	-0.022	a
	F-C-C	103.5	104.0	105.2	105.8	0.5	1.7	2.3	a
	F-C-C	110.2	106.6	107.6	108.9	-3.6	-2.7	-1.3	a
84: CHF <sub>2</sub> CF	F-C	1.313	1.312	1.315	1.297	0.000	0.003	-0.016	a
	F-C	1.362	1.375	1.367	1.349	0.013	0.006	-0.013	a
	F-C	1.365	1.379	1.374	1.357	0.015	0.010	-0.008	a
	F-C-C	104.7	105.2	106.3	107.0	0.5	1.6	2.3	a
	F-C-C	107.8	107.9	110.0	110.7	0.1	2.2	2.8	a
	F-C-C	113.5	113.4	115.3	116.2	-0.1	1.8	2.7	a
85: CF <sub>3</sub> CF	F-C	1.343	1.367	1.369	1.354	0.024	0.027	0.011	a
	F-C	1.306	1.303	1.306	1.289	-0.003	-0.001	-0.017	a
	F-C	1.345	1.362	1.363	1.347	0.017	0.018	0.002	a
	F-C	1.345	1.362	1.363	1.347	0.017	0.018	0.002	a
	F-C-C	115.9	118.1	118.2	117.2	2.2	2.3	1.3	a
	F-C-C	104.8	104.8	105.9	106.2	-0.1	1.0	1.3	a
	F-C-C	107.2	110.8	112.0	111.2	3.6	4.9	4.0	a
	F-C-C	107.2	110.8	111.9	111.2	3.6	4.8	4.0	a
87: CHFCH-Z	F-C	1.362	1.361	1.359	1.336	-0.001	-0.003	-0.026	a
	F-C-H	111.3	116.1	114.8	115.2	4.8	3.5	3.8	a
88: CHFCH-E	F-C	1.353	1.358	1.352	1.332	0.005	-0.001	-0.021	a
	F-C-H	111.5	115.4	114.4	114.5	3.9	2.8	3.0	a
89: CH <sub>2</sub> CF	F-C	1.318	1.316	1.318	1.298	-0.002	0.001	-0.020	a
	F-C-C	130.8	131.6	132.2	133.4	0.9	1.5	2.6	a
90: CHF <sub>2</sub> CF-Z	F-C	1.358	1.357	1.354	1.332	-0.001	-0.004	-0.026	a
	F-C	1.312	1.309	1.308	1.289	-0.003	-0.003	-0.023	a
	F-C-H	114.4	117.3	116.5	116.6	2.9	2.0	2.2	a
	F-C-C	132.0	133.1	134.2	136.3	1.1	2.2	4.3	a
91: CHF <sub>2</sub> CF-E	F-C	1.351	1.353	1.345	1.326	0.002	-0.006	-0.024	a
	F-C	1.316	1.312	1.311	1.291	-0.004	-0.005	-0.024	a
	F-C-H	114.4	116.9	116.3	116.3	2.5	1.9	1.9	a
	F-C-C	128.8	130.1	131.2	132.3	1.3	2.5	3.6	a
92: CF <sub>2</sub> CH	F-C	1.333	1.343	1.346	1.328	0.010	0.013	-0.005	a
	F-C	1.327	1.337	1.337	1.320	0.010	0.010	-0.006	a
	F-C-C	125.7	125.1	126.1	125.4	-0.6	0.3	-0.3	a
	F-C-F	109.1	109.8	107.9	109.1	0.8	-1.2	0.0	a

... Continued ...



Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
93: CF <sub>2</sub> CF	F-C	1.330	1.341	1.343	1.325	0.011	0.012	-0.005	a
	F-C	1.326	1.336	1.334	1.318	0.010	0.008	-0.008	a
	F-C	1.315	1.307	1.305	1.286	-0.008	-0.010	-0.029	a
	F-C-C	125.2	124.8	125.4	125.1	-0.4	0.3	0.0	a
	F-C-F	111.6	111.3	109.9	110.8	-0.3	-1.8	-0.8	a
	F-C-C	129.7	131.4	133.1	134.8	1.7	3.4	5.1	a
95: CH <sub>2</sub> FC	F-C	1.398	1.402	1.392	1.372	0.004	-0.006	-0.026	a
	F-C-H	109.0	108.1	107.5	108.1	-1.0	-1.5	-0.9	a
96: CHF <sub>2</sub> C	F-C	1.368	1.380	1.373	1.355	0.012	0.005	-0.013	a
	F-C	1.368	1.380	1.374	1.355	0.012	0.005	-0.013	a
	F-C-H	109.2	109.8	109.3	109.2	0.6	0.0	0.0	a
	F-C-F	107.5	102.6	101.9	103.7	-4.9	-5.6	-3.8	a
97: CF <sub>3</sub> C	F-C	1.343	1.365	1.366	1.351	0.021	0.022	0.007	a
	F-C	1.343	1.365	1.366	1.351	0.021	0.023	0.008	a
	F-C	1.357	1.365	1.366	1.350	0.008	0.009	-0.007	a
	F-C-C	114.9	114.5	112.4	114.1	-0.4	-2.5	-0.8	a
	F-C-F	107.6	104.7	104.6	104.6	-2.9	-3.0	-3.0	a
	F-C-F	109.2	105.8	103.7	105.7	-3.4	-5.5	-3.5	a
99: HCCF	F-C	1.294	1.300	1.296	1.277	0.006	0.002	-0.017	a
100: FCCF	F-C	1.300	1.301	1.296	1.278	0.001	-0.004	-0.022	a
	F-C	1.300	1.301	1.296	1.278	0.001	-0.004	-0.022	a
102: CHFC	F-C	1.342	1.355	1.344	1.322	0.013	0.002	-0.020	a
	F-C-H	116.9	115.0	113.6	114.6	-1.9	-3.3	-2.2	a
103: CF <sub>2</sub> C	F-C	1.325	1.343	1.342	1.324	0.018	0.017	-0.001	a
	F-C	1.325	1.343	1.342	1.324	0.018	0.017	-0.001	a
	F-C-C	123.0	124.9	126.2	125.3	1.9	3.3	2.3	a
	F-C-F	114.0	110.1	107.5	109.4	-3.9	-6.5	-4.6	a
105: CCF	F-C	1.286	1.286	1.285	1.263	0.000	-0.002	-0.023	a
107: CHF <sub>2</sub> CO	F-C	1.366	1.357	1.349	1.328	-0.010	-0.017	-0.038	a
	F-C-H	117.4	119.5	118.6	118.7	2.1	1.2	1.3	a
108: CF <sub>2</sub> CO	F-C	1.338	1.340	1.344	1.321	0.002	0.005	-0.018	a
	F-C	1.338	1.340	1.344	1.321	0.001	0.005	-0.017	a
	F-C-C	122.6	122.7	111.5	123.1	0.1	-11.1	0.5	a
	F-C-F	114.8	114.6	101.7	113.8	-0.2	-13.1	-1.0	a
112: CH <sub>2</sub> FCHO	F-C	1.391	1.389	1.379	1.359	-0.002	-0.011	-0.032	a
	F-C-C	109.2	107.1	107.7	109.5	-2.1	-1.5	0.3	a
113: CHF <sub>2</sub> CHO	F-C	1.367	1.373	1.370	1.352	0.006	0.002	-0.016	a
	F-C	1.367	1.373	1.370	1.352	0.006	0.002	-0.016	a
	F-C-C	107.9	107.4	108.5	108.9	-0.5	0.6	1.0	a
	F-C-C	107.9	107.3	108.5	108.9	-0.6	0.6	1.0	a
114: CF <sub>3</sub> CHO	F-C	1.333	1.353	1.353	1.338	0.020	0.020	0.005	a
	F-C	1.350	1.360	1.362	1.346	0.011	0.013	-0.003	a
	F-C	1.350	1.360	1.363	1.346	0.011	0.013	-0.003	a
	F-C-C	112.2	114.7	114.9	114.3	2.5	2.7	2.1	a
	F-C-C	109.2	111.1	112.2	111.2	1.9	2.9	2.0	a
	F-C-C	109.2	111.1	112.1	111.3	1.9	2.9	2.0	a

... Continued ...

Tabelle 3.52: (continued)

Bindungslängen [Å] und Winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
115: CH3CFO	F-C	1.368	1.370	1.368	1.351	0.003	0.000	-0.017	a
	F-C-C	110.0	111.6	113.2	112.4	1.6	3.2	2.3	a
116: CH2FCFO	F-C	1.377	1.379	1.367	1.349	0.002	-0.009	-0.027	a
	F-C	1.366	1.371	1.368	1.350	0.005	0.001	-0.017	a
	F-C-C	109.6	106.9	107.2	108.8	-2.7	-2.4	-0.7	a
	F-C-C	107.9	109.7	112.1	111.0	1.8	4.2	3.1	a
117: CHF2CFO	F-C	1.352	1.365	1.361	1.344	0.012	0.008	-0.008	a
	F-C	1.354	1.363	1.355	1.338	0.009	0.002	-0.016	a
	F-C	1.363	1.366	1.361	1.343	0.003	-0.003	-0.021	a
	F-C-C	108.9	108.0	108.4	108.8	-0.9	-0.5	0.0	a
	F-C-C	108.5	110.1	114.3	113.1	1.6	5.8	4.6	a
118: CF3CFO	F-C	1.333	1.351	1.351	1.336	0.018	0.017	0.003	a
	F-C	1.345	1.354	1.351	1.334	0.009	0.006	-0.011	a
	F-C	1.342	1.352	1.355	1.339	0.010	0.012	-0.004	a
	F-C	1.342	1.352	1.355	1.339	0.010	0.012	-0.004	a
	F-C-C	110.3	111.5	112.0	111.1	1.2	1.6	0.8	a
120: CH2FCO	F-C	1.381	1.382	1.369	1.351	0.001	-0.012	-0.030	a
	F-C-C	110.3	107.5	107.9	109.7	-2.8	-2.4	-0.6	a
	F-C	1.357	1.366	1.359	1.344	0.010	0.003	-0.013	a
	F-C	1.357	1.366	1.359	1.344	0.010	0.003	-0.013	a
	F-C-C	107.3	107.4	108.1	108.7	0.1	0.8	1.4	a
121: CHF2CO	F-C	1.357	1.366	1.359	1.344	0.010	0.003	-0.013	a
	F-C	1.357	1.366	1.359	1.344	0.010	0.003	-0.013	a
	F-C-C	107.3	107.4	108.1	108.7	0.1	0.8	1.4	a
	F-C-C	107.3	107.4	108.1	108.7	0.1	0.8	1.4	a
	F-C	1.340	1.361	1.360	1.346	0.020	0.020	0.006	a
122: CF3CO	F-C	1.339	1.352	1.352	1.337	0.013	0.013	-0.002	a
	F-C	1.339	1.352	1.352	1.337	0.013	0.013	-0.002	a
	F-C-C	112.7	112.9	113.4	112.8	0.2	0.7	0.1	a
	F-C-C	108.5	111.1	111.4	110.9	2.6	2.9	2.4	a
	F-C-C	108.5	111.1	111.4	110.9	2.5	2.9	2.3	a
125: C3F6	F-C	1.343	1.355	1.357	1.343	0.013	0.015	0.001	a
	F-C	1.347	1.355	1.358	1.344	0.008	0.011	-0.003	a
	F-C	1.347	1.355	1.358	1.344	0.008	0.011	-0.003	a
	F-C	1.348	1.348	1.345	1.326	0.000	-0.003	-0.022	a
	F-C	1.319	1.331	1.331	1.314	0.012	0.013	-0.004	a
	F-C	1.320	1.331	1.332	1.316	0.011	0.012	-0.004	a
	F-C-C	111.6	111.2	112.4	112.1	-0.4	0.7	0.4	a
	F-C-C	110.4	112.0	112.9	112.3	1.7	2.6	1.9	a
	F-C-C	110.4	112.1	113.0	112.3	1.7	2.6	1.9	a
	F-C-C	113.5	115.1	116.0	116.0	1.6	2.5	2.5	a
126: C3F6H	F-C	1.344	1.357	1.359	1.344	0.013	0.015	-0.001	a
	F-C	1.339	1.351	1.354	1.339	0.012	0.014	-0.001	a
	F-C-C	125.1	124.3	125.0	124.6	-0.8	0.0	-0.5	a
	F-C-C	122.3	124.8	125.6	125.1	2.5	3.3	2.8	a

... Continued ...

**Tabelle 3.52:** (continued)

Bindungslängen [Å] und winkel [°]									
Molekül	Variable	Exp	Berechnet			Fehler			Lit.
			OM1	OM2	OM3	OM1	OM2	OM3	
	F-C	1.348	1.358	1.361	1.347	0.011	0.014	-0.001	a
	F-C	1.381	1.377	1.371	1.352	-0.004	-0.010	-0.029	a
	F-C	1.333	1.335	1.336	1.318	0.002	0.003	-0.014	a
	F-C	1.337	1.337	1.339	1.322	0.000	0.002	-0.015	a
	F-C-C	110.3	111.6	112.7	112.0	1.2	2.4	1.7	a
	F-C-C	111.5	113.3	114.1	113.6	1.8	2.6	2.0	a
	F-C-C	109.4	110.8	112.0	111.0	1.4	2.6	1.6	a
	F-C-C	107.0	106.4	107.1	108.3	-0.6	0.1	1.3	a
	F-C-C	114.8	116.2	117.5	117.7	1.3	2.7	2.9	a
	F-C-C	113.4	115.1	115.6	115.9	1.7	2.1	2.4	a
Bindungslängen									
Mean error						0.0085	0.0060	-0.0110	
Mean absolute error						0.0123	0.0116	0.0141	
Number of comparison						214	214	214	
Bindungswinkel									
Mean error						0.25	0.56	1.00	
Mean absolute error						1.73	2.18	1.78	
Number of bond angles						212	212	212	

a N.L.Haworth, M.H.Smith, G.B.Bacsokay, and J.C.Mackie, J.Phys.Chem. A 104, 7600 (2000).

**Tabelle 3.53:** Standardbildungsenthalpien für Validierungssatz C:  
OM1, OM2 und OM3 Ergebnisse.

Standardbildungsenthalpien [kcal/mol]									
Molekül	G3	Berechnet			Fehler			$\delta_{exp}$	Lit.
		OM1	OM2	OM3	OM1	OM2	OM3		
Hydrogen fluoride	-65.1	-69.7	-63.3	-64.9	-4.6	1.8	0.2	0.2	a
F2 molecule	0.0	12.3	0.6	-1.7	12.3	0.6	-1.7	0.0	a
CF4	-223.0	-217.2	-220.5	-215.8	5.8	2.5	7.2	0.3	a
COF2	-152.7	-137.1	-136.5	-138.7	15.6	16.2	14.0	0.4	a
NF3	-31.6	-36.2	-30.1	-35.9	-4.6	1.5	-4.3	0.3	a
F2O	5.9	16.2	5.3	7.2	10.3	-0.6	1.3	0.5	a
C2F4 (F2C=CF2)	-157.4	-157.6	-156.3	-155.9	-0.2	1.1	1.5	0.7	a
CF3CN	-118.4	-126.7	-125.5	-125.1	-8.3	-7.1	-6.7	0.7	a
Difluoromethane	-107.7	-104.7	-110.9	-106.6	3.0	-3.2	1.1	0.4	a
Trifluoromethane	-166.6	-160.6	-166.8	-162.0	6.0	-0.2	4.6	0.8	a
Vinyl fluoride	-33.2	-33.1	-32.3	-32.2	0.1	0.9	1.0	0.4	a
Acetyl fluoride	-105.7	-99.8	-100.0	-103.0	5.9	5.7	2.7	0.8	a
1,3-Difluorobenzene	-73.9	-74.2	-72.9	-74.0	-0.3	1.0	-0.1	0.2	b
1,4-Difluorobenzene	-73.3	-73.9	-72.6	-73.5	-0.6	0.7	-0.2	0.2	b
Fluorobenzene	-27.7	-28.8	-27.3	-28.0	-1.1	0.4	-0.3	0.3	b
C2F6	-321.3	-316.3	-317.5	-310.5	5.0	3.8	10.8	0.8	b
CF3	-111.3	-120.4	-123.2	-122.6	-9.1	-11.9	-11.3	0.5	b
... Continued ...									

**Tabelle 3.53:** (continued)

Standardbildungsenthalpien [kcal/mol]									
Molekül	G3	Berechnet			Fehler			$\delta_{exp}$	Lit.
		OM1	OM2	OM3	OM1	OM2	OM3		
Mean error					2.08	0.79	1.17		
Mean absolute error					5.45	3.48	4.05		
Number of comparison					17	17	17		

a L.A.Curtiss, K.Raghavachari, P.C.Redfern, and J.A.Pople, J.Chem.Phys. 106, 1063 (1997). b G3 values derived from the data in: L.A.Curtiss, K.Raghavachari, P.C.Redfern, V.Rassolov, and J.A.Pople, J.Chem.Phys. 112, 7374 (2000).

**Tabelle 3.54:** Standardbildungsenthalpien für Validierungssatz C: MNDO, AM1 und PM3 Ergebnisse.

Standardbildungsenthalpien [kcal/mol]									
Molekül	G3	Berechnet			Fehler			$\delta_{exp}$	Lit.
		MNDO	AM1	PM3	MNDO	AM1	PM3		
Hydrogen fluoride	-65.1	-59.7	-74.3	-62.7	5.4	-9.2	2.4	0.2	a
F2 molecule	0.0	7.3	-22.5	-21.7	7.3	-22.5	-21.7	0.0	a
CF4	-223.0	-214.2	-225.7	-225.1	8.8	-2.7	-2.1	0.3	a
COF2	-152.7	-138.6	-146.3	-141.6	14.1	6.4	11.1	0.4	a
NF3	-31.6	-34.2	-40.0	-24.4	-2.6	-8.4	7.2	0.3	a
F2O	5.9	18.2	10.5	-4.8	12.3	4.6	-10.7	0.5	a
C2F4 (F2C=CF2)	-157.4	-175.7	-175.1	-168.2	-18.3	-17.7	-10.8	0.7	a
CF3CN	-118.4	-113.2	-119.5	-115.0	5.2	-1.1	3.4	0.7	a
Difluoromethane	-107.7	-111.8	-116.1	-103.8	-4.1	-8.4	3.9	0.4	a
Trifluoromethane	-166.6	-163.8	-172.5	-162.0	2.8	-5.9	4.6	0.8	a
Vinyl fluoride	-33.2	-34.5	-34.0	-28.6	-1.3	-0.8	4.6	0.4	a
Acetyl fluoride	-105.7	-96.5	-98.8	-98.7	9.2	6.9	7.0	0.8	a
1,3-Difluorobenzene	-73.9	-70.9	-67.9	-63.3	3.0	6.0	10.6	0.2	b
1,4-Difluorobenzene	-73.3	-71.0	-68.0	-63.3	2.3	5.3	10.0	0.2	b
Fluorobenzene	-27.7	-25.2	-23.3	-20.2	2.5	4.4	7.5	0.3	b
C2F6	-321.3	-299.6	-313.2	-317.8	21.7	8.1	3.5	0.8	b
CF3	-111.3	-137.0	-142.8	-132.1	-25.7	-31.5	-20.8	0.5	b
Mean error					2.50	-3.91	0.57		
Mean absolute error					8.62	8.82	8.34		
Number of comparison					17	17	17		

a und b gleiche Referenzen wie in der Tabelle 3.54

**Tabelle 3.55:** Ionisierungsenergien (kcal/mol) für Validierungssatz C: OM1, OM2 und OM3 Ergebnisse.

Ionisierungsenergien [kcal/mol]								
Molekül	Exp	Berechnet			Fehler			Lit.
		OM1	OM2	OM3	OM1	OM2	OM3	
F+	401.7	372.7	393.4	398.8	-29.0	-8.3	-2.9	a
FH+	369.9	350.5	362.7	373.9	-19.4	-7.2	4.0	a
CF2+	263.3	251.5	256.6	253.8	-11.8	-6.7	-9.5	a
CH3F+	287.6	289.6	284.5	288.8	2.0	-3.1	1.2	a
CH3OF+	261.5	262.0	259.2	266.9	0.5	-2.3	5.4	a
HOF+	293.1	282.0	292.1	294.7	-11.1	-1.0	1.6	a
Mean error					-11.48	-4.76	-0.04	
Mean absolute error					12.28	4.76	4.10	
Number of comparison					6	6	6	

a L.A.Curtiss, K.Raghavachari, P.C.Redfern, and J.A.Pople, J.Chem.Phys. 106, 1063 (1997).

**Tabelle 3.56:** Elektronenaffinitäten für Validierungssatz C. OM1, OM2 und OM3 Ergebnisse.

Elektronenaffinität [kcal/mol]								
Molekül	Exp	Berechnet			Fehler			Lit.
		OM1	OM2	OM3	OM1	OM2	OM3	
F	78.4	49.7	70.4	75.7	-28.7	-8.0	-2.7	a
CF2	4.1	-0.5	7.6	8.0	-4.6	3.5	3.9	a
OF	52.4	35.7	55.0	57.2	-16.7	2.6	4.8	a
HCF (1A')	12.5	1.8	11.1	12.6	-10.7	-1.4	0.1	a
Mean error					-15.19	-0.86	1.52	
Mean absolute error					15.19	3.87	2.87	
Number of comparison					4	4	4	

a L.A.Curtiss, P.C.Redfern, K.Raghavachari, V.Rassolov, and J.A.Pople, J.Chem.Phys. 110, 4703 (1999).



# Kapitel 4

## Zusammenfassung

Die vorliegende Arbeit befasst sich hauptsächlich mit quantenchemischer Programmentwicklung. Zwei Themenkreise wurden bearbeitet: Zum einen wurde untersucht, inwieweit die Technik des automatischen Differenzierens bei quantenchemischen Programmen eingesetzt werden kann, um automatisch einen Code zur analytischen Berechnung von Energiegradienten zu erzeugen. Zum anderen wurden in einem grösseren neuen Programmsystem systematische und stochastische Methoden implementiert, um Parameter bei komplizierten Modellierungsaufgaben möglichst effizient und automatisch optimieren zu können. Diese beiden allgemeinen Fragestellungen wurden im Rahmen der semiempirischen Quantenchemie angegangen, um die dort verfügbaren Methoden und Programme zu erweitern und zu verbessern.

Im ersten Teil wurde mit Hilfe der ADIFOR-Software ein analytischer Gradienten-Code für alle semiempirischen SCF Verfahren erzeugt, welche im MNDO97 Programm implementiert sind. Die Richtigkeit und die numerische Präzision dieses Codes wurden durch ausgedehnte Vergleiche bestätigt. Die ADIFOR-generierten Routinen für den analytischen Gradient sind im Falle der semiempirischen Standardverfahren (MNDO, AM1, PM3) relativ langsam verglichen mit existierenden handkodierten Versionen, aber sie sind schneller als ihre numerischen Pendanten. Für neuere semiempirische Verfahren mit Orthogonalisierungskorrekturen (OM1, OM2, OM3) bietet der ADIFOR-erzeugte Code erstmals die Möglichkeit, den Gradienten analytisch zu berechnen. Die integral-direkte Version des ADIFOR-basierten Codes ist der konventionellen Version klar vorzuziehen, da sie viel weniger Kernspeicher beansprucht und nur wenig höhere Rechenzeiten erfordert.

Im zweiten Teil wurde das Programmsystem `gps1` (global parameter searcher, version 1) entwickelt, welches es erlaubt, die Parametrisierung von semiempirischen Verfahren zu automatisieren und effizienter zu gestalten. Implementiert wurde sowohl eine iterative systematische Suche in Kombination mit lokalen gradientenbasierten Optimierungen günstiger Parametersätze als auch eine sto-

chastische Suche mittels genetischer Algorithmen. Bei den beiden durchgeführten Anwendungen erwies sich die erste Alternative jeweils als deutlich günstiger; dies lag vor allem daran, dass jeweils hochwertige Startparameter zur Verfügung standen und somit eher eine lokale als eine globale Optimierung gefragt war. Beim ersten Anwendungsbeispiel ging es um eine testweise Reparametrisierung des etablierten MNDO Verfahrens; erwartungsgemäss führen die optimierten neuen MNDO-Parameter (Elemente H, C, N, O) zwar zu merklichen statistischen Verbesserungen bei den untersuchten Eigenschaften und Molekülklassen, aber nicht zu wirklichen qualitativen Fortschritten, so dass eine Abkehr von den etablierten MNDO-Parametern nicht geboten erscheint. Bei der zweiten Anwendung wurden die bislang nicht verfügbaren Parameter von Fluor für die OM2 und OM3 Verfahren hergeleitet. Diese Optimierungen erwiesen sich als erfolgreich, weil OM2 und OM3 mit den neuen Parametern (ebenso wie OM1) deutlich bessere Ergebnisse für Fluorverbindungen liefern als die Standardverfahren. Daher können die neuen Fluorparameter zum allgemeinen Gebrauch freigegeben werden.

Insgesamt haben sich die in beiden Bereichen durchgeführten Entwicklungen bei den bisherigen Tests und Anwendungen bewährt. Von unmittelbarem Nutzen sind der analytische Gradienten-Code für die OM $x$ -Methoden und das Programmsystem `gps1` zur Optimierung von Parametern, welches innerhalb der Abteilung bereits bei anderen Projekten eingesetzt wird. Auch wenn die im Rahmen dieser Arbeit gewonnenen Erfahrungen beim automatischen Differenzieren und beim Optimieren sich konkret auf die Methoden der semiempirischen Quantenchemie beziehen, so ist zu erwarten, dass sie auch in anderen Bereichen der Theoretischen Chemie hilfreich sein können.



# Anhang A

## Programm-Listing

---

```
1  PROGRAM diff3
2  !
3  ! program which calculates the effects of finite precision
4  ! by differentiating a function with 10 different step sizes.
5  ! Done for single and double precision. The test function
6  ! is  $F(X) = 1/x$ 
7  !
8  ! The affix _pm denotes a plusminus in short term
9  !
10 IMPLICIT NONE
11
12 ! Declare parameters
13 INTEGER, PARAMETER :: single = SELECTED_REAL_KIND(p=6,r=37)
14 INTEGER, PARAMETER :: double = SELECTED_REAL_KIND(p=13)
15 !
16 ! List of local variables:
17 REAL(KIND=double) :: ans           ! True (analytic) answer
18 REAL(KIND=double) :: d_ans,d_ans_pm ! Double precision answer
19 REAL(KIND=double) :: d_fx         ! Double precision  $F(x)$ 
20 REAL(KIND=double) :: d_fxdx,d_fxdx_p,d_fxdx_m ! Double precision  $F(x+dx)$  resp.  $F(x-dx)$ 
21 REAL(KIND=double) :: d_dx         ! Step size
22 REAL(KIND=double) :: d_x         ! Location to evaluate  $dF(x)/dx$ 
23 INTEGER           :: i           ! Index variable
24 REAL(KIND=single) :: s_ans,s_ans_pm ! Single precision answer
25 REAL(KIND=single) :: s_fx         ! Single precision  $F(x)$ 
26 REAL(KIND=single) :: s_fxdx,s_fxdx_p,s_fxdx_m ! Single precision  $F(x+dx)$  resp.  $F(x-dx)$ 
27 REAL(KIND=single) :: s_dx         ! Step size
28 ! REAL(KIND=single) :: s_x =0.20E0 ! Location to evaluate  $dF(x)/dx$ 
29 REAL(KIND=single) :: s_x         ! Location to evaluate  $dF(x)/dx$ 
30 !
31 d_x = SQRT(1.D0/42.D0)
32 s_x = SQRT(1.D0/42.D0)
33 !
34 ! Print headings.
35 WRITE(*,' (" DX           Exakt   SP mono   SP two   DP mono           DP two   ")')
```

```
36 !
37 ! Calculate analytic solution at x=0.20.
38   ans = 1.D0 / d_x**2
39 ! Calculate answer from definition of differentiation
40   DO i = 1,15
41   !
42   ! Get delta x.
43     s_dx = 1.0 / 10.0**i           ! single precision
44     d_dx = 1.D0 / 10.D0**i       ! double precision
45   !
46   ! Calculate single precision answer one side.
47     s_fxdx = -1. / (s_x + s_dx)
48     s_fx   = -1. / (s_x)
49     s_ans  = ( s_fxdx - s_fx ) / s_dx
50   !
51   ! Calculate single precision answer two side.
52     s_fxdx_p = -1. / (s_x + s_dx)
53     s_fxdx_m = -1. / (s_x - s_dx)
54     s_ans_pm = (s_fxdx_p - s_fxdx_m) / (2 * s_dx)
55   !
56   ! Calculate double precision answer.
57     d_fxdx = -1.D0 / ( d_x + d_dx )
58     d_fx   = -1.D0 / (d_x)
59     d_ans  = ( d_fxdx - d_fx ) / d_dx
60   !
61   ! Calculate double precision two side
62     d_fxdx_p = -1.D0 / ( d_x + d_dx )
63     d_fxdx_m = -1.D0 / ( d_x - d_dx )
64     d_ans_pm = ( d_fxdx_p - d_fxdx_m ) / (2 * d_dx)
65   !
66     WRITE (*,101) d_dx, ans, s_ans,s_ans_pm, d_ans,d_ans_pm
67     101 FORMAT (1X, ES10.3, F12.7, F12.7, F12.7, F22.14, F22.14)
68   !
69   END DO
70 !
71 END PROGRAM
```

---

# Anhang B

## Beispiel für den Vorwärtsmodus

Gegeben sei die differenzierbare Funktion

$$f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R} \text{ mit } f(x) = \frac{x_1 \cdot x_2 - 7}{x_1 + x_2}$$

wobei  $D = \{x \mid x = (x_1, x_2) \in \mathbb{R}^2, x_1 + x_2 \neq 0\}$ . Gewünscht ist  $f'(3, 8)$ . Wir definieren die Funktionen  $f_1, f_2, \dots, f_7 : D \rightarrow \mathbb{R}$  durch

$$\begin{aligned} f_1(x) &= x_1 \\ f_2(x) &= x_2 \\ f_3(x) &= 7 \\ f_4(x) &= f_1(x) \cdot f_2(x) \\ f_5(x) &= f_4(x) - f_3(x) \\ f_6(x) &= f_1(x) + f_2(x) \\ f_7(x) &= f_5(x) / f_6(x) \end{aligned}$$

Offensichtlich ist  $f_7 = f$ . Wir verwenden hier *RAT* als Funktion, die zu einem Typ  $\omega \in +, -, \cdot, /$  und zu Paaren  $A = (a(x), a'(x))$  und  $B = (b(x), b'(x))$  das Paar  $R = (r(x), r'(x))$  berechnet, oder kurz:

$$R \leftarrow \text{RAT}(\omega, A, B)$$

Nun betrachten wir die Paare  $Y_k = (f_k(x), f'_k(x))$  für  $k = 1, 2, \dots, 7$ . Mit  $x = (3, 8)$  sind die Paare  $Y_1, Y_2, Y_3$  bekannt. Die folgenden Paare  $Y_4, Y_5, Y_6, Y_7$  können schrittweise mit *RAT* berechnet werden:

$$\begin{aligned} Y_1 &\leftarrow (3, [1, 0]) \\ Y_2 &\leftarrow (8, [0, 1]) \\ Y_3 &\leftarrow (7, [0, 0]) \\ Y_4 &\leftarrow \text{RAT}(\cdot, Y_1, Y_2) = (24, [8, 3]) \\ Y_5 &\leftarrow \text{RAT}(-, Y_4, Y_3) = (17, [8, 3]) \\ Y_6 &\leftarrow \text{RAT}(+, Y_1, Y_2) = (11, [1, 1]) \\ Y_7 &\leftarrow \text{RAT}(/, Y_5, Y_6) = \left(\frac{17}{11}, \left[\frac{71}{121}, \frac{16}{121}\right]\right) \end{aligned}$$

Berechnung explizit für  $Y_7$ : Sei

$$r = \frac{a}{b} \text{ und } r' = \frac{a' - r \cdot b'}{b}.$$

Somit ergeben sich folgende Werte:

$$\frac{71}{121} = \frac{8 - \frac{17}{11} \cdot 1}{11} \text{ resp. } \frac{16}{121} = \frac{3 - \frac{17}{11} \cdot 1}{11}$$

Somit ist  $f(3, 8) = \frac{17}{11}$  und  $f'(3, 8) = [\frac{71}{121}, \frac{16}{121}]$ . Für jede explizit gegebene rationale Funktion  $f$  und jedes zulässige Argument  $x$  können wir  $f'(x)$  mit *RAT* schrittweise berechnen. Man beachte, dass eine explizite Formel für die Ableitung  $f'$  an keiner Stelle der Berechnung verwendet wurde.

# Anhang C

## Programmbeschreibung von Parma

Parma steht für das Akronym 'Programm für automatisches Erzeugen von Z-Matrizen'. Das Programm ist in Fortran f90 geschrieben und weist ca. 1700 Zeilen auf. Es ist in der Lage, beliebig lange Molekülketten zu generieren. Die Molekülgeometrien (Bindungslängen, Bindungswinkel sowie Diederwinkel) sind per Parameteranweisungen definiert und können zufällig um einen bestimmten Betrag (wiederum per Parameteranweisungen gelöst) variieren. Das Programm ist derart konzipiert, dass es zu Beginn interaktiv einige Fragen stellt und anschliessend die entsprechenden Moleküle kreiert.

Folgende Optionen sind bisher implementiert, die Reihenfolge entspricht ungefähr den zu beantwortenden Fragen von Parma:

### **Molekülart**

Bisher wurden Alkyl- sowie Alkinketten implementiert. Desweiteren stehen Polypeptide (nur Glycin) zur Verfügung. Weitere Ergänzungen sind ohne grosse Programmkenntnisse einfach möglich. Die Molekülgeometrien werden aufgrund der vordefinierten Bindungslängen (etc.) in der Parameteranweisung gebildet. Ein Zufallsfaktor kann den entsprechenden Geometrieparameter innerhalb eines Faktors variieren. Dies kann vermieden werden, indem der Zufallsfaktor auf Null gesetzt wird.

### **Anzahl der C-Atome**

Eine Integerzahl definiert die Länge der Molekülkette. Für Alkinketten muss die Anzahl der C-Atome gerade sein.

### **Anzahl der Files**

Es kann entweder nur ein File mit der gewünschten Länge der Molekülkette erzeugt werden (z.B. 'C4-Alkyl'), oder es werden auch Files für alle weiteren kürzeren Molekülketten (also C3, C2 und C1-Alkyl) kreiert.

**Koordinatentyp**

Interne oder kartesische Koordinaten können erzeugt werden, interne Koordinaten jedoch nur für maximal 99 Atome.

**Input- und Outputfiles**

Die Files sind MNDO99 konform, es werden Standardoptionen für MNDO99 gesetzt. Um die Moleküle zu visualisieren, können auch MOLDEN-Files erzeugt werden. Molden ist ein Visualisierungsprogramm, welches von vielen Plattformen unterstützt wird.

**Tests**

Sind die entsprechenden Zufallszahlen nicht auf 0 gesetzt, so kann es vorkommen, dass sich bei langen Molekülketten die einzelnen Einheiten zu nahe kommen ('Wollknäuel'). Ein einfacher Algorithmus probiert, die jeweiligen Einheiten in einem vernünftigen Abstand zu halten. Es ist aber nicht auszuschliessen, dass der Algorithmus fehlschlägt.

# Anhang D

## Technische Anmerkungen zum Programm `gps1`

Im folgenden sollen die Funktionalität und die Optionen von `gps1` kurz beschrieben werden. Für genauere Erläuterungen sollte der Source-Code zu Rate gezogen werden. Das Programm ist in Fortran f90 geschrieben und beinhaltet ca. 36000 Zeilen. Während der Entwicklungsarbeit (ca. 3 Jahre) wurde `gps1` mittels `forcheck` [62] jeweils auf syntaktisch korrekten Code geprüft und gegebenenfalls korrigiert. Das Akronym `gps1` steht für **g**lobal **p**arameter **s**earcher **v**ersion **1**.

### D.1 Kompilation

Der Code sollte syntaktisch sauber sein und kann daher einfach mit folgendem Befehl kompiliert werden: `f90 -o gps1 gps1.f90`. Das Programm erzeugt viele Shell-Skripts, welche zum Teil Spezialzeichen hinsichtlich Fortran enthalten, erwähnenswert sind `&` (Ampersand) und `\` (Backslash). Wenn es mit speziellen Flags kompiliert werden soll, so ist darauf zu achten, dass `'-assume nobackslash'` in den Kompilierungsoptionen enthalten ist. Während der Entwicklungszeit wurden die Compiler von Fujitsu (Linux) und Compaq (Tru64 UNIX) verwendet. Folgende spezielle Optionen wurden für den Fujitsu-Compiler unter Linux benutzt: `f90 -Ha -Hs -He -Hu -Ec -Ei -Es -Eu -static-flib -o gps1 gps1.f90`. Hauptsächlich handelt es sich um Optionen, welche beim Überlauf von Arrays oder bei inkonsistenten Vergleichen von REAL-Zahlen das Programm mit einem entsprechenden Fehler abbrechen. Für nähere Erläuterungen sei auf das Manual des entsprechenden Compilers verwiesen. Im folgenden wird immer davon ausgegangen, dass das aus `gps1.f90` erzeugte Executable den Namen `gps1` trägt.

### D.2 Programmaufruf

Das Programm `gps1` kann mit folgenden Argumenten aufgerufen werden:

- a) `./gps1 --help`
- b) `./gps1 --create inputfile from pp42.inp`
- c) `./gps1 inputfile` oder `./gps1 inputfile > outputfile`
- d) `./gps1 integer`

Option **a)** erzeugt ein File `gps1.txt`, welches die möglichen Steuerparameter für `gps1` beschreibt. Option **b)** erzeugt ein File namens `inputfile` für einen `gps-run`. Dabei wird ein übliches Parametrisierungsfile für `pp99` (in diesem Fall `pp42.inp`) als Vorlage verwendet, und es werden die Default-Werte für `gps1` gesetzt. Option **c)** gibt den Befehl wieder, mit dem ein `gps-run` gestartet wird. Die übliche Umleitung der Standardausgabe in ein Outputfile ist natürlich möglich. Option **d)** ruft spezielle Subroutinen auf. Als Integerwerte können Zahlen zwischen 6 und 28 angegeben werden (z.B. `./gps1 13`). Diese Subroutinen sind in `gps1.txt` anhand der Variable `'iRunJob'` genauer beschrieben. Hauptsächlich handelt es sich um Hilfsprogramme, welche die entstehenden Outputfiles in LaTeX-kompatible Tabellen und Plots übersetzt. Dies ist sehr praktisch, wenn es darum geht, diverse Moleküleigenschaften für verschiedene Parametersätze miteinander zu vergleichen.

Es kann vorkommen, dass `gps1` neu gestartet werden muss; mögliche Gründe sind meistens Hardware-Fehler oder Stromausfälle, welche `gps1` abstürzen lassen. Dazu muss im Inputfile die Option `iRestartGrid = 1` gesetzt werden, zusätzlich muss ein Restartfile zu Verfügung gestellt werden. Danach kann das Programm wiederum über den Befehl `'./gps1 inputfile'` neu gestartet werden. Das Programm überprüft, ob das Inputfile komplett und sinnvoll ist, zudem gibt es Warnungen verschiedenster Art aus, bei Bedarf können verschiedene Optionen interaktiv gewählt werden.

Es hat sich bewährt, `gps1` lokal und nicht vom Fileserver aus aufzurufen. Der Vorteil liegt darin, dass in diesem Fall ein Ausfall des Fileservers (und sei es auch nur für ein paar Sekunden) `gps1` nicht stört. Des weiteren empfiehlt es sich, `gps1` vom Scratch-Directory aus zu starten, da ansonsten das automatische Backup viele redundante Daten speichern muss.

### D.3 Programmkonzept

`gps1` ist derart konzipiert, dass es - nach gewissen Kriterien - automatisch Inputfiles mit einem zugehörigen Satz an Parametern für das eigentliche Parametrisierungsprogramm `pp99` erstellt. Diese Inputfiles werden von `pp99` bearbeitet; das heisst, `pp99` wird innerhalb von `gps1` aufgerufen und berechnet resp. optimiert dann die Fehlerquadratsumme (SSQ) für alle erstellten Inputfiles mit den jeweiligen Parametern. Dieser Aufruf geschieht durch den in Fortran üblichen Befehl:

```
CALL SYSTEM ("pp99 < inputfile > outputfile")
```



Diese explizite Trennung zwischen dem SSQ-berechnenden resp. SSQ-minimierenden Programm (pp99) und dem externen Steuerprogramm (gps1) wurde vorgenommen, um ein möglichst stabiles Laufverhalten beider Programme zu gewährleisten. In seltenen Fällen kann es nämlich vorkommen, dass pp99 aufgrund gewisser Fehlerbedingungen die Berechnung abbricht, und gps1 kann dann als externes Programm geeignet darauf reagieren. Prinzipiell ist es natürlich auch möglich, dass in gps1 bislang noch nicht entdeckte Fehler auftreten, welche bei einer Abkopplung von pp99 leichter zu lokalisieren sind als in einem grossen monolithischen Code. Im folgenden werden weitere Leistungsmerkmale von gps1 vorgestellt.

### D.3.1 Programmübersicht

Die Abbildung D.1 gibt eine Übersicht zu gps1 und dessen Funktionen. Die einzelnen Punkte werden nachfolgend kurz erklärt.

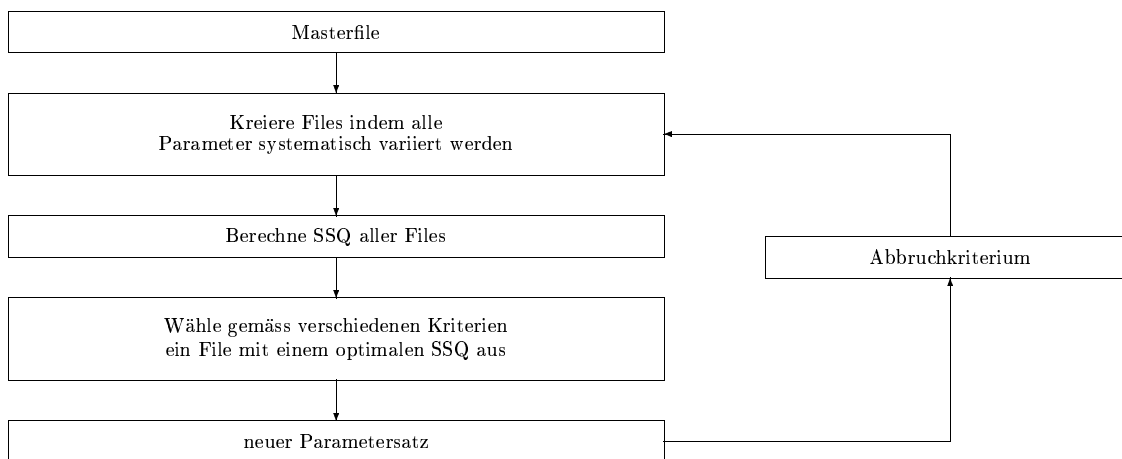


Abbildung D.1: Programmübersicht zu gps1

### D.3.2 Masterfile

Das erste Masterfile für einen gps-run ist ein pp99-konformes File. Ausgehend von diesem File kann entweder von Hand oder durch einen Aufruf an der Konsole (`./gps1 -create gps.inp from pp.inp`) ein Masterfile für gps1 erzeugt werden. Dieses File beinhaltet alle notwendigen Steuerparameter, welche für einen gps1-run notwendig sind. Die jeweiligen Flags sind im File `gps1.txt` genauer erklärt. Die wichtigste Sektion des Masterfiles enthält die zu optimierenden Parameter. Bei jedem dieser Parameter ist jeweils eine Zeile hinzugefügt, in welcher die obere und untere Grenze des durchsuchten Wertebereichs spezifiziert wird.

### D.3.3 Erzeugung von Inputfiles

Ausgehend von den gegebenen Startwerten der zu optimierenden Parameter (inklusive der unteren resp. oberen Grenzen) werden neue Parameter generiert und entsprechende Inputfiles erstellt. Dabei werden zwei Optimierungsstrategien unterstützt, **grid-search** und **ga-search**, welche schon in den Kapiteln 3.4.1 und 3.4.2 vorgestellt wurden und daher hier nur kurz diskutiert werden.

Die **grid-search** sucht den Parameterraum ab, indem die zu optimierenden Parameter innerhalb eines bestimmten Bereiches systematisch variiert werden. gps1 bildet dann entweder a) alle möglichen Kombinationen (unter Umständen eine sehr grosse Zahl an Files) oder b) eine eingeschränkte Zahl von Kombinationen, indem jeweils nur ein Parameter variiert wird und die anderen Parameter ihren Startwert behalten, wodurch die Zahl der erzeugten Files drastisch reduziert wird. Die grid-search geschieht iterativ, indem der SSQ-Wert zunächst für alle generierten Files berechnet wird und dann, gegebenenfalls nach einer zwischengeschalteten partiellen SSQ-Minimierung, das File (resp. dessen Parametersatz) mit dem besten SSQ-Wert im nächsten Schritt zum Kreieren neuer Inputfiles herangezogen wird.

Mittels der genetischen Suche (**ga-search**) wird der Parameterraum stochastisch abgesucht. Dazu werden die Parameter gemäss genetischen Operatoren (siehe dazu auch das Kapitel 3.3.3.6) modifiziert und die zu den neuen Parametersätzen gehörenden SSQ-Werte berechnet. Ausgehend von diesen SSQ-Werten werden nun im Gegensatz zur grid-search mehrere 'gute' Parametersätze für das Kreieren neuer Inputfiles verwendet.

### D.3.4 SSQ-Berechnung

Der SSQ-Wert eines Parametersatzes setzt sich aus vielen Beiträgen zusammen. pp99 bietet die Möglichkeit, die SSQ-Werte in Gruppen zusammenzufassen, z.B. hinsichtlich bestimmter Eigenschaften wie Bindungslängen oder Dipolmomente. Desweiteren besteht die Möglichkeit, diese Gruppen weiter in Untergruppen aufzuteilen, z.B. Bindungslängen von Dreifachbindungen oder Doppelbindungen. Diese Informationen werden automatisch in ein File (fort.93) geschrieben.

gps1 bietet die Möglichkeit (Variable `issq > 0`), diese Teil-SSQ-Werte ebenfalls zu verarbeiten. Dazu muss der Benutzer ein Referenzfile (fort.93) mit statistischen Kenndaten für einen Referenz-Parametersatz in einem Directory namens 'dssq' zur Verfügung stellen. Ein solches File fort.93 erhält man dadurch, dass man über einen pp99-Aufruf mit der Inputoption 'IPARM = 1' ein File fort.19 kreiert, welches bei einer Auswertung mit einem Hilfsprogram (eva99) das gewünschte File (fort.93) liefert. gps1 erzeugt für jeden Parametersatz ein File 'fileproperties', welches alle relevanten statistischen Kenndaten enthält, die jedoch defaultmässig deaktiviert sind (eine Null am Ende jeder Zeile). Durch Setzen einer 1 werden diese Kenndaten (insbesondere die mittleren absoluten Fehler) jeweils mit den Referenzwerten ver-

glichen (File fort.93 im Directory dssq). Durch Setzen einer 2 wird zusätzlich auf die richtige Zahl an Molekülen getestet, um Probleme bei der zugrunde liegenden SSQ-Berechnung in pp99 auszuschliessen (z.B. keine SCF-Konvergenz, wodurch das betreffende Molekül keine Beiträge zum SSQ-Wert leistet und dadurch einen zu niedrigen Gesamt-SSQ-Wert vortäuscht); falls solche Probleme auftreten sollten, wird der betreffende Parametersatz von gps1 verworfen.

Nach diesen Vorbereitungen wählt gps1 den Parametersatz mit dem niedrigsten neuen SSQ-Wert aus. Der Benutzer kann dabei per Input (issq > 0) die Bedingung stellen, dass gewisse von ihm gewählte statistische Kenndaten (Teil-SSQ-Werte) besser sind als im Referenzdatensatz (fort.93). Gelingt dies nicht, kann er Parametersätze auch zur Auswahl zulassen, deren Teil-SSQ-Werte die betreffenden Referenzwerte (fort.93) um einen gewissen Prozentsatz übersteigen, wobei die maximale prozentuale Abweichung in einer zweiten Spalte innerhalb des Files 'fileproperties' anzugeben ist (üblicherweise 5 % resp. Faktor 1.05).

## D.4 Jobberechnung

Die zeitaufwendigsten Schritte bei der Durchführung der Parametrisierung sind sicherlich die Berechnungen der jeweiligen SSQ-Werte. Deshalb ist es mehr als vernünftig, diese Berechnungen auf mehrere Prozessoren zu verteilen. Dies kann auf zwei verschiedene Arten geschehen:

- Queue** Die jeweiligen Inputfiles werden automatisch an die Queue gesandt, welche derzeit (Sommer 2002) aus ca. 80 Prozessoren mit mindestens 1 GB RAM pro CPU besteht. Auf allen Maschinen läuft pp99, welches die Inputfiles verarbeitet und Outputfiles erzeugt. Bei Bedarf werden die optimierten Parameter weiter benutzt, um beispielsweise fort.19 Files zu generieren und Auswertungen mittels eva99 zu machen. Desweiteren ist es möglich, die Outputfiles zu komprimieren, um Plattenplatz zu sparen. Diese Aufgaben werden bereits auf den betreffenden Queue-Maschinen erledigt, um die lokale Maschine nicht mit zu vielen weiteren Aufgaben zu belasten.
- Lokal** Die Inputfiles können auch lokal auf einer einzigen Maschine berechnet werden. Im Fall einer Mehrprozessormaschine kann eine entsprechende Anzahl an Jobs parallel laufen.

Diese beiden Arbeitsweisen können über eine Inputoption (icalonqueue) angesteuert werden. Eine Kombination beider Vorgehensweisen wurde nicht implementiert, da dies einen sehr grossen Aufwand an Administration erfordert hätte. Für einen guten Durchsatz ist die Parallelverarbeitung über die Queue empfehlenswert.

## D.5 Diversa

Im folgenden wird kurz auf einige weitere praktische Features von gps1 eingegangen. Die verschiedenen Punkte sind alphabetisch geordnet.

- **checker** Es kann vorkommen, dass Files während des Kopierens zwischen dem Fileserver und den diversen Computern verloren gehen, möglicherweise wegen Überlastung des Netzwerkes [101]; die genauen Gründe sind jedoch unklar geblieben, weil solche Fehler praktisch nicht reproduzierbar waren. Sie führten in der ersten Implementierung von gps1 zu einem Zustand, in dem das Programm endlos auf diese Files wartete. Um ein solches Verhalten zu verhindern, kontrollieren jetzt diverse Subroutinen, ob Files innerhalb einer vernünftigen Zeitspanne ankommen. Falls nicht, werden diese Tests mehrfach wiederholt (typischerweise in Sekundenabständen), bevor gps1 eine erneute SSQ-Berechnung für den betreffenden File initiiert.
- **checkvars1** gps1 besitzt die Möglichkeit, gewisse gps-Steuer-Variablen aus dem laufenden Betrieb heraus zu ändern. Es kann sich im Verlauf einer Parametrisierung z.B. herausstellen, dass gewisse Queue-Optionen ineffektiv sind oder dass man den Debug-Print aktivieren möchte. Durch Ändern der Variablen innerhalb des Files checkvars1, das beim Start von gps1 erstellt wird, kann das Laufzeitverhalten von gps1 bis zu einem gewissen Grad beeinflusst werden.
- **Diskspace** Es ist möglich, die Outputfiles der SSQ-Berechnungen auf der Harddisk abzuspeichern. Dies kann bei sehr vielen Files zu Platzproblemen führen. Deshalb wurde eine Subroutine eingebaut, welche die jeweiligen Outputfiles derart zusammenfasst, dass nur noch die Startparameter und - wenn die Optimierung erfolgreich war - die optimierten Parameter gespeichert sind. Zudem ist es möglich, die Files mit entsprechenden Programmen (z.B. gzip, compress) zu komprimieren.
- **eva99** Bei Bedarf kann eine Option gewählt werden, welche automatisch eine mndo99-artige SSQ-Berechnung mit den zu den jeweiligen Inputfiles gehörenden Parametersätzen veranlasst (IPARM = 1). Hierdurch wird ein Ergebnis-File (fort.19) erzeugt, welches dann durch ein Auswertungsprogramm (eva99) genauer analysiert werden kann. Dieses Programm erstellt für den jeweiligen Parametersatz eine systematische Übersicht der Ergebnisse (Referenzwerte versus berechnete Werte) und die entsprechenden statistischen Kenndaten.
- **./gps1 Integerzahl** Viele kleinere und grössere selbständige Programme wurden während dieser Arbeit geschrieben. Wenn möglich, sind diese in gps1 integriert. Sie dienen hauptsächlich zur Auswertung der zahlreichen Files, die während einer Parametrisierung entstehen. Diese Programme sind interaktiv, d.h. der Benutzer wird in der Regel aufgefordert, die für die Auswertung erforderlichen Files oder Directories zu benennen. Danach erledigt gps1 gemäss der eingegebenen Integerzahl die gewünschten Aufgaben.

- **inputfile** gps1 wird durch ein Inputfile gesteuert. Das Inputfile ist formatfrei, zudem können die Steuervariablen an beliebigen Stellen stehen. Es müssen nicht alle Steuervariablen für gps1 definiert sein, fehlende Angaben werden durch Default-Werte ergänzt. Das Programm bricht ab, wenn einige der allerwichtigsten Variablen fehlen.
- **inputchecker** Einige Inputoptionen von gps1 können sich gegenseitig ausschliessen. In solchen Fällen macht das Programm den Benutzer auf diese Fehler aufmerksam und bricht, wenn nötig, den Start ab.
- **multifiles** Bei einer systematischen Versuche, in der für jedes File nur eine einzige SSQ-Berechnung erfolgt und welche meistens innerhalb weniger Sekunden abgeschlossen ist, entsteht ein grosser Overhead, wenn alle diese kleinen Jobs separat an die Queue gesandt würden. Deshalb ist es möglich, viele kleine Jobs zusammenzufassen und gemeinsam abzuarbeiten. Dazu wird zuerst lokal ein spezielles Jobskript geschrieben und dieses dann an die Queue gesandt. Im Fehlerfall ist hier besondere Sorgfalt geboten, weil beim Abbruch eines solchen Jobskripts meist noch nicht alle dort enthaltenen kleinen Jobs berechnet worden sind. Daher überprüft eine Kontrollroutine, ob wirklich alle kleinen Jobs ordnungsgemäss beendet wurden, und falls nicht, so wird das entsprechende File nochmals an die Queue gesandt.
- **Namenskonvention** Um die vielen Input- und Outputfiles zu unterscheiden, ist jedem Namen eine Zahl zur eindeutigen Identifizierung beigefügt. Folgendes Schema gemäss Figur D.2 wird verwendet:
  - a ist der Name für das File, welcher 5 Buchstaben lang sein muss, b kenn-

**Abbildung D.2:** Namenskonvention

```
felix.001002000000003
|---| |-||-|-----|
a   b c   d
```

zeichnet den retrial-loop (nur GA), **c** den Zähler für die aktuelle Generation (GA) oder für den grid-search Durchgang, und **d** bezeichnet die Zahl des Files innerhalb des entsprechenden Loops von **c**.

- **parallel** Werden die SSQ-Berechnungen lokal ausgeführt, so kann durch entsprechendes Definieren der Variablen 'iparallel' von Mehrprozessormaschinen Gebrauch gemacht werden. Dies bedeutet lediglich, dass 'iparallel' unabhängige pp99-Jobs laufen. Eine Parallelisierung innerhalb von gps1 wurde nicht betrieben, da bei weitem die meiste Rechenzeit für die SSQ-Berechnungen in pp99 verbraucht wird.
- **pp99** Es ist möglich, das systemweite pp99-Standardprogramm zu benutzen oder eine modifizierte lokale Version von pp99 zu verwenden. Dies erlaubt es den verschiedenen Benutzern, individuell weiterentwickelte Versionen von pp99

einzusetzen. Bei Verwendung verschiedener Plattformen (z.B. Linux und Unix) muss der Benutzer die entsprechenden pp99-Executables bereitstellen, und das Programm spricht dann beim Verwalten der Queue automatisch das richtige Executable an.

- **restart** Es kann vorkommen, dass gps1 die Berechnung abbricht. Die Gründe liegen (meistens) nicht in Programmfehlern bei gps1, sondern viel eher in nicht abfangbaren Fehlern, wie z.B. dem Zusammenbrechen des Netzwerkes. Deshalb ist eine Restart-Prozedur eingebaut, welche das Fortsetzen der Berechnung ermöglichen soll. Dazu muss im gps-Inputfile ein Flag zur Aktivierung der Restart-Routinen gesetzt werden. Der Benutzer wird dann interaktiv nach bestimmten Files gefragt. Diese Restart-Prozedur kann sowohl für den grid-search als auch für den ga-search angewählt werden.
- **Systemanforderung** gps1 alloziert dynamisch soviel Memory wie benötigt. Dies ermöglicht es, gps1 auch auf Maschinen mit wenig Memory laufen zu lassen. Während der Laufzeit beansprucht gps1 wenig an Systemzeit, vorausgesetzt, dass die Files auf der Queue berechnet werden. Die Prozessorauslastung durch gps1 beträgt typischerweise weniger als 1%. Dieser Wert ändert sich, wenn gps1 Files lokal schreibt oder liest. Generell verbringt gps1 jedoch die meiste Zeit damit, auf pp99-Outputfiles zu warten.
- **Validierung** gps1 kann Validierungsrechnungen mittels pp99 unterstützen und automatisieren. Dazu muss der Benutzer die Variable 'ivaler' auf einen Wert ungleich Null setzen und den gewünschten semiempirischen Ansatz via 'ivalidmethods' definieren. Des weiteren muss er innerhalb des Programmes gps1 eine Subroutine geringfügig erweitern (SUBROUTINE valer), welche den für eine Validierung benötigten externen Parametersatz (File fort.14) erstellt.

## D.6 Technische Probleme

Hier sollen kurz einige technische Probleme bei der Anwendung von gps1 dokumentiert werden, die im Laufe dieser Arbeit gelöst werden mussten.

- **argument list too long** Dieses Problem tritt auf, wenn in einem Directory mehr als 1023 Files vorhanden sind und der Benutzer versucht, diese mit einer Wildcard (\*) zu behandeln. Abhilfe schafft der Umweg über ein 'find'-Kommando, siehe dazu auch die entsprechende man-page.
- **disk full** Eine Disk kann voll erscheinen, obwohl sie es nicht ist. Dazu muss man wissen, dass UNIX Files und Directories einen 'Node' zuweist und dass es für deren Anzahl eine obere Grenze gibt [101]. Ist diese überschritten, wird die Harddisk von UNIX als voll angesehen. Abhilfe schafft ein Höhersetzen dieser Grenze.

- **lost files** Wie bereits oben beschrieben, kann es passieren, dass Files während des Kopierens oder auch während des Sendens an die Queue verloren gehen. Dieser nicht reproduzierbare Fehler erwies sich als sehr hartnäckig. Er wurde im Produktionsbetrieb durchschnittlich etwa einmal pro Tag beobachtet. Die einzige Lösung bestand in aufwendigen Tests, ob die betreffenden Befehle korrekt ausgeführt wurden, nötigenfalls mit einer Wiederholung dieser Befehle.

## D.7 Schlussbemerkungen

Die obigen technischen Anmerkungen zu `gps1` sind sicher unvollständig. Weitere Optionen können dem Source-Code und den dort vorhandenen Kommentaren sowie der Inputbeschreibung `gps1.txt` (siehe D.2) entnommen werden. Der Status von `gps1` liegt zwischen dem Entwicklungsstadium und Anwendungsstadium. Die Weiterentwicklung von `gps1` ist wünschenswert und wird betrieben, aber das Programm steht bereits im jetzigen Zustand anderen Benutzern zur Verfügung und ist bereits bei anderen Projekten innerhalb der Gruppe im Einsatz.





# Literaturverzeichnis

- [1] GAUSS, J.: *Molecular Properties*. In: GROTEENDORST, J. (Herausgeber): *Modern Methods and Algorithms of Quantum Chemistry*, Band 3 der Reihe *NIC Series*, Seiten 509–560, Jülich, 2000.
- [2] PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING und B. P. FLANNERY: *Numerical Recipes in Fortran 77*. Press Syndicate of the University of Cambridge, 1992.
- [3] PULAY, P.: *Analytical derivative methods in quantum chemistry*. *Adv. Chem. Phys.*, 67:241–286, 1987.
- [4] PULAY, P.: *Analytical Derivative Techniques and the Calculation of Vibrational Properties*. In: YARKONY, D. (Herausgeber): *Modern Electronic Structure Theory*, Band 2, Seiten 1191–1248. World Scientific, Singapore, 1995.
- [5] YAMAGUCHI, Y., Y. OSAMURA, J. D. GODDARD und H. F. SCHAEFER: *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*. Oxford University Press, New York, 1994.
- [6] PATCHKOVSKII, S. Doktorarbeit, Universität Zürich, 1997.
- [7] PATCHKOVSKII, S. und W. THIEL: *Analytical First Derivatives of the Energy in the MNDO Half-Electron Open-Shell Treatment*. *Theoret. Chim. Acta*, 93:87–99, 1996.
- [8] PATCHKOVSKII, S. und W. THIEL: *Analytical First Derivatives of the Energy for Small CI Expansions*. *Theor. Chem. Acc.*, 98:1–4, 1997.
- [9] PATCHKOVSKII, S. und W. THIEL: *Analytical Second Derivatives of the Energy in MNDO Methods*. *J. Comp. Chem.*, 17(11):1318–1327, 1996.
- [10] FISCHER, H.: *Wissenschaftliches Rechnen - Eine Einführung in das Scientific Computing*, Seiten 53–104. J. Herzberger, Akademie-Verlag, Berlin, 1995.
- [11] BISCHOF, C., A. CARLE, P. KHADEMI und A. MAUER: *ADIFOR 2.0: Automatic Differentiation of Fortran 77 Programs*. *IEEE Computational Science & Engineering*, 3(3):18–32, 1996.

- [12] BISCHOF, C., A. CARLE, P. HOVLAND, P. KHADEMI und A. MAUER: *ADIFOR2.0 User's Guide (Revision D)*. Technical Memorandum ANL/MCS-TM-192, Argonne National Laboratory, 1998.
- [13] BISCHOF, C. und M. BÜCKER: *Computing Derivatives of Computer Programs*. In: GROTENDORST, J. (Herausgeber): *Modern Methods and Algorithms of Quantum Chemistry*, Band 3 der Reihe *NIC Series*, Seiten 315–327, Jülich, 2000.
- [14] FAGAN, M. und A. CARLE: *Adifor 3.0 Overview*. Technischer Bericht CAAM-TR00-03, Rice University, Department of Computational and Applied Mathematics, 2000.
- [15] SHIRIAEV, D.: *ADOL-F Automatic Differentiation of Fortran Codes*. In: BERZ, MARTIN, CHRISTIAN H. BISCHOF, GEORGE F. CORLISS und ANDREAS GRIEWANK (Herausgeber): *Computational Differentiation: Techniques, Applications, and Tools*, Seiten 375–384. SIAM, Philadelphia, PA, 1996.
- [16] *The TAPENADE tutorial*. <http://www-sop.inria.fr/tropics/>.
- [17] GIERING, R. und T. KAMINSKI: *Recipes for Adjoint Code Construction*. ACM Transactions on Mathematical Software, 24(4):437–474, 1998.
- [18] BEDA, L. M., L. N. KOROLEV, N. V. SUKHIKH und T. S. FROLOVA: *Programme zum Automatischen Differenzieren für die Maschine BESM (in Russisch)*. Technischer Bericht, Akademie der Wissenschaften der UdSSR, Moskau, 1959.
- [19] RALL, L. B.: *Automatic Differentiation: Techniques and Applications. (Lecture Notes in Computer Science 120)*. Springer-Verlag, Berlin, 1981.
- [20] GRIEWANK, A.: *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, PA, 2000.
- [21] M. FAGAN, RICE UNIVERSITY, HOUSTON, TEXAS: *persönliche Mitteilung*.
- [22] B. LANG, RWTH AACHEN: *persönliche Mitteilung*.
- [23] THIEL, W.: *Program MNDO97, University of Zürich*, 1999.
- [24] DEWAR, M. J. S. und W. THIEL: *Ground States of Molecules. 38. The MNDO Method. Approximation and Parameters*. J. Am. Chem. Soc., 99:4899–4907, 1977.
- [25] DEWAR, M. J. S., E. ZOEBISCH, E. F. HEALY und J. J. P. STEWART: *A New General Purpose Quantum Mechanical Model*. J. Am. Chem. Soc., 107:3902–3909, 1985.

- [26] STEWART, J. J. P.: *Optimization of Parameters for Semiempirical Methods. I. Method.* J. Comput. Chem., 10:209–220, 1989.
- [27] STEWART, J. J. P.: *Optimization of Parameters for Semiempirical Methods II. Applications.* J. Comput. Chem., 10:221–264, 1989.
- [28] THIEL, W. und A. A. VOITYUK: *Extension of the MNDO Formalism to d orbitals: Integral Approximation and Preliminary Numerical Results.* Theor. Chim. Acta, 81:391–404, 1992.
- [29] THIEL, W. und A. A. VOITYUK: *Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group.* J. Phys. Chem., 100:616–629, 1996.
- [30] DEWAR, M. J. S. und D. A. LIOTARD: *An Efficient Procedure for Calculating the Molecular Gradient using SCF-CI Semiempirical Wavefunctions with a Limited Number of Configurations.* J. Mol. Struct. (Theochem), 206:123–133, 1990.
- [31] KOLB, M. und W. THIEL: *Beyond the MNDO Model: Methodical Considerations and Numerical Results.* J. Comput. Chem., 14:775–789, 1993.
- [32] WEBER, W. und W. THIEL: *Orthogonalization Corrections for Semiempirical Methods.* Theor. Chem. Acc., 103:495–506, 2000.
- [33] STEWART, J. J. P.: *Semiempirical Molecular Orbital Methods.* In: LIPKOWITZ, K. B. und D. B. BOYD (Herausgeber): *Reviews in Computational Chemistry*, Band 1, Seiten 45–81. VCH Publishers, New York, 1990.
- [34] STEWART, J. J. P.: *Parametrization of Semiempirical MO Methods.* In: RAGUÉ SCHLEYER VON (Herausgeber): *Encyclopedia of Computational Chemistry*, Band 3, Seite 2000. John Wiley & Sons, 1998.
- [35] VENKATASUBRAMANIAN, V. und A. SUNDARAM: *Genetic Algorithms, Introduction and Applications.* 3:1599, 1998.
- [36] DEVILLERS, J.: *Genetic Algorithms in Molecular Design.* Academic Press, London, 1996.
- [37] GLOVER, F. *ORSA J. Comput.*, 1:190–206, 1989.
- [38] GLOVER, F. *ORSA J. Comput.*, 2:4–32, 1990.
- [39] HOLLAND, J. H.: *Adaption in Natural and Artificial Systems.* MIT Press, Cambridge, MA, 1992.
- [40] GOLDBERG, D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, Reading, MA, 1989.

- [41] RECHENBERG, I.: *Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*. Frommann Holzboog, Stuttgart, 1973.
- [42] SCHWEFEL, H.-P.: *Numerical Optimization of Computer Models*. Wiley, Chichester, 1981.
- [43] FOGEL, L. J., A. J. OWENS und M. J. WALSH: *Artificial Intelligence through Simulated Evolution*. Wiley, New York, 1966.
- [44] FOGEL, D. B.: *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. IEEE Press, Piscataway, NJ, 1995.
- [45] DARWIN, C.: *The Origin of Species*. Dent Gordon, London, 1973.
- [46] LUCASIUS, C. B. und G. KATEMAN. *Trends. Anal. Chem.*, 10:254–261, 1991.
- [47] HIBBERT, D. B. *Chemom. Intell. Lab. Syst.*, 19:277–293, 1993.
- [48] LUCASIUS, C. B., M. L. BECKERS und G. KATEMAN. *Anal. Chim. Acta*, 286:135–153, 1994.
- [49] GOLDBERG, D. E. und C. H. KUO. *J. Comput. Civil. Eng.*, 1:128–141, 1987.
- [50] ANDROULAKIS, I. P. und V. VENKATASUBRAMANIAN. *Comput. Chem. Eng.*, 15:217–228, 1991.
- [51] SYSWERDA, G. und J. PALMUCCI: *in Proceedings of the 4th International Conference on Genetic Algorithms*. Seiten 502–508. Morgan Kaufmann, San Mateo, CA, 1991.
- [52] DANDEKAR, T. und P. ARGOS. *Protein Eng.*, 5:637–645, 1992.
- [53] DANDEKAR, T. und P. ARGOS. *J. Mol. Biol.*, 236:844–861, 1994.
- [54] GLEN, R. C. und A. W. R. PAYNE. *J. Comput.-Aided Mol. Design*, 9:181–202, 1995.
- [55] ROGERS, D. und A. J. HOPFINGER. *J. Chem. Inf. Comp. Sci.*, 34:854–866, 1994.
- [56] MICHALEWICZ, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, New York, 1996.
- [57] BÄCK, T.: *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, 1995.
- [58] GLEAN, R. C. und A. W. R. PAYNE. *J. Comput.-Aided Mol. Design*, 9:181–202, 1995.

- [59] SYSWERDA, G.: *Handbook to Genetic Algorithms*. L. Davis, Van Nostrand Reinhold, New York, 1989.
- [60] HAUPT, R. L. und S. E. HAUPT: *Practical Genetic Algorithms*. A Wiley-Interscience Publication, 1998.
- [61] SHAFFER, R. E. und G. W. SMALL: *Learning optimization from nature, genetic algorithms and simulated annealing*. *Analyt. Chem.*, 69:236–242, 1997.
- [62] FORCHECK: *A Fortran Verifier and Programming Aid*, Leiden University, Niederlande, 1999. Version V12.78.
- [63] WEBER, W.: *Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Implementierung, Parametrisierung und Anwendungen*. Doktorarbeit, Universität Zürich, 1996.
- [64] KOLB, M.: *Ein neues semiempirisches Verfahren auf Grundlage der NDDO-Näherung: Entwicklung der Methode, Parametrisierung und Anwendungen*. Doktorarbeit, Bergische Universität - Gesamthochschule Wuppertal, 1991.
- [65] SCANO, P. und C. THOMSON. *J. Comp. Chem*, 12:172, 1991.
- [66] KOLB, M. und W. THIEL: *Beyond the MNDO Model: Methodical Considerations and Numerical Results*. *J. Comput. Chem.*, 14:775, 1993.
- [67] WEBER, W. und W. THIEL: *Orthogonalization corrections for semiempirical methods*. *Theor. Chem. Acc.*, 103:495, 2000.
- [68] SCHOLTEN, M. Doktorarbeit, Universität Düsseldorf, 2003.
- [69] KOLLMAR, C. und M. C. BÖHM: *An analysis of the zero differential overlap approximation. Towards an improved semiempirical MO method beyond it*. *Theor. Chim. Acta*, 92:13, 1995.
- [70] BRUIJN, S. DE. *Stud. Phys. Theor. Chem.*, 21:251, 1982.
- [71] GREGORY, A. R. und M. N. PADDON-ROW: *Ability of CNDO-2 and INDO MO Methods to Cope with Nonbonded Interactions*. *J. Am. Chem. Soc.*, 98:7521, 1976.
- [72] BASILEVSKY, M. V., A. G. SHAMOV und V. A. TIKHOMIROV: *Transition State of Diels-Alder Reaction*. *J. Am. Chem. Soc.*, 99:1369, 1977.
- [73] CARAMELLA, P., K. N. HOUK und L. N. DOMELSMITH: *Dichotomy between Cycloaddition Transition-States Calculated by Semiempirical and Ab-Initio Techniques*. *J. Am. Chem. Soc.*, 99:4511, 1977.

- [74] SPANGET-LARSEN, J.: *The alternant hydrocarbon pairing theorem and all-valence electrons theory. An approximate LCOAO theory for the electronic absorption and MCD spectra of conjugated organic compounds, part 2.* Theor. Chim. Acta, 98:137, 1997.
- [75] POPLE, J. A., D. P. SANTRY und G. A. SEGAL: *Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures.* J. Chem. Phys., 43:S129, 1965.
- [76] FISCHER-HJALMARS, I.: *Zero Differential Overlap in  $\pi$ -Electron Theories.* Adv. Quant. Chem., 2:25, 1965.
- [77] BRUIJN, S. DE: *Analysis of the Inadequacies of Some Semi-empirical MO Methods as Theories of Structure and Reactivity.* Int. J. Quant. Chem., 25:367, 1984.
- [78] BROWN, R. D. und K. R. ROBY: *Approximate Molecular Orbital Theory for Inorganic Molecules I. Analysis of Possible Integral Approximations II. Methods of Evaluating Basic Parameters.* Theoret. chim. Acta (Berl.), 16:175,194, 1970.
- [79] CHANDLER, G. S. und F. E. GRADER: *A Re-Examination of the Justification of Neglect of Differential Overlap Approximations in Terms of a Power Series Expansion in  $S$ .* Theoret. chim. Acta (Berl.), 54:131, 1980.
- [80] SPANGET-LARSEN, J.: *On Bridging the Gap Between Extended Hückel and NDO Type LCAO-MO Theories.* Theoret. chim. Acta (Berl.), 55:165, 1980.
- [81] FISCHER-HJALMARS, I.: *Deduction of the Zero Differential Overlap Approximation from an Orthogonal Atomic Orbital Basis.* J. Chem. Phys., 42:1962, 1965.
- [82] ROBY, K. R.: *On the Justifiability of Neglect of Differential Overlap Molecular Orbital Methods.* Chem. Phys. Lett., 11:6, 1971.
- [83] ROBY, K. R.: *Fundamentals of an Orthonormal Basis Set Molecular Orbital Theory.* Chem. Phys. Lett., 12:579, 1972.
- [84] COOK, D. B., P. C. HOLLIS und R. MCWEENY: *Approximate ab initio calculations on polyatomic molecules.* Mol. Phys., 13:553, 1967.
- [85] COFFEY, P. und K. JUG: *Semiempirical Molecular-Orbital Calculations and Molecular Energies - New Formula for Beta Parameter.* J. Am. Chem. Soc., 95:7575, 1973.
- [86] COFFEY, P.: *Potential Energy Integrals in Semiempirical MO Methods.* Int. J. Quant. Chem., 8:263, 1974.

- [87] FILATOV, M. J., O. V. GRITSENKO und G. M. ZHIDOMIROV: *CNDO-S<sup>2</sup>-a semiempirical SCF MO method for transition metal organometallics*. Theor. Chim. Acta, 72:211, 1987.
- [88] NANDA, D. N. und K. JUG: *SINDO1. A Semiempirical SCF MO Method for Molecular Binding Energy and Geometry I. Approximations and Parametrization*. Theoret. chim. Acta (Berl.), 57:95, 1980.
- [89] JUG, K., R. IFFERT und J. SCHULZ: *Development and Parametrization of SINDO1 for Second-row Elements*. Int. J. Quant. Chem., 32:265, 1987.
- [90] DEWAR, M. J. S. und N. L. SABELLI: *SPO (Split-Orbital) Method and its Application to Ethylene*. J. Chem. Phys., 34:1232, 1961.
- [91] DEWAR, M. J. S. und N. L. SABELLI: *The s.p.o. (split-p-orbital) method II. Further definition and application to acetylene*. Band 264 der Reihe A, Seite 431. Proc. R. Soc. London, 1961.
- [92] DEWAR, M. J. S. und N. L. SABELLI: *Split-Orbital (SPO) Method 3. Relationship to other MO Treatments and Application to Benzene, Butadiene, and Naphthalene*. J. Phys. Chem., 66:2310, 1962.
- [93] KLOPMAN, G.: *A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules*. J. Am. Chem. Soc., 86:4550, 1964.
- [94] OHNO, K.: *Some Remarks on the Pariser-Parr-Pople Method*. Theoret. chim. Acta (Berl.), 2:219, 1964.
- [95] KOLLMAR, C. und W. THIEL. unpublizierte Arbeiten.
- [96] KRAUSS, M. und W. J. STEVENS: *Effective Potentials in Molecular Quantum Chemistry*. Annu. Rev. Phys. Chem., 35:357, 1984.
- [97] STEVENS, W. J., H. BASCH und M. KRAUSS: *Compact effective potentials and efficient shared-exponent basis sets for first- and second-row atoms*. J. Chem. Phys, 81:6026–6033, 1984.
- [98] HAWORTH, N. L., M. H. SMITH, G. B. BACSKAY und J. C. MACKIE: *Heats of Formation of Hydrofluorocarbons Obtained by Gaussian-3 and Related Quantum Chemical Computations*. J. Phys. Chem. A, 104:7600–7611, 2000.
- [99] CURTISS, L. A., K. RAGHAVACHARI, P. C. REDFERN und J. A. POPLE: *Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation*. J. Chem. Phys., 106:1063, 1997.

- [100] CURTISS, L. A., K. RAGHAVACHARI, P. C. REDFERN und J. A. POPLER: *Assessment of Gaussian-3 and density functional theories for a larger experimental test set*. J. Chem. Phys., 112:7374, 2000.
- [101] H. LENK, MPI FÜR KOHLENFORSCHUNG, MÜLHEIM, DEUTSCHLAND: *persönliche Mitteilung*.



# Lebenslauf

## PERSÖNLICHE DATEN

**Name:** Steiger Rainer  
**Geburtsdatum:** 02.01.1969  
**Nationalität:** Schweiz  
**E-mail:** steiger@rainer.ch

## AUSBILDUNG

**08.76 - 07.90** Primarschule und Mittelschule in Chur, Schweiz  
Matura Typus C

**10.90 - 07.91** **Biologiestudium, Universität Zürich, Schweiz**

**10.91 - 10.93** **Chemiestudium, Universität Zürich, Schweiz**  
Juli 1992: 1. Vordiplom  
Oktober 1993: 2. Vordiplom

**10.93 - 10.95** **Universität Zürich, Schweiz**  
Studium der Organischen Chemie, Nebenfach  
Biochemie und Umweltnaturwissenschaften

**10.95 - 04.96** **Universität Zürich, Schweiz**  
Diplomarbeit betreut von Prof. Dr. J. A. Robinson

**10.96** **Diplom in Organischer Chemie**

**01.97 - 12.97** **Universität Zürich, Schweiz**  
Doktorarbeit in Theoretischer Chemie, betreut von Prof. Dr. W. Thiel

**01.98 - 08.99** **Systemadministrator in der Gruppe von Prof. W. Thiel**

**09.99 - 12.02** **Max-Planck-Institut für Kohlenforschung, Mülheim, Deutschland**  
Doktorarbeit, betreut von Prof. Dr. W. Thiel.

## STIPENDIEN

**9.99 - 12.02** **Promotionsstipendium der Max-Planck-Gesellschaft**